

POLITEXT

Francesc Tarrés Ruiz

Sistemas audiovisuales

1 - Televisión analógica y digital

EDICIONS UPC

Primera edición: junio de 2000

Diseño de la cubierta: Manuel Andreu

© Francesc Tarrés, 2000

© Edicions UPC, 2000
Edicions de la Universitat Politècnica de Catalunya, SL
Jordi Girona Salgado 31, 08034 Barcelona
Tel.: 934 016 883 Fax: 934 015 885
Edicions Virtuals: www.edicionsupc.es
E-mail: edicions-upc@upc.es

Producción: CPET (Centre de Publicacions del Campus Nord)
La Cup. Gran Capità s/n, 08034 Barcelona

Depósito legal: B-30.958-2000
ISBN: 84-8301-393-2
ISBN Obra completa: 84-8301-401-7

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del copyright, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

Presentación

Los sistemas de comunicación audiovisual han experimentado, desde los inicios de la televisión en blanco y negro, una constante evolución tecnológica encaminada a mejorar tanto la calidad como la cantidad de los servicios. No obstante, debe tenerse en cuenta que el elevado número de usuarios y el coste de los equipos terminales suponen una considerable inercia comercial que plantea serias dificultades a la introducción de nuevos sistemas. Por ello, la evolución tecnológica casi siempre se realiza de forma progresiva, manteniendo cierto grado de compatibilidad con los sistemas precedentes e introduciendo de forma paulatina mejoras y servicios adicionales. La introducción de información de color compatible con los sistemas en blanco y negro o la adición de canales de audio estereofónico constituyen claros ejemplos de esta evolución progresiva. Más recientemente, la rápida implantación de la televisión digital por vía satélite se ha obtenido sin que ello suponga un coste excesivo para el usuario, decodificando la señal digital en el extremo receptor y convirtiéndola a los formatos analógicos convencionales, para los que ya estaban preparados los equipos reproductores. También de forma progresiva, se están introduciendo nuevos receptores que admiten tanto entradas analógicas como digitales.

Esto significa que, al menos durante cierto tiempo, coexistirán los formatos analógicos y los digitales a los que, seguramente, se incorporaran nuevas características como la alta definición, los canales de retorno para interactuar con el usuario, el acceso a servicios audiovisuales o multimedia a través de internet, etc. Aunque en un futuro próximo es probable que dejen de transmitirse señales de televisión analógicas, los receptores deberán seguir manteniendo éstos formatos para proporcionar cierto grado de compatibilidad con los reproductores de VHS, cámaras de vídeo y otros equipos que actualmente se disponen y que el usuario deseará mantener.

Este texto pretende proporcionar una perspectiva general de los principios tecnológicos en los que se basan los sistemas de comunicación audiovisual, teniendo en cuenta los principales formatos analógicos y digitales que actualmente comparten el mercado de equipos y sistemas no profesionales. Debido a la extensión y complejidad de algunos temas, hemos considerado oportuno dividir los contenidos en dos volúmenes que pueden seguirse de forma totalmente autónoma e independiente. En éste primer volumen se tratarán únicamente los aspectos relacionados con la captura, el tratamiento y la transmisión de las imágenes, centrándonos principalmente en los sistemas de televisión analógica y digital y en las cámaras de vídeo. En un segundo volumen se desarrollarán los sistemas de audio y las comunicaciones audiovisuales en sistemas o redes informáticas. Aunque los contenidos se agrupan, principalmente, en torno a las aplicaciones de televisión (audio y vídeo) y la codificación de vídeo en sistemas informáticos, los conceptos expuestos son válidos y generalizables a otros formatos

audiovisuales domésticos como el DVD (*Digital Versatile Disc*), el Minidisc, los reproductores MP3, los sistemas de videoconferencia por internet o RDSI, etc. En cualquier caso, en todo momento se proporcionará información detallada de los aspectos comunes y las diferencias más significativas entre los formatos más conocidos.

El primer capítulo es una introducción general a los sistemas de comunicación visual en la que se desglosan las etapas necesarias para convertir una escena real en una señal que represente la información contenida en esa escena. Para ello, se describen las principales características del sistema visual humano, que nos permiten discernir aquellos parámetros físicos que realmente contribuyen en la percepción de la escena de los que no aportan información significativa. Todos los elementos desarrollados en este primer tema pueden considerarse básicos y se aplican tanto a los sistemas analógicos como a los digitales. Se examinan con detalle aspectos diversos como la descomposición en componentes de color RGB, la transmisión de la información mediante componentes de luminancia y señales diferencia de color, el número de imágenes por segundo, los sistemas entrelazados y no entrelazados, las relaciones de aspecto de la pantalla, el número de líneas, el ancho de banda de las señales y los principios de la televisión analógica y digital. Se incluyen numerosas referencias históricas y tecnológicas con el objetivo de justificar la procedencia de los parámetros más habituales en los sistemas de televisión.

Los sistemas de televisión analógicos se desarrollan en los dos capítulos siguientes. El primero se dedica a los sistemas en blanco y negro, prestando especial atención a las señales de sincronismo y a los principios de representación de imágenes mediante tubos de rayos catódicos. También se analizan, a nivel de diagrama de bloques, los subsistemas del receptor dedicados a recuperar la información de sincronismo a partir de la señal de vídeo compuesto. El estudio de los sistemas de color compatibles se centra en la descripción del NTSC y el PAL. En ambos casos se presentan abundantes justificaciones analíticas que pueden obviarse sin que se produzca una pérdida de continuidad en la comprensión de los aspectos más tecnológicos.

La televisión digital y los principios de compresión de imágenes se describen en el capítulo 4. El lector interesado exclusivamente en los sistemas digitales puede abordar directamente este tema, en el que sólo se utilizan los elementos y conceptos desarrollados en el primer capítulo. El estudio incluye un análisis exhaustivo de las diferentes alternativas para la compresión de imágenes con pérdidas y sin pérdidas. Se describen las peculiaridades del estándar JPEG para la compresión de imágenes estacionarias y se profundiza en los estándares MPEG-1 y MPEG-2 para las secuencias de imágenes en movimiento. La parte final se dedica al estándar DVB, que define la transmisión de televisión en formato digital.

El último capítulo se dedica a las cámaras de vídeo. Se proporcionan los elementos de óptica geométrica necesarios para comprender los conceptos de enfoque, *zoom* y las relaciones entre iris, obturador y profundidad de campo. También se describen las principales unidades fotométricas y se presentan los principios físicos de los sensores CCD para la captura de imágenes.

Barcelona, abril de 2000

Índice

1 Elementos básicos de los sistemas de comunicación visual

1.1	Introducción.....	15
1.2	Cadena básica de un sistema de comunicación visual	16
1.3	De la escena a la señal de televisión.....	19
1.4	Componentes de color RGB	21
1.4.1	La luz y el color.....	21
1.4.2	Percepción del color en el sistema visual humano.....	24
1.4.3	El ojo humano	25
1.4.4	Células sensibles: conos y bastones.....	26
1.4.5	Estímulos de los fotorreceptores: sensación de color.....	31
1.4.6	Tríadas RGB en <i>displays</i> de TV	32
1.4.7	Sistemas de coordenadas para la representación del color: principios de colorimetría.....	33
1.4.8	Normalización de las coordenadas de color: coeficientes triestímulo.....	35
1.4.9	Cambio de sistemas de coordenadas.....	35
1.4.10	Sistema de coordenadas con primarios monocromáticos	36
1.4.11	Sistema de coordenadas basado en los fósforos del receptor NTSC.....	37
1.4.12	Triángulo de <i>Maxwell</i> y diagrama cromático	39
1.4.13	Representación gráfica de la mezcla de colores	41
1.4.14	Situación de los colores en el triángulo de <i>Maxwell</i> y en el diagrama de cromaticidad.....	43
1.4.15	Algunos comentarios sobre la representación gráfica de los colores	48
1.4.16	La señal de luminancia	50
1.4.17	Sistemas de coordenadas XYZ.....	52
1.4.18	Obtención de las componentes de color: filtros dicroicos.....	55
1.4.19	Transmisión de la información de color: señales diferencia de color	57
1.5	Proyección de las escenas en el plano de imagen.....	64
1.5.1	Sistemas para la representación de imágenes 3D	65
1.5.2	Mecanismos de visión tridimensional.....	66
1.5.3	Sistemas estereoscópicos	70
1.5.4	Sistemas autoestereoscópicos.....	77
1.5.5	<i>Displays</i> volumétricos	82
1.5.6	Hologramas y <i>displays</i> holográficos	85

1.6	Limitación de la imagen: tamaño del sensor y relación de aspecto	89
1.7	Muestreo temporal de las imágenes	98
1.7.1	Un poco de historia: del <i>Phenakistoscope</i> al <i>Cinématographe</i>	100
1.7.2	La persistencia, el parpadeo y el fenómeno <i>phi</i>	102
1.7.3	Frecuencia de imagen.....	104
1.7.4	<i>Aliasing</i> temporal en las imágenes	102
1.8	Muestreo espacial de las imágenes: filas	108
1.8.1	Antecedentes de la televisión.....	112
1.8.2	Número de líneas	116
1.8.3	Agudeza visual	117
1.8.4	Distancia de visionado	119
1.8.5	Número mínimo de líneas	120
1.8.6	Selección del número de líneas.....	121
1.8.7	Número de líneas en los sistemas de alta definición	123
1.8.8	Entrelazado de las imágenes.....	123
1.8.9	Problemas derivados del entrelazado de las imágenes	127
1.8.10	Inserción de señales de sincronismo en la señal de televisión	127
1.8.11	Exploración entrelazada de las imágenes en la cámara	129
1.8.12	Receptores de 100 Hz.....	130
1.8.12	Ancho de banda aproximado de la señal de televisión	132
1.8.14	Resolución y factor de resolución en un sistema de televisión	137
1.8.15	Inserción de la información de color	140
1.8.16	Sistemas de vídeo en componentes	141
1.8.17	Sistemas de vídeo compuesto	143
1.9	Muestreo espacial de las imágenes: filas y columnas	145
1.9.1	Conceptos básicos.....	145
1.9.2	Sistemas de adquisición.....	149
1.9.3	Formatos de televisión digital.....	150

2 La señal de televisión en blanco y negro

2.1	Introducción.....	159
2.2	Príncipios de funcionamiento del tubo de rayos catódicos	160
2.2.1	Cátodo, filamento y rejilla.....	160
2.2.2	Rejillas de aceleración y enfoque.....	161
2.2.3	Deflexión del haz	161
2.2.4	Pantalla.....	162
2.2.5	Forma del <i>raster</i>	162
2.2.6	Modulación del haz mediante la señal de luminancia	162
2.2.7	Príncipio de funcionamiento de los tubos de color	164
2.3	Corrección gamma	167
2.4	Señales de deflexión del haz	170
2.4.1	Caso ideal	170
2.4.2	Señales de deflexión del haz: aproximación al caso real	171
2.5	Sincronismo de línea	173

2.6	Sincronismo de campo	180
2.6.1	Sincronización de los generadores de barrido vertical	188
2.7	Especro de la señal de televisión en blanco y negro.....	190

3 La señal de televisión en color: sistemas compatibles

3.1	Introducción.....	197
3.2	Las señales de color en los sistemas NTSC y PAL	198
3.2.1	Señales diferencia de color en el NTSC y PAL	199
3.2.2	Modulación de las señales diferencia de color en los sistemas NTSC y PAL	207
3.3	El sistema NTSC.....	209
3.3.1	Selección de la frecuencia portadora de la señal de croma.....	209
3.3.2	Patrones de interferencia en el sistema NTSC	213
3.3.3	Modificación de la frecuencia de imagen.....	219
3.3.4	Demodulación de las componentes I y Q: caso ideal	222
3.3.5	Demodulación con errores de fase	224
3.3.6	Inserción de la salva de la portadora de color	225
3.3.7	Demodulación de las componentes de color: banda lateral superior vestigial	228
3.3.8	Demodulación directa de las componentes U y V.....	233
3.3.9	Diagrama de bloques de un codificador NTSC.....	235
3.3.10	Diagrama de bloques del decodificador NTSC.....	238
3.3.11	Separación de las señales de luminancia y croma	239
3.3.12	Regeneración de la portadora de color	246
3.3.13	Problemas del sistema NTSC	247
3.4	El sistema PAL	250
3.4.1	Corrección de la fase mediante la inversión de una de las componentes de croma	250
3.4.2	Diagrama de bloques de un decodificador de croma en PAL-S.....	252
3.4.3	Diagrama de bloques de un decodificador de croma en PAL-D	253
3.4.4	Selección de la frecuencia portadora de color.....	258
3.4.5	Especro de la señal PAL.....	262
3.4.6	Generación de sincronismos en el sistema PAL	264
3.4.7	La señal de salva en el sistema PAL	265
3.4.8	Diagramas de bloques de un codificador y un decodificador PAL	267
3.5	Normas utilizadas en la transmisión de señales de TV analógicas en color.....	268

4 Televisión digital

4.1	Introducción.....	273
4.2	Compresión de imágenes: necesidad y conceptos básicos.....	276
4.3	Medida de la información de una fuente	284
4.3.1	Entropía de una fuente de mensajes	286
4.4	Códigos de longitud variable	287

4.4.1	Códigos de <i>Huffman</i>	288
4.4.2	Modificaciones del código de <i>Huffman</i>	291
4.5	Otros tipos de códigos	293
4.5.1	Codificación por longitud de series (<i>Run Length Encoding-RLE</i>).....	293
4.5.2	Codificación LZW	294
4.5.3	Codificación aritmética	295
4.6	Codificación y predictores.....	296
4.7	La transformada coseno discreta.....	298
4.7.1	Definición de la transformada coseno unidimensional.....	301
4.7.2	Extensión de la transformada coseno a dos dimensiones	305
4.7.3	Codificación por zonas.....	311
4.7.4	Codificación por umbral.....	313
4.8	Codificación de imágenes estacionarias: el estándar JPEG	314
4.8.1	JPEG: modo secuencial base (<i>baseline</i>)	316
4.8.2	Codificación de imágenes con múltiples componentes	324
4.8.3	Modo progresivo.....	325
4.8.4	Calidades subjetivas obtenidas con el JPEG.....	327
4.9	Codificación de secuencias de vídeo: MPEG	329
4.9.1	Uso de la redundancia temporal para la compresión de vídeo	332
4.9.2	Compensación de movimiento.....	335
4.9.3	Algoritmos para la estimación de los vectores de movimiento.....	340
4.9.4	Tipos de imágenes en el MPEG	347
4.9.5	Orden de transmisión de las imágenes	349
4.9.6	Algunos ejemplos con secuencias y codificadores ideales	351
4.9.7	Diagramas generales del proceso de codificación y decodificación de imágenes.....	352
4.9.8	Estructura jerárquica de la trama MPEG	354
4.9.9	Algunos detalles sobre la codificación de las imágenes del tipo I.....	355
4.9.10	Algunos detalles sobre la codificación de las imágenes del tipo P.....	357
4.9.11	Detalles sobre la codificación de las imágenes del tipo B	359
4.9.12	Control de la tasa de bits	361
4.9.13	Tratamiento de imágenes entrelazadas en el MPEG-2	362
4.9.14	Restricción de parámetros, niveles y perfiles.....	366
4.10	Multiplexación de secuencias MPEG.....	369
4.10.1	Estructura del paquete básico (<i>Packetized Elementary Stream</i>)	370
4.10.2	La trama de programa (<i>Program Stream</i>).....	372
4.10.3	La trama de transporte.....	372
4.10.4	Información específica de programas (<i>Program Specific Information – PSI</i>)	375
4.10.5	Multiplexado de paquetes elementales en la trama de transporte.....	376
4.11	Principios del DVB	378
4.11.1	Estándar DVB-S	379
4.11.2	Estándar DVB-C	380
4.11.3	Estándar DVB-T	380

5 Cámaras

5.1	Introducción.....	381
5.2	Diagrama de bloques de una cámara	382
5.3	Principios de óptica geométrica y formación de imágenes	386
5.3.1	Lentes, grupo óptico y lente ideal	387
5.3.2	Formación de imágenes con lentes ideales	391
5.3.3	Angulo de visión.....	394
5.3.4	Regulación de la cantidad de luz: diafragma y obturador.....	397
5.3.5	Profundidad de campo.....	402
5.4	Fotometría	408
5.4.1	Intensidad luminosa: la candela	409
5.4.2	El flujo luminoso: lumen	413
5.4.3	La iluminancia: lux	414
5.5	Sensores de imagen de estado sólido.....	416
5.5.1	Sensores CCD matriciales	420
5.5.2	Sensores CCD de transferencia de cuadro (<i>Frame Transfer CCD</i>).....	421
5.5.3	Sensores CCD de transferencia de cuadro divididos (<i>Split Frame Transfer CCD</i>).....	422
5.5.4	Sensores CCD de transferencia interlínea (<i>Interline Transfer</i>).....	423
5.5.5	Dispositivos de inyección de carga (<i>Charge Injection Devices CID</i>)	425
5.5.6	Sensores de píxel activo	426
5.5.7	Captura de imágenes en color con un único sensor.....	427
5.6	Sistemas automáticos de control de la señal de vídeo	429
5.6.1	Zoom óptico y zoom digital.....	429
5.6.2	Mecanismos de estabilización de imagen.....	429
5.6.3	Sistemas de autoenfoque	430
5.6.4	Ajuste del balance de blancos.....	434
5.6.5	Modos de exposición y efectos	435
5.7	Tipos de cámaras.....	436
5.7.1	Cámaras de estudio	436
5.7.2	Cámaras de campo (<i>Electronic News Gathering – Eng</i>)	437
5.7.3	Cámaras de circuito cerrado de TV o de vídeo vigilancia	437
5.7.4	Cámaras industriales y de visión artificial	437
5.7.5	Cámaras de vídeo domésticas	439
5.7.6	Cámaras para aplicaciones médicas y científicas.....	439
5.7.7	<i>WebCams</i>	439
	Bibliografía	441
	Índice alfabético.....	443

Bibliografía

ABRAMSON, A. *The story of television: 1880 to 1941*. Ed. McFarland.

BENOIT, H. *Televisión digital*. Madrid. Paraninfo. 1998.

BRUCH, W. "The PAL colour TV system basic principles of modulation and demodulation". *NTZ Communications Journal 3*, pp 255-268. 1964.

GIBSON, J; BERGER, T. *Digital compression for multimedia. Principles and Standards*. Morgan Kaufmann. 1998.

GONZALEZ, R; WOODS, J. *Digital Image Processing*. 2 Ed. Addison-Wesley. 1992

HILL, R. *A first course in coding theory*. Oxford. Clarendon Press. 1986.

ISO 10918. *Digital compression and coding of continuous tone still images (JPEG)*. ISO/IEC.

ISO 11172-1, 11172-2, 11172-3. *Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbps. (MPEG-1 video, audio, system)*. ISO/IEC.

ISO 13818-1, 13818-2, 13818-3. *Coding of moving pictures and associated audio (MPEG-2 video, audio, system)*. ISO/IEC.

JACKSON, K.J; TOWNSEND, G.B. *Television and video engineer's reference book*. London. Butterworth-Heinemann. 1991.

NETRAVALI, A., HASSELL, B. *Digital Pictures. Representation, compression and standards*. New Jersey. 2 Ed. Plenum Press. 1995.

RAO, K.R; HWANG, J.J. *Techniques and Standards for image, video and audio coding*. Prentice Hall. 1996.

RUIU, D., et al. *Testing digital video*. Amsterdam. Hewlett-Packard. 1997.

SAWYER, PRONK. *Cámaras digitales*. Madrid. Editorial Paraninfo.

SINCLAIR, I. *Birth of the box: the story of television*. Sigma Press. 1995.

SOLARI, S.J. *Digital video and audio compression*. McGraw-Hill. 1997.

TAYLOR, J. *DVD Demystified*. McGraw-Hill. 1998.

TORRES, L; CASAS, J.R; LLEIDA, E. *Sistemas analógicos y digitales de televisión*. Barcelona. Colección Politex. Edicions UPC.

UMBAUGH, S. *Computer Vision and Image Processing*. Prentice Hall. 1998.

Índice alfabético

- 100 Hz, receptores, 130
4:2:0, 154
4:2:2, 150
4:4:4, 150
64-QAM, 380
- A**
Aberración, 389
 cromática, 389
 de coma, 389
 de curvatura, 389
 esférica, 389
Acceso condicional, 18, 275
Acomodación, 68
Advanced Photo System, 395
Agudeza visual, 117
Aliasing temporal, 107
Alta definición, 96
 formatos digitales, 157
Anaglyph, 70
Análisis de umbral, 313
Anamórficas, lentes, 93
Ancho de banda, 132
 calidad subjetiva, 137
 señal TV, 132
Ángulo
 de observación, 123
 de visión, 90, 394
 sólido, 409
 vertical de observación, 120
Ánodo, 161
Antena. 18
Apertura numérica, 399
Aproximaciones sucesivas, 325
- APS, 395
Arj, 277
Armat, Thomas, 102
Arrastre de línea, 131, 127
ASA, 408
Astigmatismo, 29, 389
ATM, 329
Audio
 estéreo, 273
 multicanal, 273
Autoenfoque, 430
 por CCD lineal, 432
 por infrarrojos, 430
 por ultrasonidos, 430
AVI, 155, 334
- B**
Baird, J., 116
Balance de blancos, 434
Banda IV, 271
Banda lateral superior PAL, 264
Banda V, 271
Banda visible, 22
Barkaunsen, 173
Barras de color, 201
Barrido
 horizontal, 170
 vertical, 170
Baseline, JPEG, 316
Bastones, 26
BBC, 95
Betacam, 141
Bidireccionales, 347
Binario natural, 293

- Bits, 285
Black-matrix, 166
Block matching, 341
Blooming, 425
Bloque de imagen, 325, 341, 355
Bobinas de deflexión, 170
Borrado de campo, 180
Braun, 114
Brillo, 24
Bruch, 250
Búsqueda
 en tres etapas, 344
 jerárquica, 344, 346
 logarítmica, 344
- C**
- C de Landolt, 117
C.I.E, 36
Cable, televisión, 18, 198
Calidad *broadcasting*, 329
Calidades subjetivas, JPEG, 327
Cámara, 381
Cámara
 reflex, 385
 diagrama de bloques, 382
 exploración entrelazada, 129
 de campo, 437
 de estudio, 436
 industrial, 437
 lineal, 438
 progresiva, 129, 438
Camcorders, 439
Camera Control Unit (CCU), 436
Canal en espera, 132
Canales
 guía, 273
 temáticos, 273
Candela, 409, 410
Cañón electrónico, 114
Cañones
 en delta, 165
 en línea, 165
Capa
 de bloque, 355
 de grupo de imágenes, 355
 de imagen, 355
- de macrobloque, 355
de secuencia, 354
de *slice*, 355
Capacidades commutadas, 253, 257
Carey, G.R, 112
CAT, 375
Cátodo, 160
Cátodos de estado sólido, 160
CATV, 18, 19
CCD, 149, 253, 257, 381, 416
 de línea, 419
 de transferencia interlínea, 423
 de tres fases, 416
 matriciales, 420
CCTV, 384
CCU, 436
CD-ROM, 277, 329, 347
Charge Coupled Device, 381, 416
Charge Injection Devices, CID, 425
Chlorolabe, 30
CIE, 412
CIF, 156, 331
Cinemascope, 93
Cinématographe, 101
Cinepack, 334
Cinerama, 93
Círculo separador de sincronismos, 183
Closed GOP, 350
Codecs, 335
Codificación
 aritmética, 293, 295
 con pérdidas, 281
 de canal, 274
 de fuente, 274
 de imágenes, 314, 357
 diferencial, 279, 283
 por umbral, 313
 por zonas, 311
 RLE, 293
 secuencial, JPEG, 316
Codificador PAL, 267
Código
 externo, 379
 Gray, 293
 interno, 379
 de encriptación, 275

- de Huffman, 288
de Huffman de desplazamiento, 292
de longitud variable, 287
protección errores, 275
truncados, 291
Coeficiente de correlación, 296
Coeficientes
 AC, 317
 cromáticos, 40
 DC, 317
 de alterna, 317
 de predicción, 279
 transformados, 307
 triestímulo, 35
COFDM, 380
Colores monocromáticos, 22
Commission Internationale de L'Eclairé, 36
Common Intermediate Format, 156
Compact disc, 273
Compactación de la energía, 300
Compatibilidad, 57
 directa, 197
 indirecta, 58, 197
Compatible, 197
Compensación de movimiento, 335
Compensación de movimiento bidireccional, 348
Componentes de color, 21
Compresión
 con pérdidas, 277
 intra-trama, 338
 sin pérdidas, 277
Configuración FIFO, 419
Conos, 26
Contraste, 147
Coordenadas XYZ, 50, 52
Córnea, 25
Coroide, 26
Corrección gamma, 167
Correlación cruzada, 341
Corriente de oscuridad, 424
Cristalino, 26
Cromaticidad uniforme, 55
Crominancia, 197
 PAL, 199
Cruce de color, 257
- D**
- d.p.i., 149
D1, 141
D2, 141
D2-MAC, 116, 141
Daguerre, Louis Jacques Mandé, 100
Daltonismo, 32
DCT, 300
 Bidimensional, 317
De France, H, 250
Decodificador PAL, 267
Decodificadores de satélite, 198
Deflexión, 161, 165, 170
 Horizontal, 170
Della Porta, Giovanni Battista, 68
Delta frames, 334
Demultiplexación, 331
Desplazamientos fraccionales, 343
Detección del sincronismo, 176
DFT, 304
Diafragma, 397
Diagrama cromático, 39
Dickson, W.K.L, 101
Diferencia de color, 197
Digital 8, 384, 439
Digital Scan, 131
Digital Still Cameras, 385
Digital Versatile Disc, 329
Digital Video Broadcasting, 274, 329, 378
Digital Video Interactive, 154
Dirección del gradiente conjugado, 344
Disparo controlado, 179
Disparo directo, 179
Displays holográficos, 85
Displays volumétricos, 82
Dispositivos de acoplamiento de carga, 416
Distancia de visionado, 119
Distancia focal, 387, 392
Distorsión de cojín, 162
Distorsión geométrica, 389
Doble cortinilla, 402
DTS, 371
DVB, 116, 274, 329, 378
DVB-C, 380
DVB-S, 379
DVB-T, 380

DVCam, 141

DVCPro, 141

DVD, 116, 350

DVD-Video, 329, 347

DVI, 154

E

Eastman, George, 100

EBU, 38

EDTV, 20, 331

Efecto

Hannover, 253

Kell, 127, 135

EIT, 376

Electronic News Gathering, 437

Elementary stream, 369

Elipses de Mac-Adam, 54

Enfoque pasivo, 431

ENG cameras, 437

Enhanced definition TV, 20

Ensamblaje de bloques, 341

Entrelazado de las imágenes, 123

Entropía, 286

de la fuente, 286

EOB, 322

Error

absoluto, 341

cuadrático, 341

de fase, 251

Erythrolabe, 30

ES audio, 369

ES data, 369

ES, MPEG, 369

Espectador

activo, 80

pasivo, 80

Espectro PAL, 262

Espectro señal de televisión, 190

Estabilización de imagen, 429

Estereorradianes, 409

Estereoscopio de Brewster, 70

Estroboscopio, 107

Estructuras de predicción, 297

Euclides de Alejandría, 68

Euroconector, 19, 143, 198

Exploración

en *zigzag*, 320

entrelazada, 129

progresiva, 331

Exposímetro, 408

F

Factor

de Kell, 136

de resolución, 137

de *zoom*, 429

Fatiga ocular, 81

Fenómeno *phi*, 98, 102, 103

FFT, 300

Filamento, 160

Filtro

digital, 132

predictor, 279

Filtros

de color, 427

dicroicos, 55

Flicker, 102, 330

Flujo luminoso, 413

Fly-wheel, 189

Foco, 387

Formato 4:2:0, 154

Formato

de la Academia93

panorámico, 96

SIF, 330

Cinematográficos, 93

Fósforos, 165

en delta, 33

Fotografía

35 mm, 395

digital, 385

electrónica, 149

de miniatura, 35 mm, 386

Fotogramas

clave, 334

diferencia, 334

Fotometría, 408

Fotómetro, 408

Fourier, 298

Fourney, 379

Fóvea, 28

Fracción de Weber, 147

Frame

- differencing*, 334
- transfer CCD*, 421

Frecuencia

- de imagen, 104
- de línea, 181
- espacial, 312
- subportadora, 258

Fuente de mensajes, 284

Funciones cosenoïdales, 308

G

Gabor, D, 85

Gama de colores, 48

Gamma, 164

Generación

- de barrido horizontal, 185
- de sincronismos PAL, 264

Generador de efectos, 384

Generadores de barrido vertical, 188

Goodwin, Hannibal, 100

GOP, 347

Grandes angulares, 395

Group of pictures, 347

Grupo

- cerrado de imágenes, 350
- de imágenes, 347
- óptico, 383, 386

H

H-261, 275, 340

H-263, 275, 328

Hannover, efecto, 253

Hart, 298

HDTV, 331

Helmholtz, Hermann L. F. von, 29

HHR, 331

Hi-8, 384, 439

High level, 368

High profile, 368

High-1440 level, 368

Hipermetropía, 29

Histograma, 280

Holograma, 85

Huffman, 281, 320

Huffman, códigos, 288

I

Iconoscopio, 114

Iluminancia, 414

Imagen

- B, 339
- de referencia, 337, 341
- I, 339
- P, 339

Imágenes

- 3D, 65
- base, 308
- intra, 316
- reales, 388
- virtuales, 388

IMAX, 94

IMAX-3D, 73

Impulso de sincronismo de campo, 181, 184

Impulsos

- de pre-igualación, 181
- de sincronismo de línea, 184

Indeo 3.2, 334

Intensidad del haz, 163

Intercalación de líneas, 125

Interleaving, 325

Interline Transfer CCD, 423

Interline twiter, 127

International Commission of Illumination, 412

Interpolación de líneas, 132

Intra-trama, 343, 347

IRD, 18

Iris, 397

ITU 601, 150, 276, 330

ITU-601 4:2:0, 332

ITU-656, 153

J

JBIG, 315

Joint Photographic Experts Group, 315

JPEG, 275, 278, 298, 314

- 2000, 300, 315, 328

Lossless, 315

LS, 315

modo jerárquico, 325

modo progresivo, 325

modo secuencial base, 316

K

Karhunen-Loeve, 298

Kell, efecto, 127

Kell, R.D., 134

Key frame, 334

Kinemacolor, 102

Kinetograph, 101

Kinetoscope, 101

L

Lemple, 294

Lente, 387

convergente, 387

divergente, 387

ideal, 390

anamórfica, 93

Line crawl, 127

Line Flywheel, 179

Línea activa, 128

Líneas

NTSC, 251

PAL, 251

Lohscheller, 317

Lossless compression, 277

Lossy, compression, 277

Low level, 367

Lumen, 413

Lumière, Hnos, 101

Luminancia, 50, 415

Lux, 414

Luz blanca, 22

LZW, 293, 294

M

Macrobloque, 341, 355

Macrobloques,

entrelazado, 363

tipos B, 359

tipos P, 358

Maculea lutea, 28

Main Level, 368

Main Profile, 368

Main Profile, Main Level, 369

Markov, 311

Máscara, 162, 165

zonal, 313

Matiz de color, 24

Matrices de cuantificación, 320

Memorias digitales, 253

Mensajes tipo B, 322

Mesópica, 27

Método base, JPEG, 316

Mezcla de colores, 41

MiniDV, 384, 439

Miopía, 29

MJPEG, 332

Modo

jerárquico, JPEG, 316

progresivo, JPEG, 316

sin pérdidas, JPEG, 316

Modos de exposición, 435

MOSFET, 416

Motion-JPEG, 332

Moving Pictures Experts Group, 329

[MP@ML](#), 154, 369

MPEG, 275, 278, 298

MPEG-1, 155, 275, 329, 340, 341

MPEG-2, 116, 154, 275, 329

MPEG-2

sistemas, 274

imágenes entrelazadas, 362

MPEG-4, 300, 328, 329

MPEG-7, 329

MPEG-Systems, 369

Muestreo

espacial, 108

temporal, 98

MultiCam 1300AF, 434

Múltiples cortinillas, 402

Multiplexación, 331

Multiplexación vídeo-audio, 274

Multiplexor MPEG, 376

MUSE, 95, 141

Muybridge, Eadweard, 100

N

National Television System Committee, 198

NHK, 95

NICAM, 269

Nipkow, P, 113

NIT, 376

Nivel de blanco, 176

- Nivel de negro, 176
Nivel de sincronismo, 176
Niveles MPEG, 367
Norma G, 269
Normas TV CCIR, 268
NTSC, 63, 197, 198
Número de líneas, 116
 alta definición, 123
 selección, 121
 mínimo, 120
Números f, 399
- O**
Obturador, 397, 399
 doble cortinilla, 401
 electrónico, 74, 383
Ojo de pez, 395
- P**
Packet identification code, 374
Packetized elementary stream, 370
Packs (MPEG), 372
Pagar por
 canal, 18, 273
 ver, 18
 programa, 273
PAL, 63, 197, 250, 258
PAL D, 252
PAL S, 252
Palabra código, 290
Palabras tipo A, 322
PALplus, 96, 98, 123
Pantalla, 162
Paquetes elementales, 370
Paquetes, MPEG, 372
Parallax, 67, 80
Parámetro p, 407
Parpadeo, 102
PAT, 375
Patrón
 interferente, 145, 260
 interferente PAL, 258
Pay per channel, 18, 273
Pay per view, 18, 273
PCM, 273
PCR, 374
- Perfiles MPEG, 367
Persiana veneciana, 253
Persistencia, 98, 102
PES, MPEG, 370
Phase alternating line, 199
Phase locked loop, 179
Phenakistoscope, 100
Picture elements, 146
PID, 374
Pistas monocularares, 66
Píxel, 146, 156
Plano de imagen, 391
Plano focal, 387, 393
PLL, 179
PMT, 375
Poder de resolución, 117, 389
Polaroid, 71
Pórtico
 anterior, 175
 posterior, 176
Post-igualación, 181
Predicción
 backward, 339, 340
 bidireccional, 339
 forward, 339
 inter-trama, 296
 intra-trama, 296
Predictor, 279
 diferencial, 296
Presentation time stamp, 371
Primarios monocromáticos, 36
Primarios NTSC, 37
Prisma dicroico, 427
Profundidad de campo, 402
Program
 asignation table, 375
 clock reference, 374
 map table, 375
 specific information, 375
 stream, 370, 372
PSI, 375
PTS, 371
Punto ciego, 28
Pupila, 25
Puppet effect, 85
Purkinje, 27

Q

QCIF, 156
QPSK, 379
QSIF, 156

R

R/S, 321
Radiodifusión terrena, 198
Raster, 162
RCA, 164
Receptor decodificador integrado, 18
Receptores de 100 Hz, 130
Red de distribución, 17
Reed-Solomon, 379
Reflex, 385
Registro magnético, 384
Rejilla, 160, 161
Relación de aspecto, 89, 91
 16:9, 95
 4:3, 95
Relación de entrelazado, 124
Resolución, 137
Resolución en color, 256
Restricción de parámetros, 366
Retina, 26
Retrocompatibilidad, 58, 197
RLE, 293, 320
Rosing, Boris, 114
Rotación de fase, 251
RS-232, 385
Run length encoding, 293
Run/Size, 321

S

Salva PAL, 265
Saturación de color, 24, 45
Scanners, 149
SDT, 376
SECAM, 63, 197, 250, 269
Secuencia 8 campos, 261
Selección espectral, 325
Senlecq, C., 112
Sensación de color, 31
Sensor de imagen, 383
Sensores de píxel activo, 426
Señales de barrido, 178

Señales

 de color, 198
 de prueba, 180
 de sincronismo, 127
 diferencia de color, 57, 198

Serrado entre líneas, 362

Showscan, 105
SIF, 155, 331
Significant Pel Area, 330
Simple profile, 368

Sinapsis, 26

Sincronismo
 de campo, 180
 de línea, 173

Sistema

 NTSC, 63
 PAL, 63, 250
 UVW, 55

Sistemas

 autoestereoscópicos, 77
 de autoenfoque, 430
 digitales, color, 64
 estereoscópicos, 70
 tridimensionales, 65
Skipped macroblock, 357
Slice, 355
SMATV, 378
Smear, 420, 421
SMPTE, 37, 95
Snellen, patrón de, 118
SNR scalable, 368
Sobremodulación, 200

Society of Motion Picture and Television Engineers, 37

Source intermediate format, 155
SPA, 330
Spatial scalable, 368
Split Frame Transfer CCD, 422
Stanford, Leland, 100
Stream de vídeo, 331
Stuffing, 374
Subimágenes, 297
Supportadora de audio, 261
S-Vídeo, 19, 143

T

Tabla de

asignación de programas, 375

descripción de servicio, 376

hora y fecha, 376

información de red, 376

información de sucesos, 376

cuantificación, 317

Huffman, JPEG, 324

Lohscheller, 317

Talbot, W.H.F, 100

Talbot-Plateau, ley de, 102

Tamaño del sensor, 89, 394

Tarjetas digitalizadoras, 149

Tasa de bits, MPEG, 361

TDT, 376

Técnica de umbral, 314

Telecine, 106

Telefunken, 250

Teleobjetivo, 395

Teletexto, 180

Televisión digital, 273

por cable, 275

Televisión en color, 197

Televisión interactiva, 18, 378

Televisión por satélite, 18

Teoría tricromática, 29

Tiempo de

exposición, 397

línea, 128

Tonalidad de color, 45

Tono de color, 24

Trama de

programa, 370

transporte, 370, 372

vídeo, 331

Trama elemental MPEG, 369

Trama MPEG, 354

Transferencia de cuadro, CCD, 421

Transformación diferencial, 279

Transformada

coseno, 298, 299, 337

wavelet, 328

Transformadas

bidimensionales, 305

ortonormales, 303

Transmultiplexor MPEG, 378

Transport stream, 370

Tríadas

rectangulares, 165

RGB, 32

Triángulo de Maxwell, 39

Triniton, 33, 167

Tubo de

cámara, 116

rayos catódicos, 32, 114, 160

TV de alta definición, 96

U

Unidad de control de cámara, 436

Unidades fotométricas, 408

USB, 385

V

VCO, 179

1 Elementos básicos de los sistemas de comunicación visual

1.1 Introducción

En este capítulo pretendemos proporcionar una perspectiva general de los distintos elementos que integran un sistema de comunicación de imágenes en movimiento. Expondremos, desde un punto de vista muy general y utilizando en algunos casos razonamientos intuitivos, los distintos subsistemas que intervienen en la codificación de una escena en una señal eléctrica y cómo a partir de éstas se puede reproducir una aproximación a la escena original. Nos concentraremos únicamente en los sistemas que intervienen en la generación de la señal de banda base de los distintos sistemas de televisión, suponiendo, por tanto, que el lector está suficientemente familiarizado con las nociones básicas relativas a modulación y radiodifusión que permiten la transmisión a distancia de estas señales.

El objetivo final es múltiple. Por una parte queremos presentar una perspectiva general de la tecnología usada en los sistemas de representación de imágenes en movimiento actuales, sus tendencias futuras y sus antecedentes históricos. También queremos proporcionar las bases teóricas y la filosofía general de los sistemas de televisión, preparando al lector para poder abordar en capítulos posteriores los detalles concretos de cada sistema. Por ello, se incluyen aspectos tan diversos como los fundamentos del sistema visual humano, los principios de colorimetría, la definición de diversos parámetros básicos en la señal de televisión como son el número de líneas, el número de imágenes por segundo, su ancho de banda, resolución, entrelazado, etc. Se presentan nociones básicas sobre sistemas de televisión digital y de alta definición y se introducen diversas alternativas que actualmente se están considerando para la representación de imágenes en tres dimensiones. La presentación de estos conceptos es en algunos casos rigurosa y completa por cuanto, o bien se trata de nociones básicas que tendrán una importante incidencia en capítulos posteriores, o bien son aspectos parciales que no serán cubiertos posteriormente. En otros casos, la presentación tiene un carácter eminentemente introductorio por cuanto se trata de aspectos que serán profundamente desarrollados en otros capítulos. Así, el lector observará que dedicamos varias páginas a diversos aspectos relacionados con el sistema visual humano o a la descomposición de la luz, pues haremos constantes referencias a ellos en futuros capítulos.

Aspiramos a que el lector identifique la función de cada uno de los elementos, así como las limitaciones que estos introducen en la calidad de la señal que finalmente se reproduce y, sobre todo, que sea consciente de que el estado actual de estos sistemas (y con toda probabilidad, su evolución futura) es consecuencia de un proceso evolutivo en el que cualquier innovación tecnológica ha

intentado introducirse en el mercado de una manera gradual, manteniendo en lo posible la compatibilidad y la coexistencia entre varios formatos para garantizar un flujo de ventas en constante aumento a las grandes empresas del sector.

El capítulo está organizado tomando como hilo argumental las distintas etapas de muestreo y proyección que se aplican a una escena para convertirla en una señal eléctrica que contiene la información necesaria para su reproducción en el receptor. Entre estas etapas se considera la descomposición de la imagen en las tres componentes de color primarias, la proyección de la escena a un plano de imagen, los límites físicos del sensor de imagen y su relación de aspecto, y finalmente las distintas etapas de muestreo temporal y espacial de las imágenes. En paralelo a este proceso de conversión progresiva de la escena a una señal eléctrica, desarrollamos aspectos relativos a las características del sistema visual humano que intervienen directamente en la selección de los parámetros de muestreo, evaluando su incidencia sobre la calidad final de la imagen y las posibles pérdidas de información que de ellos se derivan. También aprovechamos este hilo argumental para introducir aspectos históricos relativos al cine o la televisión, tendencias de futuro, nuevas tecnologías, etc.

Esta estructura puede parecer en un principio una exposición un tanto desordenada, ya que no sigue el patrón cronológico usado en la mayoría de textos sobre televisión. Sin embargo, creemos que es muy enriquecedora, ya que desde los inicios se comparan distintas alternativas tecnológicas que proporcionan, sin necesidad de leer todos los capítulos, una panorámica general sobre los sistemas de televisión. Además, los distintos aspectos temáticos que iremos desarrollando surgen de una manera natural y sin pérdida de continuidad. Creemos que la alternancia entre aspectos históricos, analíticos, tecnológicos, visuales, etc, facilita la lectura y aumenta su interés sin que suponga una pérdida de rigor o de perspectiva histórica. En cualquier caso, hemos realizado un considerable esfuerzo para que esta organización, intencionadamente caótica, sea estrictamente progresiva y lineal en lo que se refiere a la exposición de los conceptos necesarios para la comprensión del texto.

1.2 Cadena básica de un sistema de comunicación visual

En la figura 1.1 se representa un diagrama general de un sistema de radiodifusión de señales de televisión. Los distintos elementos que forman este sistema tienen como objetivo transmitir la escena tridimensional para que pueda ser visualizada en tiempo real por múltiples receptores. Junto con la información óptica, también se transmiten una o varias señales adicionales que proporcionan información de audio sobre la escena y que serán consideradas en capítulos posteriores.

Las escenas que pretendemos transmitir pueden considerarse funciones que dependen de varias variables como la posición, el tiempo y la longitud de onda. Para poder transmitir esta información mediante sistemas de comunicación convencionales es necesario realizar distintos tipos de proyecciones y muestreos de la escena que permitan convertir esta función multidimensional en una señal que dependa de una única variable temporal. Esta señal es la que será modulada para enviarla a distintos usuarios. Es evidente que en el proceso de proyección y muestreo de la escena se sufrirán pérdidas de información que inevitablemente conducen a una reducción de la calidad de la imagen. Así, por ejemplo, en los sistemas de televisión convencional sólo se transmite una proyección plana de la escena sobre la cámara, de modo que el receptor reproducirá una imagen en la que no está

codificada la información tridimensional de la escena original. Sin embargo, el espectador será capaz de interpretar correctamente esta información espacial, pues el cerebro humano está entrenado para identificar la posición real de los objetos a partir de su tamaño relativo y de los planos ocultos.

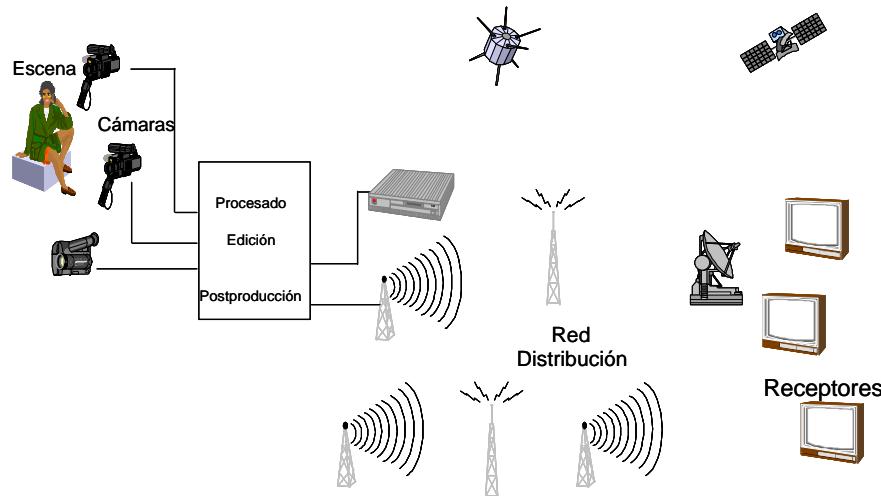


Fig. 1.1 Elementos de un sistema de comunicación visual

En el estudio de televisión intervienen distintos equipos y profesionales cuya función es la de producir, a partir de las distintas tomas de imágenes y sonido de la escena, la señal de TV que finalmente llegará al espectador. En este proceso colaboran profesionales técnicos y artísticos. Los equipos utilizados para la producción de la señal son muy variados y dependen del tipo de programa que se transmita. Los más usados son las cámaras y micrófonos, los sistemas de mezcla y acondicionamiento de señales, los equipos para la edición y postproducción de vídeo y finalmente los equipos de amplificación y transmisión de las señales a los centros de distribución. En una transmisión en directo suelen utilizarse varias cámaras y micrófonos, que son controlados desde una mesa de mezclas en la que se seleccionan los planos y tomas microfónicas que formaran la base del programa. Estas señales son editadas posteriormente para introducir títulos, rótulos y efectos, y finalmente se envían a los sistemas de radiodifusión. Algunos de estos equipos pueden estar integrados en unidades móviles si el lugar donde se producen las escenas así lo requieren. En este caso, la señal se suele enviar al estudio central, donde es reeditada y se le añaden algunos títulos y efectos adicionales. También tienen un papel fundamental en el estudio de televisión los sistemas para el registro de las señales de vídeo en cinta magnética que permiten almacenar los programas en directo y facilitan la producción de documentales y programas registrados. Algunos de estos equipos serán estudiados con cierto detalle en capítulos posteriores.

La red de distribución se encarga de garantizar que las señales lleguen en buenas condiciones al área de cobertura de la emisión. Las características de esta red están determinadas por la tecnología utilizada para la transmisión de las señales pudiendo existir redes de distribución por cable, por satélite, terrestres y, en algunos casos, redes mixtas que utilizan varios de estos sistemas. La red de

radiodifusión terrestre es la que tiene actualmente mayor número de usuarios y es utilizada para la difusión de programas de ámbito nacional o local. En este caso, la distribución de la señal se realiza mediante redes de radioenlaces.

En la radiodifusión directa por satélite la señal se envía desde el centro servidor a un centro de transmisión mediante radioenlaces terrenos o cable y posteriormente es transmitida al satélite (*up-link*) que realiza una trasposición de frecuencia de la señal y la redirige a una amplia zona de cobertura sobre la tierra (*down-link*). La televisión por satélite requiere que el espectador disponga de una antena parabólica de gran ganancia, ya que el nivel de la señal recibida es pequeño debido al largo trayecto que debe recorrer sin amplificación (36000 Km en satélites geoestacionarios), y de un conversor de señales que le permita trasladar la señal del satélite a las bandas de televisión convencionales con las que puede operar su receptor. La *televisión por satélite* proporciona grandes coberturas de ámbito internacional y facilita la distribución de la señal a regiones poco pobladas o de difícil acceso geográfico, cuya cobertura mediante radioenlaces terrestres o cable sería excesivamente costosa.

Los sistemas de distribución por cable (CATV) surgieron a principios de los años 50 en Estados Unidos para proporcionar señal de televisión a regiones cuyo perfil geográfico dificultaba establecer una cobertura convencional mediante radioenlaces. Posteriormente, la CATV se extendió a zonas metropolitanas en las que la presencia de edificios de gran altura dificultaba la correcta recepción de la señal. Actualmente, este sistema de distribución está muy extendido en algunos países donde se ofrece al espectador una señal de televisión de gran calidad (la degradación de la señal por transmisión es muy baja) con un elevado número de canales y de servicios. Normalmente el proveedor del servicio distribuye una señal con más de 50 canales en los que se incorporan tanto canales convencionales (que también se transmiten vía terrestre o satélite) como canales exclusivos para los abonados del servicio. La CATV ha sido precursora en la introducción de los conceptos de *acceso condicional* a la programación, que posteriormente se han extendido a los sistemas terrestres o vía satélite y que requieren que el espectador disponga de un equipo *IRD* (receptor decodificador integrado) que le permita recomponer la información de vídeo que ha sido encriptada en la transmisión. En este entorno aparecen los conceptos de pagar por canal (*pay per channel*) y pagar por ver (*pay per view*). En el primer caso, el espectador paga una cuota mensual que le permite desencriptar toda la programación del canal, mientras que en el segundo caso, el proveedor del servicio informa de precio de cada programa y factura al abonado de acuerdo con la programación que éste ha seleccionado. El concepto de pagar por ver requiere una cierta interactividad entre el usuario y el centro servidor. La CATV también ha sido precursora de este concepto de *televisión interactiva* ya que, desde sus orígenes, ha proporcionado un canal de retorno que permite establecer una comunicación directa con el servidor. Con ello se permite que el espectador participe en encuestas de opinión pública, solicite los vídeos que desea ver o efectúe una compra inmediata de los productos anunciados (telecompra). La extensión del concepto de televisión interactiva a sistemas unidireccionales, como la televisión por satélite o terrestre, requiere que el usuario disponga de un módem telefónico que se utiliza como canal de retorno auxiliar.

En la recepción de la señal y su presentación al espectador intervienen varios subsistemas que dependen del tipo de distribución de la señal utilizado. La *antena* es el elemento encargado de recoger la energía radioeléctrica de las señales de televisión en sistemas de radiodifusión terrestre y por satélite. En el primer caso suele ser una array del tipo YAGI que se orienta hacia el repetidor más cercano y que tiene una ganancia moderada (unos 14 dB's). En el caso de televisión por satélite, suele

tratarse de una antena parabólica de gran ganancia (30 dB's) que se apunta hacia el satélite que transmite las señales y que incorpora un subsistema demodulador que convierte las señales recibidas en señales de vídeo que puedan ser interpretadas directamente por un receptor convencional. En sistemas de CATV la señal de TV se obtiene directamente del cable de suministro y también requiere que el usuario disponga de un demodulador que acondicione las señales al receptor. En el caso en que el usuario esté abonado a canales de televisión de acceso condicional, también es necesario un decodificador o descriptrador de la información. Los receptores actuales suelen admitir, dependiendo del país, uno o varios de los formatos analógicos convencionales (NTSC, PAL, SECAM, PALplus) y con varios tipos básicos de entradas de vídeo: la señal de radiofrecuencia, la señal RGB, la señal de vídeo compuesto o banda base y el formato SVideo. La primera se corresponde con las señales de vídeo moduladas que son recibidas en la antena de un sistema de radiodifusión terrestre. Estas señales son sintonizadas y demoduladas por el propio receptor, pudiendo memorizar la posición de varios canales. El resultado de esta modulación es la señal de vídeo compuesto o señal de vídeo en banda base que contiene la información de luminancia y de color. Las señales RGB son señales de vídeo ya demoduladas procedentes de un magnetoscopio doméstico o de un demodulador de CATV o vía satélite y son introducidas en el receptor a través de una entrada específica denominada *euroconector*. Finalmente, la señal S-Vídeo es un formato utilizado en magnetoscopios domésticos y cámaras portátiles en la que la información de brillo y de color se proporcionan en cables separados. Actualmente, muchos fabricantes de televisores están incorporando demoduladores digitales en el propio receptor, por lo que se espera que muy pronto sea posible recibir las señales de radiodifusión digital sin necesidad de utilizar un decodificador externo.

El objetivo de este capítulo es definir, desde un punto de vista puramente conceptual, los distintos elementos que tienen una incidencia directa en la obtención de una señal eléctrica que proporcione información sobre la escena. Se presentará una perspectiva general sobre las distintas alternativas posibles y se definirá la señal de vídeo en blanco y negro como primera aproximación a un sistema de televisión. La descripción detallada de otros conceptos como la introducción de la información de sincronismo y de color, la televisión digital y de alta definición se realizará con detalle en capítulos posteriores.

1.3 De la escena a la señal de televisión

En el caso más general, una escena puede ser representada como una función de 5 variables independientes:

$$\epsilon(x, y, z, t, \lambda) \quad (1.1)$$

donde x, y, z representan la posición de un punto genérico del espacio que radia con un determinado espectro radioeléctrico visible λ y que, posiblemente debido al movimiento de los objetos, se modifica a lo largo del tiempo t.

El principio en el que se basan todos los sistemas de televisión actuales consiste en hacer un muestreo sobre esta función de cinco variables en el modo adecuado para obtener una función unidimensional (señal temporal) que pueda ser transmitida utilizando los sistemas de modulación convencionales.

Algunos sistemas de televisión se han apartado de este criterio general y han propuesto la transmisión simultánea de varias señales que son recomuestas en el receptor. Entre estos sistemas cabe destacar una de las primeras tentativas de televisión en color y algunos sistemas de televisión con definición mejorada (*EDTV-Enhanced Definition TV*) propuestos a principios de los noventa. En estas últimas propuestas, se pretendía transmitir una señal de televisión convencional junto un canal auxiliar, que al ser demodulado por el receptor permitiría obtener información sobre líneas adicionales de la imagen que no se transmiten en un sistema estándar de televisión. No obstante, ninguna de estas alternativas ha sido explotada comercialmente ni tampoco se prevé que sean usadas en el futuro.

En el proceso de conversión de la escena a una función temporal intervienen distintos procesos de proyección y muestreo que de uno u otro modo supondrán una pérdida parcial de información respecto al contenido de la escena original. En la selección de estos parámetros de muestreo y proyección deben tenerse en cuenta las características del sistema visual humano, que es el que finalmente decidirá si la calidad de la imagen es adecuada. En consecuencia, al diseñar un sistema de televisión, es necesario conocer con detalle las características de la visión humana, ya que nos permitirán determinar qué partes de la información no requieren ser transmitidas por cuanto no aportan ninguna mejora subjetiva sobre la calidad de la imagen. Con ello, el ancho de banda de la señal transmitida se reducirá al estrictamente necesario para poder reproducir imágenes de calidad aprovechando al máximo el espectro radioeléctrico.

En las secciones sucesivas presentaremos por una parte los procesos de proyección y muestreo de la escena que se definen en cualquier sistema de televisión y por otra las características visuales que permiten ajustar los parámetros de estos procesos. Consideraremos únicamente aquellos aspectos de la visión que tengan una incidencia directa sobre los parámetros de muestreo de la escena, dejando algunos aspectos, como la interpretación detallada del movimiento, para capítulos posteriores en los que estos factores tengan una repercusión directa en la definición del sistema de televisión.

A continuación resumimos las distintas etapas de muestreo y su función básica:

- a) *Separación en componentes de color:* Toda la información contenida en la variable λ puede representarse mediante tres componentes discretas que corresponden a los colores primarios rojo, verde y azul. Este proceso de muestreo no representa una pérdida de información aparente en el sistema visual humano. Analizaremos con detalle los mecanismos de visión y percepción del color, su representación analítica y los principios físicos usados para la descomposición del color en tres componentes.
- b) *Proyección plana de la imagen:* La escena se proyecta mediante un sistema óptico sobre un plano de imagen. Esta proyección representa una pérdida significativa de información espacial y reduce las tres variables espaciales de la escena a las dos variables de la imagen. En esta sección consideraremos diversas propuestas para la representación de imágenes en tres dimensiones.
- c) *Límites del sensor:* La imagen de la escena sólo se considera dentro de los límites del sensor por lo que sus variables espaciales están acotadas dentro de estos límites. Veremos que el tamaño del sensor y su relación de aspecto tienen una incidencia directa sobre la resolución del sistema y la integración del espectador en la escena.

- d) *Muestreo temporal*: Las imágenes pueden presentarse al espectador como una secuencia de fotogramas. Si esta secuencia es suficientemente rápida no puede distinguirse de la información original. La variable temporal, de naturaleza continua, puede sustituirse por una secuencia de imágenes sin pérdida aparente de información. Consideraremos los fenómenos de persistencia y parpadeo de imágenes en la retina.
- e) *Muestreo espacial (Líneas)*: Cada una de las imágenes se descompone en un número finito de líneas. Si este número es suficientemente elevado, el espectador será incapaz de percibir la diferencia con la imagen original. Los sistemas de televisión analógicos surgen de manera natural a partir de este muestreo.
- f) *Muestreo espacial (Retícula)*: En este caso la imagen se muestrea tanto en el sentido horizontal como en el vertical. Las muestras obtenidas constituyen la base de los sistemas de televisión digitales.

1.4 Componentes de color RGB

La primera transformación que se realiza de la escena original consiste en sustituir su dependencia con la variable continua longitud de onda por tres muestras correspondientes a los colores primarios rojo, verde y azul:

$$\begin{aligned}\mathbf{e}_R(x, y, z, t) &= \mathbf{e}(x, y, z, t, \mathbf{I}_R) \\ \mathbf{e}_G(x, y, z, t) &= \mathbf{e}(x, y, z, t, \mathbf{I}_G) \\ \mathbf{e}_B(x, y, z, t) &= \mathbf{e}(x, y, z, t, \mathbf{I}_B)\end{aligned}\tag{1.2}$$

El principio en el que se basa esta descomposición es que a partir de la combinación adecuada de los tres colores primarios puede obtenerse cualquier sensación de color. De este modo, se reduce la dimensionalidad de la escena de una función de cinco variables a tres funciones de cuatro variables. Podemos entender este proceso como un muestreo de la escena en la variable λ . Las características de la visión humana nos permiten asegurar que usando únicamente tres muestras de esta variable, correspondientes a los colores rojo, verde y azul adecuadamente ponderados, podremos representar todos los colores sin que ello suponga ninguna pérdida de información visual.

Para comprender mejor este fenómeno debemos considerar diversos aspectos relacionados con la naturaleza de la luz y la percepción de los colores en el sistema visual.

1.4.1 La luz y el color

La luz puede considerarse, para los fenómenos de visión que nos ocupan en este texto, como una onda electromagnética con longitudes de onda comprendidas entre 780 nm y 380 nm. En la figura 1.2 se representa la situación en el espectro radioeléctrico de las ondas visibles y los colores asociados a las distintas longitudes de onda. La longitud de onda más larga corresponde al color rojo y la más corta al violeta. Más allá del margen visible se sitúan las frecuencias correspondientes al infrarrojo y al

ultravioleta. El color blanco está formado como la superposición de todas las frecuencias del espectro visible mientras que el negro supone que no existen componentes espirituales en la banda visible. Los colores indicados en el margen visible se corresponden con los que se obtendrían para una luz formada por una única frecuencia pura y se denominan colores monocromáticos. Las tonalidades rojas tienen longitudes de onda que van desde los 760 nm hasta los 650 nm, las tonalidades amarillas están situadas en el margen de 630 nm a 560 nm, las verdes de 540 nm a 500 nm, las azules de 500 nm a 420 nm y las violetas de 420 nm a 400 nm.

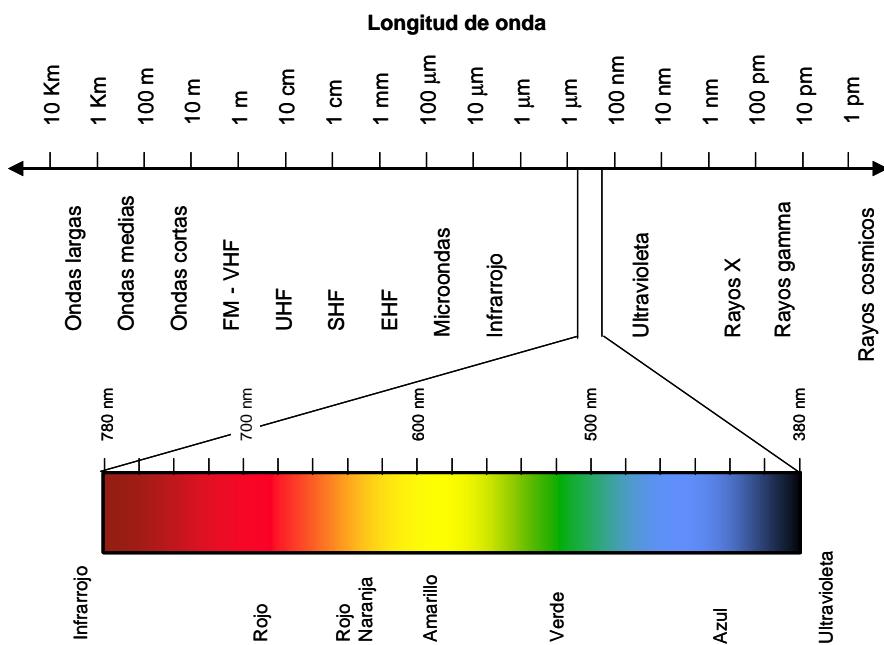


Fig. 1.2 Espectro de radiofrecuencia y frecuencias del visible

En la figura 1.3 se representa un experimento que permite descomponer una luz blanca, es decir una luz con un espectro plano en todo el margen visible, en los distintos colores monocromáticos. El experimento consiste en hacer pasar la luz blanca a través de un prisma que, al tener un coeficiente de refracción dependiente de la frecuencia, desvía cada componente frecuencial en una dirección distinta.

Los colores monocromáticos se corresponden con los colores del arco iris. Algunos colores, como el púrpura o el magenta, sólo pueden obtenerse mediante la superposición de distintas componentes espirituales. Así, el magenta es una sensación coloreada obtenida a partir de la combinación del rojo y el azul.

En general, el color observado por el sistema visual humano depende de dos factores: el espectro de la luz incidente y las características de reflexión y absorción de las ondas visibles de los objetos. La composición espectral de la luz que recibe un espectador puede expresarse como el producto del espectro de la luz incidente por el coeficiente de reflexión de los objetos:

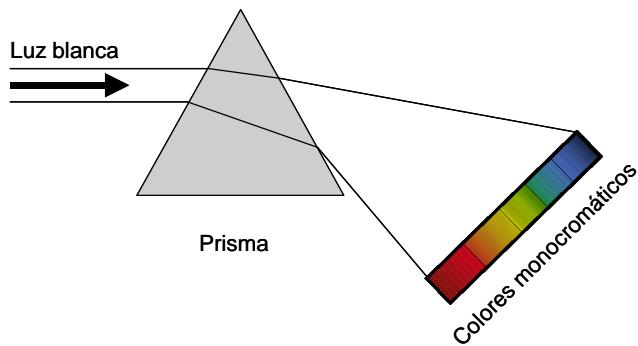


Fig. 1.3 Descomposición de la luz blanca en colores monocromáticos

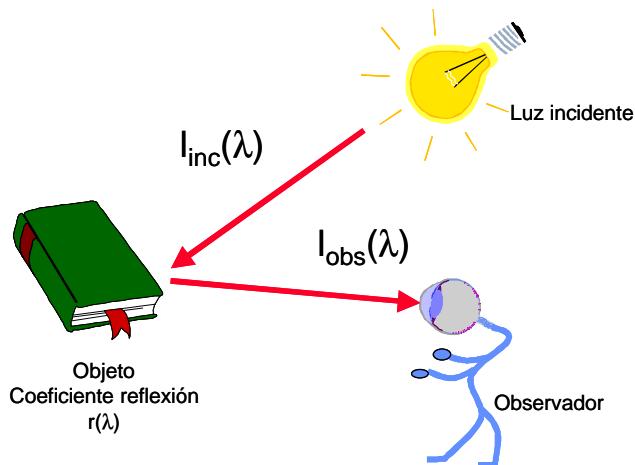


Fig. 1.4 Espectro de luz incidente y espectro observado

$$l_{obs}(\mathbf{I}) = l_{inc}(\mathbf{I}) \cdot r(\mathbf{I}) \quad (1.3)$$

donde $l_{obs}(\lambda)$ es el espectro de la luz observada, $l_{inc}(\lambda)$ el de la luz incidente y $r(\lambda)$ el coeficiente de reflexión de los objetos (ver figura 1.5). De acuerdo con esta ecuación, cuando la luz incidente es blanca (la luz solar es una buena aproximación a la luz blanca), el espectro de la luz reflejada por los objetos tiene la misma forma que su coeficiente de reflexión ya que podemos considerar que $l_{inc}(\lambda) = Kte$. Así, un objeto cuyo coeficiente de reflexión sea próximo a la unidad en la región correspondiente al color verde y prácticamente cero para el resto de frecuencias será observado como de color verde cuando la luz incidente sea blanca. Sin embargo, cuando incida sobre él una luz monocromática de color amarillo, absorberá toda la energía incidente, adoptando una apariencia de color negro. En la figura 1.5 se ilustra como un mismo objeto puede observarse de color amarillo

cuando la luz incidente es blanca, rojo cuando el espectro de la luz está situada en la región de baja frecuencia del visible, o verde cuando la iluminación corresponde a un espectro situado en la región de alta frecuencia. En general, pues, el color de los objetos depende de las características espetrales de la iluminación a la que están sometidos. Cuando nos referimos, de una manera genérica, al color de un objeto suele entenderse que hablamos del color que observaremos cuando esté expuesto a una luz blanca.

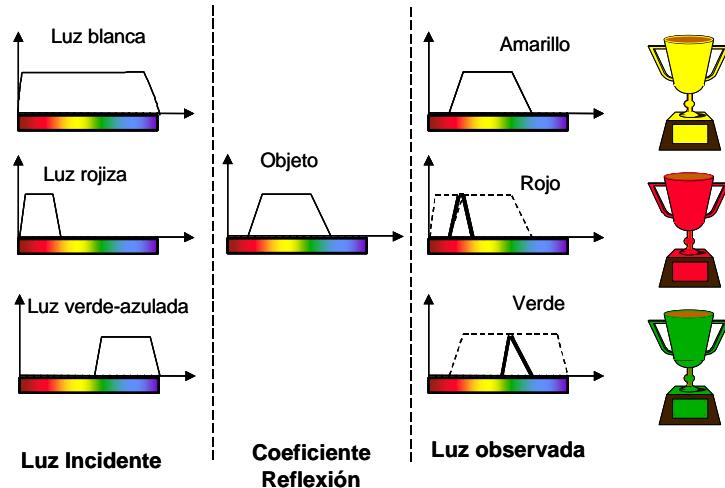


Fig. 1.5 El color de los objetos depende de las características de la luz incidente

1.4.2 Percepción del color en el ojo humano

El sistema visual humano es capaz de discriminar un elevado número de colores. Su resolución en colores monocromáticos depende de la frecuencia y puede estimarse que en las bandas del azul, verde y amarillo pueden distinguirse dos colores siempre que su longitud de onda difiera en al menos 1 nm. En las regiones extremas de la banda visible, correspondientes al rojo y al violeta, la capacidad de discriminación de colores se reduce, y puede estimarse que se requiere una diferencia de 10 nm para distinguirlos. Estas medidas son aproximadas y representan un valor medio para sujetos sin problemas de visión. Se obtienen proyectando colores monocromáticos en las dos mitades de una pantalla en la que diversos espectadores indican si son capaces de diferenciar los dos colores. Desde un punto de vista cualitativo pueden definirse tres atributos para describir el color: el brillo, el matiz o tono y la saturación o pureza. El brillo indica la intensidad de la luz que percibe el ojo. El matiz o tono es la propiedad que permite clasificar los colores como distintos, distinguiendo entre el rojo, verde, amarillo, etc. La saturación indica en qué grado un color se acerca o se aparta del blanco. Así, podemos hablar de un rojo más o menos claro entendiendo que todos ellos corresponden a una misma tonalidad roja. Diremos que un color es puro cuando no contiene blanco. Veremos que es posible establecer medidas cuantitativas directamente relacionadas con estos atributos.

Además de los colores monocromáticos, el sistema visual humano es capaz de distinguir otros tonos de color formados por la superposición de varias frecuencias y apreciar colores con mayor o menor

'pureza' dependiendo del espectro final de la luz incidente. En general, la luz procedente de los objetos no es casi nunca un color espectralmente puro, sino una mezcla compuesta de muchas longitudes de onda distintas.

1.4.3 El ojo humano

En la figura 1.6 se representa una sección transversal de un ojo humano en el que pueden apreciarse los principales elementos que intervienen en el proceso de formación y captación de imágenes. El glóbulo ocular tiene una forma aproximadamente esférica, de unos 20 mm de diámetro medio que está recubierta por un tejido fibroso de gran resistencia denominado membrana esclerótica. Esta membrana es opaca en sus superficies posterior y lateral. El segmento frontal de la membrana esclerótica recibe el nombre de córnea y es transparente para permitir el paso de la luz hacia el interior de la cavidad ocular. La córnea actúa como elemento protector de las partes interiores del ojo y presenta una curvatura superior al resto del glóbulo ocular. Justo debajo de la membrana esclerótica hay una capa de tejido vascular, denominada coroides, que es la responsable de proporcionar nutrientes a todos los elementos del ojo. La coroides está fuertemente pigmentada para ayudar a reducir la cantidad de luz exterior que entra en el ojo y absorber la luz difundida en su interior.

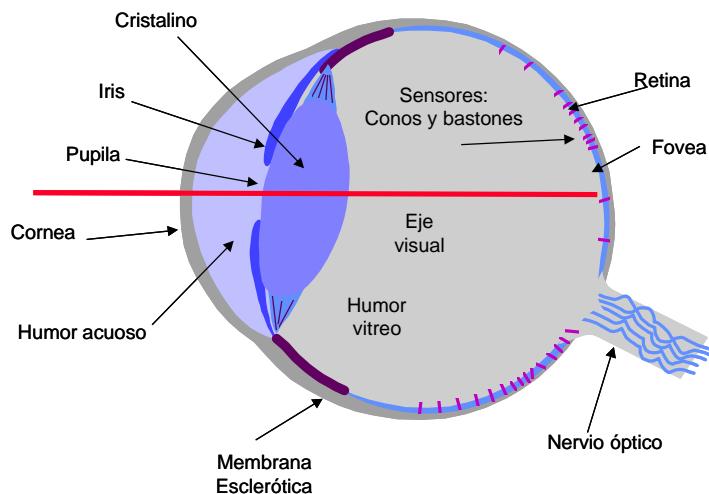


Fig. 1.6 Sección transversal del ojo humano

En el interior de la córnea se sitúa la cámara anterior, que está rellena con un líquido denominado humor acuoso. En el interior de esta cavidad está situado el iris, que es una cortina muscular, de forma aproximadamente circular, que permite regular la cantidad de luz que incide en la parte interior del ojo. El color de este músculo determina el color de los ojos de una persona. La luz incide a través de la pupila, que es una abertura circular situada en el centro del iris. El diámetro de la pupila se regula mediante el iris, que deja al descubierto una superficie mayor o menor de ésta de acuerdo con las condiciones de luz ambientales, permitiendo que los sensores de luz situados en el interior de la cavidad ocular puedan operar en un amplio margen dinámico de condiciones de iluminación. El

tamaño de la pupila depende únicamente de las condiciones de iluminación externa y se controla de forma automática sin que intervenga la voluntad del observador. Su diámetro puede oscilar entre unos 8 mm (en el caso de poca iluminación) hasta unos 1.5 mm (para condiciones de gran iluminación) por lo que la superficie total del orificio por el que entra la luz puede modificarse en un factor de aproximadamente 28, dependiendo de las condiciones de luz exteriores.

Detrás de la pupila se encuentra el cristalino. Su función básica es la de actuar como una lente que enfoca las imágenes en la retina, que es la superficie interior del ojo en la que se encuentran los sensores de luz. La forma del cristalino se controla a través del cuerpo ciliar, una estructura muscular que forma parte de la coroide y que permite variar la distancia focal y la posición de esta lente. El cristalino se adapta para que los objetos sobre los que el sistema visual quiere centrar su atención permanezcan enfocados en la retina. La luz infrarroja y la ultravioleta son absorbidas, en su mayor parte, por las proteínas que forman la estructura del cristalino, ya que en cantidades excesivas podrían dañar el ojo. La cámara situada entre la pupila y el cristalino se denomina cámara posterior y está rellena con el humor acuoso.

Entre el cristalino y la retina se encuentra una cámara rellena de un líquido gelatinoso y transparente que se denomina humor vítreo. Este líquido proporciona un equilibrio de presiones para que el globo ocular mantenga su forma esférica.

La retina es una capa de tejido nervioso que cubre aproximadamente las dos terceras partes del ojo interno. En ella se encuentran las células sensibles a la luz que proporcionan la información básica de visión que posteriormente será analizada e interpretada en el cerebro. De hecho, la retina puede interpretarse como una ramificación del cerebro ya que sus tejidos nerviosos tienen la misma constitución, y está conectada al mismo a través del nervio óptico. La retina es por tanto la parte más importante del ojo, pudiendo considerar que la función del resto de elementos es, por una parte proporcionarle nutrientes, y por otra enfocar la luz en la región apropiada. Es un tejido de cierta complejidad, formado por varias capas, de las cuales únicamente una capa intermedia contiene células sensibles a la luz. Estas células reaccionan a las distintas longitudes de onda incidentes generando impulsos nerviosos. La luz debe atravesar las capas superficiales hasta alcanzar estas células. La capa más interna de la retina está formada por complejas interconexiones entre las células (sinapsis) en las que se combina la información recibida por estos elementos sensibles y se dirige hacia el nervio óptico. Esta información es posteriormente procesada en la corteza visual del cerebro donde se analizan e interpretan las imágenes.

1.4.4 Células sensibles: conos y bastones

Existen dos tipos de células sensibles a la luz, que se denominan, como consecuencia de su aspecto físico externo, conos y bastones. Los bastones son elementos cilíndricos, con un diámetro inferior a 1 μm , que terminan con una ligera curvatura y que presentan una elevada sensibilidad a la luz. Son muy abundantes en animales con visión nocturna, y en el hombre son los responsables de la visión en condiciones de baja luminosidad, también denominada visión escotópica. Su respuesta muestra dependencia de la frecuencia de la luz incidente, teniendo un máximo de sensibilidad para frecuencias próximas a los 500 nm. La señal que transmiten al cerebro no proporciona información sobre la composición espectral de la luz, de modo que una iluminación monocromática de 500 nm con baja

intensidad puede provocar la misma respuesta que una iluminación de 600 nm de mayor intensidad, siendo ambas señales indistinguibles por los órganos de visión. Los bastones, por tanto, no proporcionan información sobre el color de los objetos.

Los conos presentan una menor sensibilidad a la luz y sólo se activan en condiciones de iluminación normal o diurna (también denominada, visión fotópica). Son células de aspecto cónico con un diámetro medio que oscila entre 1 μm y 5 μm . Presentan una sensibilidad máxima en la longitud de onda de 555 nm y proporcionan información básica sobre la composiciónpectral de la luz por lo que son los responsables de la percepción del color. Las curvas de sensibilidad de los conos y los bastones se representan, de manera aproximada y en tanto por ciento respecto al máximo, en la figura 1.7. Estas curvas se obtienen proyectando en una pantalla dos luces monocromáticas de distinta frecuencia y amplitud y pidiendo que el observador indique cuándo considera que ambas tienen el mismo brillo. En caso en que la intensidad de la luz sea elevada (visión fotópica), el espectador observará las dos señales con colores distintos, pero deberá decidir basándose únicamente en la sensación de brillo que le producen las imágenes.

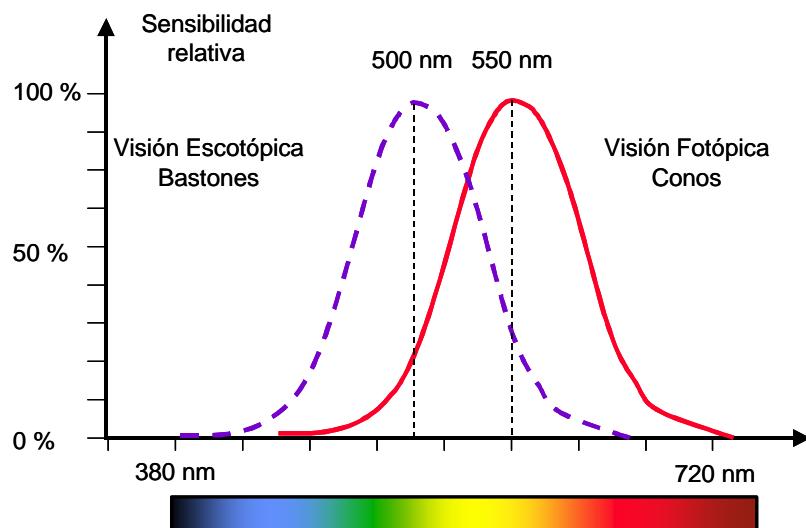


Fig. 1.7 Curvas de sensibilidad espectral de los conos y los bastones

El desplazamiento relativo que existe entre las curvas de visión escotópica y fotópica se conoce con el nombre de desplazamiento de *Purkinje* y es el responsable de algunos fenómenos psicofisiológicos de cambio de luminosidad aparente que se producen al anochecer. Así, al oscurecer, las tonalidades rojas tienden a perder brillo mientras que las azules parecen aumentar su luminosidad debido al desplazamiento existente entre las curvas de sensibilidad correspondientes a los conos y a los bastones. Este efecto es debido a que, al reducirse la iluminación, la visión cambia gradualmente de fotópica a escotópica de manera que la sensibilidad del ojo al color rojo disminuye (volviéndose más oscuro) mientras que la sensibilidad al azul aumenta (apareciendo más brillante). Esta región intermedia entre la visión fotópica y la escotópica, en la que operan simultáneamente los conos y los bastones, se conoce con el nombre de visión mesóptica.

La distribución de los conos y los bastones en la retina no es uniforme. Los primeros tienen una concentración máxima en la zona central de la retina, denominada fóvea central, donde se sitúa el eje de visión. Mantienen una concentración elevada en las proximidades del eje visual, en una región circular de aproximadamente 1 cm de diámetro denominada *macula lutea*, que rodea la fóvea central. Los bastones se sitúan principalmente en la periferia de la retina, teniendo una concentración prácticamente nula en la fóvea.

En la figura 1.8 se representa de forma aproximada la densidad de conos y bastones en la retina respecto al eje visual; se observa que los bastones presentan una concentración máxima en una región circular que forma un ángulo de unos 20° respecto al eje visual y que prácticamente son inexistentes en la fóvea. La distribución de los bastones es aproximadamente simétrica respecto a la fóvea, exceptuando el denominado punto ciego, que corresponde al punto de la retina de donde parte el nervio óptico, sobre el que no se encuentran ni conos ni bastones. Se estima que el número aproximado de conos en individuos con visión normal es de unos 7 millones de sensores, mientras que el de bastones es del orden de los 120 millones.

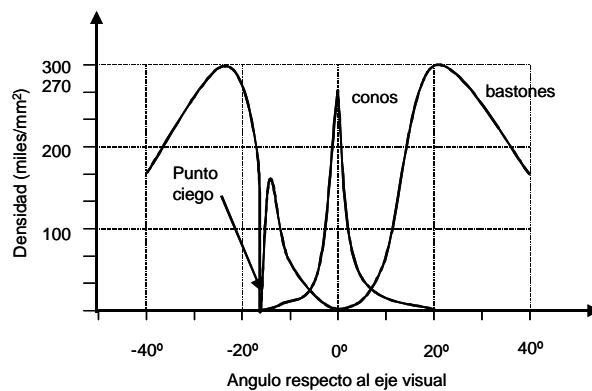


Fig. 1.8 Distribución y densidad de los conos y los bastones en la retina

En condiciones de iluminación normal o alta, la percepción de la visión es debida fundamentalmente a los conos, ya que los bastones, de mayor sensibilidad, se saturan y no proporcionan información al cerebro. En condiciones de poca iluminación, la pupila se dilata permitiendo que la imagen quede enfocada sobre la región periférica de la retina, obteniendo una respuesta debida fundamentalmente a los bastones. Los conos, al recibir poca energía lumínica, no se excitan, por lo que las imágenes que se perciben no contienen información sobre el color de los objetos. Por ello, en condiciones de baja iluminación, es difícil distinguir el color de los objetos, ya que los conos no llegan a excitarse y la visión es fundamentalmente en blanco y negro. Además, es en la región próxima a la fóvea donde quedan correctamente enfocadas las imágenes, por lo que la percepción del detalle de los objetos no será buena si los sensores situados en esta zona no responden. Por ello, resulta difícil leer en condiciones de baja iluminación. En otras palabras, podríamos decir que la frase '*De noche, todos los gatos son pardos*', queda ampliamente justificada a partir de la distribución y la respuesta de los elementos sensores de luz en el ojo.

Algunos problemas de visión como la miopía o la hipermetropía se producen como consecuencia de un incorrecto enfoque de las imágenes en la fóvea. En el primer caso, el cristalino enfoca las imágenes dentro de la región del humor vítreo, antes de que incidan sobre la retina, mientras que en el segundo caso, el punto de enfoque de las imágenes queda más allá de la retina. En ambos casos, las imágenes que se forman en la retina quedan desenfocadas. Esto explica también que el problema de la miopía se acentúe en condiciones de baja iluminación, puesto que en este caso, el desenfoque de la imagen en la periferia de la retina es mayor, con lo que la resolución de las imágenes cuando solo están operativos los bastones es considerablemente inferior a la que se obtiene en la región próxima a la fóvea. El astigmatismo es un problema visual que se origina como consecuencia de una deformación de la córnea.

La percepción del color se debe a que existen tres tipos distintos de conos que presentan una respuesta dependiente de la longitud de onda de la luz incidente. La sensación de color se produce cuando el cerebro interpreta las señales enviadas por cada uno de estos elementos. La existencia de estos tres tipos de receptores fue postulada en 1802 por Thomas Young, aunque no fue verificada experimentalmente hasta 1964 por biofísicos americanos. El postulado de Young fue desarrollado teóricamente por Hermann L. F. von Helmholtz y se conoce como la teoría tricromática de Young-Helmholtz. Esta teoría se fundamentaba en el fenómeno de mezcla de colores previamente descrito por Newton, quien había realizado el experimento de descomposición de la luz blanca en colores monocromáticos descrito en el apartado anterior. Newton también observó que si algunos colores monocromáticos se volvían a hacer pasar a través de un prisma, podían obtenerse, como resultado de su mezcla, sensaciones de color distintas a las de los colores monocromáticos. La mezcla de pigmentos elementales también era utilizada desde antiguo en pintura para obtener distintas tonalidades y matices de color a partir de pocos colores básicos.

En términos modernos, existen tres tipos de conos con respuestas frecuenciales distintas dentro del margen visible y cuya máxima sensibilidad se corresponde aproximadamente con los colores rojo, verde y azul. Las curvas de sensibilidad para cada uno de los tipos de conos se representan esquemáticamente en la figura 1.9. Obsérvese que la curva correspondiente al receptor azul está ampliada en un factor 20 para que pueda compararse con las curvas asociadas a los receptores rojo y verde, ya que estos últimos tienen una mayor sensibilidad.

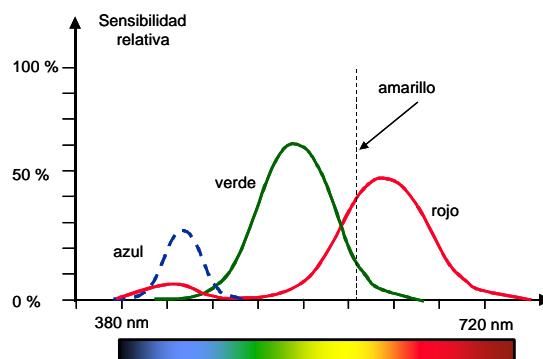


Fig. 1.9 Sensibilidad espectral de los diferentes tipos de conos. La sensibilidad del color azul está multiplicada por un factor 20 para que su amplitud sea comparable a la de los otros sensores

Estas curvas pueden inducir a cierta confusión cuando se comparan con la curva de sensibilidad espectral asociada a la visión fotópica presentada en la figura 1.7. En este último caso, si bien los mecanismos de combinación de las señales recibidas por los distintos tipos de conos no son conocidos con exactitud, podríamos decir que la curva de visión fotópica se correspondería aproximadamente con la suma de cada una de las curvas asociadas a los distintos receptores. Recuérdese que la curva de visión fotópica se obtiene pidiendo al observador que valore únicamente el brillo de la imagen y que no tenga en cuenta su color, por lo que incorpora la valoración conjunta de cada uno de los receptores al brillo global de la imagen.

La percepción del color de los objetos depende del tipo de conos que se activan al incidir un determinado estímulo luminoso en la retina. Así, por ejemplo, cuando la luz incidente es una componente monocromática correspondiente al amarillo, se activarán los receptores rojo y verde (véase la gráfica de la figura 1.9) que la corteza visual interpretará como una sensación de amarillo.

Si bien la teoría tricromática de la existencia de tres tipos de sensores lumínicos en la retina ha sido verificada por distintos experimentos, aún no está completamente claro cómo se transmiten estas señales al cerebro ni como éste las interpreta. La presencia de los tres tipos de conos se ha podido comprobar con distintas pruebas y experimentos. Una de estas experiencias consiste en examinar la luz reflejada por la retina bajo diversas condiciones de iluminación. Cuando la iluminación es de baja intensidad, de manera que puede considerarse que solo están activos los bastones, la luz reflejada tiene un espectro en el que se detecta una fuerte absorción de las componentes azules (longitudes de onda próximas a 500 nm). Esto es debido a que los bastones absorben preferentemente estas componentes frecuenciales, como se deduce de la curva de sensibilidad escotópica.

En condiciones de iluminación diurna, donde puede considerarse como buena aproximación que sólo están operativos los conos, la luz reflejada por la retina muestra una fuerte absorción energética en las longitudes de onda correspondientes al rojo y al verde debido a la presencia de los pigmentos *erythrolabe* y *chlorolabe* existentes en los conos sensibles a estas frecuencias. La absorción en la componente azul es más difícil de detectar debido a la menor sensibilidad de estos elementos. Otra experiencia con la que se ha conseguido verificar la presencia de estos tres tipos de sensores consiste en examinar a través del microscopio una escisión de la retina humana, observando que los conos adoptan los colores correspondientes a los tres colores primarios postulados en la teoría tricromática. Recientemente se ha postulado la existencia de hasta 7 tipos de conos que son sensibles a distintas longitudes de onda. No obstante, esta hipótesis no está completamente verificada ni aceptada, ya que sólo los 3 tipos principales tienen una densidad importante en la retina. Por lo que respecta a los sistemas de televisión no parece necesario hacer uso de estos resultados para obtener una buena calidad de imagen.

La densidad de conos en el ojo humano es muy superior al número de señales que se envían al cerebro a través del nervio óptico. De hecho, el número de fibras del nervio óptico es sólo del orden de decenas de miles mientras que el número de conos es del orden de millones. Está plenamente aceptado que una parte importante del procesamiento de señales implícito en el sistema de visión humano se realiza directamente en la retina, combinando las señales sensadas por los conos para enviar al cerebro información preprocesada. Esta etapa del sistema visual se denomina visión de bajo nivel y no está claro hasta qué punto se trata de una interconexión de señales hereditaria o si se realiza un cierto aprendizaje en las primeras etapas después del nacimiento.

1.4.5 Estímulos de los fotoreceptores: sensación de color

La conclusión más relevante de este mecanismo de visión es que la sensación de color depende únicamente de las señales recibidas por los conos y transmitidas al cerebro. Los órganos de visión serán incapaces de diferenciar dos imágenes que produzcan el mismo patrón de excitación en los conos. Hemos visto que una luz monocromática pura correspondiente al color amarillo provocará una cierta respuesta de los conos correspondientes al rojo y al verde, ya que ambas curvas muestran cierta sensibilidad a esta frecuencia. La sensación de color obtenida podrá reproducirse combinando una luz roja y una luz verde que produzcan la misma respuesta en los conos que la luz amarilla original, siendo imposible distinguir, para el sistema visual, la composición espectral de ambos estímulos.

Si el objeto radia un espectro visible arbitrario $C_i(\lambda)$, podemos calcular la respuestas de cada uno de los sensores construyendo un modelo de percepción del color basado en las curvas de sensibilidad de la figura 1.9. Según este modelo, el estímulo obtenido por los distintos tipos de receptores puede expresarse como:

$$\begin{aligned}\alpha_R(C_i) &= \int_{\lambda_{\min}}^{\lambda_{\max}} S_R(\lambda) \cdot C_i(\lambda) d\lambda \\ \alpha_G(C_i) &= \int_{\lambda_{\min}}^{\lambda_{\max}} S_G(\lambda) \cdot C_i(\lambda) d\lambda \\ \alpha_B(C_i) &= \int_{\lambda_{\min}}^{\lambda_{\max}} S_B(\lambda) \cdot C_i(\lambda) d\lambda\end{aligned}\quad (1.4)$$

donde $\alpha_R(C_i)$, $\alpha_G(C_i)$, $\alpha_B(C_i)$ representan la respuesta de cada uno de los receptores al espectro visible $C_i(\lambda)$ y $S_R(\lambda)$, $S_G(\lambda)$, $S_B(\lambda)$ corresponden sus curvas de sensibilidad espectral. Los límites de integración, λ_{\min} y λ_{\max} son las longitudes de onda extremas en el margen visible.

Atendiendo a este modelo, podemos concluir que dos espectros visibles distintos $C_1(\lambda)$ y $C_2(\lambda)$ producirán la misma sensación de color en el observador siempre que:

$$\begin{aligned}\alpha_R(C_1) &= \alpha_R(C_2) \\ \alpha_G(C_1) &= \alpha_G(C_2) \\ \alpha_B(C_1) &= \alpha_B(C_2)\end{aligned}\quad (1.5)$$

Es evidente que esta condición puede verificarse sin necesidad de que los dos espectros $C_1(\lambda)$ y $C_2(\lambda)$ sean iguales. Así por ejemplo, para que las componentes α_R coincidan, basta con que las áreas bajo las curvas $S_R(\lambda)C_1(\lambda)$ y $S_R(\lambda)C_2(\lambda)$ sean iguales, lo que puede cumplirse aun cuando los dos espectros sean distintos.

La conclusión directa de este resultado es que, debido a que el sistema visual humano no identifica el espectro visible completo de los objetos, sino que únicamente utiliza estas tres componentes ponderadas, es posible que objetos con espectros completamente distintos puedan producir las mismas sensaciones de color. Esto justifica que todos los sistemas de representación de imágenes utilicen

únicamente tres muestras del espectro de la escena sin que ello represente una pérdida de información desde el punto de vista del espectador.

Esta idea es el principio básico sobre el que se basa la síntesis de los colores: para reproducir el color de un objeto que tiene un espectro $C_1(\lambda)$ basta con sintetizar un espectro $C_2(\lambda)$ que produzca los mismos estímulos visuales que el original. Veremos los detalles de esta idea en el siguiente apartado.

Resumiendo, podemos obtener cualquier sensación de color combinando los colores primarios para que la respuesta de los tres tipos de conos coincida con la que se obtiene con los estímulos del mundo real, de modo que cualquier estímulo luminoso, sea cual sea su composición espectral, puede ser sustituido por una mezcla adecuada con los colores primarios rojo, verde y azul sin que sea perceptualmente distingible la diferencia entre ambos estímulos. Así pues, la escena original $\epsilon(x,y,z,t,\lambda)$ puede sustituirse, sin pérdida aparente de información para un observador humano, por tres funciones que representan las tres componentes de color de la escena tal y como habíamos postulado en la ecuación 1.2.

Visión dicromática

El daltonismo es un defecto de la visión relacionada con la carencia de respuesta de uno de los tres tipos de conos. Alrededor de un uno por ciento de los varones son dicromáticos, es decir, utilizan sólo dos de los colores primarios para la percepción del color. Los individuos con ceguera en la componente roja son incapaces de distinguir entre las tonalidades rojas, naranjas, amarillas y verdes resultantes de la mezcla entre los colores rojo y verde. Todas estas tonalidades las observan como verdes. La ceguera en la componente roja tampoco permite discriminar entre las tonalidades anteriores, aunque en este caso se observan como rojas. Estos dos tipos de daltonismo sólo pueden diferenciar entre los colores verde, rojo y amarillo basándose en su brillo, por lo que tienen dificultades para distinguir los colores de un semáforo o de distintos elementos habituales en seguridad vial. La luz blanca la observan coloreada debido a la falta de respuesta en uno de los colores primarios. La ceguera en la componente azul es más rara y sólo afecta a un individuo de cada 50000. En este caso pueden distinguirse las distintas tonalidades entre el rojo y el verde y la capacidad global de discriminación de colores es bastante superior a los otros dos tipos de daltonismo.

1.4.6 Tríadas RGB en *displays* de TV

En sistemas de televisión en color las imágenes se descomponen en sus tres colores básicos mediante filtros ópticos cuyos principios de funcionamiento describiremos en un apartado posterior. La señal que se envía al receptor contiene información sobre las tres componentes de color pero no describe el contenido espectral completo de los objetos. Las imágenes se presentan al espectador en un *display*, que normalmente es un tubo de rayos catódicos (o más recientemente *displays* de cristal líquido), en el que se superponen los tres colores básicos y a partir de los cuales, el sistema visual recompondrá la información de color completa. En el caso de los tubos de rayos catódicos, la imagen que se presenta al espectador se obtiene iluminando tres tipos distintos de fósforos que emiten en longitudes de onda próximas a los colores básicos. Los fósforos son de tamaño reducido y se encuentran distribuidos

uniformemente en toda la superficie de la pantalla de modo que al observarlos desde cierta distancia el ojo integra sus contribuciones observando la mezcla de colores y no sus colores individuales.

En la figura 1.10 se representa un esquema de la distribución de los fósforos en la pantalla de un receptor. Cada grupo de tres fósforos se denomina tríada y, para el caso representado en la figura, los fósforos tienen forma aproximadamente circular y su distribución es tal que la tríada tiene una forma triangular. Existen otras posibles formas y distribuciones, como en el caso de los tubos de imagen del tipo Trinitron, en las que los fósforos se distribuyen en bandas verticales y que permiten aprovechar una mayor área de iluminación en la pantalla. En el capítulo dedicado a los *displays* analizaremos con cierto detalle las distintas alternativas existentes en el mercado y discutiremos sus ventajas e inconvenientes.

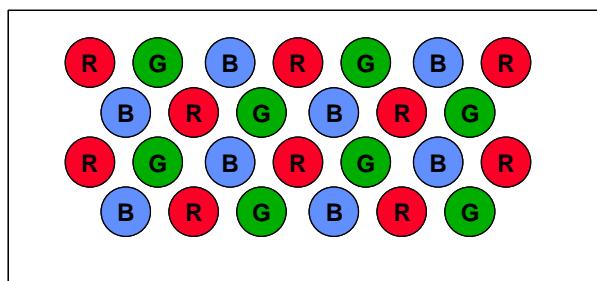


Fig. 1.10 Disposición en delta de los fósforos en un display

Los colores exactos de los fósforos son muy importantes. En principio, eligiendo un determinado conjunto de tres fósforos, cuyas radiaciones estén centradas aproximadamente en las frecuencias de los colores rojo, verde y azul, pueden obtenerse casi todas las sensaciones de color. Para obtener un determinado color bastará con ajustar el nivel de radiación de cada fósforo para que activen los conos de la retina en la proporción adecuada. Sin embargo, si se sustituye uno de los fósforos por otro cuyo espectro de radiación no sea exactamente igual, la sensación de color se modifica apreciablemente, siendo necesario recalcular las intensidades de radiación de todos los fósforos. Por ello, en los sistemas de televisión debe establecerse *a priori* las características de radiación y los colores básicos de los fósforos para que todos los receptores vean los mismos colores.

1.4.7 Sistemas de coordenadas para la representación del color: principios de colorimetría

Hemos visto que es fundamental definir sistemas de coordenadas que permitan representar los colores sin ambigüedad y con arreglo a algún convenio preestablecido. Con ello podremos cuantificar la información de color y estableceremos las bases para transmitir esta información a los receptores. La medida, cuantificación y representación gráfica de los colores y sus posibles combinaciones se conoce con el nombre de colorimetría.

La definición de un sistema de coordenadas para cuantificar los colores exige especificar con exactitud las características espectrales primarios, de modo que un color pueda identificarse como una triplete de números que establece en qué proporción debe intervenir cada uno de estos primarios en su

síntesis. Una vez fijado el conjunto de primarios, la obtención de la tripleta de coordenadas asociadas a un determinado espectro $C_1(\lambda)$ requiere la resolución de un sistema de lineal ecuaciones. El planteamiento del problema es el siguiente:

- Supongamos que $p_R(\lambda)$, $p_G(\lambda)$, $p_B(\lambda)$ representan los espectros de radiación asociados a cada uno de los tres colores considerados como primarios y denotemos como β_R , β_G , β_B las coordenadas asociadas a un color $C_1(\lambda)$ representado sobre el anterior conjunto de colores primarios.
- El color que obtenemos ponderando los primarios mediante estas coordenadas es:

$$C_2(\lambda) = \beta_R \cdot p_R(\lambda) + \beta_G \cdot p_G(\lambda) + \beta_B \cdot p_B(\lambda) \quad (1.6)$$

- Es evidente que si las coordenadas anteriores representan correctamente el color, los estímulos producidos por $C_1(\lambda)$ y $C_2(\lambda)$ en los tres tipos de receptores visuales deben coincidir. Esto implica que deben verificarse las siguientes ecuaciones:

$$\begin{aligned} \alpha_R(C_1) &= \alpha_R(C_2) = \beta_R \cdot \alpha_R(p_R) + \beta_G \cdot \alpha_R(p_G) + \beta_B \cdot \alpha_R(p_B) \\ \alpha_G(C_1) &= \alpha_G(C_2) = \beta_R \cdot \alpha_G(p_R) + \beta_G \cdot \alpha_G(p_G) + \beta_B \cdot \alpha_G(p_B) \\ \alpha_B(C_1) &= \alpha_B(C_2) = \beta_R \cdot \alpha_B(p_R) + \beta_G \cdot \alpha_B(p_G) + \beta_B \cdot \alpha_B(p_B) \end{aligned} \quad (1.7)$$

donde hemos aplicado que la relación integral que define el cálculo de la respuesta de los receptores de la retina es lineal.

- Nótese que los términos de la derecha son escalares que, una vez conocidos el espectro del color a sintetizar $C_1(\lambda)$ y las curvas de sensibilidad de cada uno de los receptores, podrían ser calculados a partir de la ecuación 1.4. Los términos $\alpha_X(p_Z)$ corresponden a ecuaciones análogas que también podrían calcularse a partir del espectro visible de los primarios y las curvas de sensibilidad. En este caso, las expresiones vendrían dadas por:

$$\alpha_X(p_Z) = \int_{\lambda_{\min}}^{\lambda_{\max}} S_X(\lambda) \cdot p_Z(\lambda) d\lambda \quad (1.8)$$

donde los subíndices X y Z se utilizan para denotar cualquiera de las componentes R,G,B.

- La ecuación 1.7 puede expresarse en forma matricial como:

$$\begin{bmatrix} \alpha_R(C_1) \\ \alpha_G(C_1) \\ \alpha_B(C_1) \end{bmatrix} = \begin{bmatrix} \alpha_R(p_R) & \alpha_R(p_G) & \alpha_R(p_B) \\ \alpha_G(p_R) & \alpha_G(p_G) & \alpha_G(p_B) \\ \alpha_B(p_R) & \alpha_B(p_G) & \alpha_B(p_B) \end{bmatrix} \cdot \begin{bmatrix} \beta_R \\ \beta_G \\ \beta_B \end{bmatrix} \quad (1.9)$$

que en forma vectorial escribiremos:

$$\vec{\alpha} = \mathbf{P} \cdot \vec{\beta} \quad (1.10)$$

donde \mathbf{P} representa la matriz asociada a los colores primarios que definen este sistema de coordenadas, y los vectores α y β los estímulos de los sensores y las coordenadas del color respectivamente.

Esta relación algebraica simple tiene un gran interés teórico, puesto que establece que la relación entre las coordenadas de color y los estímulos recibidos por los conos es lineal (siempre de acuerdo con el modelo simplificado que estamos utilizando). Esto permite establecer dos resultados importantes:

- 1) Las coordenadas de un color obtenido como superposición de dos espectros $C_1(\lambda)$ y $C_2(\lambda)$ se corresponden con la suma de las coordenadas asociadas a cada uno de los espectros.
- 2) Si la amplitud del espectro incidente aumenta en un factor γ , las coordenadas del color asociado también quedan afectadas por este mismo factor.

Estos dos resultados permiten que podamos operar con las coordenadas de los colores como si se tratara de vectores algebraicos. La suma de dos tripletas de coordenadas representará la combinación de dos colores y el producto de una triplete por un escalar, un aumento o disminución de luminosidad, ya que, en este caso, la composición espectral de la luz no se modifica.

1.4.8 Normalización de las coordenadas de color: coeficientes triestímulo

Es habitual trabajar con coordenadas de color normalizadas, de manera que el color blanco corresponda a la triplete $(1,1,1)$. Con ello se consigue que las tres componentes tengan órdenes de magnitud parecidos, simplificando la interpretación de las posibles mezclas de colores. El proceso de normalización puede considerarse como un ajuste de amplitud de los espectros primarios que se utilizan como patrón.

Los sistemas de coordenadas en los que el color blanco queda representado con todas las componentes igual a la unidad son los más utilizados y reciben el nombre de componentes triestímulo.

1.4.9 Cambio de sistema de coordenadas

Veremos que existen diversos sistemas de referencia para expresar cuantitativamente los colores. El uso de uno u otro sistema de coordenadas depende de la problemática que estemos considerando, por lo que suele ser habitual tener que realizar cambios de coordenadas para expresar los coeficientes triestímulo en diferentes sistemas.

El cambio de coordenadas se realiza, en todos los casos, multiplicando el vector de coeficientes por una matriz 3×3 . En efecto, supongamos que \mathbf{t} y \mathbf{t}' son los coeficientes triestímulo que representan a un mismo color en dos sistemas distintos que vienen caracterizados por las matrices \mathbf{Q} y \mathbf{Q}' :

$$\begin{aligned}\bar{\alpha}(C) &= \mathbf{Q} \cdot \vec{\mathbf{t}} \\ \bar{\alpha}(C) &= \mathbf{Q}' \cdot \vec{\mathbf{t}}'\end{aligned}\tag{1.11}$$

La relación entre los coeficientes triestímulo en los dos sistemas viene dada por:

$$\vec{t}' = (\mathbf{Q}')^{-1} \cdot \mathbf{Q} \cdot \vec{t} = \mathbf{R} \cdot \vec{t} \quad (1.12)$$

donde \mathbf{R} representa la matriz de cambio de sistema de coordenadas. Evidentemente, la relación inversa se obtiene multiplicando por la inversa de esta matriz:

$$\vec{t} = \mathbf{R}^{-1} \cdot \vec{t}' \quad (1.13)$$

Es evidente que aunque sólo hemos deducido estas ecuaciones para primarios normalizados (coeficientes triestímulo), el tipo de relación se mantiene aun cuando estemos considerando un sistema de coordenadas no normalizado.

Estas transformaciones lineales entre distintos sistemas de coordenadas tienen una gran importancia desde el punto de vista tecnológico, pues, en principio, permiten desvincular el tipo de fósforos utilizados en recepción de las señales transmitidas. Así por ejemplo, si un determinado fabricante decide introducir un nuevo tipo de fósforos en sus receptores, sólo requerirá realizar la transformación lineal que relaciona su sistema de reproducción con el codificado en la transmisión para adaptarse a las características de los nuevos elementos.

1.4.10 Sistema de coordenadas con primarios monocromáticos

En 1931 el C.I.E. (*Commission Internationale de L'Eclairage* – comité internacional dedicado a la establecer normativas sobre la representación del color) propuso utilizar como sistema de coordenadas para representar los colores un conjunto de primarios monocromáticos. Las longitudes de onda de estos primarios son de 700 nm, 546,1 nm y 435,8 nm para el rojo, verde y azul respectivamente.

Al tratarse de colores monocromáticos, pueden expresarse analíticamente como:

$$\begin{aligned} p_R(\lambda) &= \delta(\lambda - 700 \cdot 10^{-9}) \\ p_G(\lambda) &= \delta(\lambda - 546,1 \cdot 10^{-9}) \\ p_B(\lambda) &= \delta(\lambda - 435,8 \cdot 10^{-9}) \end{aligned} \quad (1.14)$$

donde $\delta(\lambda)$ representa la función Delta de Dirac, de modo que en este caso, los coeficientes de la matriz \mathbf{P} asociada a este sistema de coordenadas representan muestras sobre las curvas de sensibilidad de los tres tipos de receptores de la retina. En efecto, podemos escribir:

$$\alpha_X(p_Z) = \int_{\lambda_{\min}}^{\lambda_{\max}} S_X(\lambda) \cdot p_Z(\lambda) d\lambda = \int_{\lambda_{\min}}^{\lambda_{\max}} S_X(\lambda) \cdot \delta(\lambda - \lambda_Z) d\lambda = S_X(\lambda_Z) \quad (1.15)$$

de modo que particularizando los valores de X y Z a las distintas componentes de color R, G y B, obtenemos una matriz \mathbf{P} en la que cada fila está formada por tres muestras de la curva de sensibilidad de un tipo de conos en las frecuencias asociadas a los colores considerados como primarios:

$$\mathbf{P} = \begin{bmatrix} S_R(\lambda_R) & S_R(\lambda_G) & S_R(\lambda_B) \\ S_G(\lambda_R) & S_G(\lambda_G) & S_G(\lambda_B) \\ S_B(\lambda_R) & S_B(\lambda_G) & S_B(\lambda_B) \end{bmatrix} \quad (1.16)$$

Este conjunto de primarios se normaliza respecto a un blanco de referencia que tiene un espectro plano en todo el margen visible. Los coeficientes triestímulo obtenidos por esta normalización se denotaran como R_C , G_C y B_C .

1.4.11 Sistema de coordenadas basado en los fósforos del receptor NTSC

La selección de los fósforos para el sistema NTSC se realizó en 1953 teniendo en cuenta los materiales existentes en aquella época y las tendencias futuras. Los colores seleccionados, si bien están centrados sobre las regiones roja, verde y azul del espectro visible, no fueron puramente monocromáticos por cuanto el rendimiento luminoso de este tipo de materiales es muy bajo y hubiera sido necesario aplicar grandes potencias de señal para excitarlos adecuadamente. En este caso los primarios están normalizados para que, con los tres componentes triestímulo igual a la unidad, se obtenga un color aproximadamente blanco que se utiliza como referencia. Este blanco de referencia es ligeramente distinto que el definido por el CIE. Los coeficientes triestímulo utilizados en este sistema de coordenadas se denotan como R_N , G_N y B_N .

La relación matricial entre las componentes del receptor NTSC y los primarios monocromáticos del C.I.E. viene dada por:

$$\begin{bmatrix} R_N \\ G_N \\ B_N \end{bmatrix} = \begin{bmatrix} 0.842 & 0.156 & 0.091 \\ -0.129 & 1.320 & -0.203 \\ 0.008 & -0.069 & 0.897 \end{bmatrix} \cdot \begin{bmatrix} R_C \\ G_C \\ B_C \end{bmatrix} \quad (1.17)$$

Evidentemente, la relación inversa puede obtenerse invirtiendo la matriz de cambio de coordenadas.

A pesar que la tecnología de fabricación de fósforos ha evolucionado considerablemente desde 1953, este sistema de coordenadas sigue siendo uno de los más utilizados para referirse a las señales que se envían a través de cualquier sistema de televisión. De hecho, veremos que aunque tanto en el sistema NTSC como en el PAL, las señales de información de color no se corresponden directamente con las componentes R_N , G_N y B_N , siempre se suelen expresar en función de este sistema de coordenadas. Debido a su gran uso, siempre que nos refiramos de manera genérica a las componentes R, G, B, sin hacer referencia a ningún subíndice, entenderemos que se trata de este sistema de coordenadas.

Desde 1953 hasta la aparición de los primeros sistemas de televisión en color en Europa, el desarrollo de materiales luminiscentes evolucionó considerablemente, por lo que la EBU (European Broadcasting Union) definió unos colores para las tríadas ligeramente distintos a los utilizados en el sistema NTSC. Estas tríadas, aunque cubren algo peor los matices de color en la gama del verde-azul, permiten obtener una luminosidad de pantalla superior a la del sistema NTSC, proporcionando una excelente sensación de plenitud de color. También el SMPTE (Society of Motion Picture and Television Engineers) ha definido un nuevo conjunto de primarios cuyo propósito es adaptarse a las

características de los diversos tipos de fósforos que se utilizan actualmente en el propio sistema NTSC. Sin embargo, a pesar de estas actualizaciones, sigue siendo práctica habitual referirse al sistema de fósforos NTSC original. Denotaremos como R_p , G_p , B_p y R_s , G_s , B_s los coeficientes triestímulo asociados a los sistemas de coordenadas definidos por la E.B.U. y el S.M.P.T.E. respectivamente. Actualmente, cada fabricante suele utilizar fósforos con características ligeramente distintas, por lo que, en cualquier caso, siempre debe realizarse un cambio de variables en el propio receptor.

En la figura 1.11 se representan los valores de los coeficientes triestímulo correspondientes al sistema NTSC necesarios para obtener los distintos colores monocromáticos. La gráfica indica, para cada longitud de onda, los valores teóricos con los que deberían excitarse los fósforos para reproducir cada color monocromático. La principal consecuencia de este resultado es que existe toda una gama de longitudes de onda que requerirían utilizar una componente negativa, lo que evidentemente impide generar estos colores, ya que los fósforos sólo admiten excitaciones positivas. En consecuencia, la gama cromática que puede reproducirse con estos fósforos no es completa y algunos colores sólo pueden ser aproximados. Este problema, como veremos, es de carácter general para todo tipo de fósforos y de primarios reales.

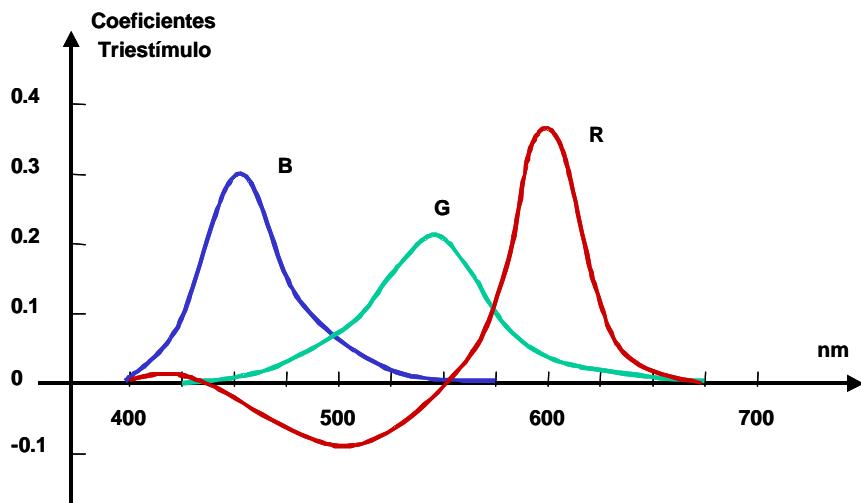


Fig. 1.11 Coeficientes triestímulo NTSC para la obtención de colores monocromáticos

La selección de un tipo u otro de fósforo permite cubrir gamas cromáticas distintas y obtener matices distintos con mayor o menor brillo en la pantalla del reproductor, pero en ningún caso cubrir completamente todos los colores. Este resultado puede parecer en principio decepcionante, pero se trata de un problema casi exclusivamente de carácter conceptual, pues la gama de colores que pueden obtenerse es suficientemente amplia como para que el espectador no sea consciente de esta deficiencia.

El C.I.E. desarrolló un sistema de coordenadas, denominado XYZ, basado en unos primarios artificiales (inexistentes) que permite expresar de modo analítico toda la gama de colores utilizando tan sólo coeficientes triestímulo positivos. Antes de examinar los detalles de este sistema, cuyo uso es también bastante frecuente, veremos algunas alternativas para representar gráficamente los colores en el sistema R_N , G_N , B_N . Estas técnicas de representación son generalizables a cualquier sistema de coordenadas por lo que suprimiremos el subíndice en las componentes triestímulo.

1.4.12 Triángulo de Maxwell y diagrama cromático

Nuestro objetivo ahora es poder representar los distintos colores sobre una superficie plana de modo que en cada punto del plano podamos asignarle un color cuyas componentes triestímulo sean fácilmente identificables. El problema fundamental es que estas componentes forman un espacio tridimensional, por lo que no es posible una representación conjunta del vector y su color asociado sobre el plano. En la figura 1.12 se representan esquemáticamente distintas componentes triestímulo como vectores en un espacio tridimensional. Es importante observar que la información de tonalidad de un color está codificada en la dirección del vector, ya que, como hemos visto, el módulo representa tan sólo el brillo o intensidad de la luz, cambiando todas las componentes de forma proporcional cuando se produce un cambio de magnitud del vector siempre que se mantenga su dirección.

Podemos aprovechar esta característica para normalizar, según algún criterio, las componentes triestímulo de modo que se pierda la información de magnitud de los vectores. Con ello, perderemos una de las variables por lo que ya será posible obtener una representación plana de los colores.

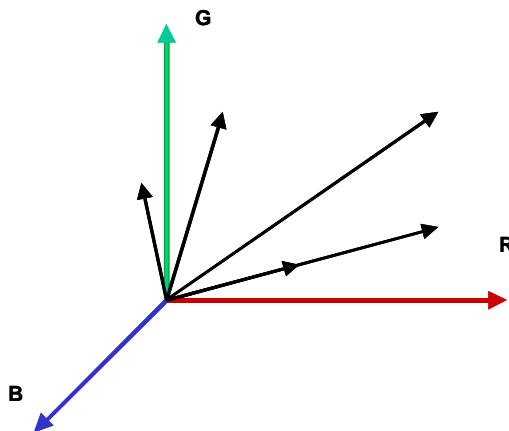


Fig 1.12 Representación de las componentes triestímulo en un espacio tridimensional

La normalización más común consiste en escalar las componentes triestímulo de manera que su suma siempre sea igual a la unidad. Esta normalización puede parecer extraña desde un punto de vista matemático, ya que lo más usual sería utilizar una normalización respecto al módulo, de modo que todos los vectores resultantes tuvieran módulo unitario. Sin embargo, esta filosofía tampoco tiene

mucho sentido en colorimetría por cuanto, aunque el módulo y el brillo están relacionados, éste último depende también de la dirección, por lo que dos vectores con módulos iguales no tienen por qué tener el mismo brillo. Esto puede parecer algo confuso, pero téngase en cuenta que únicamente hemos demostrado que un cambio de módulo supone un cambio en la intensidad de la luz y no en su tonalidad, por cuanto se mantienen las contribuciones relativas de cada componente. Un cambio en la dirección supone un cambio tanto de la tonalidad como de la intensidad. Es fácil comprender que se produzca este cambio en la intensidad de la luz si tenemos en cuenta que la sensibilidad del ojo es distinta para cada componente, de modo que cuando varía su contribución relativa también varía su sensación lumínica. Finalmente, recuérdese que nuestro propósito es representar las distintas tonalidades de color en un plano, sin interés específico en que todos ellos tengan el mismo brillo. La normalización que se utiliza es simple y permite relacionar rápidamente las componentes resultantes.

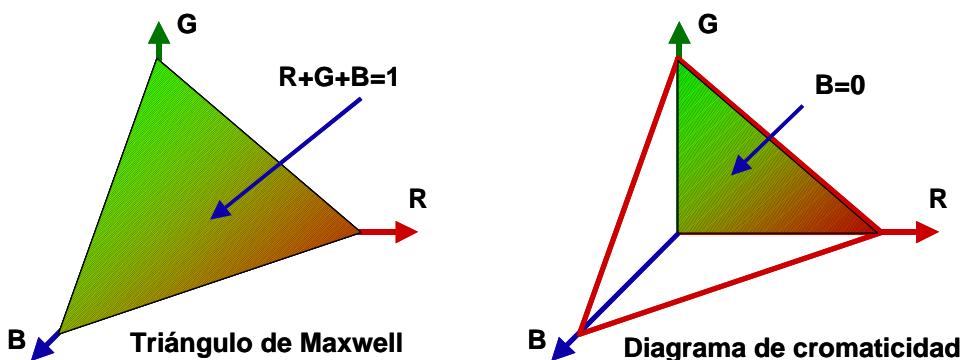


Fig. 1.13 Triángulo de Maxwell y diagrama cromático sobre el espacio de coeficientes triestímulo

Los coeficientes resultantes de esta normalización se denominan coeficientes cromáticos y se denotan con letras minúsculas:

$$r = \frac{R}{R + G + B}; \quad g = \frac{G}{R + G + B}; \quad b = \frac{B}{R + G + B} \quad (1.18)$$

Es evidente, por la propia definición de estos coeficientes, que su suma se mantiene siempre igual a la unidad por lo que uno de ellos es redundante y puede calcularse en función de los otros dos:

$$r + g + b = 1 \quad (1.19)$$

Estas nuevas variables nos ofrecen dos alternativas para representar gráficamente las distintas tonalidades de color sobre unos planos que se conocen con el nombre de triángulo de Maxwell y diagrama cromático. Ambos planos se representan gráficamente en la figura 1.13 sobre el espacio tridimensional de los componentes triestímulo y corresponden con los planos $R+G+B=1$ y $B=0$ respectivamente. En cada punto del triángulo de Maxwell se representan los colores obtenidos con los coeficientes triestímulos que verifican que la suma de sus componentes es igual a la unidad. El diagrama cromático, cuyo uso es más frecuente que el del triángulo de Maxwell, puede considerarse como una proyección del triángulo de Maxwell sobre el plano $B=0$. Dadas las dos componentes (r, g)

de un color en el diagrama cromático, pueden calcularse con facilidad las componentes triestímulo sin más que imponer la condición de que la suma de las tres componentes debe ser igual a la unidad: (r , g , $1-r-g$).

1.4.13 Representación gráfica de la mezcla de colores

En la figura 1.14 se representa una mezcla entre dos colores C_1 y C_2 en los tres sistemas de representación que estamos utilizando. En el caso de componentes triestímulo, la suma de los colores se corresponde con una suma de vectores convencional y se representa exclusivamente para ayudar a comprender el resultado de esta combinación en los otros dos sistemas de representación en los que, como resulta evidente a partir del dibujo, el color resultante se encuentra en la recta que une los colores originales. La posición exacta dentro de esta recta depende de la participación de cada color en la mezcla, situándose más cerca del color cuya suma de componentes triestímulo sea mayor.

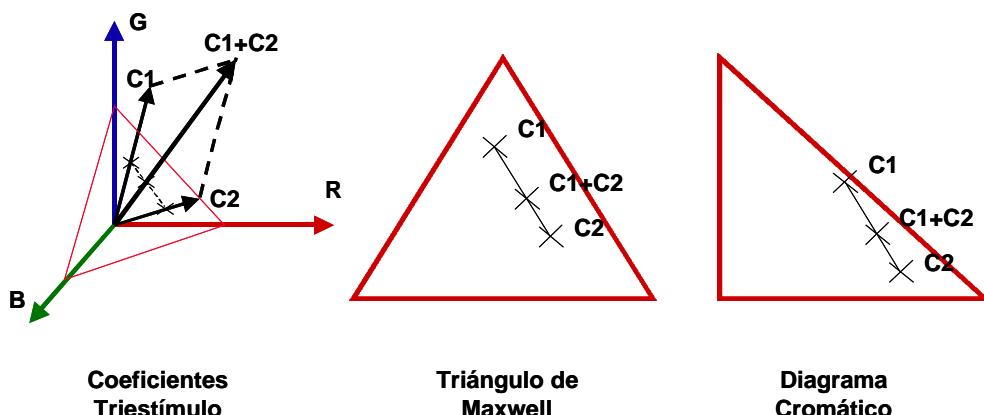


Fig. 1.14 Representación gráfica de la mezcla de colores

Determinemos la posición exacta del valor de la mezcla para el caso del diagrama cromático. Supongamos que las componentes triestímulo de los colores originales vienen dadas por:

$$\begin{aligned} C_1 &= (R_1, G_1, B_1); \\ C_2 &= (R_2, G_2, B_2); \end{aligned} \quad (1.20)$$

Las componentes cromáticas de estos dos colores son:

$$\begin{aligned} C_{1C} &= (R_1 / L_1, G_1 / L_1) \\ C_{2C} &= (R_2 / L_2, G_2 / L_2) \end{aligned} \quad (1.21)$$

donde hemos definido:

$$\begin{aligned} L_1 &= R_1 + G_1 + B_1 \\ L_2 &= R_{21} + G_2 + B_2 \end{aligned} \quad (1.22)$$

Las componentes triestímulo y las cromáticas de la mezcla de los dos colores vienen dadas por:

$$\begin{aligned} C_{1+2} &= (R_1 + R_2, G_1 + G_2, B_1 + B_2); \\ C_{1+2,C} &= \left(\frac{R_1 + R_2}{L_1 + L_2}, \frac{G_1 + G_2}{L_1 + L_2} \right) \end{aligned} \quad (1.23)$$

Este resultado se representa gráficamente en la figura 1.15, donde se indica explícitamente que la mezcla se sitúa en un punto de la recta que une los dos colores a una distancia proporcional a L_1 del color C_2 y L_2 del color C_1 . Esto justifica la idea intuitiva de que la mezcla se sitúa más o menos cerca de los colores originales dependiendo del grado de participación de éstos.

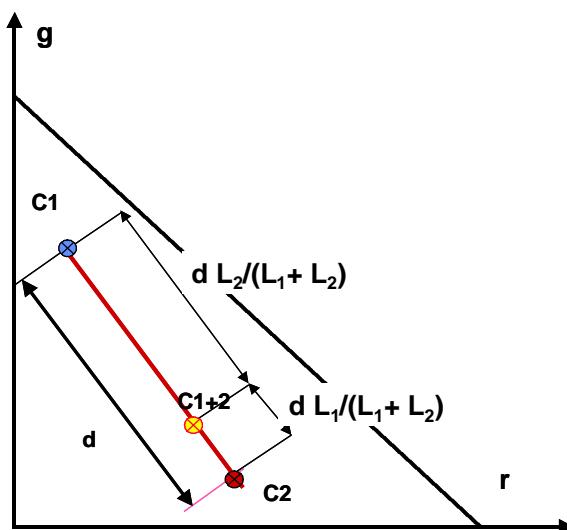


Fig. 1.15 Procedimiento para la mezcla de colores en el diagrama cromático

Comprobaremos este resultado analíticamente sólo para una de las componentes cromática. Supongamos que r_1 , r_2 y r_{1+2} son las componentes rojas correspondientes a los colores C_1 , C_2 y C_{1+2} . De acuerdo con la gráfica de la figura 1.15, hemos de comprobar que:

$$r_{1+2} = r_1 + \frac{(r_2 - r_1)}{L_1 + L_2} \cdot L_2 \quad (1.24)$$

donde $r_2 - r_1$ representa la distancia entre las dos componentes r .

Sustituyendo $r_1 = R_1/L_1$ y $r_2 = R_2/L_2$ en la parte derecha de la igualdad anterior obtenemos:

$$r_1 + \frac{(r_2 - r_1)}{L_1 + L_2} \cdot L_2 = \frac{R_1}{L_1} + \left(\frac{\frac{R_2}{L_2} - \frac{R_1}{L_1}}{L_1 + L_2} \right) \cdot L_2 = \frac{R_1 L_1 + R_1 L_2 + R_2 L_1 - R_1 L_2}{L_1 \cdot (L_1 + L_2)} = \frac{R_1 + R_2}{L_1 + L_2} \quad (1.25)$$

que concuerda con la expresión de la componente cromática roja obtenida en la ecuación 1.23.

El resultado obtenido es válido para cualquier componente cromática, y por lo tanto el mismo procedimiento puede aplicarse para el cálculo gráfico de la mezcla de colores en el triángulo de Maxwell. La figura 1.16 ilustra como aplicar el método en este caso.

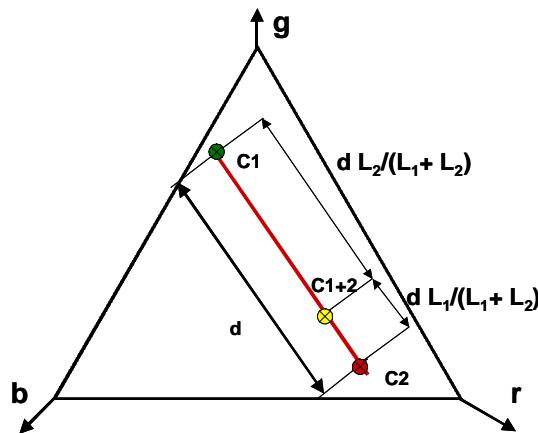


Fig. 1.16 Procedimiento gráfico para la mezcla de colores en el triángulo de Maxwell

1.4.14 Situación de los colores en el triángulo de Maxwell y en el diagrama de cromaticidad

Una vez conocido cómo se obtiene una mezcla de colores sobre el triángulo de Maxwell, resulta fácil ver dónde se situarán los distintos colores que resulten de la combinación de los primarios. Evidentemente, los vértices del triángulo se corresponden con los propios colores primarios, que en este caso serán los correspondientes a los fósforos del sistema NTSC. Por ello, la línea base del triángulo, que une los primarios rojo y azul, contiene la gama de rojos, magentas y azules conseguida mediante la combinación de estos dos primarios. Análogamente, en el lado izquierdo del triángulo aparece la gama de azules, cianos y verdes obtenidos al combinar los primarios azul y verde y en el lado derecho las tonalidades rojas, amarillas y verdes resultado de combinar los dos primarios restantes.

El resultado obtenido por la intervención conjunta de los tres colores primarios puede derivarse gráficamente descomponiendo esta mezcla de colores en dos partes, tal y como se indica en la figura 1.17. En primer lugar podemos calcular el resultado de la combinación entre los primarios rojo y azul, lo que produce un color situado sobre la línea base del triángulo de Maxwell (C_X). Este color puede mezclarse posteriormente con el verde, obteniendo el resultado mostrado en la figura e indicado como

C_Y. De acuerdo con este procedimiento, es evidente que combinando los tres colores básicos podemos desplazarnos a través de todo el triángulo limitado por estos colores.

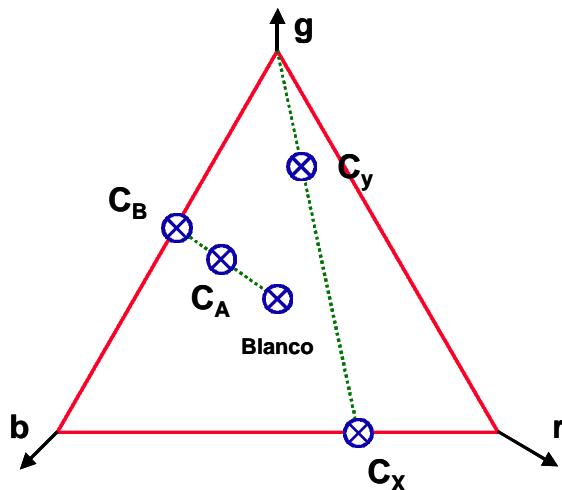


Fig 1.17 Métodos para la situación de los colores en el triángulo de Maxwell

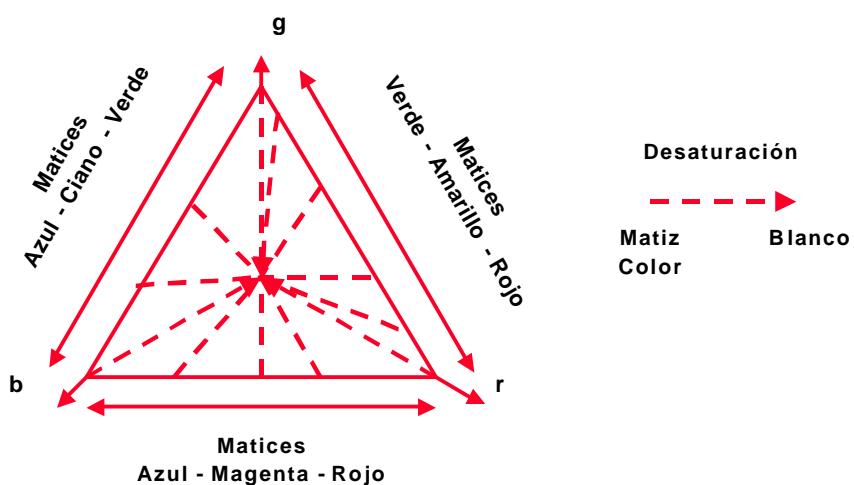


Fig. 1.18 Líneas de matiz y saturación de color en triángulo de Maxwell

Una alternativa a este procedimiento que ayuda a comprender mejor la procedencia de los colores y en qué grado interviene cada primario en su formación consiste en trazar la recta que pasa desde el blanco al color cuya composición queremos analizar y prolongarla hasta que intersecte con uno de los lados del triángulo. El blanco corresponde a las coordenadas cromáticas (1/3,1/3,1/3), por lo que está situado en el baricentro del triángulo de Maxwell. En la figura 1.17 se muestra como la recta que pasa

por el blanco y el color C_A , intersecta los lados del triángulo en el color C_B . Esto significa que C_A puede obtenerse añadiendo una cantidad adecuada de blanco al color C_B , el cual, al estar situado sobre la línea que limita el triángulo, puede obtenerse como una mezcla en la que intervienen únicamente los dos primarios situados en los vértices de ese lado del triángulo.

Esta última interpretación nos permite considerar que los colores están distribuidos según su tonalidad y saturación en direcciones distintas. Así, los lados del triángulo corresponden a distintas tonalidades o matices de color obtenidos como la combinación de dos primarios, mientras que las líneas que partiendo de uno de los lados del triángulo van hacia el baricentro (posición del blanco) corresponden a versiones de la misma tonalidad más o menos saturadas de blanco, dependiendo de su proximidad al centro. Esta distribución de los colores se indica esquemáticamente en la figura 1.18.

En la figura 1.19 se simulan los colores obtenidos por la combinación de los primarios en los lados del triángulo y también se ilustra la degradación al blanco que experimentan distintas tonalidades conforme nos aproximamos al centro del triángulo.

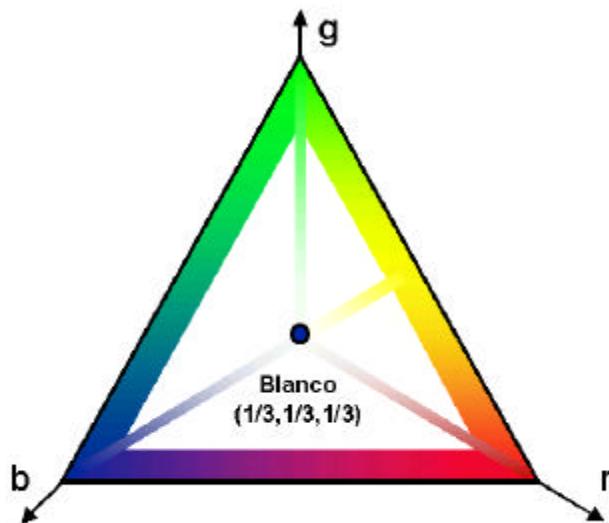


Fig. 1.19 Simulación de la distribución de los colores en el triángulo de Maxwell

Los resultados que hemos obtenido para el triángulo de Maxwell pueden generalizarse directamente para el diagrama de cromaticidad sin más que considerar que éste último puede entenderse como una proyección del primero sobre el plano $b=0$. Por ello, todas los procedimientos geométricos para calcular la combinación entre dos o más colores se mantienen, con la salvedad de que en este caso las coordenadas del color blanco corresponden al punto $(1/3, 1/3)$ del plano. En las figura 1.20 y figura 1.21 se representan la distribución de los colores obtenida en este sistema de representación.

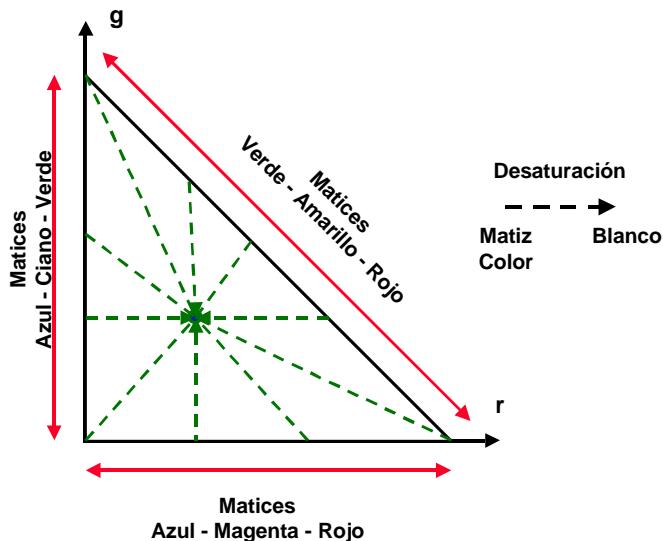


Fig. 1.20 Líneas de matiz y saturación de color en el diagrama de cromaticidad

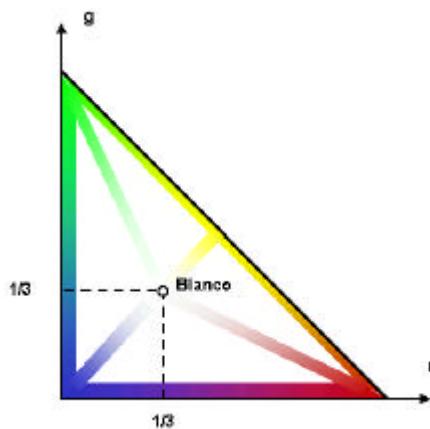


Fig. 1.21 Simulación de colores en el diagrama de cromaticidad

Ejemplo: Cálculo de los coeficientes triestímulo a partir del diagrama de cromaticidad

Consideremos un ejemplo de cómo pueden derivarse las componentes triestímulo de un color que tenemos situado sobre el diagrama cromático, tal y como se indica en la figura 1.22, en el punto $(1/6, 1/3)$. Si no se dispone de ninguna información adicional que permita estimar el brillo, sólo podremos calcular la participación relativa de cada uno de los primarios en la formación del color,

quedando indeterminada su amplitud. En nuestro ejemplo disponemos de las coordenadas r y g, y por lo tanto, podemos estimar la componente b que vendrá dada por:

$$b = 1 - r - g = 1/2 \quad (1.26)$$

De acuerdo con ello, cualquier color cuyas componentes triestímulo sean:

$$(R, G, B) = k \cdot (1/6, 1/3, 1/2) \quad (1.27)$$

siendo k una constante arbitraria, quedará proyectado sobre el punto (1/6, 1/3) del diagrama de cromaticidad. Supongamos ahora que disponemos de la información adicional que este color se ha obtenido añadiendo una cantidad (0.8, 0.8, 0.8) de blanco a una combinación entre los primarios azul y verde. Esta información adicional nos permitirá deducir los componentes triestímulo exactos.

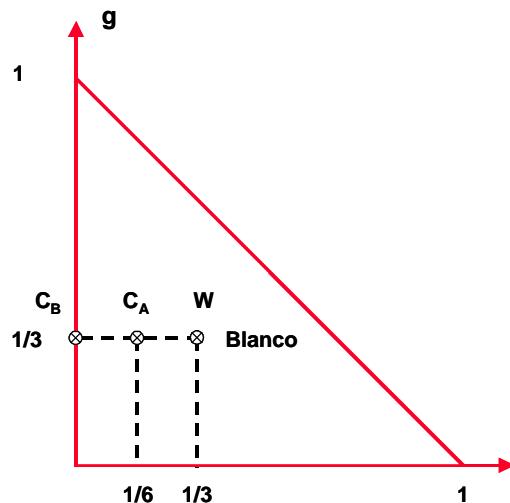


Fig. 1.22 Ejemplo para relacionar componentes cromáticas y triestímulo

El color C_A , de acuerdo con el esquema de la figura 1.22, debe corresponder a una mezcla entre los primarios azul y verde que produce el color C_B . Además, como C_A se encuentra a la misma distancia de C_B que del blanco (W), podemos deducir que la suma de las componentes triestímulo de C_B y del blanco deben coincidir. Es decir:

$$L_1 = L_2 = 0.8 + 0.8 + 0.8 = 2.4 = R_B + G_B + B_B \quad (1.28)$$

Las coordenadas del color C_B en el diagrama de cromaticidad son: (0, 1/3), por lo que sus coeficientes triestímulo deben poder expresarse como:

$$(R_B, G_B, B_B) = k \cdot (0, 1/3, 2/3) \quad (1.29)$$

Esto nos permite deducir que el valor de la constante k debe ser de 2,4. Una vez conocidas las componentes de C_B , debemos añadir la cantidad de blanco especificada para obtener el color C_A :

$$(R_A, G_A, B_A) = 2.4 \cdot (0, 1/3, 2/3) + (0.8, 0.8, 0.8) = (0.8, 1.6, 2.4) \quad (1.30)$$

Este ejemplo sugiere un procedimiento directo para extraer sistemáticamente la cantidad de blanco existente en cualquier color, obteniendo así un nuevo color como combinación de dos únicos primarios cuyo matiz es igual al original pero que está completamente saturado. Así, es evidente que los colores con componentes triestímulo $(0.3, 0.6, 0.8)$ y $(0, 0.3, 0.5)$ tienen el mismo matiz, aunque en el primero intervienen los tres primarios, mientras que en el segundo sólo intervienen el verde y el azul, siendo por tanto una versión más saturada.

1.4.15 Algunos comentarios sobre la representación gráfica de los colores

El diagrama de cromaticidad facilita comprender el efecto de un cambio de primarios en la gama de colores que pueden reproducirse. En la figura 1.23 se ilustra el efecto de sustituir los primarios RGB originales, situados en los vértices del triángulo, por un nuevo grupo de primarios. Es evidente, a partir de las reglas de mezcla de colores que hemos desarrollado, que la gama de colores reproducibles estará limitada por el triángulo cuyos vértices son las posiciones de estos nuevos primarios.

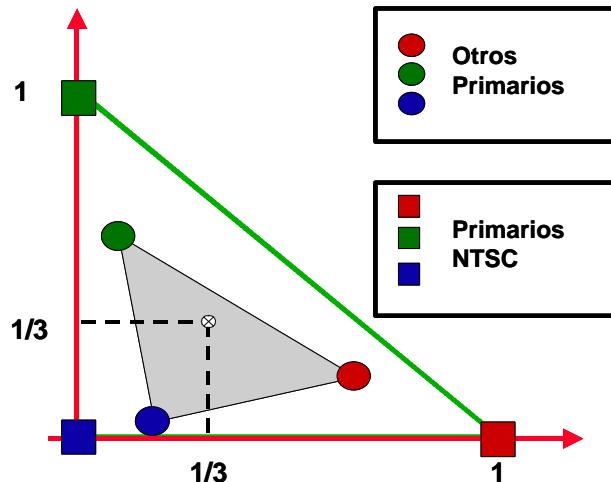


Fig. 1.23 Efecto de la selección de nuevos primarios sobre un diagrama referido a los primarios antiguos

En el caso representado en la figura 1.23, esta región aparece sombreada y el nuevo grupo de primarios puede reproducir menos colores que el conjunto original, ya que éste segundo triángulo está completamente comprendido dentro del primero. Esto no es necesariamente así, ya que es posible elegir unos nuevos colores base que tengan alguna de las componentes negativas, con lo que pueden existir nuevas regiones no cubiertas por el sistema original.

Desde un punto de vista analítico, un cambio de sistema de coordenadas supone identificar la región sombreada correspondiente a los nuevos primarios con un nuevo triángulo rectángulo, con dos de sus lados de longitud unidad y situados sobre los ejes de coordenadas. De esta forma, queremos indicar que un cambio de coordenadas en los coeficientes triestímulo supone también una modificación de las coordenadas cromáticas, con el consiguiente cambio en el triángulo de Maxwell y el diagrama de cromaticidad. La forma de estos dos últimos permanece inalterada, aunque la gama de colores en su interior variará dependiendo del conjunto de primarios elegido. Así pues, cualquier referencia al diagrama de cromaticidad debe aclarar cuál es el conjunto de primarios al que nos estamos refiriendo.

Consideremos ahora, sobre el diagrama de cromaticidad correspondiente a los fósforos del sistema NTSC, cuál es la posición teórica en la que se encuentran los colores monocromáticos. Para ello, podemos utilizar la información proporcionada por la gráfica de la figura 1.11, en la que se indica en qué grado participa cada uno de los primarios para la obtención de los colores monocromáticos. El procedimiento consiste en identificar los valores R_N , G_N , B_N asociados a una determinada longitud de onda y posteriormente calcular las coordenadas cromáticas (r , g), representado este punto sobre el diagrama de cromaticidad. La secuencia de coordenadas obtenida se representa en la figura 1.24, donde observamos que los colores monocromáticos están situados sobre una línea que tiene forma semielíptica y que está situada en el exterior de la gama de colores reproducible con los primarios NTSC. Este resultado no debería sorprender al lector puesto que ya hemos comentado anteriormente que con este conjunto de primarios no es posible reproducir toda la gama de colores. De hecho, la gráfica indica que no puede reproducirse con exactitud ninguno de los colores monocromáticos.

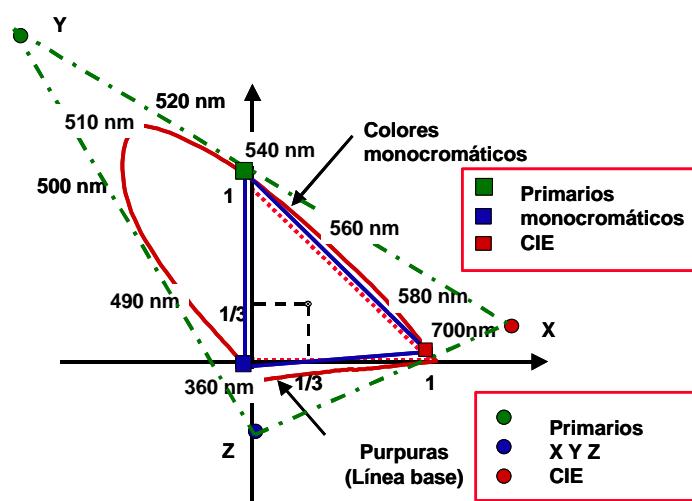


Fig. 1.24 Representación de los primarios CIE monocromáticos y XYZ en el diagrama de cromaticidad asociado a los primarios RGB del NTSC

La línea recta que une los extremos del espectro visible incluye toda la gama de púrpuras, que aun siendo colores puros (no contienen blanco), no están asociados a una única componente espectral. Esto significa que existe toda una gama de tonalidades que no pueden obtenerse a partir de la

descomposición espectral simple de la luz y que por lo tanto no están presentes en el arco iris. Los límites de esta superficie semielíptica incluyen todas las posibles tonalidades de color y las líneas rectas que parten desde el blanco hasta un punto de la curva representan los distintos grados de saturación que puede tener cada tonalidad. Así pues, esta región semielíptica incluye toda la gama cromática de colores.

También es importante notar que los primarios NTSC no están situados sobre la curva correspondiente a los colores monocromáticos debido a que, como ya hemos comentado, se prefirió elegir unos fósforos que no fueran espectralmente puros, pero que proporcionaran una mejor sensación de brillo en la pantalla del reproductor. Por el contrario, si representamos los primarios del sistema de coordenadas monocromáticas definidas por el CIE, veremos que éstos están situados justo sobre esta curva, en las longitudes de onda correspondientes a 700 nm, 546.1 nm, y 435.8 nm. En la gráfica se sitúan aproximadamente estos primarios y se observa que la gama de tonalidades que pueden obtenerse con estos dos sistemas de coordenadas son distintas y que, en ninguno de los dos casos puede conseguirse toda la gama de colores utilizando exclusivamente componentes triestímulo positivas.

Hemos comentado que el CIE definió el sistema de coordenadas XYZ, basado en unos primarios artificiales, que permitía cubrir toda la gama de colores utilizando únicamente coeficientes positivos. La posición de estos primarios en el diagrama de cromaticidad asociado a las componentes RGB del sistema NTSC se representa en la misma figura 1.24. Obsérvese que, necesariamente, los primarios de este sistema de coordenadas deben estar situados sobre los vértices de un triángulo que cubra toda la curva correspondiente a los colores monocromáticos y la gama de tonalidades púrpuras. Como esta superficie tiene una forma semielíptica, es preciso que los vértices del triángulo estén situados en su exterior, en puntos del espacio en los que no está definido ningún color real. Por ello, decimos que estos primarios son artificiales, ya que únicamente tienen sentido desde un punto de vista matemático y no muestran una correspondencia directa con ningún color real.

La selección de las posiciones de estos primarios se realizó utilizando dos criterios. Por una parte, el área limitada por estos triángulos debía ser lo más reducida posible y, por otra, una de las componentes triestímulo debía corresponderse directamente con la sensación de brillo producida por el color. La componente elegida para representar el brillo fue la Y, lo que como veremos, define un plano sobre el que debe situarse uno de los primarios. Los otros dos primarios se seleccionaron para que la región semielíptica correspondiente a toda la gama cromática tuviera un área máxima al representarla en el diagrama de cromaticidad de este sistema de referencia. Antes de definir las relaciones analíticas entre el sistema XYZ y otros sistemas de coordenadas examinaremos con cierto detalle cómo se mide y cuantifica el brillo de un color.

1.4.16 La señal de luminancia

Hemos visto que el brillo es un atributo subjetivo del color que nos permite identificar la luminosidad aparente de los objetos. Decimos que dos objetos tienen el mismo brillo cuando, independientemente de la tonalidad de sus colores, producen en el observador la misma sensación de luminosidad. Este concepto está íntimamente ligado a los sistemas de representación de imágenes en blanco y negro, ya

que en éstas se intenta reproducir la misma sensación de brillo que en la escena original utilizando distintas graduaciones de gris.

En los sistemas de televisión en blanco y negro se utiliza una señal denominada luminancia cuyo nivel es proporcional a la sensación de brillo. Esta señal se obtiene utilizando sensores de imagen, filtros ópticos y sistemas de corrección de nivel de señal cuya respuesta en frecuencia conjunta tiene una forma parecida a la curva de sensibilidad del ojo para visión diurna. La señal de luminancia o, simplemente, la luminancia, es la versión cuantitativa de la sensación de brillo.

Podemos obtener la expresión de la luminancia en función de los primarios de un determinado sistema de representación de color sin más que considerar la sensibilidad lumínica del ojo para cada uno de los colores base. Este cálculo puede realizarse determinando el área existente bajo la curva resultante de ponderar la curva de sensibilidad fotópica (figura 1.7) por el espectro correspondiente a cada uno de los primarios normalizados. Así, la sensación de brillo que produce cada uno de los primarios viene dada por:

$$\begin{aligned}\eta_R &= \int_{\lambda_{\min}}^{\lambda_{\max}} q_R(\lambda) \cdot S_f(\lambda) d\lambda \\ \eta_G &= \int_{\lambda_{\min}}^{\lambda_{\max}} q_G(\lambda) \cdot S_f(\lambda) d\lambda \\ \eta_B &= \int_{\lambda_{\min}}^{\lambda_{\max}} q_B(\lambda) \cdot S_f(\lambda) d\lambda\end{aligned}\quad (1.31)$$

Estos coeficientes se normalizan para que su suma sea igual a la unidad con lo que se obtiene la aportación relativa de cada coeficiente triestímulo en la luminancia. La expresión de la luminancia en función de los coeficientes triestímulo viene dada por:

$$Y = \frac{\eta_R}{\eta_R + \eta_G + \eta_B} \cdot R + \frac{\eta_G}{\eta_R + \eta_G + \eta_B} \cdot G + \frac{\eta_B}{\eta_R + \eta_G + \eta_B} \cdot B \quad (1.32)$$

En el caso de utilizar los primarios del sistema NTSC obtenemos:

$$Y = 0.299R_N + 0.587G_N + 0.114B_N \quad (1.33)$$

Obsérvese que los coeficientes obtenidos indican el grado de participación de cada uno de los primarios en la sensación de brillo y que la normalización de estos coeficientes supone que la luminancia asociada al blanco de referencia es la unidad.

Si utilizamos los primarios monocromáticos del CIE, la luminancia viene dada por:

$$Y = 0.177R_C + 0.812G_C + 0.011B_C \quad (1.34)$$

En este caso, los coeficientes se corresponden directamente con los valores de la curva de sensibilidad fotópica, convenientemente normalizados, en las longitudes de onda asociadas a los primarios.

1.4.17 Sistema de coordenadas XYZ

El sistema XYZ suele usarse con frecuencia como sistema patrón para realizar cálculos de colorimetría y para comparar las gamas de colores que pueden obtenerse con distintos conjuntos de primarios. Su principal ventaja es que cuando se normalizan sus componentes triestímulo, se obtiene un diagrama de cromaticidad que contiene la gama completa de colores en el primer cuadrante, lo que, sin duda, facilita los cálculos gráficos. Otra de sus ventajas es, como hemos comentado, que la componente Y representa directamente la información de luminancia asociada a cada color.

Los coeficientes triestímulo del sistema de coordenadas XYZ están definidos a partir de los primarios monocromáticos del CIE mediante la relación matricial:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.812 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \cdot \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} \quad (1.35)$$

Análogamente, su relación con los primarios en el sistema NTSC viene dada por:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{bmatrix} \cdot \begin{bmatrix} R_N \\ G_N \\ B_N \end{bmatrix} \quad (1.36)$$

Evidentemente, las relaciones inversas pueden obtenerse utilizando las inversas de estas matrices. Nótese que en las expresiones anteriores la relación entre la componente Y y los primarios coincide con los valores de la luminancia obtenidos en las ecuaciones 1.33 y 1.34.

Los coeficientes triestímulo XYZ pueden ser normalizados respecto a su suma obteniendo los coeficientes cromáticos x, y, z:

$$x = \frac{X}{X + Y + Z}; y = \frac{Y}{X + Y + Z}; z = \frac{Z}{X + Y + Z} \quad (1.37)$$

Utilizando estos coeficientes normalizados, podemos obtener el triángulo de Maxwell y el diagrama de cromaticidad en función de estas nuevas variables. Éste último se representa en la figura 1.25 donde observamos que la superficie semielíptica correspondiente a toda la gama de colores ha sido normalizada y ahora está situada en el cuadrante positivo. Nótese que los colores reproducidos sólo son aproximados por cuanto la técnica de representación, tanto si se visualizan en la pantalla del ordenador como en un papel de color, está basada en la mezcla de unos colores primarios que no pueden cubrir toda la gama cromática. Esta aproximación es más que evidente en el supuesto de que el lector disponga de una copia en blanco y negro del diagrama cromático.

En el caso del papel en color, los principios de formación y mezcla de colores son algo distintos de los que estamos considerando en este texto, pues, como en la pintura, los pigmentos que proporcionan la información de color deben considerarse como elementos absorbentes de determinadas componentes espectrales en vez de elementos emisores. Por ello, este tipo de mezcla se denomina subtractiva y utiliza unos colores primarios distintos de los que se utilizan en una mezcla aditiva como la que se

produce en un monitor de televisión. Sin embargo, los principios generales sobre la limitación de la gama de colores reproducibles siguen siendo válidos. Siempre es posible, dados unos colores básicos con los que se realiza la mezcla subtractiva, definir un conjunto de primarios que mezclados aditivamente cubran su misma gama cromática. Por ello, es frecuente representar sobre el diagrama de cromaticidad en componentes (x,y), las regiones de colores que pueden obtenerse cuando se combinan un grupo determinado de colores primarios, independientemente de si se usan mezclas aditivas (televisión, paneles electroluminiscentes) o subtractivas (pintura, impresión, fotografía, cine).

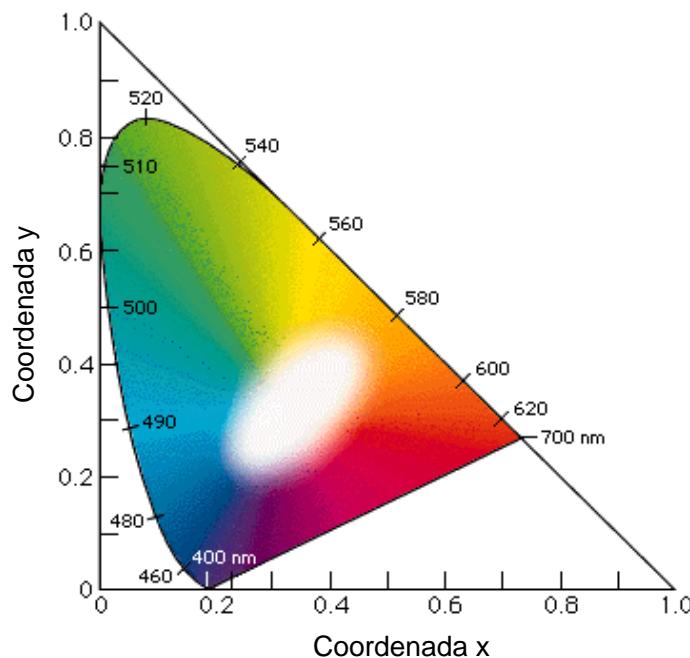


Fig. 1.25 Diagrama cromático en x,y. Simulación de la gama de colores completa

En la figura 1.26 se indican aproximadamente las posiciones de los primarios para el sistema monocromático del CIE, los fósforos NTSC y una gama cromática típica de película fotográfica. Es importante notar que la gama de colores que puede obtenerse con una película fotográfica es más limitada que la que proporcionan los primarios de televisión. Por ello, cuando se transmiten por televisión programas registrados directamente en formato de vídeo, se observa una mayor riqueza cromática que cuando se pasan películas procedentes de un registro fotográfico. La gama cromática proporcionada por una película depende de los materiales fotosensibles utilizados, de manera que es importante seleccionar adecuadamente estas características en función del tipo de colores que queramos enfatizar en la escena. Así, en fotografía, es habitual seleccionar la marca Kodak cuando el tipo de paisaje incluye tonalidades rojas, marrones, naranjas, etc., pues toda esta gama está muy bien cubierta por este carrete. La marca Fujifilm cubre muy bien las tonalidades verdes y Agfa las azules y rojas.

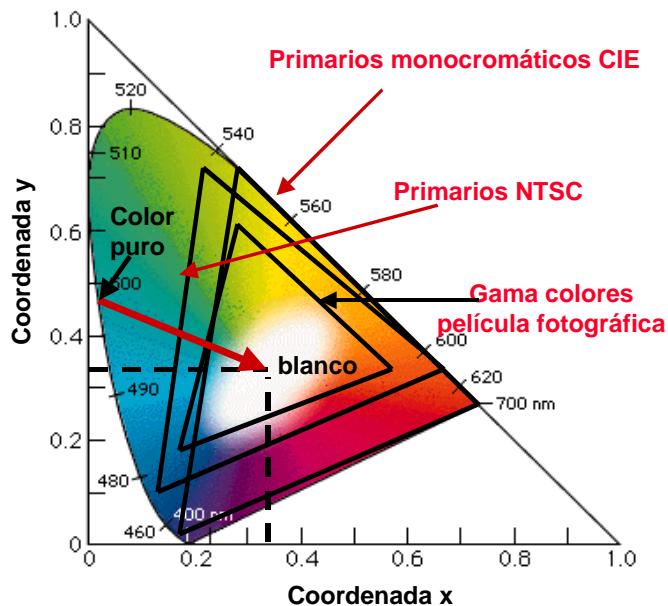


Fig. 1.26 Gamas de colores cubiertas por los distintos tipos de primarios

En la misma figura 1.26 también hemos representado la posición del blanco de referencia cuyas coordenadas son (1/3, 1/3). La línea recta que une un punto del contorno de la superficie con el blanco corresponde a colores que tienen la misma tonalidad pero distinto grado de saturación mientras que el contorno de la superficie representa todas las tonalidades puras (sin blanco).

La capacidad de discernir colores del sistema visual humano no es uniforme cuando nos desplazamos a través de la superficie cromática. Así, en la región del azul pueden distinguirse muchos más colores que en la del verde. Para indicar gráficamente sobre el diagrama de cromaticidad la capacidad del sistema visual para discernir colores, suelen utilizarse las elipses de Mac-Adam. Una elipse de Mac-Adam indica que los colores que están en su interior son indistinguibles para un observador con visión normal. El procedimiento utilizado para determinar la elipse de Mac-Adam, cuyo centro está situado sobre un determinado color C_1 , consiste en presentar a un espectador diversos colores próximos a C_1 , marcando el contorno de la elipse sobre aquellos en que se aprecian diferencias de tono o de matiz y que presentan una distancia mínima con el color de referencia. De acuerdo con este procedimiento, podemos considerar que los colores contenidos dentro del contorno serán indistinguibles, de modo que cuanto menor sea el área de una elipse, mayor es la sensibilidad del ojo para la percepción de colores en esa región.

En la figura 1.27 se han representado aproximadamente algunas elipses de Mac-Adam sobre el diagrama de cromaticidad con el único objetivo de obtener una imagen gráfica de la capacidad para distinguir los colores. Las elipses se han representado ampliando los ejes en un factor 10 para que puedan apreciarse claramente las direcciones en las que el sistema visual presenta mayor sensibilidad.

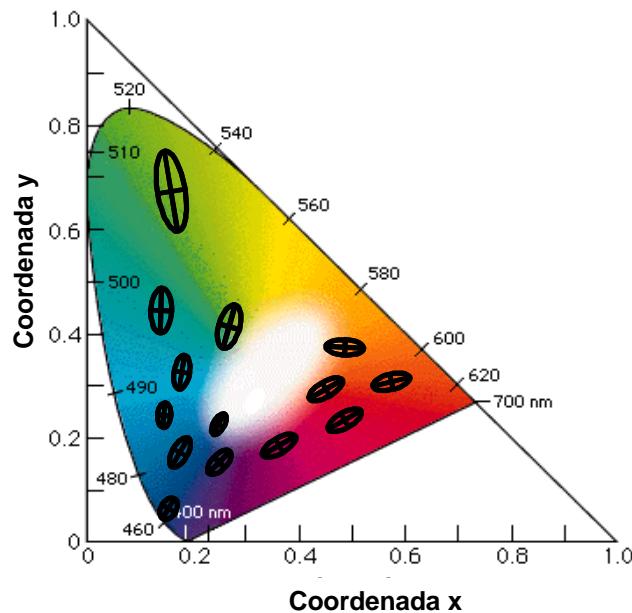


Fig. 1.27 Representación esquemática de las elipses de Mc-Adam

En 1960 el CIE definió un nuevo sistema de coordenadas basado en el sistema xyz con el objetivo de que las elipses de Mc-Adam quedaran representadas con círculos del mismo tamaño independientemente de su posición. Este sistema de coordenadas se denota como el sistema UVW y se denomina el sistema de cromaticidad uniforme. Es un sistema que se utiliza con frecuencia en la calibración y ajuste de los colores proporcionados por un *display*. La relación entre las coordenadas normalizadas u, v y las coordenadas x e y es:

$$\begin{aligned} u &= \frac{4x}{-2x + 12y + 3} \\ v &= \frac{6y}{-2x + 12y + 3} \end{aligned} \quad (1.38)$$

1.4.18 Obtención de las componentes de color: filtros dicroicos

La separación de las tres componentes de color de la escena se realiza en la cámara mediante un subsistema denominado separador óptico o filtro dicroico. Estos filtros presentan unos coeficientes de reflexión y de refracción que dependen de la frecuencia de manera que pueden desviar cada componente de color en direcciones distintas. En la figura 1.28 se representa un esquema de separador óptico en el que se observa que a través de distintas reflexiones y refracciones de las imágenes puede conseguirse separar las tres componentes de color y dirigirlas hacia distintos sensores de imagen.

Cada filtro dicroico refleja una de las componentes de color y permite la transmisión del resto, por lo que mediante dos filtros, uno que refleje la componente azul y otro que refleje la roja, pueden obtenerse las tres componentes de la imagen. Una vez separadas las componentes se utilizan filtros de ajuste adicionales (indicados en la figura 1.28) para garantizar que las señales se ajustan al sistema de representación de colores utilizado, que normalmente será el sistema de fósforos NTSC.

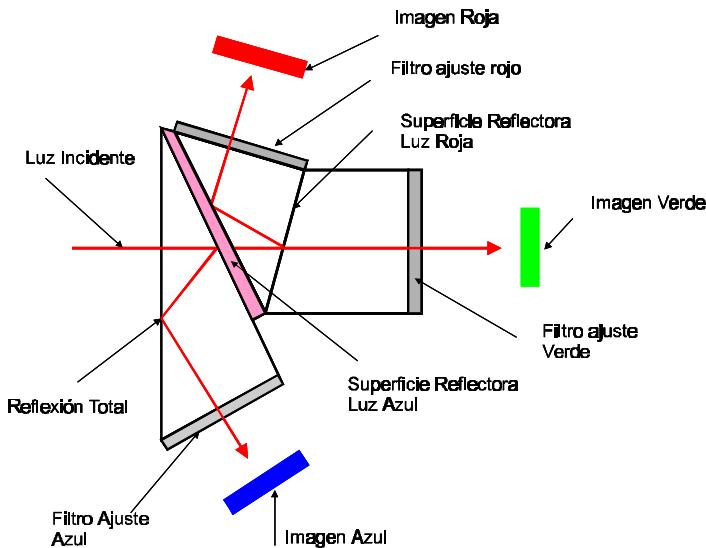


Fig. 1.28 Descomposición de la imagen en componentes de color mediante filtros dicroicos

El principio de funcionamiento de estos filtros se esquematiza en la figura 1.29 para una de las superficies reflectantes de una componente de color. Sobre una base de vidrio con un índice de refracción aproximado de 1,5 se deposita, normalmente por vaporización, una capa de sulfuro de zinc (índice de refracción de 2,4) o cualquier otro material con un índice de refracción elevado. El espesor de esta capa es de $\lambda/4$ de la longitud de onda a reflejar, así en el caso del color rojo el espesor de la capa es de unos 740/4 nm. Cuando la luz incide sobre el sistema se producen dos posibles cambios de medio entre el aire, el sulfuro de zinc y el vidrio. En el primer cambio de medio, del aire al sulfuro de zinc, parte de la luz se reflejará y parte se transmitirá a través del sulfuro de zinc. La parte reflejada en A sufre un cambio de fase de 180°, ya que el índice de refracción del sulfuro de zinc es superior al del aire. La parte transmitida volverá a encontrarse con un cambio de medio entre el sulfuro de zinc y el vidrio, donde parte de la energía será nuevamente reflejada. En este caso la reflexión no introduce cambio en la fase de la onda, por cuanto el índice del vidrio es inferior al del sulfuro de zinc. La onda reflejada en el punto B se propaga hasta el punto C, donde parte de la energía se transmitirá de nuevo hacia el aire. Los dos rayos reflejados D y E estarán en fase, y por lo tanto se sumarán constructivamente, para aquella longitud de onda para la que se ha diseñado el espesor de la capa. En efecto, la onda del rayo D ha sufrido un desfasaje de 180° como consecuencia de su reflexión directa mientras que la onda E, que ha seguido la trayectoria A-B-C, no ha sufrido ninguna inversión de fase en los cambios de medio, pero ha recorrido un trayecto de aproximadamente $\lambda/2$, por lo que ambos

rayos están en fase y se suman constructivamente. El sistema favorece por tanto la reflexión de la imagen en aquellas longitudes de onda que son 4 veces superiores al espesor de la capa de sulfuro de zinc. Para el resto de frecuencias, los rayos reflejados no estarán en fase, por lo que la energía refleja es considerablemente inferior. Superponiendo un cierto número de capas formadas por un índice de reflexión débil y uno fuerte sucesivamente, puede aumentarse la selectividad de la discriminación en frecuencia.

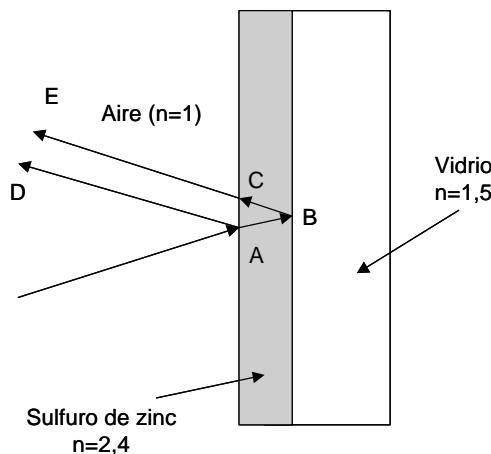


Fig. 1.29 Principio de funcionamiento de un espejo dicroico

1.4.19 Transmisión de información de color: señales diferencia de color

Para poder reproducir correctamente un color, es preciso disponer de las componentes RGB en algún sistema de coordenadas adaptado a las características de los fósforos del monitor. Si se transmitieran directamente las componentes RGB del sistema NTSC, siempre sería posible convertirlas al tipo de fósforos utilizado en nuestro receptor mediante un cambio de coordenadas, que al ser una relación lineal, puede implementarse circuitualmente mediante una red resistiva. Por otra parte, parece lógico establecer un sistema de coordenadas común en el que se realicen todas las transmisiones, dejando para el fabricante de receptores el problema de tratar estas señales para acondicionarlas al tipo de fósforos que prefiera.

Sin embargo, cuando se plantearon los primeros sistemas de televisión en color se prefirió mantener la compatibilidad con los sistemas de blanco y negro, para así realizar una transición gradual hacia el color. Esta decisión fue una cuestión puramente comercial, pues se pensó, acertadamente, que la implantación progresiva de los sistemas en color mantendría un nivel global de ventas en el sector con un aumento sostenido.

La compatibilidad exige que la señal de color pueda seguir siendo visualizada mediante un receptor monocromo con un nivel de interferencia o degradación de la información en blanco y negro imperceptible para el usuario. De este modo, cuando una emisora decidía empezar a transmitir su programación en color, podía seguir siendo vista por todo el parque de receptores en blanco y negro sin que estos apreciaran el cambio de sistema.

Un segundo requisito es la denominada compatibilidad indirecta o retrocompatibilidad, que establece que desde un receptor en color deben poder decodificarse las señales que se transmiten en blanco y negro, obteniendo las tres componentes de color iguales y en consecuencia visualizando una imagen de las mismas características que en un televisor en blanco y negro. Veremos que tanto la compatibilidad directa como la indirecta condicionan el tipo de señales que deben transmitirse para codificar el color.

Así pues, es evidente que todo sistema de televisión en color compatible requiere transmitir la información de luminancia, ya que ésta es la única componente de señal que puede ser utilizada por los receptores monocromos para decodificar correctamente la imagen en blanco y negro. Pero además de la señal de luminancia, deberemos proporcionar información adicional que permita reconstruir las componentes de color RGB que son necesarias para atacar al tubo de imagen en un receptor de color. Estas componentes adicionales deberán codificarse de manera que su efecto sobre un receptor en blanco y negro sea imperceptible. En el capítulo sobre la señal de televisión en color veremos que las señales que proporcionan la información de color están imbricadas en el espectro de la señal de blanco y negro utilizando una modulación en cuadratura cuyos parámetros se han elegido para que la interferencia que introducen en un receptor monocromo, que no está diseñado para tener en cuenta la existencia de esta información, sea lo más reducida posible.

Por otra parte, parece absurdo transmitir, junto con la información de luminancia, las tres componentes de color, ya que en este caso se emitirían cuatro señales que dependen únicamente de tres variables, lo que supondría un injustificable derroche en ancho de banda. La solución más evidente, en primera instancia, parece ser la de transmitir la luminancia y dos componentes adicionales de color. Los receptores en blanco y negro ignorarían la información de color reproduciendo directamente la luminancia, mientras que los receptores de color podrían utilizar todas las señales para obtener las tres variables independientes de color. Examinemos con cierto detalle esta posibilidad para comprobar que es una buena solución al problema de la compatibilidad directa, pero que finalmente será inviable, pues no permite la retrocompatibilidad.

Luminancia más dos componentes de color

Supongamos que en un hipotético sistema de televisión en color se transmiten, junto con la luminancia, dos componentes de color cualesquiera, por ejemplo, las componentes R_N y B_N . Es evidente que si las componentes de color se modulan de acuerdo con los criterios comentados anteriormente, su efecto sobre un receptor en blanco y negro será mínimo, por lo que éste únicamente utilizará la información de luminancia, con una degradación muy reducida debido a la presencia de las componentes de color.

El receptor en color puede interpretar correctamente la información transmitida, siendo posible recomponer las tres señales de color a partir de la luminancia y las componentes R y B. Para ello, es necesario realizar la siguiente transformación lineal:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1.703 & -0.509 & -0.194 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} Y \\ R \\ B \end{bmatrix} \quad (1.39)$$

La segunda fila de esta matriz se obtiene aislando el valor de la componente G en la expresión de la luminancia obtenida en la ecuación 1.33:

$$G = \frac{Y - 0.299 R - 0.114 B}{0.587} = 1.703 Y - 0.509 R - 0.194 B \quad (1.40)$$

relación que, al ser lineal, puede realizarse fácilmente combinando elementos sumadores y restadores mediante un circuito resistivo basado en amplificadores operacionales. En la figura 1.30 se indica esquemáticamente la obtención de las tres componentes R, G y B a partir de las señales Y, R y B que garantizan la compatibilidad directa del sistema propuesto.

Para determinar si el sistema es retrocompatible, consideremos que se transmite únicamente la señal de luminancia y veamos cuál es la información que recuperamos en un receptor en color. En este caso, las componentes de color son nulas, por lo que si aplicamos la relación matricial 1.39 obtenemos:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1.703 & -0.509 & -0.194 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1.703 Y \\ 0 \end{bmatrix} \quad (1.41)$$

Es decir, si la señal sólo contiene información de luminancia, el receptor en color presentará una imagen monocroma, pero no en blanco y negro, sino en verde. La imagen podría verse igual, pues ahora el brillo está contenido en una de las componentes de color, aunque hemos de admitir que no se trata de la solución que se espera de un sistema de televisión retrocompatible.

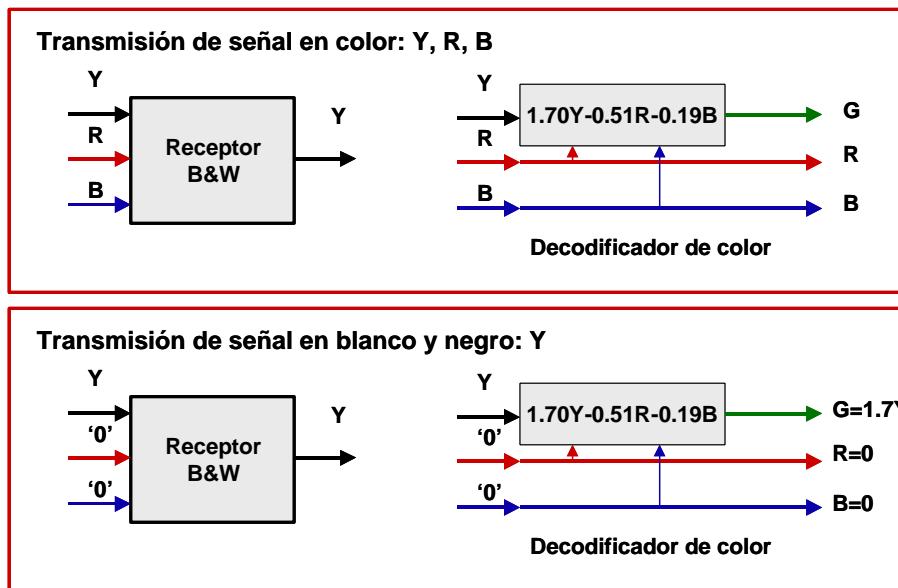


Fig. 1.30 Compatibilidad y retrocompatibilidad en un sistema Y,R,B

A partir de este resultado vemos que para que un sistema sea retrocompatible, es necesario que, cuando no se transmita información de croma, recuperemos las tres componentes de color con el mismo valor. Así, si el receptor es capaz de obtener las componentes $R=G=B=Y$, veremos la imagen con las mismas características de brillo y color que con un receptor en blanco y negro. Para obtener simultáneamente la compatibilidad directa y la indirecta, es necesario sustituir las componentes de color por las componentes diferencia de color transmitiendo las señales que consideraremos en el siguiente apartado.

Luminancia más señales diferencia de color

Se denominan señales diferencia de color a las obtenidas de restar la luminancia a cualquiera de las componentes de color. Así, las posibles señales diferencia de color son: $R-Y$, $G-Y$ y $B-Y$. Si junto con la luminancia se transmiten dos señales diferencia de color cualesquiera, es posible obtener un sistema de televisión en color que admite simultáneamente la compatibilidad directa y la indirecta.

Supongamos un sistema en el que se transmiten las señales diferencia de color $R-Y$ y $B-Y$ junto con la información de luminancia. El sistema tendrá compatibilidad directa siempre que las dos componentes de croma estén moduladas de modo que su incidencia sobre un receptor en blanco y negro sea nula (o mínima) y cuando el receptor en color pueda recuperar, a partir de las señales transmitidas, las tres componentes triestímulo.

En este último caso, requiere realizarse una transformación lineal de variables en el receptor que viene caracterizada por la matriz:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -0.509 & -0.194 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} Y \\ R-Y \\ B-Y \end{bmatrix} \quad (1.42)$$

Resulta directo comprobar la corrección de la primera y tercera fila. Para comprobar la segunda podemos escribir:

$$G = Y - 0.509(R - Y) - 0.194(B - Y) = 1.703 Y - 0.509 R - 0.194 B \quad (1.43)$$

En la figura 1.31 se indica el esquema de un decodificador de color para obtener las tres componentes RGB a partir de la luminancia y las dos señales diferencia de color.

Consideremos ahora el caso en que se transmite una señal en blanco y negro y comprobemos que las tres componentes de color serán iguales a la luminancia. En efecto, ahora las señales diferencia de color no se transmiten, por lo que estas señales serán nulas en el receptor. Si aplicamos la misma matriz de decodificación a las señales recibidas obtenemos:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -0.509 & -0.194 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} Y \\ Y \\ Y \end{bmatrix} \quad (1.44)$$

Es decir, tal y como esperábamos, obtenemos las tres componentes idénticas e iguales a la luminancia, por lo que la imagen se verá en blanco y negro y con el mismo nivel de luminancia con el que ha sido transmitida.

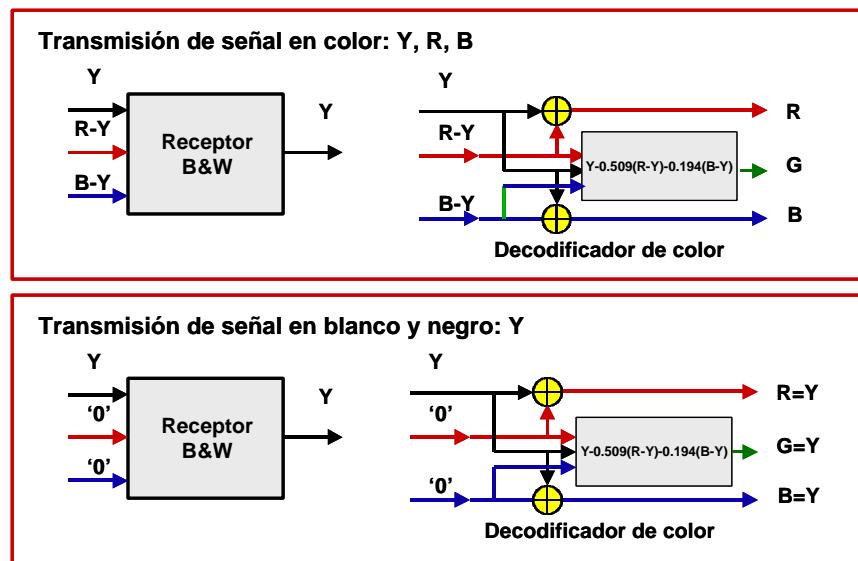


Fig. 1.31 Compatibilidad y retrocompatibilidad en un sistema Y, (R-Y), (B-Y)

Los resultados obtenidos pueden generalizarse de modo que con dos señales diferencia de color cualesquiera podamos obtener un decodificador de color compatible con la transmisión de señales en blanco y negro. No obstante, todos los sistemas de televisión utilizan siempre las dos componentes que hemos considerado, o alguna transformación lineal de las mismas, descartando la señal Y-G. La razón principal es que la señal de luminancia está más correlada con la componente verde que con las otras dos. Esto es debido a que al construir la información de luminancia la componente verde interviene prácticamente en un 60 %, mientras que las otras dos participan en un 30 % y 10 %. Es decir, gran parte de la información correspondiente a la tonalidad verde está contenida en la propia señal de luminancia. La consecuencia inmediata es que el nivel de la señal G-Y suele ser más reducido que el de las otras dos señales diferencia de color, en prácticamente la totalidad de la gama de colores, por lo que será una más sensible al ruido que en consecuencia conviene no transmitir. Otra ventaja de transmitir las señales (R-Y) y (B-Y) es que los coeficientes que se requieren en el decodificador de color son siempre menores que la unidad, por lo que éste puede sintetizarse mediante una red pasiva. Esta última característica no tiene excesiva importancia para la tecnología actual, en la que los amplificadores de señal tienen bajo ruido y son fácilmente integrables, pero tuvo cierto peso en la selección de estas componentes en los albores de la televisión en color.

La codificación de la información de color en una componente de luminancia y dos señales diferencia de color ofrece, además de la compatibilidad, ventajas adicionales, y por ello se sigue utilizando en sistemas digitales o de alta definición en los que no se busca ningún tipo de compatibilidad con los

sistemas de televisión precedentes. La principal ventaja es que el ancho de banda asignado a las señales diferencia de color puede ser considerablemente inferior al ancho de banda asignado a la señal de luminancia sin que se observe una pérdida de calidad en la imagen. La razón de esta posible reducción del ancho de banda se encuentra nuevamente en las características de percepción del color en el sistema visual humano. En condiciones de visión fotópica, los conos son responsables tanto de la percepción del brillo como del color. En la percepción del brillo colaboran todos los conos, independientemente de su tipo, por lo que la densidad equivalente de sensores de intensidad de luz en la retina es superior a la de los sensores de cada componente color. Supongamos por ejemplo un pequeño objeto de color situado a una determinada distancia del espectador. Para percibir el color del objeto se requiere que el tamaño de su imagen en la retina active, como mínimo, una región que contenga los tres tipos de conos, mientras que para percibir su presencia, a partir de su diferencia de brillo con el fondo, es suficiente con que se activen algún tipo de sensores. Una experiencia sencilla que permite comprobar esta diferencia en resolución consiste en situar un hilo de color delante de un espectador. Si la distancia es la adecuada, resulta fácil detectar la presencia del hilo, pero difícilmente podrá identificarse su color.

En definitiva, el sistema visual humano tiene mayor agudeza visual para distinguir distintos niveles de iluminación que para diferenciar colores. Por ello, las señales diferencia de color pueden enviarse con un ancho de banda menor que la luminancia, puesto que requieren transmitir menos información de color al ser el ojo menos crítico. Si se transmitieran las tres componentes RGB directamente, debería utilizarse un ancho de banda equivalente al de la luminancia en cada una de ellas, ya que, en este caso, la información de luminancia estaría distribuida en las tres componentes. Téngase en cuenta, que cuando se transmite la luminancia y las señales diferencia de color, el decodificador de color obtendrá las componentes RGB con un ancho de banda parecido al de la señal de luminancia, ya que esta señal interviene en la recomposición de todas las componentes de color. La ventaja de separar la luminancia de la información de color es que sólo la primera requerirá transmitirse con máxima resolución.

El ancho de banda con el que se codifica cada una de las señales se determina a partir de pruebas de visión subjetivas en las que el espectador puede controlar el ancho de banda asignado a cada componente hasta comprobar que un aumento de la información transmitida no mejora la calidad de la imagen.

Las señales que se transmiten en los sistemas de televisión suelen ser transformaciones de las componentes diferencia de color. Las transformaciones utilizadas dependen del sistema de televisión y pueden ser muy diversas. Así, en el sistema NTSC se transmite una versión rotada y escalada de las señales diferencia de color, mientras que en el sistema PAL únicamente se realiza un escalado previo de las señales para nivelar su margen dinámico antes de la transmisión.

En general, para cualquier combinación de estas señales diferencia de color en la forma:

$$\begin{bmatrix} Y \\ A \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r_{11} & r_{12} \\ 0 & r_{21} & r_{22} \end{bmatrix} \cdot \begin{bmatrix} Y \\ R-Y \\ B-Y \end{bmatrix} \quad (1.45)$$

puede obtenerse un decodificador de las componentes de color RGB a partir de las componentes YAB, que es compatible con la transmisión de señales en blanco y negro. Para comprobarlo, basta con observar que la matriz de decodificación puede obtenerse como:

$$\begin{aligned} \begin{bmatrix} R \\ G \\ B \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & -0.509 & -0.194 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} Y \\ R-Y \\ B-Y \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & -0.509 & -0.194 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & r_{11} & r_{12} \\ 0 & r_{21} & r_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} Y \\ A \\ B \end{bmatrix} \end{aligned} \quad (1.46)$$

de modo que cuando no se transmite información de croma ($A=0$ y $B=0$), las componentes RGB que obtenemos son todas iguales a la luminancia. Las señales A y B son, de acuerdo con la ecuación anterior, una transformación lineal de las $(R-Y)$ y $(B-Y)$ y también suelen denominarse señales diferencia de color. Sus valores dependen del sistema de televisión utilizado siendo los más habituales los que se indican a continuación.

- SISTEMA PAL:

En el sistema PAL las señales diferencia de color reciben lo nombre U y V. Están relacionadas con las $(R-Y)$ y $(B-Y)$ mediante:

$$\begin{aligned} V &= 0.877(R-Y) \\ U &= 0.493(B-Y) \end{aligned} \quad (1.47)$$

que es un caso particular de la relación matricial de la ecuación 1.45:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.877 & 0 \\ 0 & 0 & 0.493 \end{bmatrix} \cdot \begin{bmatrix} Y \\ (R-Y) \\ (B-Y) \end{bmatrix} \quad (1.48)$$

- SISTEMA NTSC.

Utiliza las señales I y Q que son una versión rotada 33° de las señales $(R-Y)$ y $(B-Y)$. Veremos la razón de esta rotación en el capítulo de sistemas de TV en color.

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.783 & -0.269 \\ 0 & 0.478 & 0.414 \end{bmatrix} \cdot \begin{bmatrix} Y \\ (R-Y) \\ (B-Y) \end{bmatrix} \quad (1.49)$$

- SISTEMA SECAM.

En este caso son también las señales $(R-Y)$ y $(B-Y)$ escaladas por otros coeficientes:

$$D_R = -1.9(R-Y); \quad D_B = 1.5(B-Y) \quad (1.50)$$

- **SISTEMAS DIGITALES.**

En los sistemas de televisión digital, se transmiten las señales (R-Y) y (B-Y) con otros factores de escalado:

$$E_{CR} = 0.713(R - Y); \quad E_{CB} = 0.564(B - Y) \quad (1.51)$$

1.5 Proyección de las escenas en el plano de imagen

Una vez la escena se ha descompuesto en los tres haces de luz con frecuencias distintas, la siguiente etapa, que también se realiza en la cámara, consiste en proyectar la escena sobre el plano de imagen. El plano de imagen es la superficie del espacio donde se situarán los elementos encargados de convertir la energía luminosa existente en cada punto del plano de imagen en una señal eléctrica. En el caso de cámaras de vídeo, los dispositivos sensores de la imagen suelen ser matrices CCD o cualquier otro tipo de tubo de imagen. En cámaras de fotografía o de cine, el elemento sensor de la imagen es la película fotográfica. En ambos casos, el elemento sensor está situado exactamente en el plano de imagen.

Desde un punto de vista analítico, la función escena queda ahora representada mediante tres funciones de 3 variables:

$$\begin{aligned} \varepsilon_R(x', y', t) \\ \varepsilon_G(x', y', t) \\ \varepsilon_B(x', y', t) \end{aligned} \quad (1.52)$$

donde las variables x' e y' indican las nuevas variables bidimensionales en las que ha quedado proyectada la escena. Alternativamente, podríamos emplear cualquier conjunto de tres funciones obtenidas como combinaciones lineales independientes de estas tres componentes. Así por ejemplo, podríamos representar la escena mediante las componentes de luminancia y diferencia de color. En general, utilizaremos la notación $\varepsilon_X(x', y', t)$ para indicar mediante el símbolo X, cualquiera de estas funciones.

Evidentemente, estas proyecciones representan una pérdida de la información espacial existente en la escena original que no podrá ser recuperada en la recepción. El espectador deberá determinar la situación espacial de los objetos en la imagen original a partir de sus tamaños relativos y de sus partes ocultas, no pudiendo reproducirse la sensación de espacio tridimensional. Sin embargo, veremos en el apartado siguiente, que aún en el caso de que deseemos reproducir sensaciones de espacio tridimensional en el espectador, es suficiente con registrar dos proyecciones planas de la imagen, con lo que la filosofía general de reducción de variables independientes se mantiene incluso en este caso.

En la figura 1.32 se representa un esquema simplificado de una cámara en color de vídeo en la que se incluyen los tres elementos básicos: grupo óptico, filtros dicroicos y sensores de imagen. El grupo óptico suele estar formado por varias lentes, cuya función es la de obtener una imagen correctamente enfocada de la escena sobre el sensor de imagen. El grupo óptico realiza una función parecida a la que

desempeñan el iris, la pupila y el cristalino en el sistema visual humano, encargándose tanto del correcto enfoque de las imágenes como de regular la cantidad de luz que incide sobre el sensor. La mayoría de grupos ópticos incorporan la función de *zoom*, que consiste en modificar la distancia relativa entre las distintas lentes para variar su distancia focal conjunta. Con ello es posible controlar el tamaño de la imagen de los objetos sobre el sensor. Así, mediante un grupo óptico de gran distancia focal es posible que objetos muy alejados del observador ocupen una región significativa del área del sensor, obteniendo la sensación de que el grupo óptico ‘acerca’ las imágenes. Análogamente, disminuyendo el poder de la lente, podemos aumentar el ángulo de visión de la cámara obteniendo la sensación de que nos alejamos de los objetos. El sistema visual humano, lamentablemente, no dispone de esta capacidad para modificar su distancia focal.

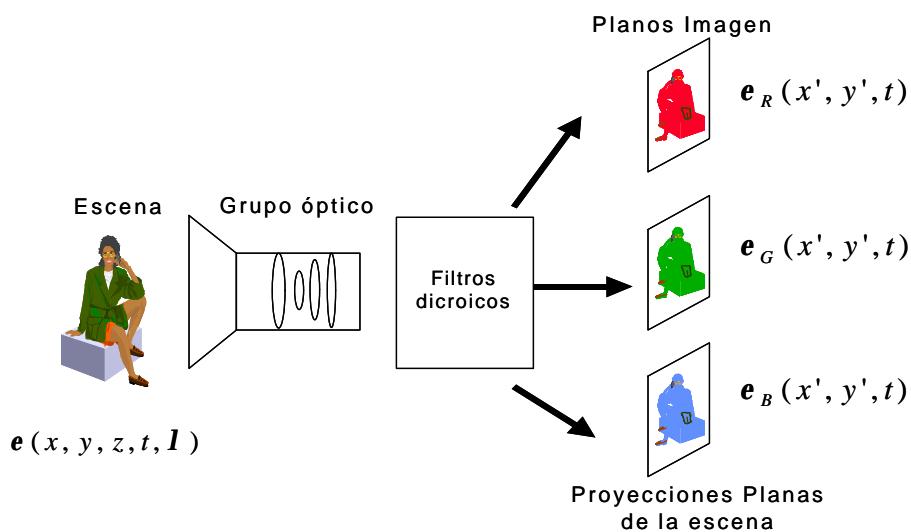


Fig. 1.32 Proceso de proyección de la escena a imágenes planas

Los filtros dicroicos, cuya función ya se ha descrito en el apartado anterior, sólo se usan en sistemas de televisión en color. En cine, fotografía o televisión en blanco y negro se utiliza un único sensor de imagen. Los filtros dicroicos suelen estar situados detrás del grupo óptico, aunque en algunos casos es posible que existan algunas lentes detrás de este separador. Es importante notar que la presencia del sensor de imagen introduce una limitación física en el tamaño de la imagen cuyos efectos discutiremos en la siguiente sección.

1.5.1 Sistemas para la representación de imágenes 3D

Recientemente ha cobrado cierto interés el desarrollo de sistemas tridimensionales, que pretenden que el espectador pueda recuperar la sensación de espacio existente en la escena original.

El interés para el desarrollo de estos sistemas abarca aplicaciones muy diversas, que van desde sistemas de gran público como el cine o la radiodifusión de señales de televisión hasta sistemas más

específicos, como el control de robots para la manipulación a distancia de brazos mecánicos en tareas de peligro, donde el operador pueda actuar sobre una imagen tridimensional del medio; la monitorización del tráfico aéreo en aeropuertos; el desarrollo de sistemas informáticos para la representación gráfica tridimensional con aplicaciones en arquitectura; diseño de interiores; sistemas para realidad virtual; video-juegos; simuladores educativos, etc.

Desde el inicio de la década de los 80, se ha detectado una gran actividad en la investigación y desarrollo de sistemas para la representación de imágenes tridimensionales y actualmente existen diversos sistemas, con filosofías de funcionamiento bastante distintas, algunas de las cuales se encuentran sólo al nivel de prototipos, aunque otras ya están siendo comercializadas con cierto éxito, sobre todo en aplicaciones de video-juegos y presentaciones audiovisuales. Probablemente la introducción de estos sistemas en la televisión será bastante tardía, ya que, si bien la mayoría de grandes empresas en el sector de la electrónica de consumo han empezado a desarrollar programas de investigación en este sentido, las soluciones no parecen por el momento triviales pues los sistemas estudiados son excesivamente complejos y caros y en algunos casos producen cansancio visual en el espectador. Además, el objetivo principal de este sector para los próximos decenios es la implantación en el mercado de los sistemas de TV digital y de alta definición, por lo que no se espera que aparezcan sistemas tridimensionales en el futuro inmediato. Sin embargo, es probable que el uso de estos sistemas empiecen a popularizarse en aplicaciones profesionales, educativas o lúdicas, en sistemas informáticos domésticos o industriales. De hecho, algunos de estos sistemas ya se comercializan actualmente y, probablemente, la principal dificultad para su implantación radica en que existe una multitud de propuestas y sistemas incompatibles que desorientan al consumidor y que se inhiben mutuamente. El objetivo de esta sección es presentar los principios básicos sobre visión tridimensional junto con las distintas propuestas que se están considerando actualmente.

1.5.2 Mecanismos de visión tridimensional

Para analizar los distintos sistemas de representación de imágenes tridimensionales es necesario comprender los mecanismos clave para la percepción del espacio en el sistema visual humano. La percepción tridimensional del espacio se realiza a través de un complejo proceso en el que intervienen conjuntamente varios factores, entre los que destacan: el uso de pistas monoculars, la convergencia y acomodación de los ojos al centrar la atención en distintos puntos del espacio, y las diferencias entre las imágenes captadas por las dos retinas debidas a la distinta perspectiva del espacio que se obtiene con cada uno de los ojos.

Las denominadas pistas monoculars permiten que seamos capaces de identificar con cierta precisión la posición de los objetos en un espacio tridimensional, aun cuando mantengamos uno de los ojos cerrados. Esta percepción del espacio hace uso de la experiencia previa en identificar objetos conocidos de los que se deducen claves, que son analizadas de manera sistemática por el cerebro para determinar su posición y orientación en el espacio. Una de las pistas más relevantes la proporcionan los objetos que quedan parcialmente ocultos por otros objetos situados más cerca del espectador. Otras se obtienen a partir del conocimiento previo del espectador sobre las formas y tamaños de objetos conocidos. Así por ejemplo, si observamos una bicicleta con un único ojo, seremos capaces de deducir la distancia aproximada a la que se encuentra a partir de su tamaño y de las dimensiones de los objetos conocidos que se encuentran a su alrededor. También podremos deducir su orientación utilizando

conceptos elementales de perspectiva visual como el tamaño relativo de cada una de sus ruedas y la forma elíptica que estas toman. Sabemos que la rueda de menor tamaño será la más alejada y que la forma de éstas será tanto más elíptica cuanto más frontalmente esté orientada la bicicleta hacia el espectador. Todas estas apreciaciones se basan en que sabemos que las dos ruedas de la bicicleta son circulares y que su tamaño es aproximadamente igual. Si construyéramos una bicicleta con ruedas elípticas de dimensiones dispares tendríamos serias dificultades para situarla en el espacio a partir de una observación monocular. Así pues, la visión monocular es una consecuencia directa sobre la experiencia de vivir en un espacio tridimensional.

Otro aspecto muy utilizado por el sistema visual para deducir la posición real de los objetos mediante el uso de un único ojo consiste en ver cómo se modifica la perspectiva de la imagen cuando movemos ligeramente la cabeza. En este caso, los objetos próximos experimentan un cambio de posición relativa importante, mientras que la posición de los objetos alejados apenas cambia. Este cambio de posición, conjuntamente con el movimiento de la cabeza, permite deducir aproximadamente la distancia a la que se encuentran los objetos. Los sistemas de representación de imágenes que permiten observar un cambio en la perspectiva de la escena cuando el espectador se desplaza se conocen como sistemas con *parallax* de movimiento.

Existen algunos programas informáticos para el diseño gráfico, o vídeo-juegos, que se autodenominan sistemas tridimensionales, que únicamente proporcionan pistas monoculares al espectador sobre un *display* plano convencional. En el entorno del diseño gráfico suele tratarse de programas que permiten trabajar y definir los objetos en un espacio tridimensional virtual permitiendo modificar con facilidad el punto de vista desde el que se observa el objeto. Probablemente el vídeo-juego 3D más popular es el *doom*, cuya filosofía ha sido adoptada por muchos otros juegos. En este caso, el usuario observa en la pantalla un mundo virtual, en el que puede desplazarse mediante el *joystick* o el teclado, proporcionándose una perspectiva natural de la escena llena de pistas monoculares. El programa modifica el punto de vista de la escena de acuerdo con los movimientos del *joystick*, admitiendo movimientos en cualquier dirección, por lo que el jugador puede inducir fácilmente la distancia a la que se encuentran los objetos. Los sistemas que únicamente utilizan pistas monoculares y representan las imágenes sobre *displays* convencionales son conocidos en el entorno de los sistemas 3D que requieren sistemas de representación específicos como sistemas de 2½ dimensiones.

La visión monocular es esencialmente plana y la percepción de los objetos en el espacio se realiza siempre utilizando pistas indirectas, por lo que no puede considerarse como una verdadera percepción tridimensional del espacio. Este tipo de visión se conoce también como visión *ciclópea*, en honor al cíclope, animal mitológico que poseía un único ojo en el centro de la cabeza, ya que el tipo de información que proporciona es el que se obtendría con un único ojo. Las pistas monoculares son la única información tridimensional que tenemos en cine y televisión convencional, puesto que las imágenes se reproducen siempre sobre un dispositivo plano. A veces, los propios movimientos de la cámara (*travellings*) se utilizan para dar al espectador una sensación subjetiva del espacio en el que transcurre la acción.

Un experimento característico para ver los límites de la percepción monocular, y que se encuentra disponible en casi todos los museos científicos, consiste en pedir a una persona que con un ojo cerrado y sin mover la cabeza toque una pequeña barra cilíndrica que está colgada del techo y próxima a él. En general se requiere realizar un buen número tentativas para tocar la barra, mientras que si el

experimento se repite con los dos ojos abiertos ésta se localiza sin dificultad. El experimento también sirve para comprobar que cuando el espectador ha adquirido información sobre las dimensiones reales de la barra, resulta mucho más sencillo localizar la posición de la barra con uno de los ojos cerrados. También se simplifica notablemente la localización de la barra si se permite que el observador desplace la cabeza. En ambos casos la percepción del espacio se realiza a través de pistas monoculares. Algunos animales, como los pájaros, tienen una notable percepción del espacio utilizando exclusivamente pistas monoculares. En estos casos, los dos ojos están situados a los lados de la cabeza, por lo que proporcionan puntos de vista completamente distintos del entorno, de modo que no puede usarse esta información para reconstruir el espacio tridimensional. Se supone que el proceso de percepción de distancias y espacio se realiza a través de rápidos movimientos de la cabeza con los que se cambia ligeramente el punto de vista de la escena. La información de los distintos puntos de vista es entonces procesada de manera secuencial por el cerebro para obtener la percepción definitiva del espacio. La diferencia fundamental entre este tipo de visión y la humana es, pues, que en el primer caso la información es procesada en serie, mientras que en el segundo se realiza en paralelo. En general, los animales que tienen los dos ojos situados en la parte frontal tienen una mejor percepción del espacio.

Los mecanismos de visión humana requieren la coordinación de los ojos interviniendo dos factores que también resultan determinantes para la percepción del espacio: la acomodación y la convergencia. Se denomina acomodación a los cambios que deben realizarse sobre el cristalino para que, al centrar la atención sobre un objeto, esté quede correctamente enfocado en la retina. De acuerdo con esto, la posición de los músculos que controlan el cristalino se modificará dependiendo de lo alejado que esté el objeto sobre el que centramos nuestra atención. Seguramente, la posición de los músculos es utilizada parcialmente por el cerebro como información adicional para la percepción espacial. Por otra parte, al fijar la atención en un objeto, es preciso modificar el eje visual de cada uno de los ojos para que éste permanezca correctamente enfocado. Esto significa que los dos ejes visuales convergen en el punto donde está situado el objeto y por lo tanto, el ángulo de convergencia de estos dos ejes constituye otra clave para determinar su distancia. Es evidente que el ángulo que formarán los dos ejes visuales será tanto mayor cuanto menor sea la distancia entre el observador y el objeto, por lo que, en principio, es posible obtener una medida subjetiva de la distancia del objeto a partir de la convergencia entre los dos ojos.

La convergencia de los dos ejes visuales y la acomodación del cristalino aportan una valiosa información para la situación de los objetos aunque no constituye seguramente la principal responsable de la percepción tridimensional del espacio, como lo constata el hecho de que podemos situar con precisión los objetos que permanecen fuera del campo de enfoque. La información más relevante que utiliza el cerebro para componer el espacio la constituyen las diferencias existentes entre las imágenes sensadas por cada una de las retinas. El filósofo griego Euclides de Alejandría, en el año 280 a.C., fue el primero en postular que la percepción de la profundidad se producía como consecuencia de en cada ojo se formaban dos imágenes ligeramente distintas de los objetos. Sin embargo, estos postulados no fueron comprobados experimentalmente hasta el año 1600 en el que el artista Giovanni Battista della Porta experimentó con varios dibujos para crear las primeras imágenes tridimensionales artificiales.

La formación de las imágenes en el espectador se esquematiza en la figura 1.33, donde se indican las disimilitudes entre las proyecciones de los puntos de una escena en las dos retinas. En este ejemplo se

supone que el espectador mantiene enfocado el punto F, por lo que los dos ejes visuales coinciden en este punto. El punto B, situado a la misma distancia del espectador, se proyecta en la retina derecha a una distancia del punto F ligeramente inferior a la que queda proyectado en la retina izquierda. En el punto A, situado más cerca del espectador, las diferencias entre cada una de las proyecciones se acentúan. Estas diferencias entre las dos imágenes permitirán situar la posición del objeto. El cálculo de la posición de los objetos a partir de sus proyecciones en la retina es un complejo problema de trigonometría que, sorprendentemente, el cerebro realiza de forma automática, resolviendo además las posibles ambigüedades que pudieran surgir de un cálculo meramente matemático por su interpretación física más plausible. Este análisis supone realizar un reconocimiento de formas previo para identificar los objetos y puntos característicos de la escena y establecer cuáles se corresponden uno a uno en las dos proyecciones. La solución al problema no siempre es única, por lo que en algunos casos surgen ambigüedades que el cerebro debe interpretar de la forma más coherente. Cuando la interpretación de la imagen no se corresponde con los objetos, sus posiciones o movimientos reales decimos que se trata de ilusiones ópticas.

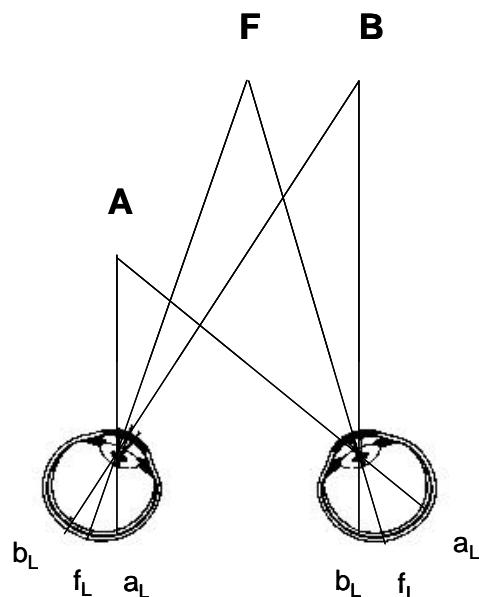


Fig. 1.33 Proyección de los objetos en las retinas de un observador

En cualquier caso, la percepción tridimensional está íntimamente ligada con las disimilitudes entre las imágenes sensadas en cada retina, por lo que cualquier sistema que pretenda reproducir las sensaciones tridimensionales deberá proporcionar al espectador esta información. Muchos de los sistemas de representación de imágenes tridimensionales se basan en separar, por algún procedimiento óptico, mecánico o electrónico dos imágenes correspondientes a dos puntos de vista distintos que se presentarán a cada ojo independientemente. Estos sistemas se conocen como estereoscópicos y son probablemente los más populares.

1.5.3 Sistemas estereoscópicos

La alternativa más utilizada para proporcionar al espectador la sensación de imágenes en tres dimensiones consiste en presentar las proyecciones planas de la escena que se obtendrían en dos puntos del espacio situados a una distancia parecida a la que separa los dos ojos (unos 6,5 cm). Para conseguir el efecto de tres dimensiones, el sistema reproductor deberá presentar al ojo derecho la imagen que se ha captado con la cámara de la derecha y al ojo izquierdo la obtenida con la cámara izquierda. La mayoría de sistemas requieren que el espectador utilice algún tipo de lentes u otros elementos ópticos que actúen como filtros separadores de las imágenes. Con ello, las imágenes proyectadas en la retina coincidirán con las que hubieran incidido en el espectador si hubiera estado en la posición de la cámara. De hecho, la sensación de espacio conseguida con esta técnica es muy espectacular y sus principios han venido utilizándose desde hace bastantes años. Actualmente, con la posibilidad de sintetizar imágenes mediante ordenador es posible presentar al espectador puntos de vista y ángulos de la escena que serían impensables en una situación real.

El primer sistema comercial para la visión de imágenes en tres dimensiones se popularizó alrededor de 1900 y estaba basado en el estereoscopio de Brewster, que había aparecido en 1847. Este equipo era un visor para uso personal que cargaba dos pares de fotografías en un sistema óptico-mecánico que permitía que cada ojo viera una imagen distinta. La fotografía estereoscópica tuvo un considerable éxito durante el primer cuarto de siglo, vendiéndose un número importante de estereoscopios y material fotográfico asociado (sobre todo colecciones de fotografías). Sin embargo, después de la segunda guerra mundial, este mercado experimentó una importante caída y prácticamente desapareció, ya que el montaje de los pares fotográficos era bastante tedioso y sólo podían ser visualizados por una única persona, por lo que no pudo competir con otras formas de entretenimiento, de gran auge, como el cine y la televisión.

En la década de los 50 se realizaron algunas producciones cinematográficas que utilizaban sistemas estereoscópicos que podían ser observados por varios espectadores simultáneamente. No obstante, sus efectos parecían poco reales y, en consecuencia, su éxito comercial fue relativamente bajo. El problema se debía a que el procedimiento utilizado para proyectar las dos imágenes en cada una de las retinas del espectador no estaba suficientemente depurado y había un importante cruce de imágenes.

El primer sistema se comercializó con el nombre de *Anaglyph* y requería que el espectador utilizase unas gafas con un cristal rojo y uno verde (que posteriormente fue azul). Las imágenes se habían obtenido mediante dos cámaras (o una cámara con dos lentes) y se proyectaban superpuestas en la pantalla. Una de las imágenes se proyectaba con una tonalidad dominante verde y la otra con una tonalidad roja, de manera que si la pantalla se observaba sin gafas se distinguía claramente que las dos imágenes estaban desplazadas. Cuando el espectador se colocaba las gafas, el cristal rojo oscurecía la imagen con esta tonalidad y dejaba pasar la verde y viceversa con el cristal verde. Por lo tanto, la señal que pasaba a través de cada ojo se correspondía con la que había captado la cámara correspondiente. En principio cada una de las imágenes eran adquiridas en blanco y negro y se proyectaban en la pantalla con el filtro correspondiente. La imagen percibida por el espectador induce sensaciones de color artificiales por cuanto en una retina se activan los conos de color rojo mientras en la otra los de color verde. El problema de este sistema es que los colores que inciden en cada ojo tienen componentes de color distintas y se producen desequilibrios de intensidad entre las dos imágenes, lo que confunde y dificulta la concentración en el espectador. Además, los filtros utilizados

en las gafas solían ser muy económicos y no separaban las dos imágenes correctamente. En algunos casos, sobre todo en fotografía, se ha intentado generalizar el sistema *Anaglyph* para percibir una sensación de color más real proyectando imágenes en color y procesándolas con filtros adecuados. No obstante, debido al filtrado de colores inherente al sistema, no pueden reproducirse todos los colores y es sumamente difícil establecer el balance de colores adecuado para que cada fotograma pueda ser visualizado con un mínimo de error de color. En la figura 1.38 se muestra un fotograma para ser visualizado mediante este sistema. El propósito de esta ilustración es que el lector aprecie el orden de magnitud de las diferencias entre las dos imágenes que se presentan a la retina. En el caso de que se disponga de una versión en color de la fotografía podrá observarse el efecto estereoscópico utilizando un filtro rojo en el ojo izquierdo y un filtro azul en el derecho ya que la imagen para la retina derecha está codificada en rojo (el filtro elimina la componente azul) y la de la izquierda en azul (el filtro elimina la roja).



Fig. 1.38 Fotograma de una escena tridimensional en el sistema *Anaglyph*

Posteriormente se utilizó el sistema *Polaroid* que se basaba en grabar dos películas con dos lentes que polarizaban la luz en ángulos opuestos. Las imágenes se superponían posteriormente en la pantalla y el espectador debía observarlas con unas gafas que incluían unas lentes polarizadoras de características similares a las utilizadas en la grabación de la película. Esta tentativa tampoco tuvo inicialmente excesivo éxito comercial y, al igual que el sistema *Anaglyph*, fue considerada como una curiosidad tecnológica utilizándose esporádicamente en algunos museos o ferias.

Estos sistemas tuvieron un importante impulso a finales de la década de los 80, en los que se mejoraron tanto los sistemas de separación de las imágenes como la precisión en el posicionamiento de las imágenes en pantalla. Actualmente existen diversos teatros especiales en donde se proyectan películas y documentales registrados en formatos parecidos y que tienen un considerable éxito comercial.

En general, estos sistemas de proyección utilizan dos filosofías distintas para la separación de las imágenes que, en ambos casos, requieren que el espectador utilice gafas especiales. El primer sistema deriva directamente del sistema Polaroid y utiliza filtros de polarización de luz verticales y horizontales. En la pantalla se proyectan simultáneamente las imágenes captadas por dos cámaras, una de las cuales se pasa a través de un filtro de polarización horizontal y la otra por uno vertical. El espectador lleva unas gafas en la que en cada ojo hay un filtro polarizador distinto, con lo que sólo se observa la imagen que le corresponde, produciéndose el efecto estereoscópico deseado. La evolución en la fabricación de materiales plásticos polarizadores de bajo coste ha permitido que este sistema proporcione una calidad aceptable y un reducido nivel de cruce entre las imágenes.

La alternativa a este procedimiento es un sistema en el que las imágenes correspondientes a cada ojo se presentan de manera alternativa al espectador, que está equipado con unas gafas de cristal líquido que actúan a modo de obturador. El cristal líquido es un material que puede ser opaco o transparente, dependiendo de la tensión externa que se aplique. Las gafas están equipadas con un controlador electrónico que aplica las tensiones apropiadas para que uno de los cristales sea opaco mientras el otro es transparente, y viceversa. El principio de funcionamiento del sistema se ilustra en la figura 1.35. Cuando se proyecta la imagen correspondiente a la cámara derecha, el cristal correspondiente al ojo derecho será transparente, mientras el del ojo izquierdo permanecerá opaco, de manera que esta imagen sólo será capturada por la retina derecha. La situación se invierte cuando se proyecta la imagen correspondiente a la cámara izquierda. Las imágenes se intercalan a una velocidad suficientemente alta para que el observador no pueda apreciar la naturaleza discontinua de la información que se presenta a cada ojo. Las gafas LCD que lleva el espectador están equipadas con un receptor de infrarrojos o de radio para que el sincronismo entre cada una de las imágenes y los obturadores de las gafas sea perfecto y no existan cruces entre imágenes. La percepción del espacio tridimensional en este sistema es bastante buena por cuanto las imágenes quedan correctamente separadas y no aparecen cruces.

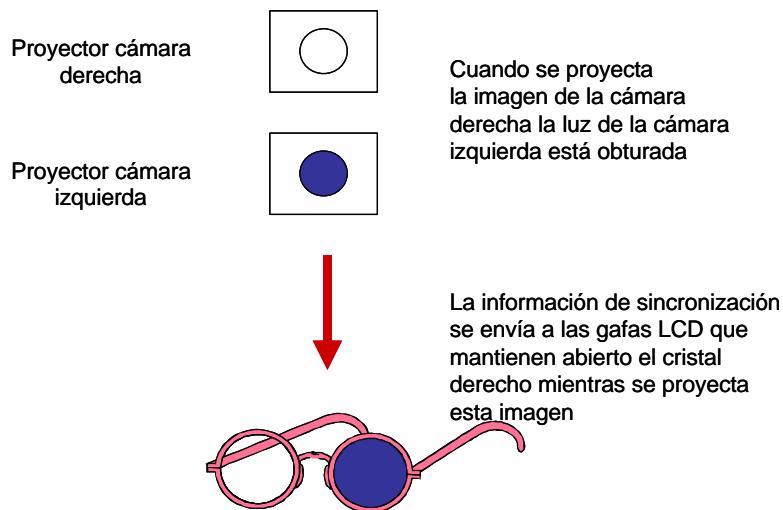


Fig. 1.35 Principio de funcionamiento de los sistemas estereoscópicos con obturadores de cristal líquido. Las imágenes proyectadas en la pantalla y las gafas están sincronizadas

El sistema de exhibición más popular es probablemente el IMAX-3D para el que existen bastantes salas de exhibición y que se basa en un formato de película de grandes dimensiones. La empresa IMAX, de origen canadiense, da soporte para la instalación de salas de exhibición y tiene un servicio de alquiler de cámaras y de equipos para la postproducción de películas en este formato. El procedimiento para la separación de las imágenes puede ser, dependiendo del teatro, mediante gafas polarizadas o mediante obturadores. En el IMAX-3D del Port Vell de Barcelona se utiliza un sistema basado en gafas polarizadas. La cámara requiere utilizar dos negativos cuyas imágenes se corresponden con las que se presentaran a cada uno de los ojos. Cada fotograma tiene una superficie tres veces superior a la del formato de 70 mm utilizado en el cine convencional por lo que la calidad de las imágenes es muy buena y puede proyectarse en pantallas de grandes dimensiones. Los fotogramas correspondientes a las dos películas se toman de manera sincronizada mediante una cámara especial (ver figura 1.36) que, junto con las lentes y la película, tiene un peso aproximado de 100 Kgs. La carga de la película en la cámara supone un laborioso proceso que requiere el empleo de un mínimo de 30 minutos. Además, cada carga sólo permite registrar un máximo de tres minutos de película tras lo cual debe repetirse el procedimiento de carga de nuevo celuloide.



Cámara IMAX 3D

Proyector IMAX 3D

Fig. 1.36 Cámara y proyector en el formato IMAX-3D

El proyector del sistema IMAX-3D intercala los fotogramas correspondientes a cada película en la pantalla. Las dos películas están arrolladas en carretes de grandes dimensiones para que puedan tener duraciones largas sin necesidad de recargarlas. Las películas se desplazan en la dirección horizontal, ya que el gran tamaño de los carretes dificultaría un desplazamiento vertical como el usado en el cine convencional. El proyector contiene dos lámparas, una para cada película, con potencias de alrededor de 15000 Vatios que deben ser enfriadas por agua. En la figura 1.36 puede observarse una fotografía de un proyector IMAX-3D donde se aprecia el gran tamaño del equipo y los carretes contenedores de la película en sentido horizontal.

La idea general de estos sistemas estereoscópicos para fotografía y cine ha sido trasplantada a otros entornos que comprenden diversas aplicaciones informáticas, prototipos de sistemas de televisión tridimensionales, video-juegos, realidad virtual, etc. Los dos procedimientos más usados para la

separación de las imágenes siguen siendo las gafas polarizadas y las de obturación de imagen mediante cristal líquido. Ambos sistemas ofrecen una calidad comparable y, por supuesto, permiten una codificación del color real que no era posible con el sistema *Anaglyph*. En los dos casos se produce una ligera pérdida de luminosidad puesto que los cristales utilizados sólo dejan pasar una fracción de la luz incidente ya que no son completamente transparentes. El cruce entre imágenes, es decir, la parte de energía que recibe el ojo derecho de la imagen correspondiente al ojo izquierdo o viceversa, es algo menor en el caso de las gafas de cristal líquido, aunque es tolerable, y siempre dependerá de la calidad del filtro, en las polarizadas.

La principal ventaja del sistema de obturación electrónica de las imágenes es que puede adaptarse fácilmente para usarse con monitores convencionales. En este caso, las imágenes correspondientes a cada ojo se alternan en la pantalla y se sincronizan con los obturadores de cristal líquido, con lo que puede obtenerse un sistema de representación en tres dimensiones modificando únicamente la secuencia de imágenes que se presenta al usuario y sin necesidad de añadir ningún tipo de *hardware* adicional. Es un sistema que puede adaptarse a sistemas de televisión o equipos informáticos sin más que adecuar la señal de vídeo para que proporcione alternativamente las imágenes correspondientes a cada ojo e incorporar un subsistema que permita la correcta sincronización de los obturadores. El principal inconveniente de esta tecnología es que al *multiplexar* temporalmente los dos canales, se reduce la frecuencia de imagen por lo que puede llegar a producirse un parpadeo significativo. Este efecto es tanto más notorio cuanto menor sea la frecuencia de actualización de imágenes en la pantalla, por lo que, si bien se adapta sin excesivos problemas a equipos informáticos con velocidades de refresco altas, plantea serias dificultades para su implantación en sistemas de televisión convencionales.

En el ámbito de los sistemas de televisión en tres dimensiones se han realizado algunos prototipos que combinan dos señales de vídeo, una para cada ojo, y que presentan las imágenes alternadas en un tubo de rayos catódicos convencional, con lo que se consigue mantener la frecuencia de refresco de imagen. A pesar de que esta filosofía requiere el uso de receptores específicos para combinar la información de los dos canales en imágenes estereoscópicas, estos sistemas serían compatibles con la TV convencional, ya que bastaría con sintonizar el canal correspondiente a uno de los ojos para que el usuario que no tuviera un receptor 3D pudiera ver las imágenes en formato convencional. La señal de vídeo del otro canal estereoscópico puede usarse para introducir la información necesaria para mantener el sincronismo de los obturadores con el monitor. Sin embargo, se supone que ninguno de estos prototipos se convertirá en un estándar por cuanto, en cualquier caso, imponen el uso de gafas especiales en el espectador, lo que se suele considerar como una restricción importante para que los sistemas tengan una buena aceptación.

Actualmente se comercializan distintos sistemas basados en obturadores de cristal líquido para aplicaciones informáticas. La integración de estos sistemas en un equipo informático es relativamente simple, y afecta, casi exclusivamente, al *software*. El sistema *VR-SimulEyes*, uno de los más populares, únicamente requiere, como se indica en la figura 1.37, conectar una pequeña extensión de *hardware* adicional en la salida del conector VGA a la que también permanece conectado el monitor. Los pares de imágenes estereoscópicas se envían secuencialmente al monitor insertando en una de las líneas visibles un código que indica a qué ojo corresponde cada imagen. Para sincronizar los obturadores con la imagen de la pantalla basta con reconocer este código y comutar las tensiones aplicadas al cristal líquido. De ello se encarga el *hardware* adicional, que está integrado en una caja

externa de reducidas dimensiones y que proporciona la señal de control de las gafas. Este *hardware* dispone de varias salidas para que puedan conectarse simultáneamente varios usuarios. La sensación de profundidad y la percepción espacial es muy aceptable y el parpadeo, si el ordenador es rápido, es poco perceptible. Si acaso, se observa sobre todo en los objetos exteriores a la pantalla y siempre puede reducirse o anularse oscureciendo la habitación. Existen otros productos de características y prestaciones parecidas y, probablemente, el verdadero problema para su implantación definitiva en el mercado es la competencia que ejercen entre ellos y la carencia de productos *software* genéricos para ellos. La mayoría de fabricantes se han dedicado a establecer colaboraciones con compañías de *software* para vídeo-juegos produciendo versiones de programas comerciales de éxito que puedan adaptarse al formato tridimensional. Nuevamente, el citado *Doom* es probablemente el programa del que se han creado más versiones estereoscópicas. Sin embargo, aún existen pocos programas de aplicaciones para trabajar con estos sistemas.

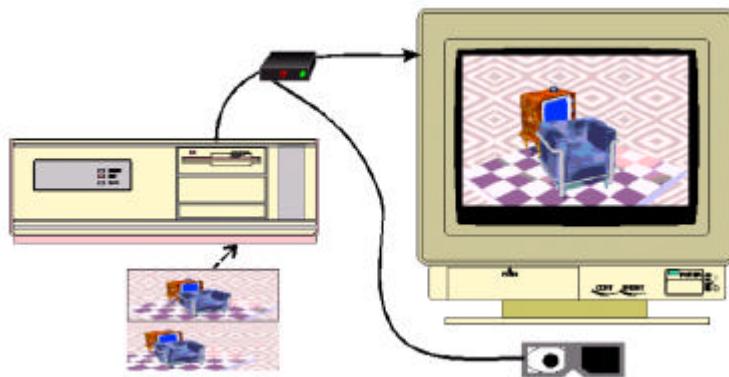


Fig. 1.37 Integración de un sistema de gafas de cristal líquido en un equipo informático

La empresa VRex dispone de una completa gama de productos para la producción de vídeo y programas informáticos en formato estereoscópico para aplicaciones domésticas y profesionales. Los elementos básicos para la presentación son los displays LCD y los proyectores que en ambos casos deben visualizarse con gafas polarizadas. La orientación principal de estos productos es la de ofrecer un soporte global para las presentaciones de empresa o docentes y la publicidad. Las cámaras son de dimensiones y peso reducidos y proporcionan las imágenes sensadas por cada uno de los grupos ópticos en una señal de vídeo NTSC o PAL convencionales en la que en las líneas pares están codificadas las imágenes correspondientes a uno de los canales y en las impares el otro. La resolución de imagen que se apreciará con cada uno de los ojos es, por tanto, la mitad de la televisión convencional, aunque este efecto no es excesivamente notorio. La principal ventaja de este sistema es que proporciona una salida estándar que puede grabarse en un magnetoscopio de cualquier formato o editarse en cualquier equipo.

Como proyector se utiliza un LCD color de alta definición en el que, sobre su superficie, están montados unos microfiltros de polarización (μ Pol) que proporcionan a la luz emitida por las líneas pares y las impares una polarización lineal o circular. De este modo, si las gafas del espectador tienen los filtros de polarización adecuados, cada uno de los ojos verá exclusivamente las líneas pares o las

impares. Estos filtros μ Pol también pueden adaptarse a ordenadores portátiles, aunque el proceso de instalación es complejo y debe realizarse en centros especializados. La tecnología actual permite que la anchura de los filtros sea de 201 μ m con lo que puede adaptarse a displays TFT de hasta 1280 × 1024 píxeles. Los filtros μ Pol también pueden adaptarse para producir copias impresas de imágenes tridimensionales. En este caso, la distribución de los polarizadores puede ser bidimensional resultando en una mejor imbricación entre los dos pares estereoscópicos.

Uno de los inconvenientes del sistema es que no puede usarse con tubos de rayos catódicos, ya que estos no ofrecen suficiente precisión para adaptar los microfiltros en la pantalla. Para la síntesis de gráficos 3D por ordenador se suministra un *software* específico que permite la edición de imágenes y finalmente intercala el contenido de los dos canales en las líneas de la pantalla.

Visores integrados

La mayoría de máquinas para vídeo-juegos de realidad virtual o simuladores profesionales utilizan la misma filosofía de los sistemas estereoscópicos, aunque el procedimiento utilizado para la separación de las imágenes es, en ocasiones, algo más sofisticado. En estos casos, suele incorporarse un visor o casco en la cabeza del usuario que incorpora dos *displays* de cristal líquido independientes, uno para cada ojo, de modo que la separación de dos imágenes es perfecta y ambas tienen una buena definición.

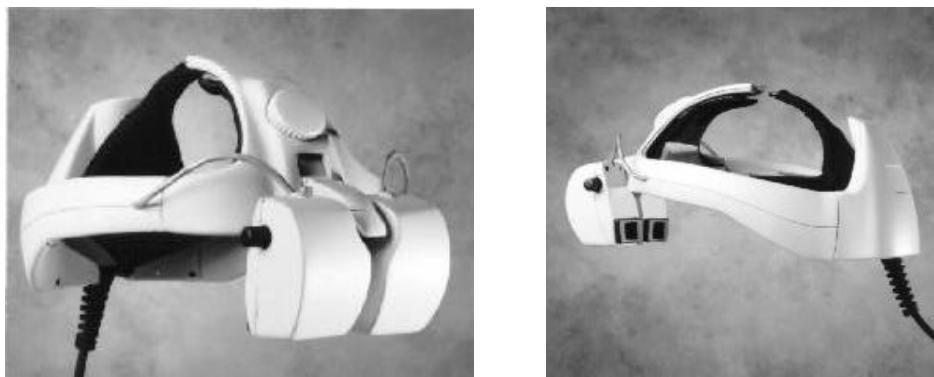


Fig. 1.38 Vistas frontal y posterior de un sistema de visión estereoscópica integrado (ProView 30)

Este tipo de sistemas se conocen como HMD, que corresponde a las siglas *Head Mounted Displays*. En algunos casos el visor puede incorporar sensores de movimiento y posición que determinan la dirección a la que el usuario dirige la cabeza e indican al ordenador que genera las imágenes que realice los cambios adecuados sobre el punto de vista de la escena de realidad virtual. En este último es fundamental que el ordenador realice los cambios del punto de vista a gran velocidad, ya que los retardos entre el movimiento de la cabeza y la actualización de las imágenes producirán una sensación de descoordinación del movimiento que inevitablemente puede conducir a la aparición de mareo y náuseas. Este tipo de sistemas tienen un elevado precio y suelen usarse en aplicaciones profesionales de carácter científico (microscopía 3D), industrial (inspección, control de procesos, CAD/CAM) o de

realidad virtual. En la figura 1.38 se presentan dos vistas de uno de estos sistemas que puede aceptar una o dos señales (para trabajar en modo mono o estéreo) tanto en formato VGA como vídeo NTSC o PAL. El sistema incorpora un sistema óptico que admite diversos ajustes para evitar la fatiga (los LCD no podrían enfocarse correctamente si no existiera óptica adicional, ya que están muy próximos a la retina). Los *displays* son del tipo TFT de color y tienen una resolución de 640×480 proporcionando un campo de visión de $24^\circ \times 18^\circ$ (Horizontal \times Vertical).

1.5.4 Sistemas autoestereoscópicos

Todos los sistemas que hemos descrito hasta el momento requieren utilizar unas gafas especiales, lo que sin duda introduce cierto grado de incomodidad en el espectador. Los denominados sistemas autoestereoscópicos continúan con la misma filosofía de presentar al usuario una visión binocular de la escena, pero sustituyen el proceso de separación de las imágenes que se realiza en las gafas por un filtrado óptico que se produce en la pantalla del propio dispositivo de presentación. En esta sección describiremos las ideas y características fundamentales de estos sistemas, de los que existen múltiples variantes tecnológicas y cuya descripción detallada excede nuestros objetivos.

Una posible configuración básica de uno de estos sistemas se esboza en la figura 1.39, donde se representa una escena genérica que es captada por dos cámaras, de modo análogo a como se realizaría en un sistema estereoscópico convencional. Las imágenes obtenidas por estas cámaras son procesadas y se combinan en una secuencia de imágenes en las que se intercalan los píxeles correspondientes a cámara, tal y como se indica esquemáticamente en la parte superior derecha de la figura 1.39. Cada una de las imágenes que se presentarán al espectador tiene una estructura de barras verticales estrechas en las que se alternan cada uno de los canales binoculares. Esta distribución queda indicada por los caracteres A y B que representan la cámara a la que corresponde cada elemento de imagen. La distancia horizontal entre dos elementos correspondientes a una misma cámara se denomina *pitch*.

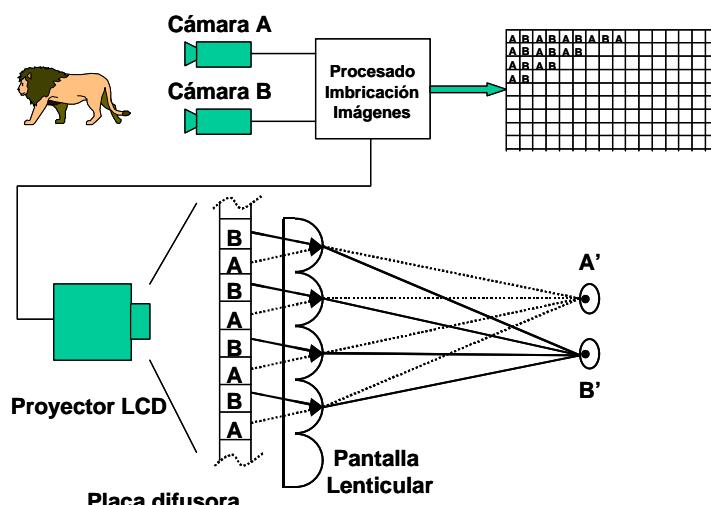


Fig. 1.39 Principios de formación de imagen en sistemas autoestereoscópicos

El resultado de esta imbricación de las cámaras se proyecta mediante un *display LCD* de alta resolución a la parte posterior de una pantalla que está compuesta por una placa difusora translúcida y una pantalla lenticular. Los detalles de su constitución física se ilustran sólo para una pequeña porción de la pantalla, en la que se ha aumentado considerablemente la escala, para poder apreciar con claridad el principio de la formación de imágenes. La pantalla lenticular está situada con gran precisión sobre la placa difusora y actúa como un elemento óptico que direcciona la luz procedente de los píxeles correspondientes a las cámaras A y B hacia los puntos A' y B', sobre los que se situará el espectador. De este modo, se produce una sensación de espacio tridimensional, ya que cada ojo sólo observa la imagen captada con la cámara que le corresponde.

La proyección de la imagen sobre una placa difusora puede sustituirse por una presentación directa sobre un LCD color del tipo TFT como los usados en ordenadores personales, o incluso, se han realizado algunos prototipos, sobre tubos de rayos catódicos. En estos últimos, sin embargo, se requiere que la deflexión del haz sea extremadamente lineal y que la distorsión geométrica de la pantalla sea muy baja para poder posicionar correctamente los elementos correspondientes a cada imagen sobre la pantalla lenticular. Aún así, por las limitaciones propias en la distorsión de los mejores tubos de rayos catódicos, la resolución obtenida es inferior a la que se obtiene con los modernos LCD. También es posible, utilizando pantallas lenticulares con una película de aluminio de alta reflexión, utilizar un sistema de proyección frontal de las imágenes.

Las pantallas lenticulares están formadas por pequeñas lentes cilíndricas distribuidas verticalmente y situadas una al lado de la otra, y aunque su uso para sistemas de vídeo autoestereoscópicos es relativamente reciente, se han venido usando desde hace años para otras aplicaciones. Probablemente, uno de los productos más conocidos que utilizan pantallas lenticulares son unas postales, que se popularizaron en la década de los 70 y que dependiendo del punto de vista desde el que son observadas representan escenas distintas. Otra aplicación típica es usarlas como pantalla para la proyección de cine o diapositivas para un grupo reducido de espectadores. Su principal ventaja respecto a otro tipo de pantallas es que son muy directivas y reflejan casi toda la luz incidente en la dirección del proyector, por lo que si los espectadores se sitúan entre la pantalla y el proyector, observan la imagen con mucha luminosidad. Sin embargo, esta alta directividad las hace inútiles en aplicaciones como el cine comercial, en las que hay espectadores que observan la pantalla desde los laterales. En la aplicación que estamos considerando, la pantalla lenticular debe estar colocada con mucha precisión respecto al sistema de presentación de las imágenes y la distancia entre las lentes debe coincidir con el *pitch* o distancia entre píxeles correspondientes a una misma imagen.

El inconveniente principal del sistema que hemos esbozado parece bastante claro: el espectador debe situarse en una determinada posición respecto a la pantalla y cualquier desplazamiento respecto a la misma representará una reducción considerable en el nivel de la imagen, con pérdida eventual de la visión binocular. Esto representa un serio problema, seguramente más grave que el uso de gafas y se han propuesto varias alternativas.

Una posible solución, adecuada para aplicaciones informáticas con un único usuario, está siendo desarrollada en el Instituto Heinrich Hertz de Berlín y se basa en una sencilla variación del sistema descrito anteriormente. La idea básica del sistema es controlar la posición de la pantalla lenticular para que cualquier movimiento del espectador se corresponda con un movimiento de la pantalla que provoque que los nuevos puntos de convergencia de las imágenes coincidan con la nueva posición de

los ojos. Para ello, se permite que la pantalla lenticular pueda desplazarse en la dirección horizontal y acercarse o alejarse ligeramente del espectador (el movimiento vertical no es necesario puesto que la pantalla es uniforme en esta dirección). Para determinar la posición de los ojos del espectador se utiliza un sistema de reconocimiento automático de formas basado en adquirir, mediante una cámara auxiliar, una imagen de los ojos del espectador, que se utiliza para realizar un seguimiento en tiempo real de sus movimientos. El sistema de reconocimiento usado actualmente tiene un retardo aproximado de unos 120 ms, por lo que es necesario utilizar un algoritmo de predicción de posición para intentar reducir este retardo.

En aplicaciones en las que se supone que pueden coexistir varios espectadores, el sistema deberá proporcionar imágenes que sean visibles simultáneamente desde distintas ubicaciones. La solución adoptada para resolver este problema consiste en captar la escena con un mayor número de cámaras, de modo que cuando el espectador se sitúe en distintas posiciones observe las imágenes captadas por cámaras distintas. En la figura 1.40 se representa esta idea para un caso simplificado en el que intervienen únicamente cuatro cámaras. El sistema representado sigue un esquema parecido al de la figura 1.39, en el que las imágenes captadas por las distintas cámaras se entrelazan espacialmente en una única imagen de acuerdo con un patrón vertical. Ahora, es posible obtener un conjunto de puntos, distribuidos en el área donde se supone que se situarán los espectadores, desde donde pueden visualizarse las imágenes correspondientes a distintas cámaras.

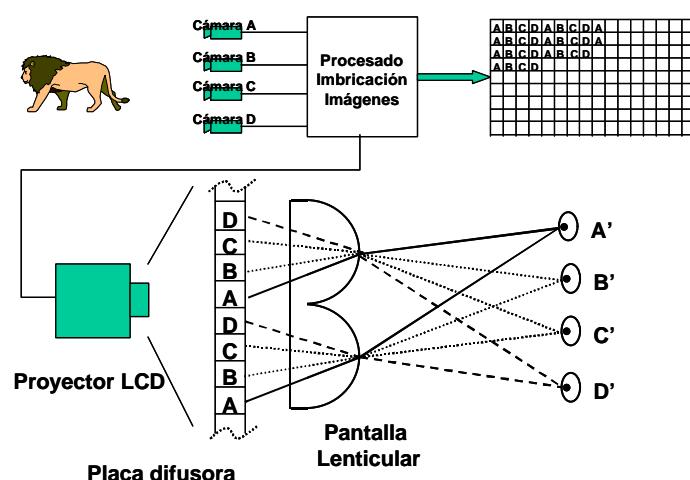


Fig. 1.40 Sistema autoestereoscópico para múltiples usuarios

Es evidente que si la resolución del proyector LCD se mantiene, al tener que multiplexar un mayor número de cámaras, el número de elementos de imagen asociados a cada una se reduce con lo que aparece una pérdida de resolución en la imagen tanto mayor cuanto mayor sea el número de cámaras implicadas. Como alternativa puede pensarse en usar un sistema basado en el uso de múltiples retroproyectores cuyo principio de funcionamiento es muy parecido al que hemos expuesto. En vez de utilizar un único proyector en el que se imbrican todas las imágenes se utiliza un proyector para cada secuencia de vídeo que ha sido capturada por una cámara distinta. La pantalla de retroproyección está

formada por dos pantallas lenticulares, una posterior y otra frontal, entre las que se encuentra una placa difusora. La pantalla lenticular posterior descompone las imágenes procedentes de los proyectores sobre la placa difusora en una componente para cada cámara, y posteriormente, la pantalla lenticular frontal realiza la operación inversa. No obstante, esta alternativa tecnológica encarece notablemente el producto, sobre todo si el número de cámaras es grande, pero la resolución dependerá exclusivamente de la calidad de los proyectores y de la pantalla lenticular.

El sistema de la figura 1.40 introduce una importante novedad respecto a los métodos binoculares que habíamos considerado hasta este momento, y es que el punto de vista de la escena depende de la posición del espectador. En efecto, cuando el espectador se sitúa entre los puntos A' y B', ve las imágenes que han sido captadas por estas cámaras, obteniendo la percepción del espacio que corresponde a esta posición. Cuando se desplaza a una posición en la que observe las imágenes correspondientes a las cámaras B' y C' percibirá un cambio de perspectiva de la escena debido a que estas dos cámaras están situadas en un punto de vista distinto. Las características del sistema de representación han cambiado drásticamente, ya que ahora no sólo percibimos la sensación de espacio tridimensional, sino que también experimentamos un cambio de perspectiva de los objetos al desplazarnos dentro del recinto en el que se nos presentan las imágenes.

Atendiendo a la posibilidad de que un desplazamiento del espectador provoque un cambio en la perspectiva de la escena, los sistemas de representación tridimensional pueden clasificarse como sistemas de *espectador pasivo* o de *espectador activo*. El primer tipo corresponde a todos los sistemas estereoscópicos que hemos considerado en la sección precedente (a excepción del visor integrado con sensores de movimiento), por cuanto en todos ellos las imágenes que se proyectan o reproducen en la pantalla son independientes de la posición del observador. Por ello, éste siempre percibirá la misma perspectiva desde cualquier posición de la sala y si se desplaza, tendrá la sensación de que la escena se desplaza con él. En otras palabras, la perspectiva que observamos es la obtenida por el par de cámaras al registrar la escena y nuestros movimientos no se corresponden con un cambio del punto de vista.

Los sistemas de espectador activo proporcionan una información más completa de la escena por cuanto no sólo nos proporcionan una imagen binocular de la misma, sino que esta información se ajusta a nuestra posición relativa en la sala. Un ejemplo característico de este tipo es el sistema autoestereoscópico que acabamos de describir. Si disponemos de distintas cámaras alrededor de la escena, separadas por una distancia igual a la separación entre los ojos, podremos proporcionar un punto de vista que depende de la posición de espectador. En un sistema de estas características, cuando el observador se desplaza experimentará cambios en el punto de vista de la escena, viendo como pueden aparecer determinadas partes de algunos objetos que, en la perspectiva previa, permanecían ocultos por otros.

En general, los sistemas de espectador activo pueden a su vez clasificarse en distintos subgrupos dependiendo del tipo de información que aporten al espectador cuando éste cambia de posición. El sistema que estamos considerando debería denominarse *parallax* de movimiento horizontal, ya que sólo aporta información de cambio del punto de vista cuando el espectador se desplaza en sentido horizontal. Nótese que debido a la simetría vertical de las pantallas lenticulares, el punto de vista de la escena no se modificará cuando el espectador se mueva en esta dirección (p.ej. agachándose). En general existen tres posibles *parallax* de movimiento: el horizontal, el vertical y el de profundidad, que se corresponden con los posibles ejes de movimiento del espectador respecto al sistema de

representación de imágenes. El *parallax* de movimiento horizontal se considera el más importante, por cuanto es el movimiento más probable del espectador. El *parallax* de profundidad es proporcionado parcialmente por casi todos los sistemas, pues si bien la perspectiva de la escena no cambia al acercarnos o alejarnos de la pantalla, el tamaño del *display* y su campo visual sí que varía, con lo que al acercarnos o alejarnos percibimos cambios significativos en el tamaño de los objetos (se aprecia principalmente en displays de dimensiones reducidas). En ocasiones es habitual en la literatura referirse al *parallax* de un sistema estereoscópico, entendiéndose como el cambio sobre el punto de vista de la escena que se obtiene al contemplar la pantalla con el ojo derecho o el izquierdo manteniendo la cabeza quieta. Es importante distinguir entre este tipo de *parallax* y el *parallax de movimiento* al que hacemos referencia en este apartado. El primero únicamente tiene en cuenta el cambio en la perspectiva debido a la posición distinta de cada uno de los ojos y es común a todos los sistemas de representación en tres dimensiones por cuanto las imágenes captadas por cada ojo deben ser distintas, mientras que el segundo considera el cambio de punto de vista debido al movimiento del observador.

Es evidente que cualquier sistema que aporte *parallax* de movimiento requiere, si se mantiene la calidad del sistema de reproducción, soportar un considerable aumento de la cantidad de información a transmitir o almacenar. El sistema que hemos descrito en la figura 1.40 no introduce ningún aumento en la cantidad de información enviada al sistema de proyección, ya que la calidad de cada una de las imágenes se ve reducida por el entrelazado de las imágenes correspondientes a cada una de las cámaras, pero la variante basada en múltiples proyectores sí que supone un considerable aumento de información y en consecuencia de ancho de banda.

Un importante problema de los sistemas autoestereoscópicos para múltiples usuarios es que la calidad con la que se observan las imágenes no es uniforme y depende de la posición del usuario. Así, cuando éste se sitúa en posiciones intermedias entre los puntos en los que teóricamente convergen las imágenes, la alta directividad de la pantalla lenticular provoca una pérdida considerable del nivel de luminosidad. Además, es posible que aparezcan cruces entre las imágenes correspondientes a distintas cámaras, perdiéndose en algunos casos el efecto tridimensional. Otra particularidad inherente al sistema es que la codificación del *parallax* de movimiento horizontal tiene una naturaleza discreta (un punto de vista por cada par de cámaras que están separadas unos 6,5 cm) por lo que, cuando el espectador se desplaza, percibirá cambios discontinuos sobre la perspectiva de la escena que producen una sensación poco real.

La fatiga ocular

Uno de los problemas comunes a todos los sistemas presentados hasta ahora es que suelen producir fatiga ocular en el espectador. Se supone que los mecanismos por los que se produce esta fatiga son debidos a la falta de consistencia entre la acomodación y la convergencia de los ojos, conceptos a los que hemos hecho referencia anteriormente. En efecto, en el mundo real, cuando centramos la atención sobre un objeto, disponemos los músculos del cristalino en la posición óptima para que éste quede enfocado correctamente y orientamos los glóbulos oculares para que los ejes visuales de los dos ojos coincidan sobre el objeto. La acomodación y la convergencia se corresponden con la posición del objeto en el espacio. Ambos ajustes se realizan sistemáticamente y sin tener conciencia de ello. No obstante, en un sistema estereoscópico, las imágenes correspondientes a cada ojo se proyectan o representan sobre una pantalla plana situada a una determinada distancia del espectador. Cuando se

separan los dos pares de imágenes, el sistema visual percibe la sensación de un espacio tridimensional en el que los objetos parecen surgir de la pantalla y encontrarse en algún punto intermedio situado entre ésta y el espectador.

La tendencia natural es, pues, la de concentrarse en determinados objetos de la escena con lo que la convergencia y la acomodación tienden a centrarse en los puntos donde aparentemente está situado el objeto. Sin embargo, la imagen sigue formándose sobre la pantalla, por lo que al acomodar los ojos al punto donde aparentemente está situado el objeto, éste aparecerá desenfocado. Para ver correctamente las imágenes y percibir la sensación tridimensional el sistema visual debe situar el punto de enfoque sobre la pantalla, aunque la posición aparente del objeto y por tanto la convergencia de los ejes visuales se produzca en una posición más cercana al espectador. Esta incoherencia entre la acomodación y la convergencia supone la aparición de una cierta fatiga ocular en el espectador que es tanto más importante cuanto mayor es la distancia entre los objetos tridimensionales y la pantalla. El problema es inherente a la propia filosofía de los sistemas estereoscópicos y en principio no tiene una solución evidente.

En las secciones siguientes consideraremos dos posibles alternativas a los sistemas estereoscópicos para la representación de imágenes tridimensionales. La principal ventaja de estos dos sistemas es que presentan al espectador prácticamente todos los tipos de información espacial que el sistema visual utiliza para la percepción del espacio, con lo que se reduce notablemente la fatiga. Estas alternativas no son las únicas que están siendo consideradas actualmente aunque, sin duda, son las que prometen resultados más alentadores y que, casi con toda seguridad, acabarán desplazando a los sistemas estereoscópicos en algunas aplicaciones.

1.5.5 Displays volumétricos

Los *displays* volumétricos son sistemas en los que se reproduce físicamente un volumen tridimensional mediante la generación de estímulos lumínicos que representan la escena. El recinto utilizado para la representación de la escena suele ser un espacio cerrado, de forma esférica, parecido al que se ilustra en la figura 1.41. La idea general consiste en, mediante algún mecanismo, generar puntos de luz en el interior del recinto que reproduzcan una imagen volumétrica virtual de la escena, de modo que el espectador disponga de prácticamente toda la información visual necesaria para la percepción tridimensional de los objetos. Los puntos de luz se denominan *voxels*, que corresponde a la abreviación de *volumetric pixel*. De acuerdo con esta filosofía, el sistema proporciona información completa sobre el *parallax* de movimiento, ya que el observador puede situarse en distintos puntos del espacio obteniendo una perspectiva distinta de la escena. Asimismo, la acomodación y la convergencia de los ojos coinciden con los puntos del espacio en los que se representa la imagen, por lo que no aparece ningún tipo de fatiga ocular. El par de imágenes estereoscópicas, responsables principales de la percepción espacial, se forman en la retina de una manera natural como consecuencia del distinto punto de vista de la escena que se obtiene con cada ojo.

La alternativa más usada en la mayoría de sistemas que se están desarrollando actualmente para la generación de los voxels dentro del recinto se representa esquemáticamente en la figura 1.42. El sistema se basa en utilizar una superficie bidimensional que se desplaza mediante rápidos movimientos periódicos barriendo todo el volumen en el que se representará la imagen. La superficie

más común suele ser una hélice que gira a gran velocidad sobre su eje según el esquema mostrado en la figura 1.42. La forma de la superficie y su naturaleza reflectante o traslúcida es muy importante, por cuanto determina las regiones que permanecerán ocultas al espectador por ser tapadas por la propia superficie. En general suele considerarse que la forma de doble hélice es adecuada para minimizar estas regiones desde cualquier posición del espectador.



Fig. 1.41 Displays volumétricos

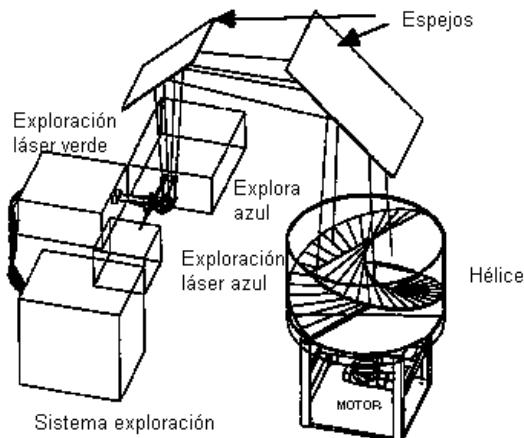


Fig. 1.42 Esquema de un display volumétrico de barrido pasivo

Los *voxels* se generan dirigiendo los pulsos de luz del láser hacia los distintos puntos del volumen cuando la hélice, que actúa como pantalla, pasa por ellos. Evidentemente, para una correcta representación de los colores es necesario usar tres rayos láser con las longitudes de onda

correspondientes a los tres colores primarios. La velocidad de la hélice debe ser suficientemente elevada para que el espectador no perciba la presencia de esta superficie. Los haces se controlan mediante moduladores acustico-ópticos, que presentan una rápida respuesta para modificar la dirección del haz utilizando un barrido aleatorio, es decir, la secuencia de puntos a los que se dirige el haz depende de la imagen a representar y sólo se barren aquellas posiciones del espacio que aportan información sobre la escena. Esto es así debido a que el número de *voxels* que deben generarse depende de la imagen que se desee producir, representando generalmente una pequeña porción del volumen total del *display*. La sincronización del sistema de exploración de los haces láser y la posición de la hélice se suele realizar mediante la lectura óptica de la posición de esta última.

Una alternativa a este procedimiento es utilizar una superficie en la que están integrados los elementos de luz, normalmente diodos electroluminiscentes (LED's), y cuya activación se controla externamente. Este procedimiento se conoce con el nombre de barrido activo y su principal ventaja es que mientras que en el sistema pasivo los *voxels* deben generarse de manera secuencial, en este caso pueden generarse simultáneamente todos los puntos de luz correspondientes a una determinada posición de la hélice. Actualmente se están considerando otras variantes que no requieren emplear una superficie de barrido y cuyo principio de funcionamiento está basado en excitar un gas fluorescente controlando el punto de excitación mediante la intersección de dos haces láser infrarrojos.

Actualmente existen diversos prototipos operativos de estos sistemas entre los que destaca un *display* de barrido pasivo desarrollado conjuntamente por el *Naval Command, Control & Ocean Surveillance Center*, RGB Technology Inc. y NEOS Technologies y que se utiliza fundamentalmente para aplicaciones de ayuda a la navegación. Este sistema utiliza una superficie helicoidal en la que se pueden representar hasta 800000 *voxels* por segundo dando un total de 40000 *voxels* por imagen si se utiliza una frecuencia de refresco de 20 Hz. Este sistema se está probando en aplicaciones de control de tráfico aéreo combinando la información de posición proporcionada por los sistemas de navegación convencional con la generación de gráficos por ordenador de los modelos de los vehículos, pudiendo llegar a representarse simultáneamente hasta 80 aviones. La generación de las imágenes por ordenador se realiza teniendo en cuenta el tipo de avión, su distancia, acimut y altura e incorporando opciones de zoom de determinadas regiones del espacio aéreo. Parece ser que este tipo de representación tridimensional simplifica notablemente la tarea del controlador y está siendo bien aceptado.

Otras posibles aplicaciones en las que parece que estos sistemas pueden tener un considerable éxito son la navegación submarina y la representación de imágenes médicas. En ambos casos se trata de utilizar el sistema para representar información de carácter tridimensional que normalmente se representa en *displays* planos. En el caso de la navegación submarina, el sistema utiliza la información proporcionada por el sonar para representar los distintos objetos detectados, su movimiento y posición relativa, la presencia de minas o torpedos, el suelo marino, etc. En las aplicaciones médicas se pretende que el sistema combine la información bidimensional obtenida por distintos sistemas de escáner del cuerpo humano para obtener imágenes tridimensionales de los órganos y partes de interés.

Si bien en las aplicaciones que hemos citado parece ser que estos sistemas pueden llegar a implantarse con éxito, es improbable que su uso se extienda a la televisión u otros sistemas parecidos, debido en primer lugar a un problema de carácter tecnológico, pero sobre todo porque parece que será un formato en el que será difícil de aceptar la presencia de actores humanos. Desde el punto de vista tecnológico el problema principal radica en que parece difícil que a corto o medio plazo puedan

aparecer sistemas de exploración tridimensional que aporten la información volumétrica necesaria para codificar una escena real. Todos los sistemas desarrollados hasta ahora se basan en la síntesis por ordenador de imágenes artificiales y en el mejor de los casos combinan fragmentos reales con imágenes sintéticas usando complejos algoritmos que requieren un tiempo de ejecución elevado.

Sin embargo, el problema principal es que la reproducción de una escena en este tipo de *displays* acentúa considerablemente el denominado *efecto muñequito* (*puppet effect*). Este efecto se presenta en general en cualquier sistema de representación tridimensional y consiste en que, para dos proyecciones de un objeto con dimensiones parecidas, la sensación de volumen y tamaño que inducen en el espectador es mucho mayor en las imágenes que se presentan con sistemas 2D que en las de 3D. No obstante, este efecto no es aún bien conocido y, probablemente, para un análisis riguroso del mismo será necesario que previamente se popularicen y perfeccionen los sistemas tridimensionales.

El problema se acentúa notablemente en los *displays* volumétricos, ya que el recinto utilizado para la representación suele ser pequeño y se enfatiza la percepción de que los actores parecen ser pequeños muñecos en un teatro de marionetas. Además, en función del punto de vista, los objetos pueden tener un cierto grado de transparencia, dejando patente la naturaleza lumínica mediante la que han sido generados y provocando una cierta sensación de ‘espectros’ en el espectador.

1.5.6 Hologramas y *displays* holográficos

El holograma de un objeto o escena es un registro plano, realizado sobre negativos fotográficos de alta resolución, de la interferencia producida entre dos haces de luz coherentes cuando la luz de uno de los haces se refleja en este objeto. Iluminando adecuadamente el registro fotográfico de la interferencia, puede reproducirse una imagen del objeto que induce en el espectador una sensación de espacio tridimensional.

El principio óptico del holograma fue propuesto por el científico húngaro D. Gabor en 1948 como un método para mejorar la resolución del microscopio electrónico. Sin embargo, esta técnica sólo cobró interés práctico a partir del desarrollo del láser en los inicios de los años 60. E.W. Leith, J. Upansieks y C.W. Stroke fueron los primeros en utilizar el holograma para el registro de imágenes en tres dimensiones utilizando fuentes de luz láser. D. Gabor recibió el premio Nobel de física por sus aportaciones teóricas a las técnicas holográficas en 1971.

El procedimiento para el registro de un holograma se realiza siguiendo un esquema como el representado en la figura 1.43. La luz emitida por un láser se descompone en dos haces mediante el uso de un espejo semitransparente. Uno de los haces, denominado haz de referencia, ilumina directamente el negativo fotográfico. El haz objeto ilumina el objeto o escena de interés y la luz reflejada y difractada por éste se dirige al negativo, donde se superpone con la luz del primer láser. La superposición entre los dos haces sobre el negativo fotográfico produce la impresión de una trama de franjas de interferencia cuyo aspecto visual no tiene ningún parecido con la imagen original. Sin embargo, sobre el negativo ha quedado registrada información de amplitud y fase sobre la escena, debido a que las franjas de interferencia indican cómo se han superpuesto en cada punto las luces procedentes de ambos haces. Así, cuando los dos haces están en fase se registra un máximo de amplitud en el negativo, mientras que cuando están en contrafase se registra un mínimo. Los niveles

de amplitud también dependen de la porción de energía que haya sido reflejada o difractada por el objeto obteniendo un patrón interferente que aporta información sobre el volumen de la escena y que puede ser recuperada mediante un proceso, que se representa en la figura 1.44, parecido al usado para la obtención del holograma. La idea general para la reconstrucción de la imagen tridimensional consiste en iluminar el registro holográfico mediante un haz láser de las mismas características que el haz de referencia. Un observador que mire a través del holograma verá, en su interior aparente, una imagen tridimensional del objeto original. Esta imagen es real (se forma también una virtual que no consideraremos), lo que significa que puede ser enfocada por sistemas ópticos para ser fotografiada con una cámara convencional o inspeccionada por un microscopio. De hecho, un método habitual para examinar cavidades dentro de un objeto que no permiten la inserción de sistemas ópticos consiste en realizar primero un holograma del mismo para posteriormente inspeccionar las zonas de interés, a la profundidad deseada, utilizando la imagen del objeto. La imagen real aparece con una curvatura invertida (pseudoscópica), aunque este problema puede solventarse realizando un segundo holograma que se registra utilizando como objeto la imagen del primer holograma.

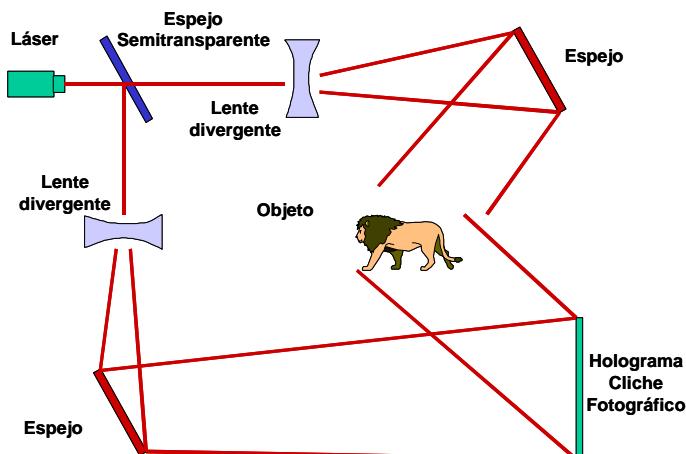


Fig. 1.43 Diagrama simplificado del registro de un holograma

Al tratarse de una representación tridimensional del objeto, el punto de vista de la imagen varía dependiendo de la posición del observador, proporcionando información de *parallax* de movimiento. Este cambio del punto de vista con la posición del espectador proporciona también una imagen distinta para cada ojo, por lo que la percepción del espacio es muy real. Además, la imagen se forma en un espacio real por lo que la acomodación y convergencia de los ojos coinciden y no produce fatiga. No obstante, el *parallax* de movimiento sólo está codificado en una determinada región que se depende de los límites del cliché fotográfico y de la zona de imagen en que se ha podido registrar correctamente la refracción.

El registro holográfico representa un ejemplo muy interesante, dentro de nuestra línea argumental, sobre el muestreo de la escena con el objeto de reducir sus variables independientes sin que se produzca una pérdida significativa de información. Hemos visto que en un holograma se puede

registrar gran parte de la información volumétrica de la escena en una única proyección plana. Además, la codificación del color es inmediata y puede realizarse en un único holograma. En efecto, para registrar la imagen en color basta con exponer el cliché sucesivamente a tres láseres monocromáticos cuyas frecuencias correspondan con los tres colores primarios. Cuando el holograma se expone simultáneamente a los tres haces de referencia, la superposición de las tres imágenes produce la sensación de color. La particularidad más interesante de este proceso es que el holograma puede registrarse en un cliché de blanco y negro pudiendo recuperar la información de color completa, lo que sin duda supone una codificación muy eficaz de la imagen.

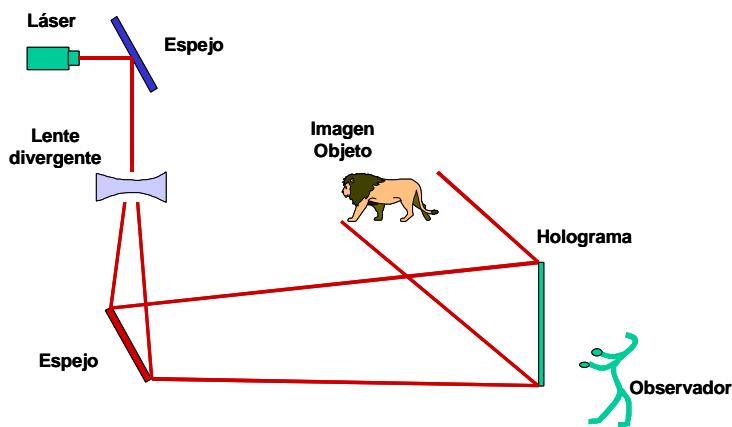


Fig. 1.44 Esquema simplificado para la reproducción de hologramas

Aunque tanto el holograma como la fotografía convencional se registran sobre un material fotosensible del mismo tipo, existen importantes diferencias entre ellos. La diferencia fundamental es que la luz usada en fotografía convencional no es monocromática y suele tener contenido espectral en casi todo el margen visible. Por ello, sobre el negativo se obtiene la superposición de un gran número de componentes incoherentes, con lo que se pierde cualquier información relativa a la fase o los retardos experimentados por las ondas al reflejarse en distintos puntos del objeto. El negativo fotográfico convencional sólo contiene información de amplitud sobre la luz de la escena, mientras que en un holograma, la interferencia producida entre los dos haces monocromáticos permite registrar tanto la información de amplitud como la de fase, siendo ésta última fundamental para recuperar una imagen tridimensional de los objetos.

Otra diferencia importante es que para realizar un registro holográfico no se requiere el uso de lentes que formen una imagen sobre la superficie del negativo, ya que basta con registrar las franjas de interferencias en una región cualquiera del espacio. En la reconstrucción de la imagen, la posición del holograma respecto a la escena durante su registro, determinará el punto de vista general de la escena. En fotografía siempre es necesario utilizar un grupo óptico que enfoque la imagen sobre el negativo, pues en caso contrario la imagen aparecerá borrosa y poco definida.

Finalmente, el registro de las franjas de interferencia requiere que el cliché tenga una resolución considerablemente superior a la requerida en fotografía convencional, ya que las franjas de interferencias suelen estar muy próximas en el espacio. Esto exige usar material fotosensible de baja

sensibilidad ya que la resolución de la película fotográfica aumenta al disminuir su sensibilidad o el grano, lo que significa, debido a que la limitación de potencia del láser, que los tiempos de exposición suelen ser elevados. En el caso en que los objetos a los que deseamos realizar una holografía tengan movimiento, es necesario modificar ligeramente el sistema de registro utilizando un láser pulsado en sustitución de uno de onda continua. El láser pulsado puede usar una potencia muy superior al de onda continua, permitiendo realizar instantáneas holográficas incluso de objetos que se muevan a gran velocidad.

La extensión de las técnicas holográficas a los sistemas de vídeo no es trivial, aunque se supone que si logran resolverse los retos tecnológicos, será probablemente el sistema utilizado en la futura televisión tridimensional. Los problemas para el registro de escenas reales son evidentes, pues no sólo se requiere reproducir unas condiciones de iluminación complejas, sino que también es necesario disponer de dispositivos electrónicos capaces de sensar las franjas de interferencia con una resolución que está muy por encima de la tecnología actual. Se estima que el número de muestras que deben tomarse de un holograma para poder reproducir correctamente las imágenes registradas, con toda la información de *parallax* existente, es de unas 10^{11} muestras, lo que supondría utilizar sensores con una resolución inimaginable en la actualidad y, en cualquier caso, exigiría emplear un elevadísimo ancho de banda para la transmisión de la señal.

A pesar de todas estas dificultades tecnológicas, se están desarrollando diversos prototipos de vídeo holográfico con resultados prometedores. Uno de los sistemas más avanzados es el Mark-II, realizado en el Massachusetts Institute of Technology (M.I.T.) por un equipo de investigación dirigido por el profesor S. Benton. Este sistema se basa en el cálculo mediante ordenador de las franjas de interferencia que producirían imágenes sintéticas. Al sintetizar las franjas de interferencia mediante modelos matemáticos, puede conseguirse una importante reducción en el número de muestras de los hologramas sintéticos restringiendo el *parallax* de movimiento a las direcciones que tienen más interés. En el Mark-II únicamente se codifica la información de parallax horizontal ya que se supone que éste será el movimiento más realizado por el espectador. Con ello se consigue reducir el número de muestras de las franjas de interferencia en un factor de aproximadamente 100.

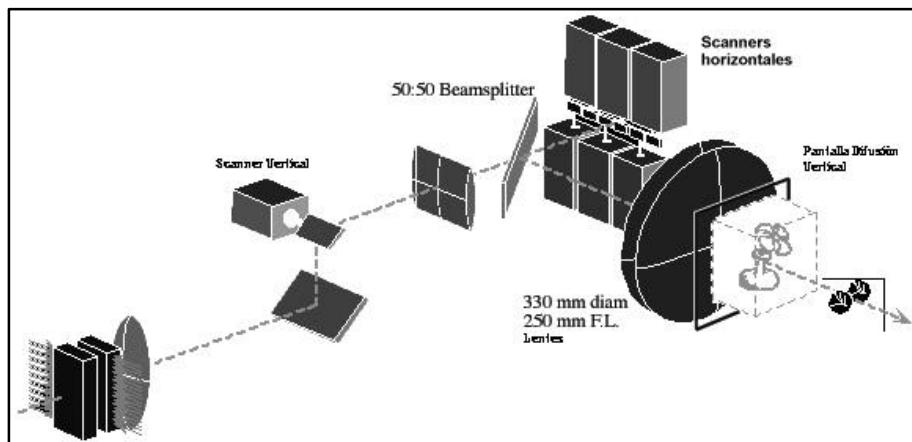


Fig. 1.45 Esquema general del display holográfico Mark II

En la figura 1.45 se muestra un esquema de este *display* en el que su configuración óptica y electrónica está adaptada para poder aplicar un máximo de paralelismo en la generación de las imágenes. La idea general del sistema es la de construir las imágenes mediante una exploración conjunta de varios haces láser cuya amplitud se modula de acuerdo con las franjas de interferencias del holograma previamente calculado. La exploración se realiza utilizando un conjunto de moduladores acustico-ópticos que barren diferentes franjas horizontales de la imagen. El sistema puede presentar imágenes de $150 \times 75 \times 150$ mm con un ángulo de visión horizontal (región con *parallax* horizontal de movimiento) de 36° y es capaz de presentar una imagen por segundo.

1.6 Limitación de la imagen: tamaño del sensor y relación de aspecto

Hemos visto que el grupo óptico se encarga de proyectar la escena sobre el plano de imagen y que sobre éste debemos situar un elemento sensor que nos proporcione información sobre las características de la luz en una región del plano de imagen. La información que nos proporciona el sensor, en aplicaciones de TV, suele ser la cantidad de luz incidente en cada punto dentro de la zona en la que está situado. Si la imagen se ha pasado a través de un filtro dicroico, pueden utilizarse, como ya se ha indicado en la figura 1.32, tres sensores para proporcionar información completa sobre los niveles de luz de cada componente de color en la imagen. Analizaremos los principios de los elementos sensores más utilizados en sistemas de televisión en el capítulo 5. En este apartado, únicamente nos interesa evaluar cómo afecta a las características globales de la imagen que la región en la que están definidos sea finita, delimitando, por tanto, una zona útil en la imagen.

Desde un punto de vista analítico, podemos escribir esta restricción como:

$$\begin{aligned}\epsilon_{FR}(x', y', t) &= \epsilon_R(x', y', t) \cdot w(x', y') \\ \epsilon_{FG}(x', y', t) &= \epsilon_R(x', y', t) \cdot w(x', y') \\ \epsilon_{FB}(x', y', t) &= \epsilon_R(x', y', t) \cdot w(x', y')\end{aligned}\quad (1.53)$$

donde $w(x', y')$ es una función que toma el valor unidad dentro de la región F en la que está situado el sensor y valor nulo fuera. En lo sucesivo designaremos a esta función como la ventana de observación.

$$w(x', y') = \begin{cases} 1 & \text{si } (x', y') \in F \\ 0 & \text{others values} \end{cases} \quad (1.54)$$

La región F es, tanto en el cine como en la televisión, una zona con forma rectangular en la que la dimensión de la componente horizontal es superior a la vertical. Podríamos encontrar algunas excepciones a esta regla general en algunas imágenes fijas (fotografías) donde el autor desea enfatizar el contenido de la escena en el eje vertical o en sistemas de imágenes con propósitos científicos en los que, en ocasiones, las imágenes tienen una forma cuadrada. Otra excepción son algunos sistemas de cine envolvente en que, como veremos, el elemento de imagen básico sigue siendo rectangular o, un sistema del que no nos ocuparemos, la transmisión de gráficos y documentos mediante imágenes como es el caso del facsímil.

La razón por la que la ventana de observación es más ancha que alta se debe básicamente a que el campo de visión del ojo humano es más grande en el eje horizontal que en el vertical. De hecho, la retina no es perfectamente circular, sino que tiene una ligera forma de ovoide que probablemente se debe a una adaptación al medio, en la que la mayoría de movimientos se producen en sentido horizontal. Si bien el campo de visión humano tiene forma elíptica, todos los sistemas de cine y televisión han elegido imágenes rectangulares, por cuanto son mucho más fáciles de tratar, tanto a lo que se refiere a su almacenamiento en una tira de película como a su exploración en sistemas de televisión. Además, el campo de visión humano tiene una forma imprecisa en la que se reduce progresivamente su resolución espacial a medida que nos acercamos a la periferia. En algunos sistemas de cine en los que se pretende reproducir las escenas con un elevado realismo, se utilizan pantallas de grandes dimensiones que pretenden cubrir todo el campo visual del espectador (de hecho algunos sistemas sitúan al espectador tan cerca de la pantalla que ésta cubre una zona superior al campo visual, por lo que el espectador debe incluso girar su cabeza para centrar la atención en determinadas partes de la imagen). No obstante, aún en estos casos, las pantallas tienen forma rectangular y la definición de la imagen es uniforme en toda su extensión.

El hecho de limitar la imagen a una ventana de observación finita es otra característica de los sistemas de televisión en la que se produce una reducción notable de la cantidad de información a transmitir. El tamaño y forma de esta ventana suelen ser parámetros que condicionan en última instancia la calidad y el realismo de las escenas transmitidas.

El tamaño de la ventana no constituye una limitación seria, en contra de lo que en un principio pueda parecer, sobre la parte de la escena que se desea transmitir. En efecto, mediante el uso de lentes adecuadas podemos hacer que la parte de interés de la escena quede perfectamente encuadrada en el área útil del sensor de imagen independientemente del tamaño que éste tenga. Para ello, dependiendo de las dimensiones físicas del sensor, necesitaremos utilizar grupos ópticos con distintas distancias focales para que los ángulos de visión de las imágenes se mantengan dentro de los límites establecidos. Analizaremos con detalle estos aspectos en el capítulo 5, pero de momento adelantemos que una óptica con una distancia focal de 55 mm en fotografía (película de 35 mm de anchura) puede producir el mismo tipo de imágenes (es decir, el mismo campo de visión) que una óptica de 24 mm para cámaras de vídeo portátiles, en las que se utiliza un sensor del tipo CCD cuyo tamaño es del orden de $\frac{1}{2}$ pulgada. Esta es una justificación de por qué los grupos ópticos de las cámaras de cine son mucho más voluminosos que los de las cámaras de vídeo portátiles y de por qué en las cámaras de vídeo domésticas una reducción del tamaño del sensor CCD supone también una reducción del volumen de la óptica. Sin embargo, no adelantemos excesivos resultados que podrán ser adecuadamente justificados más adelante y quedémonos con la idea de que el tamaño del sensor no condiciona en absoluto el encuadre de la escena, ya que esté sólo depende de la óptica.

Uno de los factores que sí están condicionados por las dimensiones físicas de la ventana de observación es la resolución de la imagen que, en general, y considerando una misma tecnología, mejora con el tamaño del sensor. Definiremos más adelante la resolución de una imagen de manera cuantitativa, pero de momento indiquemos que la resolución está relacionada con el grado de detalle o nitidez que puede llegar a representarse en una imagen. Decimos que puede llegar a representarse, ya que es posible que el director decida presentar partes de las imágenes intencionadamente desenfocadas o poco nítidas para generar determinados ambientes, aun cuando se esté utilizando un sistema con una resolución excelente. La nitidez y la resolución no son, por tanto, conceptos idénticos, aunque están

relacionados de manera que no es posible obtener imágenes extremadamente nítidas en sistemas con baja resolución.

En cine y fotografía la resolución depende de la emulsión fotosensible de la película. En general, la nitidez está limitada por el tamaño de las partículas fotosensibles que se conocen con el nombre de 'grano' de la película. Una película con mucho 'grano' indica que tiene una resolución pobre. El grano está relacionado con la sensibilidad de la película, de modo que cuanto más sensible a la luz es la emulsión fotográfica más grano presenta. Para una sensibilidad fija, la calidad de la imagen proyectada será tanto mejor cuanto mayor sea el tamaño del fotograma. Así, supongamos un mismo fotograma captado por una cámara de 16 mm y por una de 35 mm (dimensiones físicas de la anchura del fotograma en la película), con películas de la misma sensibilidad y grano, que se proyectan en una pantalla con las mismas dimensiones. Evidentemente, para que ambos fotogramas tengan el mismo tamaño en la pantalla deberemos utilizar ópticas distintas, que amplíen más el fotograma de 16 mm que el de 35 mm, por lo que también se ampliará más el grano de la primera fotografía, obteniendo una imagen final de menor resolución. Así pues, un formato de cine de 70 mm proporciona una calidad de imagen muy superior al formato estándar de 35 mm.

En cámaras de televisión ocurre un fenómeno parecido, aunque en este caso, la rápida evolución experimentada en la tecnología de los sensores CCD en los últimos años puede enmascarar parcialmente este principio para el profano. Si comparamos dos chips CCD que utilizan la misma tecnología, es decir, que integran el mismo número de elementos de imagen por unidad de superficie, es evidente que el dispositivo que disponga de mayor área tiene más elementos sensores, por lo que ofrece una mejor resolución. Ocurre, sin embargo, que podemos encontrar cámaras modernas que, utilizando un dispositivo CCD de menor área, ofrece una calidad de imagen superior a otras cámaras con un área sensora mayor, debido a que la primera utiliza una tecnología más avanzada en la que ha sido posible integrar muchos más elementos de imagen que en la segunda. No obstante, el principio sigue siendo que cuanto mayor sea el sensor, si la tecnología es comparable, mejor es la resolución de la imagen, y esto se aplica a cualquier tipo de tubos de imagen sean o no de tecnología CCD. Esto justifica nuevamente por qué las cámaras profesionales son mucho más voluminosas que las domésticas y las cámaras de cine mucho más que las de vídeo. En efecto, una cámara de vídeo profesional suele utilizar, para mejorar la resolución de las imágenes, sensores de mayor tamaño que las cámaras domésticas. Además, la distancia focal de la óptica necesaria (y en consecuencia su volumen) también aumenta con el tamaño del sensor, por lo que las cámaras profesionales tienen un mayor volumen que las domésticas. En el caso del cine, la tecnología para el registro de las imágenes es distinta y requiere un mayor tamaño de la ventana de observación, por lo que el volumen asociado a la óptica es aún más grande.

Otro aspecto ligado a la ventana de observación es la relación de aspecto de la imagen que se define como el cociente entre la anchura y la altura de la ventana. En la figura 1.46 se representa una ventana de observación genérica en el que la relación de aspecto es W:H, o bien, W/H:1.

Cuando se definieron los primeros sistemas de televisión comercial en blanco y negro, se eligió una relación de aspecto que coincidía con la utilizada en el cine. De hecho, inicialmente se pensó que buena parte de la programación de televisión estaría formada por películas previamente realizadas para el cine, por lo que resultaba conveniente que ambas relaciones de aspecto coincidieran. En la época en que se definió la televisión, la relación de aspecto utilizada en el cine era, invariablemente,

de 4:3, es decir, la altura de la pantalla es $\frac{3}{4}$ partes de su anchura. Este formato es utilizado por todos los sistemas de TV en color analógicos definidos durante los años 50 y 60: NTSC, PAL y SECAM y sólo recientemente se han propuesto algunos sistemas con una relación de aspecto distinta.

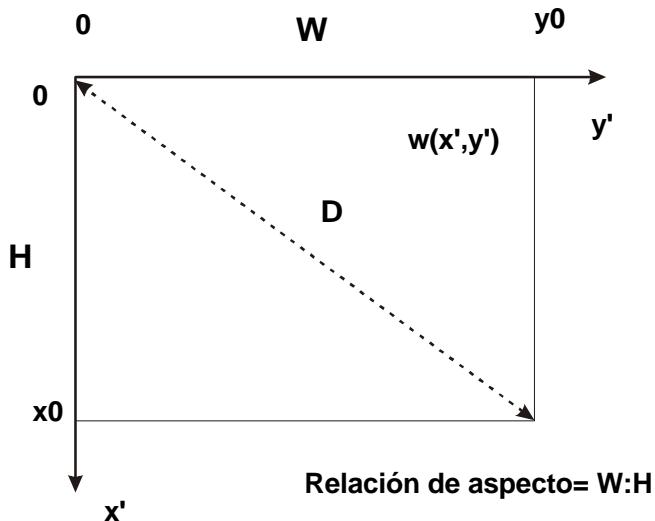


Fig. 1.46 Tamaño de la ventana de observación y relación de aspecto

Uno de los parámetros directamente relacionados con la relación de aspecto es la medida de la diagonal que se utiliza para indicar el tamaño de los monitores. Así, cuando hablamos de un monitor de 29'' nos referimos a la longitud de la diagonal de la pantalla. Dado que la suma de los cuadrados de las proporciones entre la altura y la anchura da un cuadrado perfecto, es fácil relacionar la altura, la anchura y la diagonal del monitor. En efecto, la anchura y la altura pueden determinarse como 4/5 y 3/5 partes de la diagonal respectivamente, ya que esta última mantiene un factor de proporcionalidad igual a 5:

$$D = \sqrt{(4k)^2 + (3k)^2} = 5k \quad (1.55)$$

De este modo, en un televisor de 29'', la diagonal mide $29'' \times 2,54 \text{ cm/''} = 73,6 \text{ cm}$, la altura es de $73,6 \text{ cm} \times 3/5 = 44 \text{ cm}$ y la anchura mide $73,6 \times 4/5 = 59 \text{ cm}$.

La elección de una relación de aspecto para la televisión idéntica a la que se utilizaba en el cine fue, sin duda, una decisión muy acertada por cuanto, aún hoy en día, buena parte de la programación televisiva está constituida por cintas producidas previamente para el cine. Lo que no se podía prever en el momento de definir la televisión es que las relaciones de aspecto utilizadas en el cine experimentarían un importante cambio con la introducción de los formatos panorámicos. De hecho, fue el propio éxito de la televisión lo que originó que la industria cinematográfica introdujera distintos avances tecnológicos con objeto de mejorar tanto la calidad del sonido como de la imagen para así poder competir directamente con el medio televisivo ofreciendo al espectador un entorno audiovisual de mucha mayor calidad que el que se ofrecía en la televisión.

A principios de los años 50 se empiezan a introducir formatos cinematográficos con relaciones de aspecto más anchas que la 4:3 (1,33:1). En Estados Unidos destaca el formato de la Academia, con una relación de aspecto de 1,85:1; mientras que en Europa se utilizan principalmente dos formatos con relaciones de aspecto 1,67:1 y 1,75:1, de los cuales el primero tuvo un uso más extendido. Todos estos formatos mantenían el mismo tipo de película que el formato 1,33:1, limitando la apertura vertical durante el registro de la película y utilizando lentes de menor distancia focal en la proyección para que la anchura de la proyección fuera mayor. La verdad es que todos estos formatos se adaptan mejor al campo de visión humano y permiten que éste se integre mucho más en la escena, aumentando surealismo.

En 1953 aparece el Cinemascope, cuya principal novedad tecnológica es que introduce un sistema con lentes anamórficas, que esencialmente consisten en un sistema óptico que durante el registro de la película comprime las imágenes en la dirección horizontal. Durante la proyección de la película deben usarse unas lentes inversas, que expanden la imagen en el sentido horizontal, recuperando el aspecto original de los objetos. Este procedimiento permite registrar en una película convencional un formato de proyección panorámico, aprovechando todo el negativo de la película. La relación de aspecto obtenida con este sistema es de 2,35:1. Evidentemente, aunque no es apreciable por el espectador, la resolución en sentido vertical es mayor que en sentido horizontal, ya que este último ha sido más ampliado por la lente que el primero. Si una película registrada en cinemascopio se proyectara utilizando lentes convencionales, las imágenes aparecerían alargadas y el área de proyección tendría una relación de aspecto convencional. De hecho, en muchas de las primeras películas rodadas en este formato deben de ponerse lentes convencionales durante la proyección de los títulos iniciales y los créditos finales. La razón es que en la edición de estas películas se utilizaron los equipos de titulación convencionales, ya que muchos estudios no disponían de tituladoras específicas para este formato. Por tanto, si se desea que los títulos puedan leerse correctamente, es necesario utilizar una óptica convencional al principio y al final de la película. En las producciones en las que se superponen los títulos con imágenes filmadas, estas últimas aparecen alargadas durante el inicio y final de la cinta, constituyendo una de las estampas más características de las películas en este formato.

Un año antes de la aparición del Cinemascope, Fred Waller, fotógrafo neoyorquino, propuso un formato revolucionario que utilizaba tres proyectores simultáneamente y que denominó Cinerama. La película debía grabarse con un complejo artilugio que agrupaba las tres cámaras con las que, adecuadamente sincronizadas, se obtenían las tres partes de la escena. Los proyectores estaban también sincronizados y cada uno proporcionaba 1/3 de la imagen en una pantalla muy ancha y curvada. La primera película que se produjo en este formato fue un documental de demostración titulado *This is Cinerama*. El sistema era extremadamente complejo y aunque se hizo bastante popular, tenía un coste prohibitivo. Tanto las cámaras como los proyectores debían estar perfectamente calibrados durante la grabación y la reproducción, y frecuentemente aparecían problemas de solapamiento entre las tres imágenes y de desadaptación de brillo y contraste entre las tres partes de la película. El sistema se utilizó fundamentalmente para documentales y la primera película argumental registrada en este formato, *How the West Was Won*, no apareció hasta 1962. En 1963 se abandonó este sistema que fue sustituido por el Cinerama moderno que utilizaba un sistema de lentes anamórficas parecidas al Cinemascope, registrando la información en una película de 70 mm de anchura y utilizando una relación de aspecto de 2,75:1. El formato estándar de 70 mm había sido introducido previamente en 1955 por Todd-AO Productions con las películas *Oklahoma!* (1955) y *Around the World in 80 Days* (1956). La relación de aspecto utilizada en 70 mm es de 2,20:1.

También se han realizado algunos sistemas en los que la escena se proyecta sobre pantallas circulares que envuelven completamente al espectador, con lo que se pretenden reproducir sensaciones de elevado realismo. El primer sistema de cine envolvente se instaló en 1955 en el parque de atracciones de Disneyland y combinaba once películas de 16 mm en un sistema de proyección sincronizada parecido al Cinerama original. Posteriormente el sistema se modificó utilizando nueve proyectores con películas de 35 mm. A pesar de que el espectador suele demostrar un elevado grado de satisfacción sobre la calidad y el realismo de estos sistemas, su uso se ha limitado la producción de documentales cuyas salas de exhibición se encuentran en parques temáticos, exposiciones internacionales o ferias de muestras. Probablemente, no se han explotado comercialmente ya que la propia libertad que otorgan al espectador para centrar su atención en distintas partes de la escena no facilita la exposición de una trama argumental convencional.

Algunos sistemas, como el IMAX, con superficies de pantalla extremadamente grandes (IMAX-Port Vell, 600 m²), utilizan relaciones de aspecto donde la altura de la pantalla es comparable o superior a la anchura. Esto es debido en parte a que las grandes dimensiones de la película requieren que sea pasada por el proyector en sentido horizontal y en parte a que se pretende que la zona de proyección sea superior al campo de visión del espectador, para crear así un efecto de mayor realismo. La sensación final es que el espectador debe mover la cabeza para centrar su atención en diferentes partes de la pantalla.

En la figura 1.47 se comparan las relaciones de aspecto más habituales utilizadas en el cine y en la televisión. Las gráficas se han obtenido manteniendo constante el área total de proyección para comparar únicamente los aspectos de cada sistema.

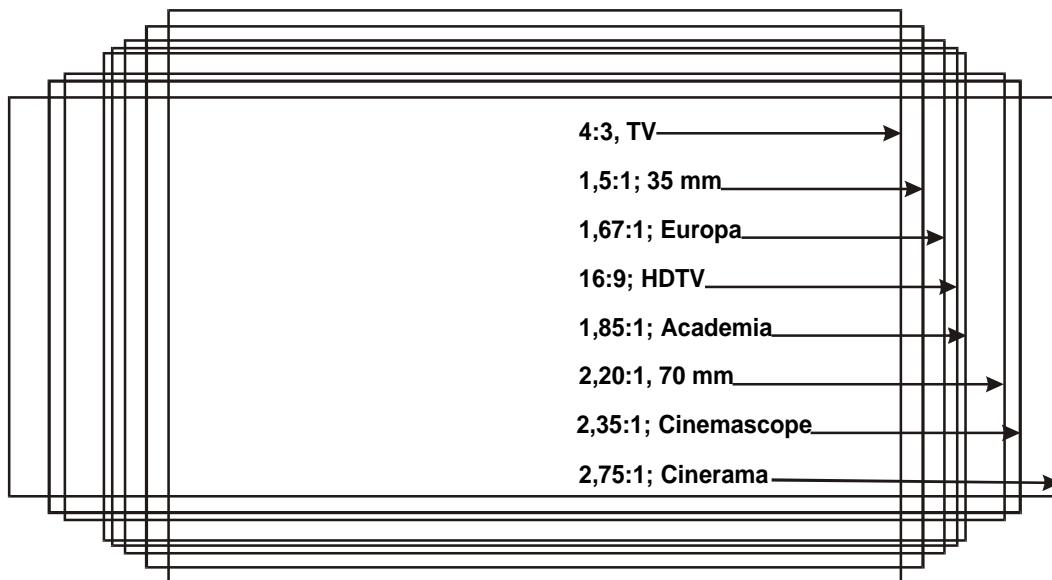


Fig. 1.47 Relaciones de aspecto más habituales utilizadas en cine y en televisión

Ante la gran diversidad de relaciones de aspecto empleadas en los distintos formatos de cine cabría preguntarse cuál es, desde un punto de vista subjetivo, la relación de aspecto ideal. La NHK, organismo de radiodifusión de Japón, durante el desarrollo del MUSE, un sistema de televisión de alta definición analógico que hoy es operativo en Japón, elaboró una serie de experimentos subjetivos para determinar cuál era la relación de aspecto ideal. La BBC y el SMPTE han desarrollado experimentos similares en los que en general se observa que las preferencias del espectador dependen del tamaño de la imagen. Se observa que cuando el área de proyección aumenta el espectador tiende a exigir un formato más panorámico, lo que justifica el uso de formatos más panorámicos en cine que en televisión. No obstante, si sigue aumentándose el área de proyección, se llega a un punto en que los formatos panorámicos saturan y son poco agradables, lo que justifica, aunque sólo en parte, el uso de relaciones de aspecto casi cuadradas en sistemas de grandes dimensiones como el IMAX.

Los resultados obtenidos por la NHK en el margen de superficies de proyección propias de un sistema de televisión se representan en la figura 1.48, observándose que la relación de aspecto que puede considerarse como ideal es de 15:9 (1,67:1), que coincide con el formato cinematográfico definido en Europa en la década de los 50. Obsérvese que los test subjetivos rechazan relaciones de aspecto muy panorámicas para las dimensiones de una pantalla de televisión. No obstante, la relación de aspecto que finalmente eligió la NHK, y que actualmente es aceptada por todos los organismos internacionales como idónea en sistemas con definición mejorada o alta definición, es de 16:9. Esta relación de aspecto es parecida a la 15:9 y ofrece importantes ventajas de compatibilidad con distintos formatos cinematográficos y con la relación 4:3 utilizada en TV convencional. En la figura 1.47 podemos comprobar que la relación 16:9 es muy parecida al formato de la Academia y al Europeo, por lo que este tipo de películas podrán adaptarse prácticamente sin utilizar zonas muertas en la pantalla a la transmisión por televisión. Además estos formatos cinematográficos son muy utilizados. Los formatos panorámicos, sobre todo el Cinerama y el Cinemascope son aún excesivamente anchos y estas películas deberán transmitirse con bandas negras en la parte superior o inferior de la pantalla (o bien cortando los laterales). No obstante, actualmente el Cinerama y el Cinemascope prácticamente ya no se utilizan por lo que, para películas nuevas el único formato que seguirá manteniendo baja compatibilidad con la televisión 16:9 será el de 70 mm.

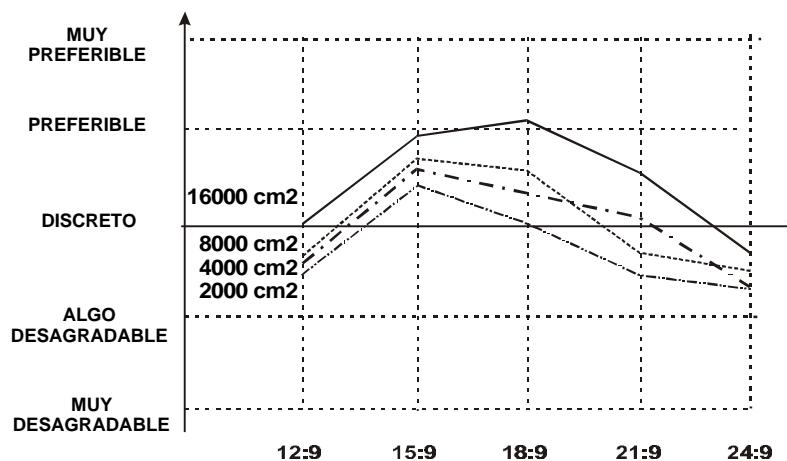


Fig. 1.48 Pruebas subjetivas sobre preferencias en las relaciones de aspecto

Otra de las razones por la que se eligió la relación 16:9 en los sistemas de TV de alta definición o sistemas de definición mejorada como el PALplus es que este formato mantiene una relación cuadrática con el formato convencional 4:3, que puede utilizarse para presentar simultáneamente varias imágenes 4:3 en una pantalla 16:9. Esta posibilidad se ilustra en la figura 1.49 y fue uno de los principales atractivos en la selección del formato. Un espectador puede visualizar simultáneamente hasta 4 programas en formato 4:3, sin que se produzca ninguna pérdida de información. El canal principal ocupa $\frac{3}{4}$ partes de la pantalla mientras en uno de los laterales se presentan tres canales auxiliares que pueden monitorizarse para ver la programación o esperar un evento concreto. No obstante, a pesar de que éste fue uno de los criterios más importantes en la selección del formato panorámico de 16:9, prácticamente ningún receptor panorámico incorpora la opción de visualizar cuatro canales simultáneamente, ya que ello supone un aumento significativo del coste del equipo al requerir cuatro sintonizadores y demoduladores distintos. En la práctica, los receptores de gama alta incorporan, a lo sumo, un doble sintonizador que permite visualizar un canal principal y sólo uno auxiliar.

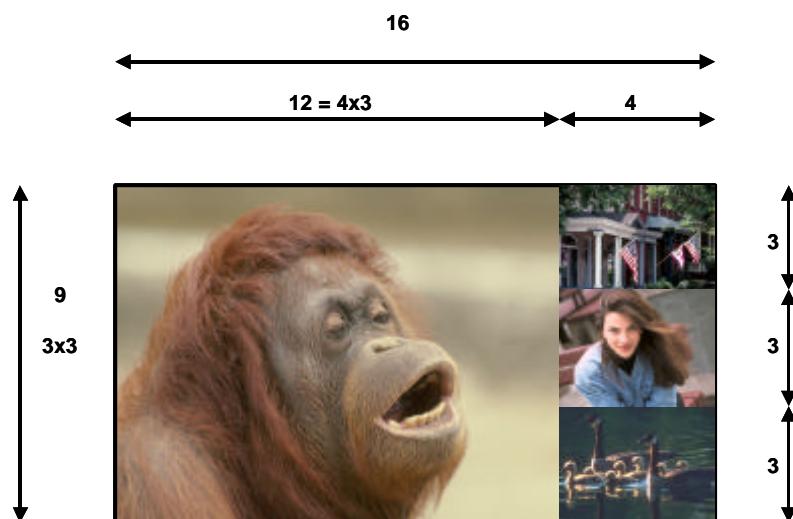


Fig. 1.49 Visualización simultánea de varios programas 4:3 en un monitor panorámico 16:9

La compatibilidad entre los dos formatos es uno de los principales problemas con los que se enfrentan los proveedores de señal de televisión a medida que empieza a popularizarse el uso de una relación de aspecto de 16:9. Los programas producidos con esta relación de aspecto deberán presentarse al telespectador que tenga un monitor 4:3 en uno de las opciones indicadas en la figura 1.50. Probablemente, el formato que más se utilizará es el de la izquierda, ya que se presenta toda la información disponible sin ninguna deformación, al precio de que el espectador ve dos barras negras en la parte superior e inferior de la pantalla. Este aspecto de la pantalla de televisión viene siendo aceptado en la transmisión de películas, pero parece difícil que sea aceptado por el televíidente en la transmisión de acontecimientos deportivos, documentales o noticiarios. El formato representado en el centro de la figura 1.50 pierde parte de la información lateral presente en la imagen 16:9, aunque

como el espectador no conoce el contenido de esta información parece que será aceptado de buen grado para algunos programas especiales como las noticias. Sin embargo, es un formato polémico, sobre todo en películas, pues se pierde parte de la información que el autor ha decidido incluir en la escena. Tampoco parece un formato que sea fácilmente aceptado en la transmisión de acontecimientos deportivos, a menos que durante la producción de los programas en formato 16:9 se tenga en cuenta que el espectador 4:3 no pierda una parte importante de la información. Finalmente, la posibilidad de la derecha es con mucho la menos aceptada, pues si bien no hay pérdida de información ni aparecen bandas negras en la pantalla, existe una deformación importante de los objetos.



**Formato 16:9
TV Alta Definición**



Formato convencional 4:3

Fig. 1.50 Opciones de presentación de imágenes 16:9 en un monitor 4:3

Un problema análogo aparece cuando se reproducen programas producidos en formato 4:3 en un televisor panorámico, problema que actualmente afecta prácticamente a la totalidad de los usuarios que disponen de este tipo de televisores. En la figura 1.51 se ilustran las distintas posibilidades que a menudo, dependiendo del tipo de televisor, puede controlar el propio televíidente. La opción de la izquierda permite ver la totalidad de la imagen 4:3 sin ningún tipo de deformación, aunque supone dejar dos bandas negras en los laterales del monitor. Es una opción que no suele gustar al usuario de un monitor panorámico porque es consciente de que se desaprovecha la inversión en el televisor panorámico, aunque es la única opción válida en determinados programas cuya información útil se distribuye a lo largo de toda la pantalla. La opción central, en la que se pierde parte de la información superior e inferior de la imagen, suele ser bastante utilizada sobre todo en noticiarios. Téngase en cuenta que la pérdida de la información superior e inferior no suele ser tan importante como la pérdida lateral que experimenta el usuario 4:3 en la transmisión de programas 16:9 ya que en la mayoría de programas la acción transcurre en el plano horizontal. La opción de la derecha consiste en deformar la imagen para que aparezca completa en toda la pantalla y es aconsejable no utilizarla.

De hecho, el problema de compatibilidad entre los formatos 4:3 y 16:9 constituye un serio problema para la implantación progresiva de sistemas panorámicos. El PALplus, un sistema evolucionado del PAL que permitía transmitir señales en formato 16:9 compatibles con el PAL convencional, tuvo serios problemas para implantarse en un mercado en el que el usuario aún no estaba totalmente dispuesto a adquirir televisores en formato panorámico, tanto por razones económicas (son bastante más caros) como por razones de compatibilidad, pues la mayoría de programas se siguen produciendo en formato 4:3. El sistema PALplus fue prácticamente abandonado con la aparición de los sistemas de televisión digitales que tienen la posibilidad de transmitir en los dos formatos. Parece que la implantación progresiva de sistemas de video como el DVD permitirá la introducción definitiva de este tipo de monitores, para obtener unas características de reproducción similares a las del cine (Home Theater Systems). Las producciones de televisión también irán progresivamente adoptando los formatos panorámicos. Probablemente, ambos formatos compartan el mercado hasta la desaparición definitiva de los formatos de transmisión analógicos como el PAL y el NTSC, que se prevé que se produzca a mediados de esta década.

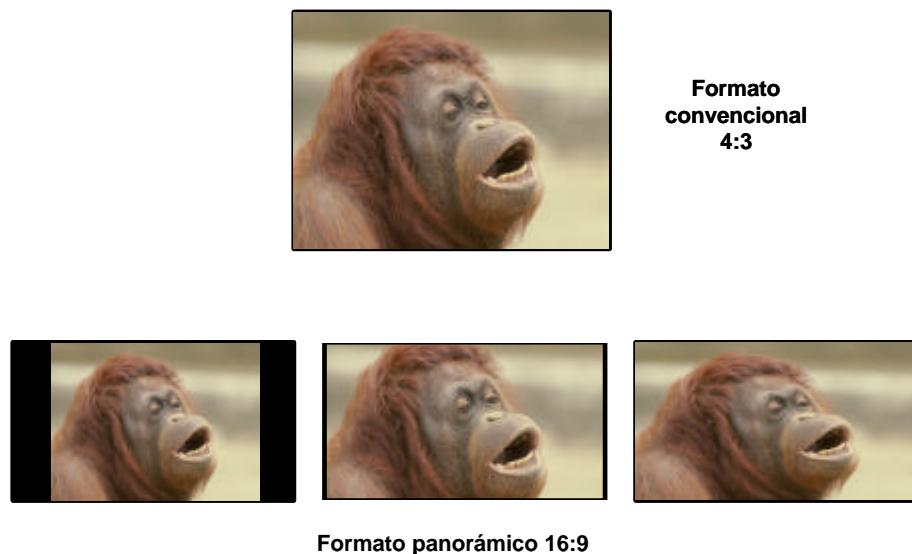


Fig. 1.51 Opciones de presentación de imágenes 4:3 en un monitor 16:9

1.7 Muestreo temporal de las imágenes

El ojo humano es capaz de percibir el movimiento de los objetos en una escena aun cuando la información visual se presente de manera discreta como una rápida sucesión de fotogramas. Este fenómeno es conocido desde la primera mitad del siglo XIX y se debe fundamentalmente a dos manifestaciones de la percepción visual conocidas como el *fenómeno phi* y la *persistencia*, cuyos principios serán descritos en esta sección.

En esencia, el muestreo temporal de la escena consiste en sustituir la función:

$$\varepsilon_{FX}(x', y', t) \quad (1.56)$$

por la secuencia de fotogramas:

$$\varepsilon_{FX}^{(k)}(x', y') = \varepsilon_{FX}(x', y', kT_0) \quad (1.57)$$

donde el subíndice FX indica que se trata de las imágenes planas definidas en el margen de la ventana de observación (F), de cualquiera de las componentes de color o luminancia (X). El superíndice k denota que se trata del k-ésimo fotograma obtenido al particularizar la imagen original en el instante kT_0 . En sistemas de TV cada una de las imágenes se denomina cuadro (*frame*). Nótese también que suponemos que las muestras temporales de la escena se toman a intervalos regulares de muestreo cuyo valor T_0 deberemos especificar para que la percepción de continuidad en el desarrollo de la escena sea apropiada. La secuencia de imágenes se representa esquemáticamente en la figura 1.52, donde se observa que la información de la escena ha sido discretizada en el tiempo y que únicamente se presentará al observador la secuencia de fotogramas obtenidos de la escena original.

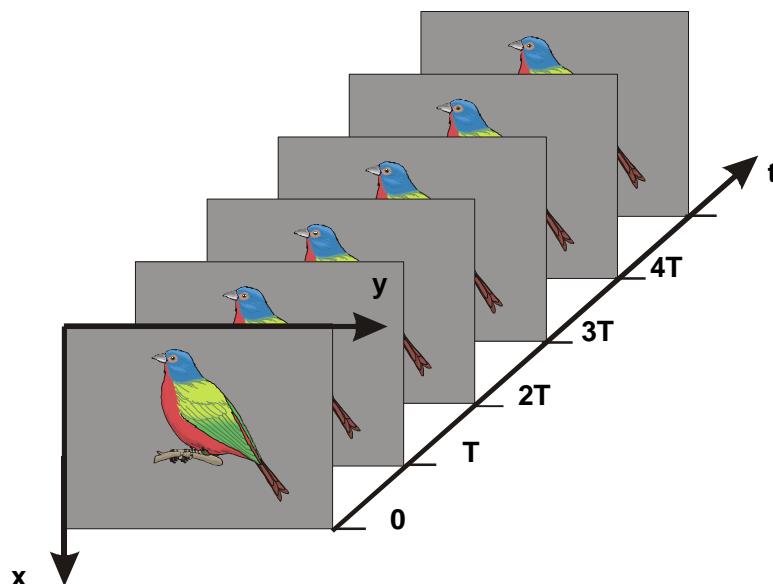


Fig. 1.52 Esquema del muestreo temporal de las imágenes

Es evidente que el muestreo de la escena definido en la ecuación 1.57 representa un caso ideal en el que es posible obtener un fotograma exacto de la escena en el instante kT_0 . En la práctica, todos los dispositivos sensores requieren que la imagen esté presente durante un tiempo determinado para que los valores de iluminación puedan ser adquiridos correctamente, por lo que los fotogramas se corresponderán con los valores promedios de la escena durante el intervalo de adquisición. Analizaremos con detalle estos efectos en el capítulo dedicado a cámaras.

1.7.1 Un poco de historia: del Phenakistoscope al Cinématographe

Los primeros artíluguos que utilizaron el principio de la rápida superposición de imágenes estacionarias para producir la sensación de movimiento datan de la primera mitad del siglo XIX. El Phenakistoscope (1832) y el Zoetrope (1834) eran juguetes ópticos que se exhibían en ferias y que se basaban en una serie de dibujos dispuestos en discos o tambores giratorios que producían la sensación de movimiento en el espectador. Estos dos ingenios, precursores del cine y la televisión, son incluso anteriores a la aparición de la fotografía, cuyos principios no fueron establecidos hasta 1839 con el desarrollo de un proceso fotográfico positivo, conocido como el daguerrotipo, elaborado por el pintor francés Louis Jacques Mandé Daguerre. Poco después el británico William H.F. Talbot desarrolla un proceso que permite obtener múltiples copias positivas a partir de un único negativo, sentando por tanto las bases para la multidifusión de reportajes fotográficos. Sin embargo, eran los albores de la fotografía y los tiempos de exposición eran de varios minutos, lo que hacía impensable sustituir los dibujos y gráficos utilizados en los discos giratorios por fotografías reales. Hacia 1870, la evolución tecnológica en la fabricación de material fotosensible ya permitía utilizar tiempos de exposición del orden de centésimas de segundo.

La primera secuencia de fotografías sobre una escena real fue realizada por el fotógrafo británico-americano Eadweard Muybridge. Este fotógrafo fue contratado por Leland Stanford, gobernador de California y gran aficionado a las carreras de caballos, para demostrar que en algún punto del galope de un caballo éste mantenía sus cuatro patas en el aire. Evidentemente se trataba de una apuesta en la que los resultados de las fotografías de Muybridge debían establecer el ganador. Después de varias pruebas y experimentos, en 1877, Muybridge consiguió tomar una secuencia de 12 fotografías del galope de un caballo. Para ello dispuso una batería de 12 cámaras fotográficas a lo largo de la pista de carreras, utilizando unos cables que estaban situados sobre la pista detectaban el paso del caballo y disparaban la cámara asociada. La secuencia de 12 fotogramas se montó sobre un disco giratorio y se proyectó, reproduciendo en movimiento del caballo en la pantalla. Stanford resultó vencedor de la apuesta.

Un experimento similar fue llevado a cabo por el fisiólogo francés Étienne-Jules Marey quien en 1882 obtuvo una secuencia de fotografías del vuelo de un pájaro. La principal innovación introducida por Marey es que usó un artílugo de su invención que permitía realizar todas las instantáneas mediante una única cámara. La cámara tenía el aspecto (y la filosofía) de un rifle y llegaba a realizar un total de 12 fotografías por segundo. Marey también montó las fotografías sobre un tambor rotatorio para sintetizar el movimiento obtenido. No obstante, tanto en el caso de Muybridge como Marey, el objetivo principal de las experiencias no era la síntesis del movimiento, sino el análisis científico de movimientos que, por su rapidez, no podían ser observados con precisión con la simple inspección visual.

En 1887, Hannibal Goodwin desarrolla un material fotosensible que puede ser enrollado en sí mismo, lo que permite realizar varias fotografías sin necesidad de sustituir el negativo. El uso de un carrete permite simplificar notablemente la cámara respecto al complejo sistema de Marey, que necesitaba cargar previamente, de manera independiente, todos los negativos. En 1888, George Eastman inicia la producción en masa de carretes para fotografía, cuyo uso se populariza rápidamente.

También en 1888, Thomas Alva Edison propone a uno de sus ayudantes de laboratorio que investigue con el objeto de desarrollar una cámara que sea capaz de registrar secuencias de fotogramas de escenas reales para posteriormente sintetizar el movimiento. Inicialmente el objetivo de Edison es que utilizar el sistema como acompañamiento visual del fonógrafo, invento que él mismo había desarrollado un año antes y que estaba teniendo un considerable éxito comercial. La propuesta de Edison es la primera en que la cámara se diseña específicamente para la síntesis del movimiento y no, como en el caso de Marey, para su análisis.

William K.L. Dickson, el ayudante al que Edison había encargado el trabajo, desarrolló una cámara que incorporaba un complejo sistema de relojería para garantizar que la frecuencia con que se tomaban los fotogramas fuera constante. Además, introdujo un sistema de perforación en los laterales del celuloide cuyo objetivo era sincronizar la toma de imágenes de la cámara con su posterior proyección. Los resultados de estos trabajos cristalizaron en el Kinetograph, que se patentó en 1893 y que podía trabajar con rollos de celuloide de quince metros de longitud tomando un total de 40 fotogramas por segundo.

Como el objetivo fundamental del sistema era la realización de películas para acompañar al fonógrafo, Edison propone a Dickson que desarrolle un dispositivo para uso individual que permita visualizar las películas obtenidas con el Kinetograph. El diseño final es un sistema de visionado personal formado por una caja con un habitáculo en el que se coloca la película, una lámpara incandescente y un obturador que permiten ver la secuencia de imágenes a través de un visor. El sistema de visionado se denomina Kinetoscope y se comercializa y distribuye fundamentalmente en hoteles, recintos feriales y lugares lúdicos.

Paralelamente, en Europa se diseñan y patentan sistemas parecidos que finalmente no son aceptados por el público. Además, Edison no realiza patentes internacionales sobre sus sistemas de cámara y visor, por lo que aparecen varias copias, con ligeras modificaciones de estos sistemas, en Europa. Una de las exhibiciones del Kinetoscope realizadas en París, inspira a los hermanos Auguste y Louis Lumière para el desarrollo del primer sistema de cámara y proyector comercialmente viable. El 28 de diciembre de 1895 se realiza la primera demostración en París del nuevo invento de los hermanos Lumière, el Cinématographe, que suele ser considerado como el inicio del cine. Es famosa la frase de los hermanos Lumière en la que aseguran que su invento es ciertamente curioso, que puede ser usado durante algún tiempo como distracción, pero que no tiene ningún futuro comercial.

La principal ventaja del Cinématographe respecto al Kinetograph es que el sistema puede usarse como cámara o como proyector y que es relativamente portátil. Mientras el Kinetograph requiere el uso de baterías y pesa casi 500 Kg, el Cinématographe se opera con el uso de una manecilla, sólo utiliza 16 fotogramas por segundo, con lo que optimiza bastante el uso de la película y tiene un peso de alrededor de 8 Kg. Las diferencias entre los dos sistemas condiciona el tipo de películas que se realizan con cada uno. En el caso del Kinetograph todas las filmaciones son escenas de circo o vodevil que se realizan en un estudio, mientras que el invento de los hermanos Lumière se utiliza fundamentalmente en la realización de reportajes y documentales filmados en exteriores. En este último caso, se envían reporteros con cámaras a todos los rincones del mundo, por lo que el Cinématographe se convierte rápidamente en un sistema muy popular que es adoptado como estándar por la mayoría de países. Tanto en un caso como en otro, las películas se presentan siempre tal cual han sido registradas, no aplicándose, por el momento ningún proceso de edición.

En Estados Unidos, el mercado de Kinetoscope se satura rápidamente y la principal fuente de ingresos para Edison es la producción de películas, creando unos estudios específicos para ello. En 1896, viendo que su sistema de proyección individual no tiene excesivo futuro, Edison adquiere los derechos para la fabricación de un proyector desarrollado por Thomas Armat y conocido como Vitascope. El nuevo proyector incorpora un mecanismo para el movimiento intermitente de la película y un sistema para el enrollado y desenrollado de la película en dos carretes durante la proyección. La primera demostración de este sistema se realiza el 23 de abril de 1986 en un teatro de Nueva York y supone una nueva filosofía para el visionado de películas en salas especiales. Durante esta época aparece en Estados Unidos un gran número de sistemas de proyección alternativos con redes de exhibición itinerantes que utilizan teatros móviles. En Europa, la incipiente industria cinematográfica británica y sobre todo la francesa empiezan a utilizar teatros fijos para la proyección de películas separándose los procesos de producción y exhibición de una forma definitiva. Pronto surgen las primeras, películas con edición, superposición de imágenes y trama argumental. El primer proceso fotográfico en color comercialmente viable para el cine se obtiene en 1908 (Kinemacolor), aunque no empieza a usarse hasta mucho más tarde. El resto de esta historia ya no tiene mucho que ver con los orígenes del muestreo temporal de imágenes, por lo que dejaremos que el lector interesado la busque en textos especializados sobre los orígenes del séptimo arte.

1.7.2 La persistencia, el parpadeo y el fenómeno phi

La sensación de continuidad que se percibe cuando se nos presenta una secuencia de imágenes a una frecuencia suficientemente rápida es debida a la acción conjunta de la persistencia de las imágenes en la retina y al fenómeno phi. La persistencia consiste en que la percepción de la imagen se mantiene durante unas fracciones de segundo después de que ha desaparecido la excitación. En la figura 1.53 se comparan dos gráficos que intentan ilustrar esta sensación visual en un experimento donde se proyectan impulsos lumínicos de amplitud constante en una pantalla. La gráfica superior representa la energía lumínica proyectada en función del tiempo, mientras la inferior representa la sensación de iluminación percibida por el sistema visual humano. Obsérvese que la respuesta del ojo no desaparece inmediatamente después de la excitación sino que se mantiene durante unos instantes. Si la frecuencia de la excitación es lenta el sistema visual es capaz de discernir entre las distintas excitaciones interpretando correctamente que la luz del proyector se apaga y se enciende alternativamente. Sin embargo, si aumentamos la frecuencia de excitación llegará un momento en que se percibe una sensación de iluminación uniforme a lo largo del tiempo equivalente a la que obtendríamos si el proyector estuviera constantemente encendido. El nivel de iluminación subjetivo que percibe coincide con el valor medio de la señal utilizada para la excitación y se conoce como la ley de Talbot-Plateau. De acuerdo con esta ley, podemos controlar el nivel de luminancia de un proyector regulando el tiempo relativo en que las imágenes se proyectan en la pantalla o permanecen ocultas, es decir, el ciclo de trabajo sin necesidad de aumentar la potencia de la fuente de luz.

Entre la región en la que discernimos correctamente el carácter pulsado del proyector y la región en la que tenemos una sensación de continuidad completa, existe un margen de frecuencias en las que se aprecia un ‘parpadeo’ de la imagen que interpretaríamos como si el nivel de iluminación del proyector no fuera uniforme. Este fenómeno se conoce en terminología inglesa con el nombre de ‘flicker’, y debido a su frecuente uso en la literatura sobre sistemas de televisión, será usado a menudo en este texto.

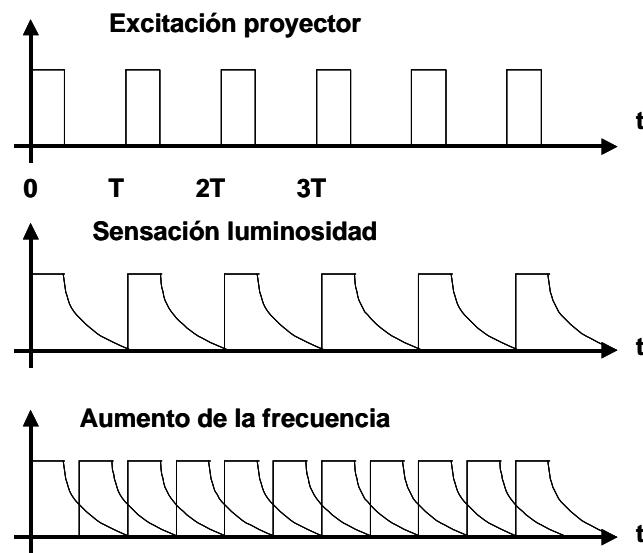


Fig. 1.53 Persistencia de las imágenes en la retina

La frecuencia a la que se empieza a desaparecer el parpadeo y las imágenes se funden en una sensación única de continuidad se denomina frecuencia crítica de *flicker* y depende del nivel de iluminación de la pantalla. Esta dependencia es de tipo logarítmico y se conoce como la Ley de Ferry-Porter:

$$f_{crítica} = 37 + 12,6 \log_{10} B \quad (1.58)$$

siendo B el nivel de brillo de la pantalla medida en *foot-lamberts*. Para tener una idea de esta magnitud diremos que con los niveles de iluminación que se utilizan en el cine o en el monitor de televisión para una visión cómoda, la frecuencia crítica es del orden de unos 40 Hz (es algo más baja en el cine debido a que las condiciones de iluminación de la sala están adecuadas al visionado de la cinta, por lo que no se requiere tanta potencia de iluminación de la pantalla).

El margen de valores de frecuencias críticas puede oscilar entre unos 60 Hz para condiciones de fuerte iluminación diurna hasta unos 4 Hz, en condiciones de iluminación nocturna y visión fotópica. Estas diferencias probablemente residen en el distinto comportamiento que presentan los conos y los bastones. En el primer caso se trata de elementos especializados en detectar niveles de iluminación elevados, por lo que se supone que una vez activados inhiben su respuesta rápidamente para estar preparados para recibir nuevos estímulos. En cambio, la respuesta de los conos es bastante lenta y puede durar hasta $\frac{1}{4}$ de segundo, por lo que si reciben un nuevo estímulo mientras están activados no se produce ninguna respuesta nueva, simplemente tardan más tiempo en desactivarse.

El fenómeno phi es el responsable principal de que el sistema visual humano sea capaz de interpolar movimientos de los que sólo dispone información fraccionada y producir la sensación de que son

continuos. Un ejemplo característico para ilustrar este tipo de respuesta del sistema visual consiste en situar dos focos de luz, uno al lado del otro, que se encienden alternativamente de modo que mientras uno está encendido el otro está apagado. Si la frecuencia a la que comutan las luces es suficientemente baja, el sistema visual interpreta correctamente la escena (alternancia del encendido de las luces). Sin embargo, si la frecuencia aumenta ligeramente (bastante por debajo del margen de *flicker*) el ojo interpretará que ‘la luz se mueve hacia delante y hacia atrás’. El mismo fenómeno es responsable de que los carteles luminosos sean interpretados sistemáticamente como un romántico ‘luces y figuras en movimiento’ en vez de, la más fría pero racional, ‘apagado y encendido de elementos electroluminiscentes’. El fenómeno se produce porque estas secuencias de imágenes parece que activan las neuronas encargadas de la detección de movimientos en las escenas, ya que la secuencia de imágenes obtenida, si bien en este caso es de naturaleza discreta, es parecida a la que se obtendría si realmente las luces se estuvieran desplazando. El mismo sistema es responsable de interpretar movimientos más complejos cuando se suceden secuencias de imágenes fijas que contienen información parcial sobre el movimiento, siendo este el principio que utilizaban los primeros discos y tambores rotatorios como el Zoetrope. Para una correcta sensación de continuidad del movimiento es necesario disponer de aproximadamente de algo más de 18 imágenes por segundo.

En este punto es necesario aclarar que aunque tanto la persistencia de la retina como el fenómeno phi intervienen conjuntamente en la percepción de continuidad de las imágenes, en sistemas como el cine o la televisión son conceptualmente distintas. El primero es el responsable de que observemos un nivel de iluminación constante, no discriminando la discontinuidad en la información, mientras que el segundo permite obtener la sensación de que los movimientos son continuos. Probablemente, una escena del cine de principios del siglo XX nos ayudaría a aclarar las diferencias entre ambos conceptos. En este caso, se presentaban 16 imágenes por segundo, con lo que aparecía un considerable parpadeo de la secuencia. Sin embargo, el movimiento podía ser interpolado con una precisión razonable. La sensación general era que los objetos se desplazaban de manera uniforme bajo una iluminación oscilante (el lector puede objetar que en muchas de las películas de esta época el movimiento parece excesivamente rápido y discontinuo). Esto es debido fundamentalmente a que las películas originales han sido convertidas directamente al formato televisivo, que al utilizar un mayor número de imágenes por segundo, produce la sensación de acelerar la acción. Consideraremos el problema de la conversión de películas de cine a señal de televisión en breve).

1.7.3 Frecuencia de imagen

Tanto en el cine moderno como en la televisión, los dos aspectos aparecen asociados al número de imágenes por segundo que se toman de la escena y al número de imágenes por segundo que se presentan al espectador; parámetros que, aunque en un principio pueda parecer que deberían de coincidir, no necesariamente debe ser así. En efecto, hemos comentado que en principio basta con una veintena de imágenes por segundo para poder estimar correctamente el movimiento de los objetos mientras que, con los niveles de iluminación utilizados en cine y televisión, se requieren más de 40 imágenes por segundo para que no aparezca el ‘parpadeo’. Una solución trivial al problema consiste en elegir un número de imágenes por segundo superior a 40, con lo que se cubren ampliamente ambos requerimientos. Sin embargo, en el cine, el precio del celuloide fue, y sigue siendo para la mayor parte de producciones con presupuesto moderado, un factor considerable en el computo total del coste de la película. Esto significa que no resulta conveniente aumentar el número de fotogramas por segundo

más allá de lo estrictamente necesario para que el sistema visual interprete correctamente el movimiento. En televisión el problema es parecido, aunque en este caso, el coste, más que al presupuesto, afecta al ancho de banda de la señal.

La solución adoptada en el cine fue ciertamente astuta y, aunque adaptada a las características del medio, se utilizó también como idea básica para reducir el ancho de banda en la señal de televisión. En el cine actual se realizan 24 fotogramas por segundo, lo que es más que suficiente para proporcionar una correcta sensación de movimiento continuado al espectador. Sin embargo, la presentación directa de estos fotogramas produciría una sensación de parpadeo notable. La solución al problema consiste en presentar dos veces cada fotograma, con lo que, al doblar la frecuencia de disparo del proyector se elimina completamente el *flicker* y se obtiene una sensación de uniformidad en la iluminación de la pantalla. La solución es simple y eficaz por cuanto basta con disponer de un obturador frente al proyector que se dispare con una frecuencia del doble de la velocidad de la cinta.

En un primer análisis parece que la solución de repetir las imágenes pueda empeorar la calidad de la reproducción, ya que al encender y apagar un mismo fotograma dos veces, el tiempo total que este permanece expuesto al espectador se reduce. Quizá podamos comprender de una manera más intuitiva este fenómeno si suponemos que realizamos el siguiente experimento: Disponemos de dos ventiladores uno con cuatro aspas y otro con dos que están girando a la misma velocidad (12,5 vueltas por segundo) a través de los cuales queremos observar una escena estacionaria. Al observar la escena por el ventilador de dos aspas notaremos un efecto de parpadeo de la escena mientras que con el de cuatro aspas este efecto desaparece. Podemos comprobar, por tanto, que aunque en el primer caso la escena aparece durante más tiempo expuesta al ojo, existe un parpadeo debido a que el número de impulsos no es suficientemente elevado.

Algunos sistemas de cine han utilizado un número mayor de fotogramas por segundo, aunque su uso no se ha generalizado. Un ejemplo son las primeras películas producidas en formato de 70 mm que utilizaron 30 fotogramas por segundo, aunque posteriormente se volvió a los 24 fotogramas convencionales. A principios de la década de los 80 apareció el sistema Showscan que utiliza 60 fotogramas por segundo. Este sistema se utiliza únicamente para documentales en salas de proyección especiales y ha tenido poco éxito.

La solución adoptada al definir los primeros sistemas de TV comerciales siguió una filosofía similar. En este caso, para simplificar el funcionamiento de algunos subsistemas, la frecuencia de imagen se hizo coincidir con la mitad de la frecuencia de red. Por tanto, en países como Estados Unidos o Japón se transmiten 30 imágenes por segundo mientras en Europa se utilizan 25. Ambos valores son suficientes para garantizar un flujo de información que permita reconstruir correctamente el movimiento de los objetos. La selección de submúltiplos de la frecuencia de red se debe principalmente a dos motivos que permitieron simplificar bastante los primeros diseños. Por una parte, la frecuencia de red se mantiene bastante estable, lo que permite que varios equipos la utilicen como patrón o señal de referencia común. Así, cuando en un estudio de TV operan simultáneamente varias cámaras es necesario que éstas realicen un muestreo sincronizado de las imágenes para que pueda comutarse entre distintas cámaras sin que se produzcan pérdidas de información. Otra ventaja derivada del uso de un submúltiplo de la frecuencia de red es que de este modo conseguía minimizarse el efecto visual que producía la interferencia de la señal de red en la pantalla de televisión.

Analizaremos estos efectos en capítulos sucesivos, aunque en cualquier caso conviene tener claro que la elección de la frecuencia de imagen se realizó en los albores de la electrónica y que probablemente, con la tecnología actual, los criterios a valorar para la selección de este parámetro hubieran sido distintos. El parpadeo de las imágenes se consiguió evitar dividiendo cada imagen en dos mitades que se transmitían una detrás de la otra utilizando una técnica de exploración conocida como entrelazado. Describiremos con detalle este concepto en la siguiente sección. Esencialmente consiste en que el receptor realice dos barridos de la pantalla por cada imagen presentando la mitad de la información en cada uno. Con ello se consigue que la frecuencia aparente de exploración de la pantalla se doble, desapareciendo el parpadeo de la imagen. Cada una de las mitades de la imagen se denomina campo o, en terminología anglosajona, *field*. Así, en Europa la frecuencia de campo es de 50 Hz mientras que en USA o Japón es de 60 Hz.

El entrelazado debe considerarse como una estrategia propuesta con el objeto de reducir la complejidad tecnológica empleada en los primeros receptores de televisión. Veremos que es responsable de algunos de los problemas de calidad de imagen que tienen los sistemas de televisión analógicos y que exige el uso de una compleja tecnología digital para su resolución en los monitores modernos.

La diferencia entre el número de imágenes por segundo que se toman en el cine y en los sistemas de TV requiere un tratamiento específico para la conversión de un formato a otro. El equipo encargado de realizar la conversión de película de cine a señal de televisión se denomina telecine. La mayoría de telecines permiten trabajar con películas en formato de 16 mm o 35 mm siendo raros los equipos que aceptan el formato de 70 mm. Normalmente, aunque se trate de películas rodadas en formato de 70 mm siempre existen copias en 35 mm, por lo que son estas últimas las que se usan para realizar la conversión de formatos. Básicamente el telecine dispone de un sensor de imagen mediante el que se convierte cada fotograma de la película en una imagen de televisión. Una vez la imagen ha sido adquirida, se separa en los dos campos y se transmite o se almacena en formato de vídeo sobre soporte magnético para su posterior transmisión. El problema básico es que el número de fotogramas por segundo en el cine es inferior al de la televisión, por lo que deben añadirse los fotogramas que faltan.

En Europa, la diferencia entre las 24 y 25 imágenes por segundo no es excesivamente grave y difícilmente se percibe por parte del espectador. Durante bastantes años, muchas emisoras europeas no disponían de equipos específicos para la conversión y estuvieron retransmitiendo películas asociando los fotogramas uno a uno. Mediante la conversión directa, una película de 120 minutos en su versión cinematográfica original, tendría una duración de poco más de 115 minutos en televisión. En otras palabras, las películas duraban un 4 % menos en TV que en el cine (este cálculo no tiene en cuenta, por supuesto, la publicidad). La conversión directa también afecta al sonido por cuanto, al tener que sincronizarse con la imagen, debe ser acelerado en la misma proporción por lo que experimenta un ligero incremento de tono que tampoco se percibe. No obstante, en los países que utilizan un formato de televisión de 30 cuadros por segundo, la conversión fotograma a fotograma de la película es totalmente inviable pues cambios temporales de un 20 % que serían perfectamente visibles y audibles. La solución adoptada en este caso depende del tipo de telecine. La más habitual consiste en repetir 6 cuadros en cada grupo de 24 fotogramas de la película original y distribuirlos de manera uniforme en el tiempo. Otra solución consiste en generar alternativamente 2 campos para los fotogramas impares y 3 para los pares. De este modo, de una secuencia de 24 fotogramas se generan 60 campos con lo que ambos sistemas funcionan sincronizados. Los telecines actuales para formato de 25 imágenes por

segundo suelen repetir el primer fotograma de cada grupo de 24 con lo que, actualmente, la duración de la película en ambos formatos coincide. La pista de audio, que en la mayoría de casos se suministra en una cinta magnética que acompaña a la película, suele sincronizarse procesando digitalmente las señales para corregir el tono o bien actuando sobre los silencios para mantener su sincronización con la señal de vídeo.

1.7.4 Aliasing temporal en las imágenes

Uno de los fenómenos que puede introducir el muestreo temporal de la escena es la aparición de efectos visuales que falseen o desvirtúen el movimiento real de los objetos. Estos efectos son debidos a que la frecuencia de muestreo elegida resulta ser insuficiente para una correcta interpretación del movimiento. Así, tal y como establece el teorema del muestreo de señales, si los movimientos de los objetos son excesivamente rápidos, pueden existir componentes de alta frecuencia que, al muestrear la escena, se solapan con otras componentes desvirtuando el movimiento real de los objetos.

Uno de los ejemplos típicos de este fenómeno aparece en muchas películas de acción y en ocasiones es incluso provocado por el director para producir en el espectador la sensación de movimiento. La escena consiste en centrar la acción de la cámara en la rueda de un carro (en películas del oeste) o de un automóvil. Cuando el vehículo arranca el espectador observa e interpreta correctamente el movimiento de la rueda, sin embargo, a medida que la velocidad aumenta se produce la sensación de que los radios de la rueda se paran o incluso giran en sentido contrario al de la marcha del vehículo. Este efecto es debido, como se ilustra en la figura 1.54, a que el ojo interpreta el movimiento de la forma más simple posible a partir de los fotogramas que se le presentan. En la figura se representan dos fotogramas consecutivos de la rueda del carro que corresponden a dos situaciones distintas pero que, inevitablemente, serán interpretadas del mismo modo, ya que las imágenes que las definen son idénticas. En la secuencia superior, la rueda del carro experimenta un giro de α rads entre dos fotogramas consecutivos, mientras que en la inferior el giro real de la rueda es de $\alpha + \pi/4$ rads. El sistema visual interpretará en ambos casos el mismo movimiento aparente de α rads ya que la simetría de la rueda no le permite distinguir entre las dos secuencias. En este ejemplo, si la velocidad angular de la rueda fuera tal que entre dos fotogramas consecutivos se produjera un giro de $\pi/4$, el espectador no apreciaría ningún movimiento, ya que todas las imágenes serían idénticas. Se produce un caso curioso cuando el ángulo de giro entre fotogramas es ligeramente inferior a $\pi/4$, ya que en este caso se interpreta que el movimiento de la rueda es en sentido contrario al real. En este último supuesto es evidente que para el sistema visual resulta mucho más factible interpretar que se ha producido un pequeño movimiento de la rueda en sentido contrario al real que un movimiento considerable en el sentido correcto.

Este efecto visual se conoce como el efecto estroboscópico y es a veces provocado para poder observar movimientos de naturaleza periódica de algunos fenómenos físicos cuya rapidez no permite observarlos en condiciones normales. El instrumento utilizado para ello se denomina estroboscopio y consiste simplemente en una lámpara (*flash*) que se ilumina a intervalos regulares y cuya frecuencia y fase de disparo puede ser controlada externamente. El estroboscopio puede utilizarse, por ejemplo, para observar detalladamente la deformación que experimenta la membrana de un altavoz cuando es sometido a distintas excitaciones periódicas. Para realizar el experimento deberemos alimentar el altavoz con una señal sinusoidal a la frecuencia en la que deseemos realizar el análisis. El sistema

debe situarse en un entorno con poca iluminación, de modo que al disparar el estroboscopio a la misma frecuencia a la que está vibrando el altavoz podremos observar una imagen estacionaria de la deformación de la membrana. Si modificamos la fase de los disparos, veremos las deformaciones progresivas que va tomando la membrana del altavoz en las distintas fases de la vibración. Incluso, si seleccionamos una frecuencia de disparo del flash ligeramente distinta a la de la vibración del altavoz podremos ralentizar el movimiento de la membrana a la velocidad deseada. Se propone que el lector se plantea que posibles frecuencias del estroboscopio utilizaría para poder observar el movimiento de una membrana que oscila a 300 Hz con una velocidad aparente de 0.1 Hz. Téngase en cuenta que por una parte es conveniente que la frecuencia del estroboscopio permita observar el movimiento sin parpadeo y por otra que la frecuencia de disparo no puede ser excesivamente elevada por la limitación propia del *flash*.

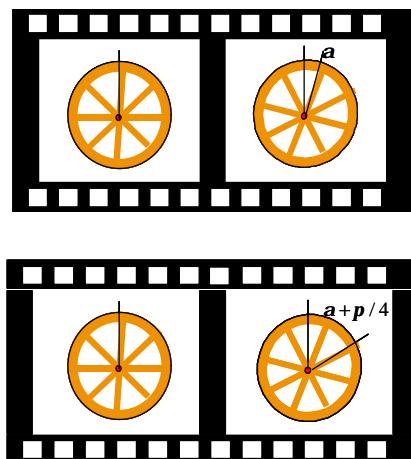


Fig. 1.54 Aliasing temporal de las imágenes. En ambas situaciones, las secuencias de imágenes obtenidas son idénticas, por lo que el sistema visual interpretará el mismo movimiento

Experiencias parecidas pueden observarse en diversos museos científicos, entre ellos el de Barcelona, en los que se incluyen diversas aplicaciones y efectos visuales obtenidos mediante la ayuda del estroboscopio. Una de las más espectaculares consiste en ‘ralentizar la caída libre de las gotas de agua’. En este caso, se utiliza una cámara cerrada en la que se provoca un goteo constante de agua. Al iluminar el flujo de gotas de agua mediante un estroboscopio podemos provocar que éste se ilumine siempre que una gota de agua pase por una posición determinada, por lo que observaremos que la gota permanece suspendida en el aire. Una ligera disminución o aumento de la frecuencia de disparo producirá la sensación de que la gota cae o sube lentamente en el medio. Pueden realizarse experimentos similares para visualizar el vuelo de insectos, analizar las vibraciones de máquinas, etc.

1.8 Muestreo espacial de las imágenes: filas

Esta es ya la última fase de muestreo de la escena antes de poder convertirla a una señal eléctrica analógica que aporte la información necesaria para su transmisión y reproducción. Además, esta es

una de las pocas etapas de muestreo y proyección de las escenas que el cine y la televisión no comparten, ya que el primer sistema finaliza sus fases de representación de la escena con el muestreo temporal comentado en el apartado anterior. De hecho, el soporte del cine es claramente bidimensional (la película es una secuencia finita de fotogramas planos), mientras que en televisión necesitamos realizar un muestreo adicional para obtener una señal unidimensional que pueda ser transmitida mediante un sistema de comunicación convencional.

La idea básica del muestreo espacial se ilustra en la figura 1.55, donde se representa un fotograma en blanco y negro que se explora en la dirección horizontal con un número de líneas predeterminado. En este caso la imagen se ha descompuesto en 25 líneas con el objeto de que pueda apreciarse el efecto de un muestreo inadecuado sobre la calidad de la imagen. El muestreo se ha simulado utilizando técnicas de procesado digital de imagen en cuyo estudio profundizaremos en capítulos posteriores. La imagen original que se muestra en esta figura está muestreada tanto en el eje vertical como en el horizontal, no obstante, el número de muestras es muy elevado, por lo que puede considerarse como si tuviera naturaleza continua.

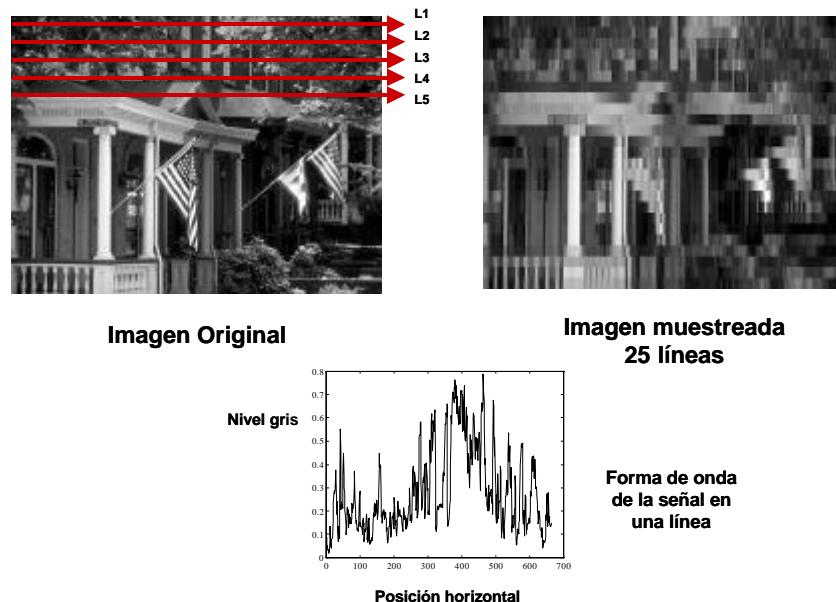


Fig. 1.55 Muestreo de una imagen por filas

Siguiendo con nuestro modelo de la escena, cada uno de los fotogramas obtenidos en la expresión de la ecuación 1.57, puede descomponerse en un número prefijado de líneas:

$$w_{FX}^{(k)}(m, y') = \varepsilon_{FX}^{(k)}(m\Delta x, y') \quad (1.59)$$

donde $m=1,..,NL$; siendo NL el número de líneas y $\Delta x=H/NL$ donde H representa la altura de la pantalla. Hemos sustituido el símbolo ε asociado a la escena por la variable w para indicar

explícitamente que la primera ha sido completamente transformada a un conjunto numerable de funciones unidimensionales. En la misma figura 1.55 se ha representado la forma de onda asociada a una de las funciones $w_{FX}^{(k)}(m, y')$, que en este caso concreto corresponde a la k-ésima imagen en blanco y negro ($X=Y$) representada en la derecha en la línea $m=17$. Para representar esta función se ha adoptado el convenio de asignar al negro un nivel de amplitud nulo y al blanco un nivel de amplitud unidad. El resto de niveles de gris está uniformemente distribuido entre estos límites.

Con esta última etapa de muestreo hemos conseguido expresar la escena en función de una única variable de naturaleza continua (y') y cuyo dominio de definición es finito (entre 0 y W , la anchura de la imagen). El resto de variables: m , k y la que indica la componente de color o luminancia X , son de naturaleza discreta y sugieren la posibilidad de transmitir la información que de ellas depende de manera secuencial, enviando la información hacia el receptor línea a línea e imagen a imagen. En el caso de transmitir imágenes en blanco y negro, la secuencia natural para enviar la información al receptor sería del tipo:

$$\dots, w_{FY}^{(k)}(m, y'), w_{FY}^{(k)}(m+1, y'), \dots, w_{FY}^{(k)}(NL, y'), w_{FY}^{(k+1)}(1, y'), w_{FY}^{(k+1)}(2, y'), \dots \quad (1.60)$$

donde se establece claramente una posible cadencia para la transmisión de líneas e imágenes.

Esta idea se esboza en la figura 1.56, donde se observa como puede construirse una señal que aporte toda la información contenida en una supuesta secuencia de imágenes en blanco y negro, encadenando todas las formas de onda de las líneas en las que se divide una imagen y repitiendo posteriormente este proceso para cada una de las imágenes. En este caso la información contenida en cada una de las líneas es del mismo tipo que la línea representada en la figura 1.55, aunque para simplificar el gráfico cada línea se ha representado simbólicamente como un pequeño hexágono alargado. Es importante observar en este punto que la idea que estamos sugiriendo para la construcción de una señal que represente a la escena es sólo una primera aproximación a la señal de televisión, pues no estamos teniendo en cuenta, por el momento, aspectos como los sincronismos o el entrelazado.

En la figura 1.56 se ha realizado, implícitamente, una concatenación de las distintas funciones de la ecuación 1.60 en el dominio temporal. Para ello, es necesario realizar un cambio de variable que convierta las funciones que dependen de una variable espacial en temporal. Es importante observar que si deseamos realizar una transmisión en tiempo real, es necesario que el tiempo que transcurre entre la transmisión de la primera línea de una imagen y la siguiente coincida con el periodo de muestreo temporal de la escena T_0 , que en el caso de la televisión europea será de 40 ms (25 imágenes por segundo). Asimismo, la duración de una línea deberá ser de T_0/N_L , donde N_L es el número total de líneas en el que se descompone la imagen. De este modo, todas las líneas pueden codificarse en el tiempo que transcurre entre dos imágenes consecutivas.

Así pues, suponiendo que redefinimos el origen de tiempo al inicio de cada línea, el cambio de variable a realizar sería:

$$y' = \frac{NL}{T_0} \cdot W \cdot t' \quad (1.61)$$

de modo que $t'=0$ corresponde con el inicio de la línea y el valor $t'=T_0/NL$ (duración de la línea) corresponde con el final de la línea o la anchura de la imagen.

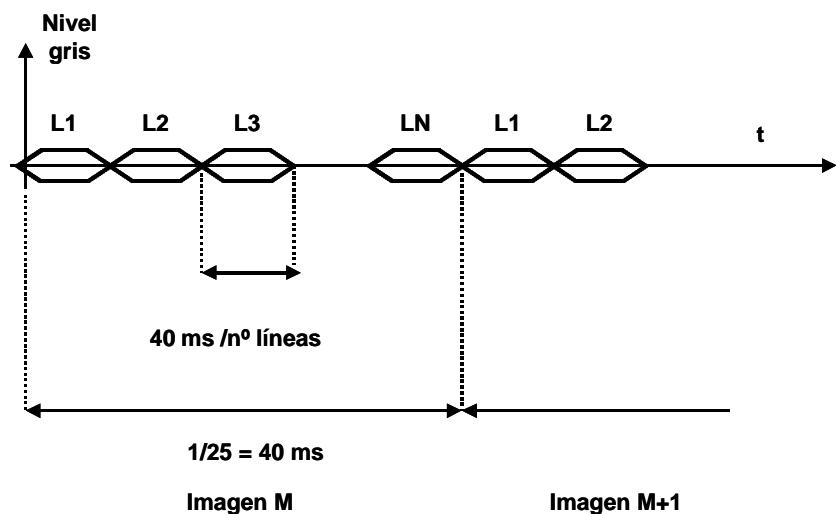


Fig. 1.56 Principio de codificación de la escena en una señal unidimensional

La expresión general de la señal en función de una única variable temporal global es algo más compleja y no es excesivamente importante, siempre el lector haya comprendido la idea intuitiva sobre la construcción de esta señal. De todos modos, para completar nuestra particular notación sobre la escena y la señal de televisión diremos que en un instante genérico t , la señal tomará el valor:

$$w_{FY}^{(k)}(m, t') \quad (1.62)$$

donde los distintos parámetros pueden calcularse en función de t de acuerdo con las siguientes expresiones:

$$(k) = \text{int}\left(\frac{t}{40 \text{ ms}}\right) + 1; \quad m = \text{mod}\left(\text{int}\left(\frac{t \cdot NL}{40 \text{ ms}}\right) NL\right) + 1; \quad (1.63)$$

$$t' = t - (40 \text{ ms}) \cdot (k - 1) - \frac{(40 \text{ ms})}{NL} \cdot (m - 1)$$

En estas expresiones $\text{int}()$ indica la parte entera, $\text{mod}(n,m)$ el resto de dividir el entero n por m y 40 ms es el periodo asociado a una frecuencia de imagen de 25 Hz. Hemos supuesto que el origen de tiempos se corresponde con la imagen $k=1$. Consideremos como ejemplo el instante de tiempo $t=41,3$ ms en un sistema con 80 líneas. En este caso, cada línea debería tener una duración de 0,5 ms, por lo que para este valor de t habrían transcurrido 1,3 ms de la segunda imagen y estaríamos a 0,3 ms del inicio de la tercera línea. Para este valor de tiempo y con $NL=80$, podemos comprobar que los parámetros de las ecuaciones anteriores valen: $k=2$, $m=3$ y $t'=0,3$ ms que concuerdan con lo previsto.

Queremos insistir en que esta idea general sobre la construcción de la señal de televisión requiere aún varios refinamientos y consideraciones sobre los distintos parámetros que pueden afectar a la calidad de la señal de televisión. A saber, cuál es el número de líneas apropiado para que el sistema tenga una

calidad aceptable, cómo afecta a esta idea el querer transmitir una señal de color, cómo se informa al receptor a qué parte de la imagen corresponde cada componente de la señal que se transmite, cómo afecta el número de líneas al ancho de banda, cómo modificar esta representación de la escena para incluir el entrelazado de los campos necesario para evitar el parpadeo de la imagen, etc. El objetivo de esta sección es dar una respuesta de carácter generalista a todas estas cuestiones. Los detalles concretos sobre formas exactas de la señal, modulaciones, etc., se dejarán para capítulos posteriores.

Sin embargo, antes de empezar a considerar estos detalles, es el momento oportuno de realizar una rápida revisión histórica sobre los orígenes de la televisión y los primeros prototipos que se consideraron para la transmisión de imágenes.

1.8.1 Antecedentes de la televisión

El primer sistema para la transmisión a distancia de imágenes en movimiento fue desarrollado en 1879 por G.R. Carey y se basaba en enviar la información recogida en un panel de fotocélulas, que actuaba como sensor de imagen, a través de múltiples cables hacia una pantalla de lámparas en la que se presentaba esta información al receptor. El prototipo original tenía un total de 2500 elementos en cada uno de los paneles, distribuidos en una matriz de 50×50 . Es evidente que la resolución obtenida era bastante baja a pesar del gran número de cables utilizado que hacía inviable la transmisión de la información a grandes distancias, ya que el número de amplificadores requeridos sería muy elevado. Por todo ello, este prototipo aún no puede ser considerado como un verdadero sistema de televisión, pues la naturaleza de la información no estaba todavía adecuada a su transmisión a través de un único canal.

En 1879, Constantin Senlecq propuso una importante modificación sobre el sistema anterior, que consistía en commutar rápidamente el canal activo de modo que sólo se transmitía simultáneamente la información recogida por un fotosensor, alternando éstos en el tiempo para transmitir toda la imagen. Esta idea se basaba en la persistencia del sistema visual humano, de manera que si el tiempo que transcurría entre activations sucesivas de una misma lámpara era pequeño, el ojo no apreciaba la interrupción de la luz.

Esta propuesta supuso el verdadero origen de la televisión, puesto que a pesar de que los subsistemas de exploración y presentación de imagen seguían teniendo serias limitaciones, la información se había adecuado a las características necesarias para su transmisión a distancia pudiendo transmitirse a través de un único canal. Considérese como ejemplo, que si se desea transmitir la información a través de un cable, la atenuación puede evitarse sólo con introducir un único amplificador en cada uno de los tramos en que se requiera restaurar el nivel de la señal. Esto sin duda simplifica notablemente la electrónica necesaria respecto a su precursor. La idea de Senlecq de transmitir la información de manera secuencial se ha mantenido en todos los sistemas de televisión surgidos posteriormente y probablemente se mantendrá para todos los sistemas que surjan en el futuro. Piénsese que los primeros sistemas de comunicación a distancia surgieron para la transmisión de datos (telegrafía) o voz (telefonía); señales que son claramente funciones de una única variable tiempo. La televisión fue el primer sistema en el que la información a transmitir dependía de varias variables y aunque seguramente hoy podrían proponerse sistemas de gran elegancia analítica para transmitir todas estas variables simultáneamente, en la práctica, la filosofía utilizada es adecuada a las características de

nuestro sistema visual y tecnológicamente viable desde hace varias décadas. La propuesta de Senlecq, a pesar de todo, no era simple para la tecnología de la época y el primer prototipo no se realizó hasta cincuenta años más tarde. Este primer prototipo usaba 16 imágenes por segundo, que se presentaban sobre un panel de 50×50 lámparas, por lo que la frecuencia de commutación del canal era de $16 \times 2500 = 40$ KHz.

La siguiente propuesta históricamente relevante surgió en 1884, cuando Pablo Nipkow propuso un sistema mecánico que permitía usar una única célula fotoeléctrica y que simplificaba significativamente el sistema de exploración de la imagen. Los sistemas de exploración mecánica tenían una resolución bastante baja y fueron rápidamente sustituidos al aparecer los primeros sistemas electrónicos. No obstante, tienen una considerable importancia histórica, ya que las primeras transmisiones de televisión se realizaron mediante estos sistemas, por lo que expondremos brevemente su principio de funcionamiento.

La primera implementación práctica del sistema propuesto por Nipkow se realizó en 1925 y su filosofía básica se esboza en la figura 1.57. La escena se enfoca mediante un sistema óptico sobre una región plana del espacio con forma rectangular detrás del cual se encuentra una célula fotoeléctrica de selenio que recoge la cantidad de luz incidente en el rectángulo. En la zona en la que se forma la imagen se inserta un disco giratorio que tiene practicados unos pequeños orificios distribuidos sobre una línea espiral como la indicada en la figura 1.57.

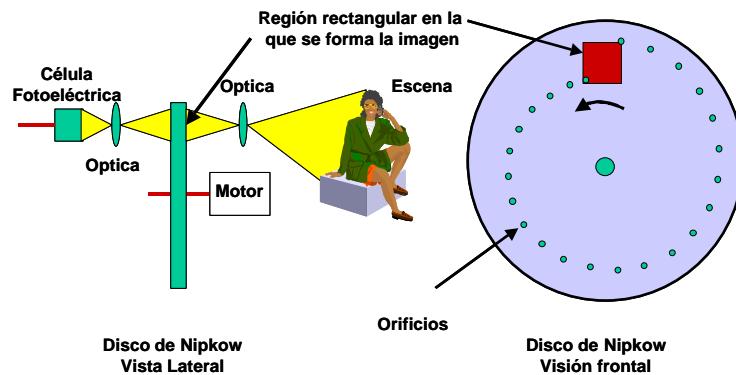


Fig. 1.57 Procedimiento mecánico para la exploración de la imagen mediante el disco de Nipkow

Estos orificios sólo dejan pasar hacia la célula fotosensible la parte de la imagen que se está explorando, ocultando el resto. Existe un segundo grupo óptico situado entre la célula de selenio y el disco que se encarga de asegurar que la luz que pasa a través del disco se concentra en la superficie sensora. La disposición de los orificios permite realizar una exploración de las líneas de la imagen tal y como se indica en la figura 1.58. Cuando el primer orificio ha pasado completamente por la primera línea del rectángulo de exploración, el siguiente orificio entra por la parte izquierda y explora la siguiente línea. La célula fotosensible, situada detrás del disco, sólo recibe la luz que deja pasarse a través de estos orificios.

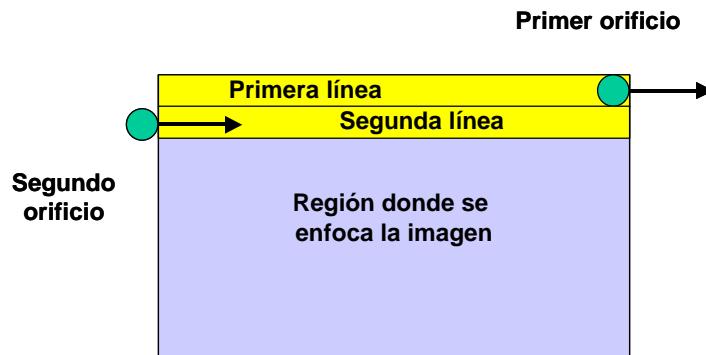


Fig. 1.58 Exploración de las líneas mediante los orificios del disco de Nipkow

La velocidad de giro del disco determina el número de imágenes por segundo y la cantidad de orificios, el número de líneas de cada imagen. El primer prototipo disponía de un total de 28 orificios y el disco giraba a 12,5 r.p.m. Nótese que la señal que se obtiene en la salida de la célula tiene unas características parecidas a la de la señal descrita en la sección anterior, por lo que este curioso mecanismo realiza directamente el muestreo temporal y espacial de la imagen y su posterior conversión a una señal temporal.

Nipkow diseñó un sistema análogo para la presentación de las imágenes. En este caso, la señal recogida por la célula sensora se utilizaba para activar una lámpara con una intensidad variable. El disco usado en la reproducción de las imágenes se encargaba ahora de proyectar la luz incidente sobre una pantalla en la que el espectador podía ver la escena. El sistema de reproducción no tuvo tanto éxito como el de exploración debido a que anteriormente ya se había desarrollado el tubo de rayos catódicos, que tenía unas prestaciones bastante superiores para la visualización de imágenes. El tubo de rayos catódicos fue inventado por Braun en 1897 y posteriormente mejorado por Wehnelt, que introdujo el cañón electrónico en 1903. Boris Rosing en 1907 fue el primero que logró usar el tubo de rayos catódicos para la presentación de imágenes de televisión.

El éxito del sistema de exploración mecánico de Nipkow se debe a que los primeros sistemas de exploración totalmente electrónica no aparecieron hasta 1932. En 1929 se realizaban transmisiones esporádicas de televisión, usando el sistema de exploración mecánico de Nipkow y transmitiendo la señal en onda media, en Alemania e Inglaterra.

La aparición del iconoscopio, desarrollado por Vladimir Zworykin en 1932, supuso un importante cambio en el sistema de exploración de la imagen mejorando significativamente la calidad. El iconoscopio utilizaba una hoja que contenía un gran número de condensadores elementales sensibles a la luz y que realizaban una función parecida a la de los receptores del sistema visual humano. Los condensadores, al ser expuestos a la luz, adquirían una carga que dependía de la intensidad de luz incidente. La lectura de la carga se realizaba utilizando un procedimiento de exploración parecido al de un tubo de rayos catódicos, de modo que la corriente generada cuando el haz se dirigía hacia un determinado punto de la hoja era proporcional a la luz incidente en esa región. El sistema, que se

representa en la figura 1.59, tenía suficiente sensibilidad para trabajar en condiciones de luz diurna normal.

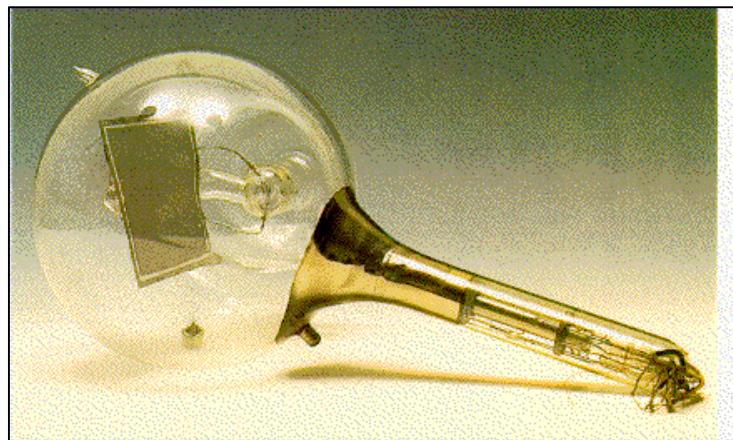


Fig. 1.59 Iconoscopio

La primera transmisión de televisión en blanco y negro fue realizada en 1936 con ocasión de la inauguración de los juegos olímpicos de Berlín. Durante este mismo año aparecen los primeros programas comerciales transmitidos por la BBC en Gran Bretaña. En Estados Unidos, las primeras transmisiones y programas regulares empezaron en 1939, con motivo de la Exposición Universal de Nueva York. En la figura 1.60 se muestra un equipo receptor de televisión del año 1936.



Fig. 1.60 Receptor de televisión del año 1936

Las primeras pruebas con señales en color se iniciaron en Estados Unidos en 1941. La calidad de las imágenes era muy baja, pero estas pruebas favorecieron el desarrollo de los sistemas de color que conducirían a la definición del sistema NTSC en 1954. Los sistemas de televisión en color europeos no aparecieron hasta los inicios de la década de los 60.

Anteriormente, en 1928, el ingeniero escocés John Baird había construido un prototipo de televisión en color que utilizaba un conjunto de tres filtros, rojo, verde y azul, que giraban mecánicamente delante de una pantalla de blanco y negro. Baird diseñó un ingenioso sistema para sincronizar las imágenes con los filtros, demostrando los principios básicos de la formación del color. El tubo de rayos catódicos para la presentación de imágenes en color fue propuesto por el alemán Werner Flechsig en 1938. El primer tubo de cámara que almacenaba la información en elementos semiconductores fue presentado por la RCA en 1952. Este tubo, denominado Vidicon, sustituyó rápidamente a sus predecesores y se convirtió en un estándar durante bastantes décadas.

El primer receptor de televisión realizado completamente con transistores fue realizado por Sony en 1960. En la década de los 60 se iniciaron las primeras transmisiones de señal de televisión vía satélite. Estas primeras conexiones se realizaban entre centros de distribución de señal para realizar programas o noticiarios de ámbito internacional. Las señales se distribuían por vía terrena a los espectadores. La radiodifusión directa de señales de televisión por satélite se introdujo hacia 1980. Hacia finales de esta misma década se presentaron los primeros sistemas de televisión de alta definición en Japón. En Europa se intentó introducir un nuevo sistema de televisión, con una definición mejorada, orientado a las transmisiones vía satélite y que permitiera una transición gradual hacia la televisión de alta definición. El sistema, que se denominó D2-MAC, no tuvo excesivo éxito y el proyecto fue abandonado. Durante la década de los 90 aparecieron los primeros sistemas de radiodifusión de señales de televisión en formato digital. El formato empleado para la codificación de las imágenes se conoce con el nombre de MPEG-2 y ha sido ampliamente aceptado. Actualmente el MPEG-2 se utiliza en el sistema europeo para la radiodifusión de vídeo digital (DVB – Digital Video Broadcasting) en todas las variantes: vía satélite, cable y terrena. También se utiliza en la codificación de vídeo y audio en sistemas de disco compacto de alta densidad como el DVD (Digital Versatile Disc). El MPEG-2 también contempla futuras evoluciones de los sistemas de televisión hacia el concepto de alta definición.

1.8.2 Número de líneas

El número de líneas en que se descompone cada imagen influye directamente en dos parámetros básicos de la señal de televisión: la calidad y grado de detalle de la imagen en el eje vertical y el ancho de banda de la señal. Es evidente que para tener una buena percepción de la imagen, el número de líneas deberá ser suficientemente elevado como para que el sistema visual no sea capaz de distinguir entre la imagen original y la imagen muestreada. La situación representada en la figura 1.55 corresponde a un claro ejemplo en el que, siempre que la imagen se observe a una distancia de lectura normal, el número de líneas parece ser insuficiente, produciéndose una considerable pérdida de calidad respecto a la imagen original. Por otra parte, es conveniente mantener el número de líneas dentro de unos límites razonables, por cuanto su aumento tiene una incidencia directa sobre el ancho de banda de la señal de TV. En efecto, podemos ver que un incremento del número de líneas supone,

de acuerdo con el diagrama de codificación de la señal de TV de la figura 1.56, una reducción del tiempo asignado a la información de cada línea, por lo que la cantidad de información a transmitir por segundo y, por tanto, el ancho de banda de la señal, aumenta. Así pues, el número de líneas deberá elegirse como un compromiso adecuado para conseguir una buena calidad en las imágenes reproducidas y a la vez mantener el ancho de banda de la señal dentro de unos límites razonables.

En la elección del número de líneas de un sistema de televisión intervienen diversos factores cuya incidencia examinaremos detenidamente. Probablemente, el factor esencial es la propia capacidad del sistema visual humano para discernir los detalles en una imagen bajo determinadas condiciones de iluminación. Es evidente, por tanto, que no tiene sentido aumentar el número de líneas de una imagen más allá de lo que el ojo es capaz de discernir, ya que en este caso, un incremento de información en la imagen no supone ninguna mejora subjetiva en su calidad.

1.8.3 Agudeza visual

La capacidad del sistema visual humano para distinguir el detalle en las imágenes suele medirse como la capacidad de discriminar dos estímulos lumínicos próximos. En general, si presentamos a un observador una retícula de líneas negras sobre un fondo blanco, éste será capaz de identificar correctamente la composición de la imagen siempre que la distancia entre las líneas negras sea superior a la resolución del sistema visual. Sin embargo, si alejamos la retícula del observador, llegará un punto en que éste será incapaz de distinguir esta imagen de la de un papel gris, ya que el ojo integrará el patrón reticular en una sensación grisácea uniforme en la que no se aprecia el detalle de su composición.

La medida de la capacidad de resolución del ojo se realiza mediante retículas de estas características bajo distintas condiciones de iluminación. El *poder de resolución* del sistema visual humano se define como el ángulo subtendido por dos líneas negras en el ojo cuando el patrón reticular está situado a la distancia límite en la que aún se aprecia el detalle de las líneas que lo forman. El inverso de este ángulo expresado en minutos de grado se define como la *agudeza visual* del ojo. Así, una agudeza visual igual a la unidad indica que el sistema visual es capaz de distinguir dos estímulos separados por un ángulo (con vértice en el ojo) igual o superior al minuto de grado. Ambas definiciones se ilustran en la figura 1.61, en donde, por simplicidad, únicamente se representan las líneas horizontales de la retícula.

La agudeza visual muestra una fuerte dependencia con la iluminación. Así, cuando tanto la retícula como el entorno tienen una fuerte iluminación diurna, la agudeza visual puede llegar a un máximo de hasta 2 unidades (es decir, 30'' de poder de resolución). Para condiciones de iluminación diurna normales, la agudeza visual se sitúa alrededor de la unidad, mientras que en condiciones de iluminación nocturna (visión escotópica) puede reducirse hasta 0,04 unidades, es decir, sólo pueden distinguirse dos líneas cuando están separadas por un ángulo de 25' (casi $\frac{1}{2}$ grado). Para las condiciones normales de visionado de la señal de televisión, suele suponerse que la agudeza visual está situada alrededor del valor unidad.

A parte de esta retícula de prueba, existen otras alternativas para medir la agudeza visual. La más usada en laboratorios es la medida a través de la llamada C de Landolt. En este caso, se proyecta una

figura circular que tiene una discontinuidad en una determinada región, por lo que adopta una forma parecida a la C. Esta figura se presenta con diferentes tamaños y rotaciones al observador, pidiéndole que identifique la posición en la que se encuentra la discontinuidad. Cuando para un tamaño arbitrario, el porcentaje de errores al apreciar la posición de la discontinuidad excede un umbral predeterminado, se considera que se ha alcanzado el límite de resolución del ojo. En este caso, el poder de resolución se toma como el ángulo subtendido desde el ojo a los dos extremos de la discontinuidad. En el caso del oftalmólogo o el optometrista, la agudeza visual se mide mediante el conocido patrón de Snellen, que consiste en una serie de hileras con letras, cuyos tamaños se reducen progresivamente. La agudeza visual del observador se obtiene para aquel tamaño de letra en el que el observador es capaz de distinguir claramente cinco letras sobre un total de seis. En este caso la medida de la agudeza visual se presenta en una forma un tanto confusa, ya que se denota como cocientes en los que el numerador indica la distancia en metros desde la que se observa patrón de letras y el denominador, la distancia a la que los detalles de las letras corresponderían a un ángulo de un minuto de grado. Según este criterio, una medida de 6/6 corresponde con una agudeza visual de una unidad y una 6/12 a media unidad, cuando el patrón de Snellen se sitúa a 6 metros del espectador.

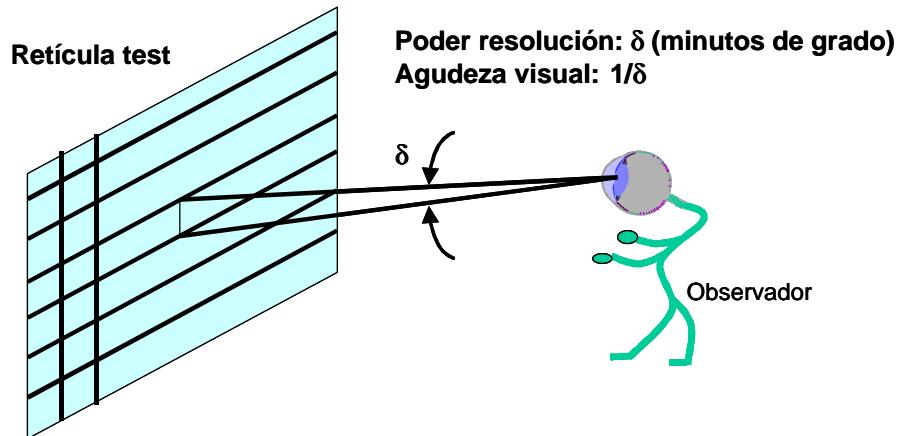


Fig. 1.61 Poder de resolución y agudeza visual

Desde un punto de vista fisiológico, la resolución del ojo está condicionada por la densidad de sensores lumínicos en la retina y por el mecanismo de transmisión de las señales a través del nervio óptico. El ángulo subtendido entre dos conos contiguos, situados en la región central la fóvea y medido desde el centro del cristalino, es aproximadamente de $30''$, lo que concuerda con la resolución máxima del sistema visual en condiciones de fuerte iluminación. Sin embargo, los mensajes que circulan a través del nervio óptico hacia el cerebro suelen ser combinaciones de respuestas de varias células sensoras, por lo que la resolución efectiva del sistema suele ser inferior a la densidad de elementos sensores. Recuérdese que el número total de mensajes enviados al cerebro es mucho menor que el número de elementos sensores, lo que indica que una parte importante del proceso y análisis de la información visual se realiza en las primeras etapas de la visión mediante la interconexión de neuronas sensibles y neuronas de proceso que se canalizan a través del nervio óptico hacia el cerebro. De hecho, parece ser que los mensajes de las células situadas en la región central de la fóvea (las que

se activan bajo condiciones de fuerte iluminación diurna) se transmiten uno a uno a través del nervio óptico, lo que justificaría que la agudeza visual conseguida cuando operan únicamente estos conos sea muy elevada. Sin embargo, a medida que los conos se alejan de esta región central, cada vez se realiza un mayor tratamiento previo de la información, con lo que disminuye la resolución a medida que empiezan a intervenir este tipo de células. El caso límite se produce en condiciones de muy baja iluminación, en la que se supone que la visión es debida a la contribución simultánea de varios bastones. Bajo este supuesto, la resolución se reduce notablemente, ya que únicamente se envían los mensajes combinados correspondientes a grupos de varios bastones, con lo que la densidad eficaz de los mismos se reduce.

1.8.4 Distancia de visionado

El número de líneas necesario para que el sistema visual no pueda distinguir entre la imagen muestreada y la original dependerá del ángulo vertical con el que se observe la imagen desde el ojo. Así, cuando observamos la imagen de la derecha en la figura 1.54 desde una distancia normal para la lectura del texto, unos 40 cm, podemos apreciar que el muestreo realizado es insuficiente, ya que el ojo percibe con claridad las discontinuidades verticales de la imagen. Sin embargo, si alejamos la figura varios metros, llegará un punto en que seremos incapaces de distinguir entre la imagen de la derecha y la de la izquierda. Esto ocurrirá cuando la distancia a la que está situado el papel es tal que el ángulo con el que se ve la imagen es de 25 minutos de grado, ya que a partir de esta distancia el ángulo subtendido por cada una de las líneas de la imagen es inferior a la resolución espacial del ojo.

En resumen, pues, el número de líneas de la imagen viene determinado, tal y como se indica en la figura 1.62, por el ángulo vertical con el que se observará la altura de la pantalla de televisión. Teniendo en cuenta que la resolución espacial media del sistema visual humano en las condiciones de iluminación de una señal de televisión es del orden de un minuto de grado, el número de líneas necesario para una correcta calidad de la imagen coincidirá con el ángulo vertical de observación de la pantalla expresado en minutos. No tendrá sentido superar este número de líneas, por cuanto supondría un aumento del ancho de banda de la señal sin ninguna mejora apreciable en la calidad de la imagen.

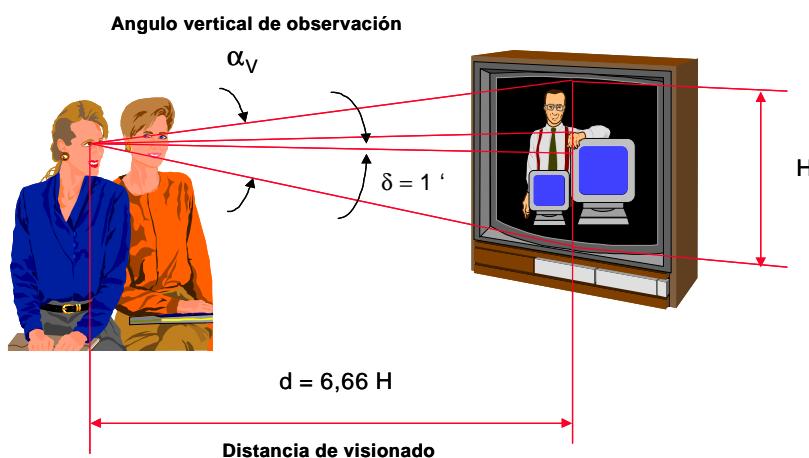


Fig. 1.62 Ángulo vertical de observación y distancia de visionado

Queda por tanto establecer cuáles son los ángulos más apropiados y agradables al espectador para observar una pantalla de televisión. En los inicios de la televisión se realizaron diversos estudios para determinar cuál sería la distancia más cómoda para ver la pantalla de televisión.

La decisión que finalmente se tomó en los primeros sistemas de televisión fue diseñarlos para que fueran observados desde una distancia que estuviera entre 6 y 7 veces la altura de la pantalla. Esta decisión se basaba en estudios estadísticos sobre la distancia a la que, libremente, se sitúa un espectador para contemplar una obra pictórica y que es, independientemente del tamaño del cuadro, de unas 4 veces la diagonal del marco. Dado que, para mantener compatibilidad con el cine, la relación de aspecto de la pantalla estaba fijada en 4:3; la relación entre la distancia de visionado y la altura de la pantalla será:

$$d \approx 4 \cdot \sqrt{H^2 + \left(\frac{4H}{3}\right)^2} = 6,66 \cdot H \quad (1.64)$$

De acuerdo con esta relación, un televisor de 28'' debería ser observado desde una distancia aproximada de 3 metros.

1.8.5 Número mínimo de líneas

Al diseñar el sistema de televisión para que sea observado desde una distancia que mantenga una relación proporcional a la altura de la pantalla, el ángulo vertical con el que se observa el monitor, y por tanto, el número de líneas que deben usarse para una correcta visualización de la imagen, son independientes del tamaño del televisor. El ángulo vertical de observación de la pantalla vendrá dado, de acuerdo con la figura 1.63, por:

$$\alpha_v = 2 \operatorname{arctg} \left(\frac{H/2}{6,66H} \right) = 8,57^\circ \quad (1.65)$$

Finalmente, teniendo en cuenta que la resolución espacial del sistema visual es del orden de 1 minuto de grado, y que, por tanto, dos líneas deberán situarse, como máximo, a una distancia angular de 1 minuto, obtenemos el número mínimo de líneas necesario para el muestreo espacial de las imágenes:

$$NL = \frac{8,57^\circ \times 60'}{1'/\text{línea}} = 514,69 \text{ líneas} \quad (1.66)$$

Este número de líneas debe tomarse como un valor aproximado que nos indica aproximadamente cuál es el valor aproximado que debe tomar este parámetro. El número de líneas visibles en el sistema PAL es algo superior al obtenido en la ecuación anterior, pero en el sistema NTSC se mantiene ligeramente por debajo. Estas diferencias no constituyen, sin embargo, ningún inconveniente en el sistema NTSC y simplemente indicaría que la distancia de observación óptima es ligeramente superior a la del PAL. De hecho, la distancia de visionado es un parámetro de diseño del sistema que el propio usuario adecua a las características de su visión y del entorno en el que observará la televisión. Así, es habitual que algunos espectadores se sitúen a distancias del televisor considerablemente inferiores a las anteriores.

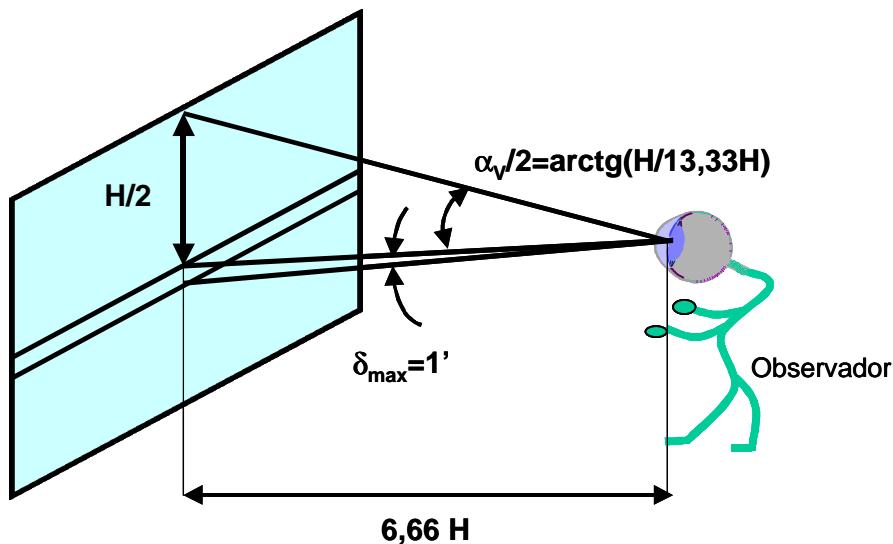


Fig. 1.63 Cálculo del ángulo vertical de visionado

1.8.6 Selección del número de líneas

En la selección definitiva del número de líneas de la señal de televisión intervienen, además de la agudeza visual discutida en los apartados anteriores, dos factores relacionados con la tecnología electrónica existente en la época en la que se definieron los primeros sistemas y que por conveniencia, simplicidad y compatibilidad se han mantenido posteriormente. El primero es que para conseguir un perfecto entrelazado entre las dos subimágenes es necesario, como analizaremos posteriormente, que el número de líneas sea impar. El segundo factor es que para facilitar los circuitos electrónicos es conveniente que pueda establecerse una relación simple entre el número de líneas y el número de imágenes por segundo. Esta relación simple se traduce en que puedan obtenerse las distintas frecuencias que intervienen en el proceso de exploración de la imagen a partir de un único oscilador global, cuya frecuencia se divide en relaciones enteras simples para obtener las distintas señales de barrido de línea, campo, imagen, etc. Para ello, como justificaremos pronto, es útil que el número de líneas de la imagen pueda expresarse como un producto de números primos cuyos valores absolutos sean relativamente bajos.

En la tabla 1.1 se proporciona la descomposición en números primos de los valores impares comprendidos entre 601 y 649 líneas. Todo este margen de valores proporciona un número de líneas adecuado a la resolución espacial del ojo. Puede observarse que de todos estos candidatos, el valor de 625 es el que tiene una descomposición con unos números primos más bajos, resultando el elegido como el número de líneas en los sistemas de televisión utilizados en Europa.

Tabla 1.1 Descomposición en números primos de los números impares comprendidos entre 601 y 649

601=601	603=67X3X3	605=11X11X5	607=607	609=29X7X3
611=47x13	613=613	615=41x5x3	617=617	619=619
621=23x3x3x3	623=89x7	625=5x5x5x5	627=19x11x13	629=37x17
631=631	633=211x3	635=127x5	637=13x7x7	639=71x3x3
641=641	643=643	645=43x5x3	647=647	649=59x11

Para justificar la idoneidad de esta selección, calcularemos las frecuencias características de una señal de televisión con una frecuencia de imagen de 25 Hz, una frecuencia de campo de 50 Hz y que utiliza un total de 625 líneas por imagen. En este caso, de acuerdo con la figura 1.56, el periodo de tiempo asignado a cada línea será:

$$T_{\text{línea}} = \frac{1/25 \text{ s / imag}}{625 \text{ lin / imag}} = \frac{1}{15625} \text{ s} = 64 \text{ ns} \quad (1.67)$$

obteniendo, por tanto, una frecuencia de línea de 15625 Hz. Con estos guarismos es fácil ver que si partimos de un oscilador de 31250 Hz, basta con pasar esta señal a través de distintos divisores de frecuencia para obtener cada una de las frecuencias de interés:

$$\text{Frecuencia de línea: } f_{\text{línea}} = f_{\text{osc}} \div 2$$

$$\text{Frecuencia de campo: } f_{\text{Field}} = f_{\text{osc}} \div (5 \times 5 \times 5 \times 5) \text{ (Pasarla a través de 4 divisores por 5)}$$

$$\text{Frecuencia de cuadro: } f_{\text{cuadro}} = f_{\text{Field}} \div 2$$

Con ello se simplifica notablemente el proceso de sincronizar las señales para la exploración vertical y horizontal de la imagen, pues todas las señales se obtienen de una misma referencia. El interés de que el número de líneas tenga una descomposición en números primos relativamente bajos radica en que es mucho más simple y fiable implementar divisores o multiplicadores de frecuencia cuando los factores de incremento o reducción son enteros bajos. En la práctica sólo se utilizan divisores de frecuencia por 2, 3, 5, 7 y raramente por 11 o por 13. En los sistemas de televisión americanos y japoneses se utilizan 525 líneas que también tienen una descomposición en números primos con características similares ($525=3 \times 5 \times 5 \times 7$). En Francia se utilizó durante algún tiempo un sistema de 819 líneas ($819=13 \times 7 \times 3 \times 3$). Este sistema ya no se usa, habiendo sido sustituido por un sistema convencional de 625 líneas.

Tabla 1.2 Características básicas de los sistemas 525/60 y 625/50

	IMAG/S	CAMP/S	LINEAS/IMAG	T _{LINEA}	F _{LINEA}	T _{IMAGEN}	F _{IMAGEN}	T _{CAMPO}	F _{CAMPO}
PAL	25	50	625	64 s	15625 Hz	40 ms	25 Hz	20 ms	50 Hz
NTSC	30	60	525	63,49 s	15750 Hz	33,33 ms	30 Hz	16,66 ms	60 Hz

En la tabla 1.2 se proporcionan los valores de frecuencia de cuadro, campo y línea para los sistemas europeo y americano. Estos valores son parámetros fundamentales de la señal de TV que conviene recordar en todo momento.

1.8.7 Número de líneas en los sistemas de alta definición

El número de líneas de los diversos sistemas de alta definición que se están proponiendo o que ya están operativos en algunos países utilizan un número de líneas que, por regla general, es el doble del que utilizan los sistemas de definición convencional. El objetivo básico de doblar el número de líneas es acercar el espectador a la pantalla para integrarlo más en la escena. Evidentemente, no tendría sentido doblar el número de líneas si el espectador hubiera de mantenerse a la misma distancia del televisor, ya que los sistemas de definición convencional están diseñados para que el número de líneas esté de acuerdo con la agudeza del sistema visual. Si situamos a un espectador delante de una pantalla de televisión a una distancia de unas 7 veces la altura de la pantalla, éste sería, en principio, incapaz de apreciar ninguna diferencia si se utilizan 625 o 1250 líneas.

En los sistemas de alta definición se pretende que el espectador perciba una mayor sensación de realismo en la escena aumentando el campo visual que ocupa la pantalla. Hemos visto que en un sistema de definición convencional, si se observa desde la distancia apropiada, el ángulo de observación vertical es de 8,57°. Análogamente, podríamos calcular el ángulo de observación horizontal que, dado que la relación de aspecto es de 4:3, sería de 11,42°. Así pues, en estos sistemas, la pantalla cubre sólo una pequeña porción del campo visual, por lo que el espectador es consciente del medio de representación y la escena pierde parte del realismo que se obtendría con un sistema que cubriera un mayor ángulo de visión, como por ejemplo ocurre con el cine.

Al doblar el número de líneas podemos acercar el espectador a la pantalla doblando, en consecuencia, su ángulo de visión vertical a unos 17°. Además, al usar una relación de aspecto con mayor panorámica, 16:9, obtendremos un ángulo de visión horizontal de aproximadamente 30°. Esta relación de ángulos de visión de la escena se adapta mejor al campo visual humano y, aunque sólo representa un 20 % de campo real, cubre la región donde se concentra la mayor parte de información captada por el sistema visual.

Es importante no confundir esta filosofía para la integración del espectador en la escena usada en los sistemas de alta definición con algunos formatos de televisión, como el PALplus, en el que se utilizan relaciones de aspecto de 16:9. En este caso, el número de líneas sigue siendo del mismo orden que en los sistemas de definición convencional, por lo que el espectador debe permanecer a una distancia aproximada de 7 veces la altura de la pantalla. Por tanto, estos sistemas si bien aumentan el ángulo de visión horizontal a unos 15°, mantienen el mismo ángulo de visión vertical, por lo que la mejora no es tan significativa.

1.8.8 Entrelazado de las imágenes

Hemos visto que para evitar el parpadeo de la pantalla es necesario aumentar la frecuencia de presentación de imágenes en el receptor. Una frecuencia de 25 imágenes por segundo resulta insuficiente, con los niveles de brillo existentes en la pantalla de un televisor, para evitar el parpadeo.

La solución adoptada consiste en realizar una doble exploración entrelazada de las líneas de cada imagen tal y como se representa en la figura 1.64. La imagen se divide en dos subimágenes o campos explorando alternativamente las líneas pares y las impares. La proximidad entre líneas consecutivas hace que el espectador integre las dos subimágenes y obtenga la sensación de que éstas se están renovando a una frecuencia doble de la real. Con ello se consigue mantener un caudal de información reducido, suficiente para interpolar correctamente el movimiento sin que aparezca el fenómeno de parpadeo.

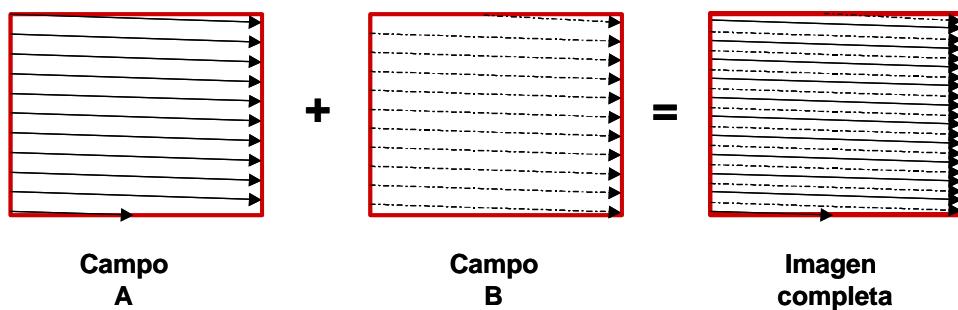


Fig. 1.64 División de la imagen en dos campos entrelazados

La necesidad de entrelazar las imágenes supone tener que modificar la estructura de la señal respecto al diagrama que hemos esbozado en la figura 1.56, ya que las líneas deberán transmitirse alternadamente y en el mismo orden en que se realiza la exploración del haz en la pantalla en el receptor.

La transmisión secuencial de las líneas sugerida en la figura 1.56 sólo puede usarse en sistemas de vídeo que no utilizan el entrelazado de los dos campos. Estos sistemas se conocen como sistemas progresivos o también como sistemas con una relación de entrelazado 1:1 (1imagen, 1 campo). Si bien la exploración progresiva no se utiliza en los sistemas actuales de televisión comercial, su uso es habitual en los monitores de ordenador. En este caso utiliza una frecuencia de imagen bastante más elevada que en televisión (72-90 Hz), por lo que no es necesario descomponer las imágenes en dos campos para evitar el parpadeo. La ventaja principal es que se eliminan todos los problemas de inestabilidad de la imagen que se producen con el entrelazado de los dos campos y que impedirían una cómoda lectura de los caracteres escritos que se presentan en un ordenador. El ancho de banda asociado a la señal de vídeo es evidentemente mayor que en un sistema entrelazado aunque, como no debe transmitirse, esto no supone una restricción importante. La exploración progresiva se ha considerado para futuros sistemas de televisión digital, aunque su implantación definitiva no será fácil, ya que todos los receptores de televisión actuales están basados en sistemas entrelazados.

Para obtener una presentación entrelazada de los dos campos con la tecnología actual, sería posible transmitir las líneas secuencialmente, almacenándolas en una memoria en el receptor, para posteriormente presentarlas de forma alternada en la pantalla. Sin embargo, cuando se desarrollaron los primeros sistemas de televisión era tecnológicamente imposible almacenar parte de la señal en el receptor, por lo que la señal transmitida debía adecuarse a las características de exploración propias del sistema de presentación de imágenes. Por todo ello, es necesario que la señal de vídeo esté

dividida en dos partes, que se corresponden con cada uno de los campos y en las que se transmiten primero todas las líneas impares y después, todas las pares. Con ello, la información se transmite en el mismo orden en que se realiza la exploración, por lo que la señal puede aplicarse directamente al tubo de rayos catódicos siempre que el sistema de barrido del haz esté correctamente sincronizado con la trama de vídeo. Esta nueva estructura se representa en la figura 1.65, donde se indica explícitamente el número de línea de pantalla que se corresponde con cada componente de la señal. En la práctica, las líneas de la señal de vídeo suelen numerarse atendiendo a su posición en la trama de señal y no a su posición en la pantalla. Esto se indica en la misma figura 1.65 manteniendo, por el momento, una notación distinta para cada tipo de numeración. En el futuro, mientras no se haga una referencia explícita a la numeración, entenderemos que las líneas están numeradas secuencialmente atendiendo a su posición en la trama de señal.

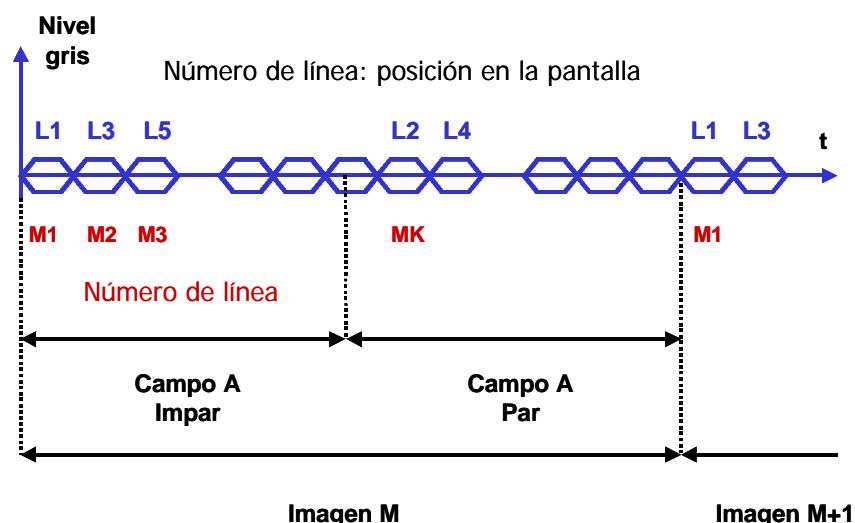


Fig. 1.65 Intercalación de las líneas en la señal de vídeo para el entrelazado de la imagen

Para un obtener un correcto entrelazado de los dos campos es necesario disponer de señales que controlen con precisión el desvío del haz en el tubo de rayos catódicos. Para ello se utilizan dos señales en forma de diente de sierra, cuya generación requiere circuitos muy simples, que controlan la desviación del haz en sentido horizontal y vertical. La necesidad de que estas señales pudieran generarse fácilmente en el receptor condicionó, en el momento de definir el formato de la señal, que el número de líneas de cada imagen fuera impar, ya que de este modo puede realizarse un cambio del campo sin interrumpir la periodicidad de las señales de barrido. En efecto, cada vez que se produce un salto de línea en la exploración de la pantalla, la señal de barrido horizontal debe producir un cambio de nivel desde su valor mínimo al máximo. Durante la exploración de la línea el nivel de la señal debe variar linealmente entre estos dos valores. Análogamente, la señal de barrido vertical evoluciona linealmente entre su valor máximo y mínimo durante la exploración de un campo y se reinicializa a su valor máximo cada vez que se empieza a explorar uno nuevo. De acuerdo con esta filosofía, si el número de líneas de la imagen es impar, cada campo tendrá asignadas un determinado número de líneas completas más media línea adicional. Con ello, tal y como se indica en la figura 1.66, cada vez

que se produce un cambio de campo, la señal de barrido horizontal puede seguir manteniendo su estructura periódica, continuando la exploración de la pantalla a partir de la mitad de la línea cuando la transición es entre un campo impar y uno par, o iniciando la exploración de una nueva línea cuando se trata de una transición entre campo par e impar.

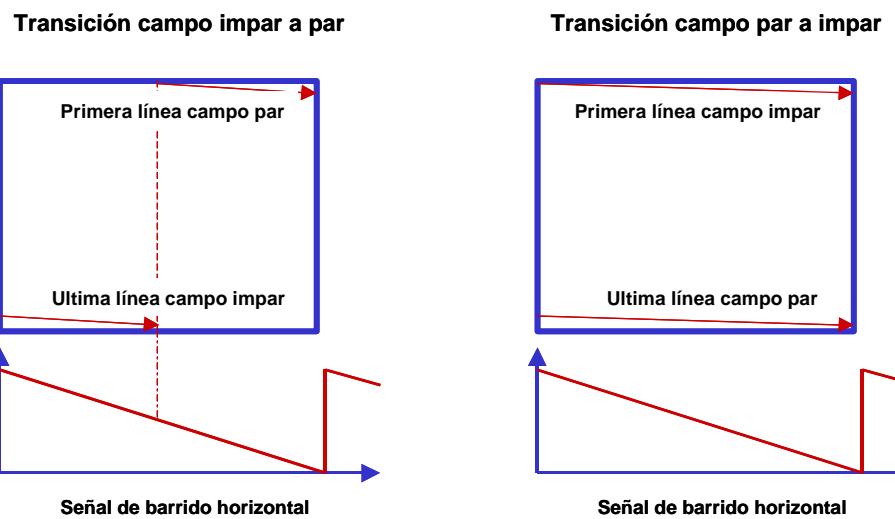


Fig. 1.66 La señal de barrido horizontal mantiene su estructura periódica en las transiciones

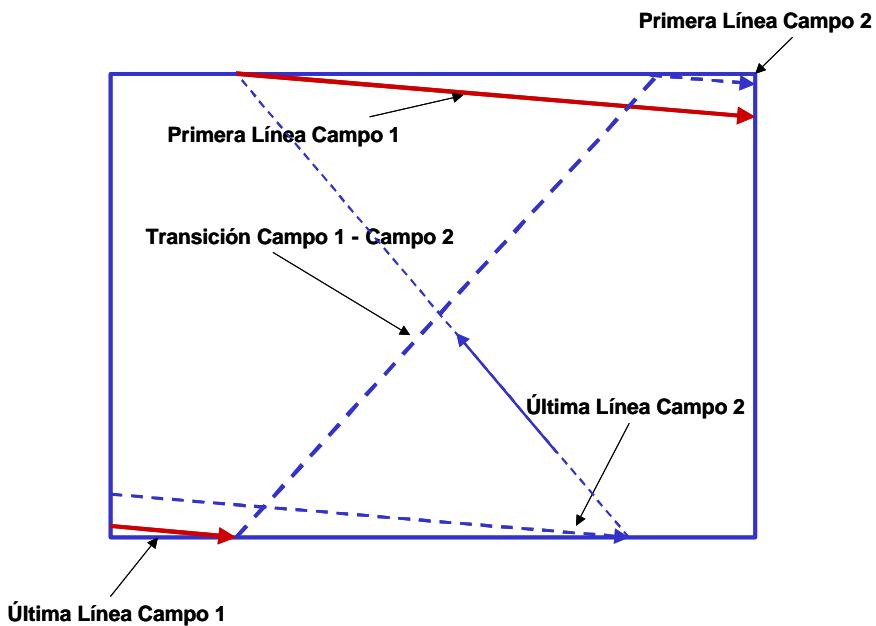


Fig. 1.67 Discontinuidades en el barrido horizontal para un número par de líneas

Esta estructura simplifica notablemente la síntesis de las señales de barrido, ya que si el número de líneas fuera par sería necesario resituar la señal de barrido horizontal cada vez que se iniciara un nuevo campo alterando su patrón de periodicidad. En la figura 1.67 se representa cómo deberían entrelazarse los campos en el caso en que el número de líneas fuera par, observándose claramente la pérdida de periodicidad en el posicionamiento horizontal del haz. En este ejemplo hemos supuesto que la primera línea del campo impar inicia su recorrido con 1/3 de la posición horizontal, mientras que la primera línea del campo par la inicia en 2/3 del extremo horizontal. Con ello, ambos campos podrían quedar perfectamente imbricados. El problema, como se observa en el dibujo, es que la última línea del campo impar finaliza en la misma posición horizontal en que ha empezado el campo, por lo que debe producirse un salto de posición horizontal antes de iniciarse el segundo campo.

1.8.9 Problemas derivados del entrelazado de imágenes

El entrelazado de los dos campos permite reducir a un nivel aceptable el parpadeo de la pantalla, aunque introduce, aun en el caso ideal en que no se produzca ningún desalineamiento de los haces durante la exploración, cierta degradación sobre la calidad de las imágenes. Los efectos más conocidos son la aparición de los denominados vibración interlínea (*interline twiter*) y arrastre de línea (*line crawl*). El primero aparece en los contornos horizontales de imágenes estacionarias como una pequeña vibración en sentido vertical del contorno debido a que en cada campo éste se representa en una posición vertical ligeramente distinta. El efecto de arrastre de línea produce la apariencia de que las líneas se desplazan verticalmente en la pantalla, debido a que cada campo se presenta al espectador en instantes de tiempo distintos lo que puede inducir, para algunas imágenes, la sensación de que los contornos se desplazan en sentido vertical, cuando en realidad la imagen permanece estacionaria.

Otro problema directamente relacionado con el entrelazado de las imágenes es el denominado efecto Kell, que reduce la resolución de las imágenes por debajo de lo que *a priori* podríamos estimar teniendo en cuenta el número de líneas. Describiremos este efecto, desde un punto de vista cualitativo, en los siguientes apartados.

1.8.10 Inserción de señales de sincronismo en la señal de vídeo

Es evidente que la estructura de la señal de vídeo, con el entrelazado de los dos campos, exige que el receptor mantenga un perfecto sincronismo entre la exploración de la pantalla y la información transmitida. Por ello, la señal propia señal de vídeo incorpora información sobre el inicio de cada línea y de cada campo. Estas componentes, convenientemente decodificadas, permiten garantizar que los circuitos que generan las señales de desvío horizontal y vertical del haz se mantengan en perfecta concordancia con la señal transmitida.

La inserción de señalización específica para identificar el inicio de las líneas y los campos impone que parte del tiempo, que en principio podría asignarse a información de imagen, deba utilizarse para estos sincronismos. Esto significa que sólo una parte del periodo de línea y de campo se dedicarán a información correspondiente a la escena, dejando el resto para incluir las señales de sincronismo. La parte del tiempo de línea dedicada a transmitir la luminancia se denomina línea activa. Análogamente, la parte del tiempo de campo que se dedica a transmitir líneas con información de vídeo se denomina

campo activo. En la figura 1.68 se representa esquemáticamente la filosofía general para la transmisión de los sincronismos.

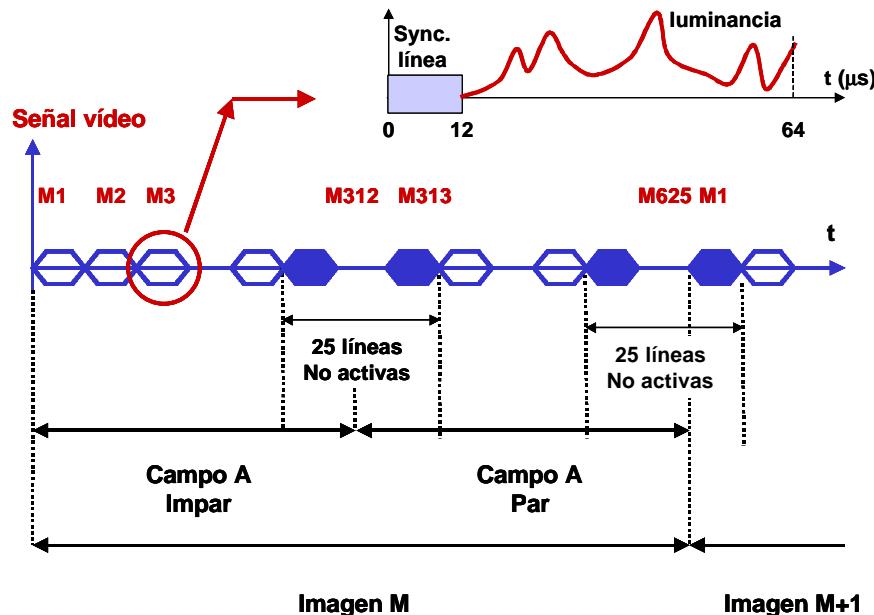


Fig. 1.68 Inserción de las señales de sincronismo en la estructura de la señal de TV

En el estándar de televisión europeo de 625 líneas y 25 imágenes por segundo, el tiempo de línea es de 64 µs, de los cuales 12 µs se dedican a la señal de sincronismo, quedando los 52 µs restantes para la imagen. La parte asignada al sincronismo, cuyo contenido analizaremos en el siguiente capítulo, permite reinicializar los circuitos de generación de señal de barrido horizontal cada vez que llega una nueva línea. El tiempo que transcurre entre el final de la parte activa de una línea y el inicio de la siguiente se utiliza para desplazar el haz desde la parte derecha de la pantalla al extremo izquierdo. Es importante que durante este tiempo, la señal aplicada al tubo sea nula, pues en caso contrario el retorno del haz se haría visible.

Para enviar el sincronismo de campo, se utilizan varias líneas consecutivas en las que se transmite una señal fácilmente separable tanto de la información de imagen como de los sincronismos de línea. En los sistemas europeos analógicos, el sincronismo de campo se transmite aprovechando 25 líneas no activas que se definen entre cada campo. Estas líneas se aprovechan además para que el haz retorne del extremo derecho inferior de la pantalla al extremo izquierdo superior y para transmitir información de teletexto o señales de prueba que se utilizan para comprobar la correcta recepción de la señal en distintos puntos del área de distribución. Dejaremos los detalles de las formas de onda de todas estas componentes para el capítulo siguiente.

Debemos observar que buena parte del tiempo en la señal de televisión analógica se dedica a la transmisión de sincronismos e información adicional que no se corresponde directamente con la

escena. En efecto, de las 625 líneas de una imagen sólo 575 son activas, por cuanto cada campo dispone de 25 líneas reservadas para sincronismos, servicios de valor añadido como el teletexto o información de control. Además, sólo un 81 % del tiempo de línea se dedica a la parte activa. Esto significa que aproximadamente un 25 % del tiempo de la señal se dedica a transmitir información que no se corresponde directamente con la imagen.

1.8.11 Exploración entrelazada de las imágenes en la cámara

Tal y como hemos planteado el muestreo temporal y espacial de las imágenes, todas las líneas correspondientes a un mismo cuadro o imagen deberían corresponder al mismo instante de tiempo. El proceso de lectura de las líneas debería realizarse una vez la imagen ha sido adquirida o memorizada en algún dispositivo sensor. La exploración de la señal obtenida en el sensor podría realizarse entrelazando las líneas, siguiendo el mismo orden en que deben transmitirse en la señal de vídeo.

Sin embargo, esta filosofía sólo se utiliza en algunos sensores de imagen de estado sólido modernos. En los tubos de cámara, la escena se expone al sensor durante un tiempo igual o inferior a la duración de un campo. La carga eléctrica almacenada durante este tiempo de exposición constituye la información que se transmitirá como uno de los campos. Posteriormente, la imagen vuelve a exponerse durante otro periodo de tiempo igual, obteniendo la información asociada al segundo campo. Con ello, las líneas correspondientes al campo par y al impar no se corresponden con la misma imagen, sino con imágenes tomadas con una diferencia de tiempos igual al periodo de campo. La razón de aplicar esta estrategia de exploración a los tubos de imagen es que la carga del sensor, después de su exposición a la luz, se reduce con el tiempo, por lo que, si mantuviéramos la carga durante todo el tiempo de cuadro, sería perceptible esta pérdida de carga en las últimas líneas. Por otra parte, de este modo, la exploración entrelazada de la imagen en la cámara se corresponde directamente con el entrelazado de los dos campos en la pantalla del televisor.

La mayoría de cámaras de vídeo actuales, basadas en sensores de estado sólido, utilizan una filosofía parecida exponiendo el sensor durante un tiempo determinado y construyendo la señal de uno de los campos a partir de las señales almacenadas. Normalmente la señal del campo A se obtiene como el promedio entre los pares de líneas (1, 2), (3, 4), (5, 6), etc; mientras que las del campo B se obtiene promediando los pares (2, 3), (4, 5), (6, 7), etc. Una vez construida la señal de uno de los campos, se anula la carga del sensor y se realiza la adquisición de la imagen para el otro campo. Con esta tecnología, las imágenes de cada uno de los campos corresponden a instantes de muestreo distintos. El efecto es muy notorio cuando existen objetos con movimiento horizontal en la escena. Si detenemos la reproducción de la secuencia en una imagen, veremos que los objetos en movimiento aparecen con dos contornos diferentes, que corresponden a las posiciones que tenían en cada uno de los campos. Generalmente, esto produce un efecto de serrado del contorno que sólo resulta molesto cuando se detiene la secuencia de vídeo para realizar el análisis de un fotograma. En modo de reproducción normal, la resolución del sistema visual de los objetos con movimiento es menor que la de los objetos estáticos por lo que, normalmente, no se aprecia la falta de definición en el contorno.

Existen algunos modelos de cámaras, denominadas cámaras progresivas, que realizan la adquisición de todas las líneas de la imagen en un único instante de muestreo. En principio estas cámaras están pensadas para aplicaciones de reconocimiento de imagen en escenas donde los objetos pueden tener

un movimiento considerable. Este tipo de cámaras también se pueden combinar con unidades de memoria dinámica para construir la señal entrelazada a partir de la información recogida. Con ello, se pueden obtener señales entrelazadas en las que los dos campos corresponden al mismo instante de muestreo. Otro caso en el que la información de los dos campos proviene del mismo instante de muestreo es en la conversión de formatos entre cine y televisión, donde cada fotograma original se explora dos veces para obtener la señal de cada uno de los campos.

1.8.12 Receptores de 100 Hz

Aunque el entrelazado de los dos campos permite reducir el parpadeo a niveles aceptables, aún es perceptible, sobre todo en regiones uniformes de la imagen con área considerable y que tienen nivel de brillo elevado. Los modernos receptores digitales permiten aumentar la frecuencia de exploración del haz hasta 100 Hz. La idea básica consiste en digitalizar los dos campos de la imagen y almacenarlos en una memoria antes de presentar la información en la pantalla. Los subsistemas de exploración de la pantalla y digitalización de campos operan en paralelo, de modo que mientras se recibe el campo par, se va actualizando la memoria de imagen asociada a este campo y simultáneamente se presenta en la pantalla la información correspondiente al campo impar obtenida previamente.

El barrido de los campos se realiza a una velocidad doble de la normal, de modo que cada exploración dura 10 ms en vez de los 20 ms asignados en un receptor convencional. Esto permite presentar dos veces la información almacenada en la memoria, aumentando la frecuencia de actualización de la pantalla a 100 Hz, con la consiguiente reducción en el parpadeo de la imagen. La secuencia de digitalización y presentación de las señales se representa en la figura 1.69.

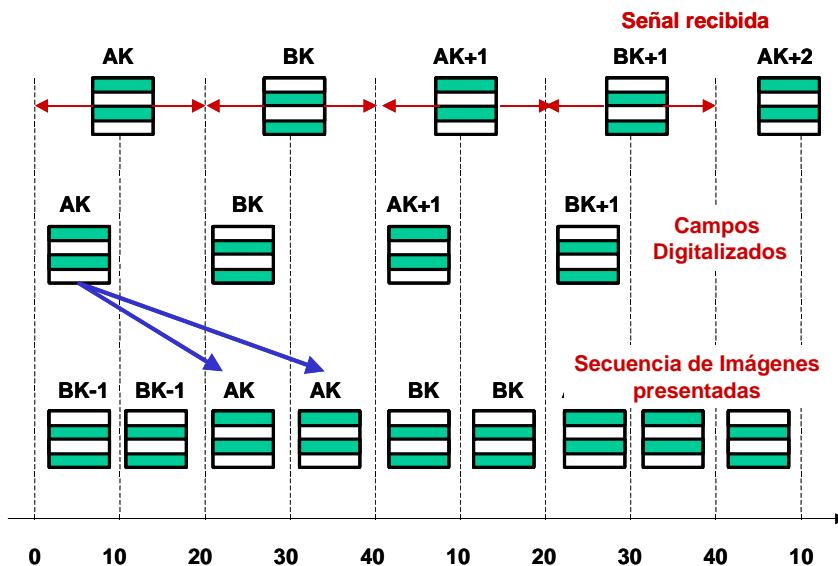


Fig. 1.69 Principio de funcionamiento de los receptores de 100 Hz

Esta nueva frecuencia de refresco de la pantalla elimina prácticamente el problema del parpadeo que aparece en las grandes áreas de la pantalla que tienen un nivel de brillo considerable. Sin embargo, los problemas de vibración interlínea y arrastre de línea se mantienen por cuanto la información de campo se sigue actualizando, como en un receptor convencional, cada 50 Hz. Una alternativa que mejora notablemente estos efectos son los receptores denominados *Digital Scan*, que son una variante del sistema 100 Hz. La secuencia de digitalización y presentación de las imágenes en estos receptores se representa en la figura 1.70. En este caso, debemos disponer de un *buffer* de memoria que permita almacenar los dos campos asociados a una imagen mientras el sistema de barrido intercala, con frecuencia de exploración doble, los campos correspondientes a la imagen anterior. La diferencia con el sistema 100 Hz es que los campos, aunque se repiten, se actualizan a una frecuencia real de 100 Hz, desapareciendo los efectos de vibración de los contornos horizontales de la imagen.

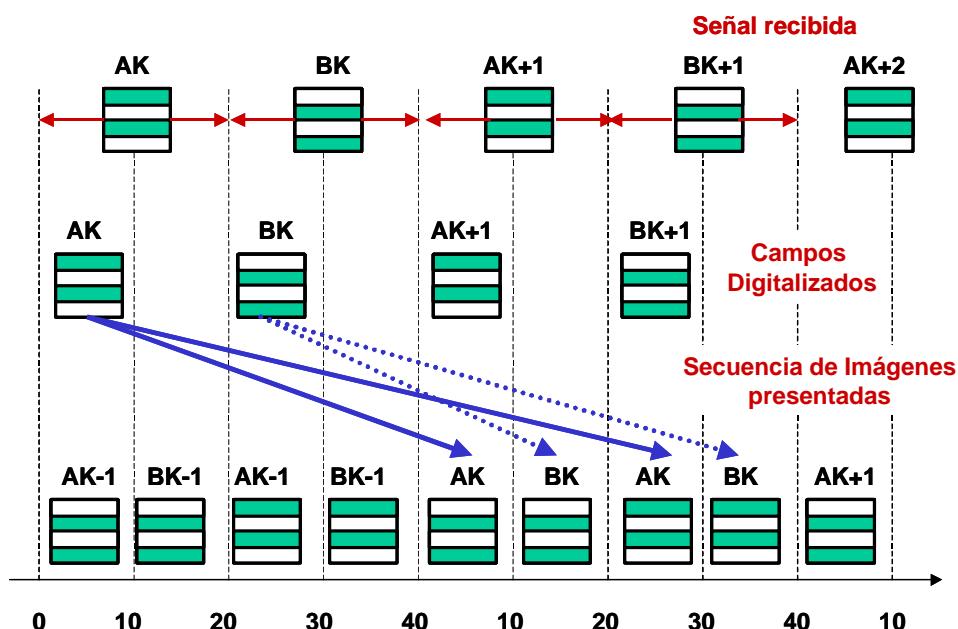


Fig. 1.70 Principio de funcionamiento del sistema Digital Scan

Podría pensarse que una posible mejora de estos sistemas consiste en almacenar los dos campos de la imagen en una misma memoria para posteriormente realizar una presentación de toda la imagen con un sistema progresivo, como en un monitor de ordenador. Las imágenes seguirían transmitiéndose a 25 Hz, pero podríamos presentarlas 2, 3 o 4 veces en pantalla para eliminar completamente el parpadeo. Todo ello sin que aparecieran los problemas de inestabilidad de imagen, propios de un sistema de presentación entrelazado. Con la tecnología de monitores y memorias digitales actuales esta solución sería factible, de manera que el paralelismo con el sistema utilizado en cine sería entonces completo. Sin embargo, el problema es ahora el tipo de exploración realizado en la cámara en la que los dos campos ya no se corresponden con una misma imagen, por lo que mantendríamos todos los problemas propios del entrelazado realizado en origen.

Normalmente, los receptores que incluyen alguna de las variantes que hemos descrito para aumentar la frecuencia de refresco también incorporan algún tipo de filtro digital que permite mejorar la calidad de las imágenes cuando las condiciones de recepción de la señal no son las adecuadas. El hecho de disponer de las muestras de los dos campos de la imagen en la memoria del receptor permite aplicar sencillas operaciones digitales que pueden eliminar algunos tipos de ruido presentes en la señal. La operación más habitual consiste en identificar zonas de la imagen que tienen un nivel o color uniforme y sustituir con su valor medio aquellas muestras aisladas que presenten un valor muy distinto a éste valor medio. Con ello se consigue eliminar los habituales picos que aparecen y desaparecen de manera errática en la pantalla cuando el nivel de señal es bajo o cuando se reproduce una señal procedente de una cinta de vídeo vieja o de mala calidad.

Recientemente han aparecido receptores que utilizan las imágenes almacenadas en memoria para realizar una interpolación de líneas inteligente. Estos receptores, aparte de aumentar el número de líneas con el que se realiza la exploración, aplican técnicas de tratamiento digital de imagen para reducir los efectos de vibración interlínea.

La memoria digital también suele aprovecharse para poder reproducir en la pantalla varios canales simultáneamente. El principio de funcionamiento de estos sistemas consiste en demodular y digitalizar dos o más canales, almacenando sus muestras en las direcciones de memoria adecuadas para que el sistema de exploración de imágenes las sitúe automáticamente sobre la pantalla. Normalmente esta opción se utiliza como canal en espera, por lo que la imagen secundaria, que sólo ocupa una pequeña porción de la pantalla, puede muestrearse a menor velocidad que el canal principal.

1.8.13 Ancho de banda aproximado de la señal de televisión

En este apartado realizaremos diversas consideraciones sobre el ancho de banda de la señal de televisión en blanco y negro. Evidentemente, el ancho de banda de esta señal está directamente relacionado con las variaciones de nivel que se producen en la luminancia. Cambios bruscos en la luminancia, que se corresponden con contornos verticales en la imagen, exigen la presencia de componentes de alta frecuencia, mientras que las regiones uniformes serán codificadas con un nivel constante de la señal, por lo que son las responsables de la existencia de componentes continuas.

Es obvio, por tanto, que el ancho de banda depende de las características de la imagen y que variará en función de ésta. En general, cuando hablamos del ancho de banda de la señal interpretaremos que nos referimos al valor mínimo necesario para que puedan transmitirse todo tipo de imágenes con una calidad aceptable. Un sistema con un ancho de banda reducido limita la presencia de cambios de nivel bruscos en la señal, por lo que, probablemente, los contornos de la imagen no quedarán bien definidos. El ancho de banda adecuado es por tanto un compromiso entre la eficiencia espectral del sistema de comunicaciones y la calidad con que pueden reproducirse las imágenes.

Para tener una primera idea sobre el orden de magnitud del ancho de banda, consideremos cuál es la frecuencia máxima que tiene sentido transmitir para imágenes estacionarias. Evidentemente, si la imagen a transmitir tuviera un nivel de gris uniforme, la luminancia tendría un valor constante en todas las líneas, por lo que la señal sólo contendría componente continua. Si la imagen que

transmitimos está formada por barras negras y blancas en sentido vertical, la luminancia deberá tomar sus valores extremos a lo largo de una línea, con lo que su frecuencia estará directamente relacionada con el número de barras verticales, tal y como se ilustra en la figura 1.71.

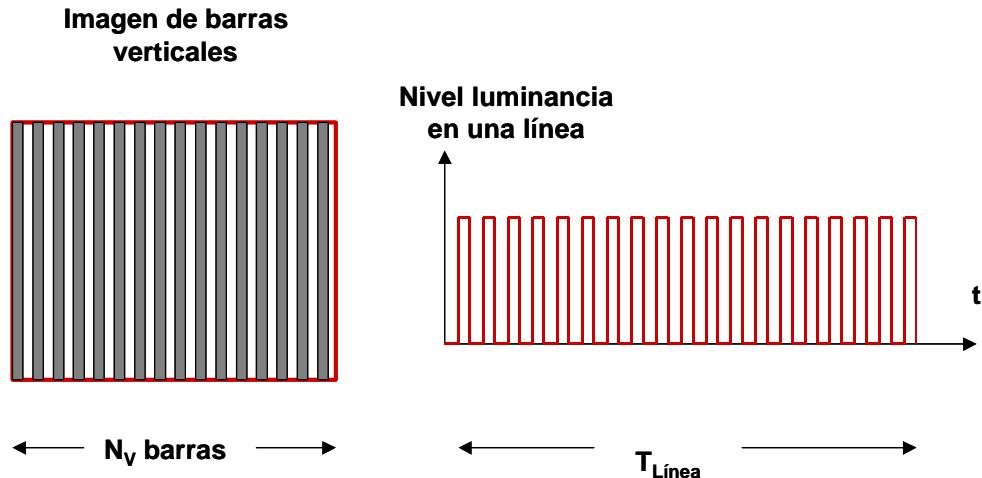


Fig. 1.71 Señal de luminancia asociada a una imagen formada por barras verticales

La relación entre la frecuencia fundamental de esta señal, el número de barras y el periodo de línea viene dada por:

$$f_0 = \frac{1}{T_0} = \frac{1}{T_{Línea}/(N_V/2)} = \frac{N_V}{2 \cdot T_{Línea}} \quad (1.68)$$

que indica que esta frecuencia fundamental aumenta de manera proporcional con el número de barras verticales. La aparición del factor 2 es debida a que un ciclo de la señal de luminancia se corresponde con una barra negra más una blanca. En principio el periodo de línea que deberíamos utilizar en esta expresión debería ser el tiempo de línea activa, por cuanto todas las barras deben codificarse en este fragmento de la línea. Sin embargo, como únicamente se trata de un cálculo aproximado del ancho de banda, normalmente se utiliza el tiempo de línea total.

Para estimar el valor máximo que puede tomar la frecuencia de esta señal deberemos considerar cuál es el número máximo de barras que tiene sentido representar en la pantalla del receptor. Es obvio que este parámetro está relacionado con la agudeza del sistema visual, que ya hemos utilizado para calcular el número de líneas de la imagen. Como la agudeza visual es igual en la dirección vertical que en la horizontal (es algo más reducida en el sentido diagonal), podemos expresar el número máximo de barras verticales en función del número de líneas en que hemos dividido la imagen. En efecto, el número máximo de barras horizontales que teóricamente podrían representarse en una pantalla sería igual al número de líneas del sistema de televisión. Para ello, deberíamos alternar el valor de la luminancia entre su valor máximo y mínimo entre cada línea.

Por otra parte, debemos tener en cuenta que la pantalla tiene una relación de aspecto distinta de la unidad por lo que, manteniendo la densidad de barras constante, el número de barras verticales será mayor que el número de líneas en un factor igual al de la relación de aspecto. Por tanto obtenemos:

$$N_{V\max} = \frac{W}{H} \cdot N_L \quad (1.69)$$

Nuevamente, N_L debería ser el número de líneas activas, aunque normalmente se sustituirá por el número total de líneas, ya que sólo se trata de una primera aproximación al problema.

Sustituyendo esta última expresión en la ecuación 1.68 obtenemos:

$$f_{0\max} = \frac{1}{2} \cdot \frac{W}{L} \cdot \frac{N_L}{T_{Línea}} \quad (1.70)$$

Podemos expresar el tiempo de línea en función del número de líneas y de la frecuencia de actualización de imágenes (f_I), con lo que obtendremos la expresión de esta frecuencia máxima en función de los parámetros básicos de un sistema de televisión.

$$f_{0\max} = \frac{1}{2} \cdot \frac{W}{L} \cdot N_L^2 \cdot f_I \quad (1.71)$$

El valor de esta frecuencia para un sistema de 25 imágenes por segundo, 625 líneas y una relación de aspecto 4:3 es de aproximadamente 6.5 MHz. En el caso del estándar americano de 30 imágenes por segundo y 525 líneas, se reduce a unos 5.5 MHz, mientras que para un hipotético sistema de alta definición analógico con 1250 líneas, 50 imágenes por segundo y relación de aspecto 16:9, tomaría un valor aproximado de unos 70 MHz. En este cálculo aproximado, debemos tener en cuenta que únicamente consideramos la frecuencia fundamental de la señal, ya que al tratarse de una señal cuadrada tendrá armónicos con un contenido frecuencial importante.

Sin embargo, el valor obtenido mediante la ecuación 1.71 suele considerarse como una cota superior del ancho de banda requerido por un sistema de televisión, ya que la señal que hemos considerado constituye un caso extremo que raramente se produce en escenas reales. Además, no hemos tenido en cuenta algunos problemas relacionados con el entrelazado de las imágenes que afectan a la resolución del sistema y que impedirían representar una imagen con estas características.

R.D. Kell analizó en 1934 la resolución vertical que podía obtenerse con un sistema de televisión experimental, estableciendo que no era posible representar un número de barras horizontales igual al número de líneas, debido, por una parte, a un problema de interferencia intrínseco al entrelazado de las imágenes, y por otra, a la propia limitación de la resolución en el monitor. La experiencia de Kell consistió en representar sobre un sistema de televisión de 100 líneas (50 líneas por campo) una alternancia entre líneas blancas y negras (barras horizontales), intentando determinar el número máximo de barras que podían visualizarse. En principio, el resultado que esperaríamos sería de unas 100 barras, que corresponde con una señal en la que todas las líneas del campo par son blancas y las del campo impar negras. Sin embargo, al presentar esta señal se obtiene una imagen gris uniforme en la que no pueden distinguirse las barras horizontales. De hecho, para poder observar la presencia de las barras es necesario reducir su número a unas 64.

Una de las razones de esta limitación es que el haz de exploración de la pantalla no es puntual, sino que tiene un grosor finito, lo que produce cierto grado de superposición entre líneas adyacentes, limitando la resolución del monitor. Este factor puede reducirse notablemente con los sistemas de enfoque electrónico del haz actuales.

El otro factor responsable de la reducción de resolución respecto a su valor esperado se debe a la propia estrategia de muestreo de la imagen que hemos utilizado como consecuencia directa de querer entrelazar los dos campos. La resolución vertical de la imagen se reduce debido a que el muestreo de la imagen en dos campos introduce repeticiones periódicas del espectro que se superponen con las componentes visibles, impidiendo representar señales de elevada frecuencia.

El problema es parecido al que se produce al muestrear señales unidimensionales con una frecuencia inferior al doble del ancho de banda. En este caso, el muestreo supone la repetición periódica del espectro de la señal analógica en múltiplos de la frecuencia de muestreo. Si esta frecuencia no es lo suficientemente elevada, el espectro original se superpone con las repeticiones periódicas provocando el fenómeno del *aliasing*. La reducción de la resolución vertical observada en el efecto Kell tiene un origen parecido, aunque en este caso debe considerarse el muestreo de funciones de varias variables.

La intercalación de dos campos, uno con líneas blancas y otro con líneas negras, podría ser interpretada de diversas maneras por el observador, constituyendo un claro ejemplo de *aliasing*. Así, una posible interpretación de la señal presentada en la pantalla es que la secuencia obtenida corresponde a una imagen estacionaria formada por líneas blancas y negras alternadas. Otra posible interpretación, coherente con la imagen percibida por el espectador, es que se trata de una imagen intermitente que pasa de ser completamente blanca a completamente negra con una frecuencia igual a la de cuadro. Finalmente, también es posible interpretar que se trata de una imagen formada por líneas horizontales blancas que, sobre un fondo negro, se desplazan en sentido vertical. Esta diversidad de posibles interpretaciones sobre el origen de la imagen indica la existencia de un fenómeno de *aliasing*, ocasionado por la propia estrategia de muestreo de la escena, que el sistema visual interpreta de la manera más plausible integrando toda la información para obtener una imagen aparentemente gris.

Es importante notar que este efecto desaparece cuando se utiliza un sistema de televisión con exploración progresiva, tanto en la cámara como en el *display*. En este caso las imágenes presentadas al espectador cumplen exactamente con los principios de muestreo adaptados a las características del sistema visual que originalmente habíamos propuesto, por lo que no se produce la pérdida de resolución vertical. Sin embargo, si sólo realizamos la exploración progresiva en uno de los subsistemas de adquisición o representación de la imagen, manteniendo el entrelazado en el otro, el efecto Kell no desaparece. Por ello, no tiene sentido utilizar un receptor que digitalice los dos campos y realice una exploración progresiva mientras se mantenga la exploración entrelazada en la cámara.

El efecto Kell suele tenerse en cuenta en la expresión de la frecuencia máxima de la señal mediante la introducción de un factor multiplicativo K , que adapta la ecuación 1.71 a la resolución vertical real del sistema:

$$f_{0\max} = \frac{1}{2} \cdot K \cdot \frac{W}{L} \cdot \frac{N_L}{T_{\text{Línea}}} \quad (1.72)$$

El valor de la constante K depende del sistema de televisión y de las características del sistema de exploración de la cámara y del *display*. En el sistema de televisión de 100 líneas utilizado originalmente por Kell tenía un valor de 0.64. En general, para sistemas entrelazados suele estar comprendido entre 0.8 y 0.6. En sistemas de exploración progresiva puede considerarse, a efectos prácticos, igual a la unidad. Con la introducción del factor de Kell, la expresión de la frecuencia máxima puede tomarse como una medida estimativa del ancho de banda requerido por un determinado sistema.

El hecho de que el factor de Kell sea la unidad para sistemas de exploración progresiva indica que no es necesario doblar el ancho de banda de un sistema entrelazado para convertirlo en uno progresivo que tenga la misma resolución. En efecto, el número de líneas horizontales que pueden visualizarse en el sistema entrelazado puede considerarse que es aproximadamente KN_L , de modo que para obtener la misma resolución en un sistema progresivo podemos reducir el número de líneas del sistema en un factor K sin pérdida de resolución aparente.

De todos modos, la expresión 1.72 sólo debe utilizarse como una medida aproximada del ancho de banda, ya que la imagen que hemos utilizado en su cálculo no puede entenderse como una de las más probables a transmitir. En general, las imágenes suelen tener un gran contenido energético en la región de baja frecuencia y poco contenido en alta frecuencia. Esto es así debido a que la información correspondiente a alta frecuencia está concentrada en los contornos, mientras que la de baja frecuencia corresponde a las regiones con niveles uniformes. Es evidente que la proporción entre las zonas de la imagen que corresponden a un contorno y las que corresponden a una región uniforme es muy favorable a estas últimas.

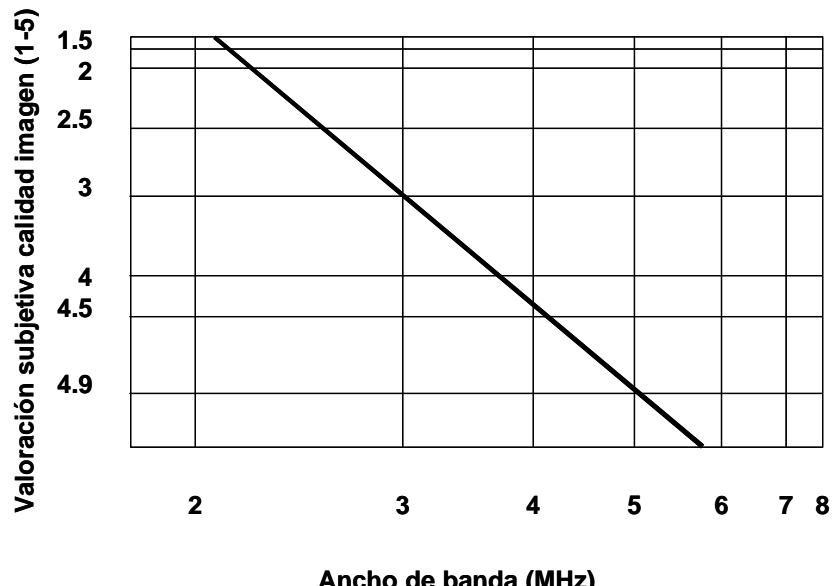


Fig. 1.72 Valoración subjetiva de la calidad de imagen en función del ancho de banda

El elemento determinante para seleccionar el ancho de banda de un sistema de televisión son las pruebas subjetivas sobre la calidad de la imagen en función de la banda de la señal. En la figura 1.72 se representan las medidas de calidad efectuadas con un gran número de espectadores en función del ancho de banda. Las pruebas fueron realizadas por el ITU-R (Comité Consultivo Internacional de Radiocomunicaciones) y se utilizan como referencia para establecer la relación entre el ancho de banda de la señal y la calidad de la imagen. Las curvas se obtienen presentando al espectador señales con distinto ancho de banda, valorando éste su calidad con una calificación entre 5 (calidad óptima) y 1 (péssima). De acuerdo con estos resultados, un ancho de banda de 5 MHz obtiene una valoración media superior al 4,5, lo que suele considerarse más que suficiente para aplicaciones de televisión comercial. En el estudio de televisión suele trabajarse con un ancho de banda superior para garantizar que la calidad de la señal no se degrada en origen. Para tener una idea general sobre esta calificación subjetiva téngase en cuenta que la valoración recibida por un sistema de vídeo doméstico como el VHS es de unos 2,5 puntos, lo que correspondería a señales de televisión de unos 2,5 MHz de ancho de banda.

1.8.14 Resolución y factor de resolución en un sistema de televisión

La resolución es un parámetro de los sistemas de televisión directamente relacionado con el ancho de banda de la señal. Se define como el número de líneas verticales que pueden representarse en una distancia igual a la altura de la pantalla. Esta definición puede parecer un tanto retorcida pero, como todo en televisión, tiene su justificación. En la figura 1.73 se ilustra gráficamente la interpretación de este parámetro.

La resolución se mide como el número máximo de líneas verticales que pueden visualizarse en una región horizontal de la pantalla igual a su altura y siempre tiene un valor entero. Es evidente que la resolución está ligada al ancho de banda global de la cadena de televisión, ya que éste determinará, como hemos visto, la frecuencia máxima que puede transmitirse por el canal y representarse en el monitor. Por otra parte, también es evidente, aunque en general no se haga una referencia explícita en el nombre del parámetro, que se trata de una medida de la resolución sólo en el sentido horizontal. La resolución en sentido vertical depende exclusivamente del número de líneas y del factor de Kell del sistema, quedando completamente especificada con estos dos parámetros. El hecho de contabilizar únicamente las líneas en una anchura igual a la altura de la pantalla se debe a que de este modo pueden compararse directamente la resolución en el sentido vertical y horizontal sin tener en cuenta la relación de aspecto.

El factor de resolución se define como el cociente entre la resolución y el ancho de banda utilizado para transmitir la señal de barras verticales asociada.

$$FR \text{ (Factor de resolución)} = \frac{R}{B_w} \quad (1.73)$$

donde R indica la resolución (en número de líneas) y Bw el ancho de banda. El factor de resolución se mide por tanto en líneas/Hz. Este parámetro es una constante que depende exclusivamente del sistema de televisión y que establece una relación directa entre el ancho de banda del canal y el número de líneas verticales que pueden visualizarse en el receptor.

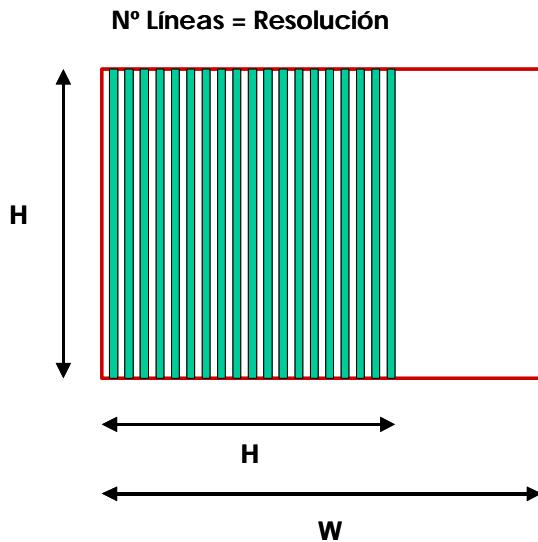


Fig. 1.73 La resolución es el número de líneas verticales en una anchura igual a la altura.

Para expresar el factor de resolución en función de los parámetros básicos del sistema de televisión consideremos un hipotético sistema en el que pueden llegar a visualizarse un máximo de N_H líneas verticales en la pantalla. La resolución del sistema será:

$$R = N_H \cdot \frac{H}{W} \quad (1.74)$$

donde H/W representa el inverso de la relación de aspecto, ya que R se corresponde con las líneas que pueden representarse en una anchura de la pantalla igual a la altura.

Por otra parte, el ancho de banda de una señal de vídeo de N_H líneas verticales se corresponde con la frecuencia máxima asociada a la señal de luminancia. El periodo de la señal puede calcularse como el cociente entre el tiempo de línea activa y el número de ciclos en una línea ($N_H/2$):

$$T_{\min} = \frac{T_{Línea activa}}{N_H/2} \quad (1.75)$$

Por lo que el ancho de banda de la señal será:

$$B_W = f_{0\max} = \frac{1}{T_{\min}} = \frac{N_H}{2 \cdot T_{Línea activa}} \quad (1.76)$$

Sustituyendo esta expresión en la ecuación 1.73 obtenemos:

$$FR = 2 \cdot T_{Línea activa} \cdot \frac{H}{W} \quad (1.77)$$

Si sustituimos estos valores para un sistema de televisión, como el estándar europeo, con una relación de aspecto 4:3 y un tiempo de línea activa de $52 \mu s$, obtenemos:

$$FR = 2 \cdot 52 \cdot 10^{-6} \cdot (3/4) (\text{líneas / Hz}) = 78 \cdot 10^{-6} \text{ líneas / Hz} \quad (1.78)$$

lo que permite establecer una relación lineal entre ancho de banda y resolución horizontal. Cada MHz adicional de ancho de banda proporciona aproximadamente 80 líneas de resolución.

Esta relación puede usarse como alternativa para estimar el ancho de banda necesario para la señal de televisión. Para un sistema como el europeo, que utiliza un total de 625 líneas, de las cuales 575 son activas y, suponiendo que el factor de Kell es del orden de 0.7, obtenemos una resolución vertical de aproximadamente 400 líneas. Es evidente que no tiene sentido que la resolución horizontal sea superior a la vertical, por cuanto la agudeza visual del ojo es la misma en ambas direcciones. Por lo tanto, un ancho de banda aproximado de 5 MHz proporciona una resolución horizontal comparable a la vertical, por lo que es suficiente para una correcta recepción de la señal.

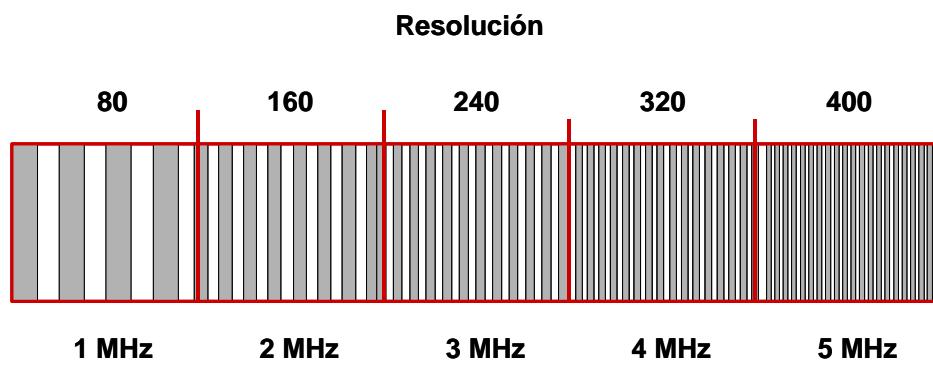


Fig. 1.74 Imagen de prueba para estimar el ancho de banda

La resolución horizontal proporciona un método simple y directo para medir aproximadamente el ancho de banda equivalente del conjunto formado por el transmisor, el canal y el receptor. Para ello basta con transmitir una señal de prueba, constituida por distintas frecuencias, cuya imagen en pantalla permite comprobar qué componentes se transmiten correctamente a través del sistema. Esta señal suele incorporarse en una zona de la carta de ajuste de la mayoría de emisoras y tiene el aspecto representado en la figura 1.74. La imagen corresponde a franjas de barras verticales de distinta resolución. Normalmente se utilizan 5 franjas con resoluciones de 80, 160, 240, 320 y 400, que se corresponden aproximadamente con las frecuencias de 1, 2, 3 4 y 5 MHz. Si en la pantalla puede apreciarse el detalle de las barras verticales en todos los fragmentos, el sistema tiene un ancho de

banda mínimo de 5 MHz. Por el contrario, si el fragmento correspondiente a una resolución de 400 líneas se visualiza como gris, significa que esta componente ha sido filtrada por el sistema de manera que el ancho de banda es inferior a los esperados 5 MHz. Análogamente, si el fragmento correspondiente a una resolución de 320 líneas no se aprecia correctamente, significa que el ancho de banda del sistema es inferior a los 4 MHz, etc.

En la figura 1.75 se representa una alternativa a esta carta de ajuste. En este caso el ancho de banda puede estimarse directamente observando la posición en la que las distintas líneas aparecen superpuestas. Esta imagen suele estar tabulada para simplificar su lectura.

Un método más preciso para estimar el ancho de banda del sistema consiste en transmitir una única línea de test que está formada por distintas frecuencias, previamente calibradas en el centro difusor, con una estructura parecida a la de la imagen de la figura 1.74. No obstante, en este caso la señal de test se transmite dentro de la estructura de la señal de televisión en una de las líneas no visibles. La medida del ancho de banda se realiza mediante monitores de forma de onda que permiten visualizar cómo se han degradado las distintas componentes frecuenciales en el punto de recepción.

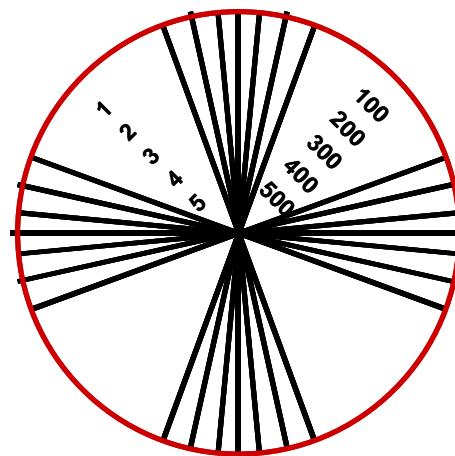


Fig. 1.75 Elemento de una carta de ajuste para estimar el ancho de banda del canal

1.8.15 Inserción de la información en color

Hasta el momento sólo hemos considerado la transmisión de una señal en blanco y negro. La inserción de la información de color puede realizarse de muy diversas maneras dependiendo del sistema de codificación del color empleado. Desde un punto de vista puramente conceptual, una posible propuesta, basada en la misma filosofía que hemos utilizado en la ecuación 1.60 para definir la señal de luminancia, podría ser la de transmitir secuencialmente las tres componentes de color R,G,B para cada una de las líneas en las que se ha dividido la imagen. De este modo, la ecuación 1.60 se generalizaría:

$$\dots, w_{FR}^{(k)}(m, y'), w_{FG}^{(k)}(m, y'), w_{FB}^{(k)}(m, y'), w_{FR}^{(k)}(m+1, y'), w_{FG}^{(k)}(m+1, y'), w_{FB}^{(k)}(m+1, y'), \dots \quad (1.79)$$

Es decir, tal y como se ilustra en la figura 1.76, este hipotético sistema de televisión multiplexaría las tres componentes de color en el tiempo, transmitiendo por tanto toda la información de color asociada a imagen.

Sin embargo, es evidente que no estamos ante una buena estrategia para la transmisión de la información de color por cuanto, tal y como hemos visto en la sección 1.4.6, siempre es recomendable transmitir la luminancia y las señales diferencia de color, tanto para obtener sistemas compatibles como para reducir el ancho de banda necesario para transmitir la información.

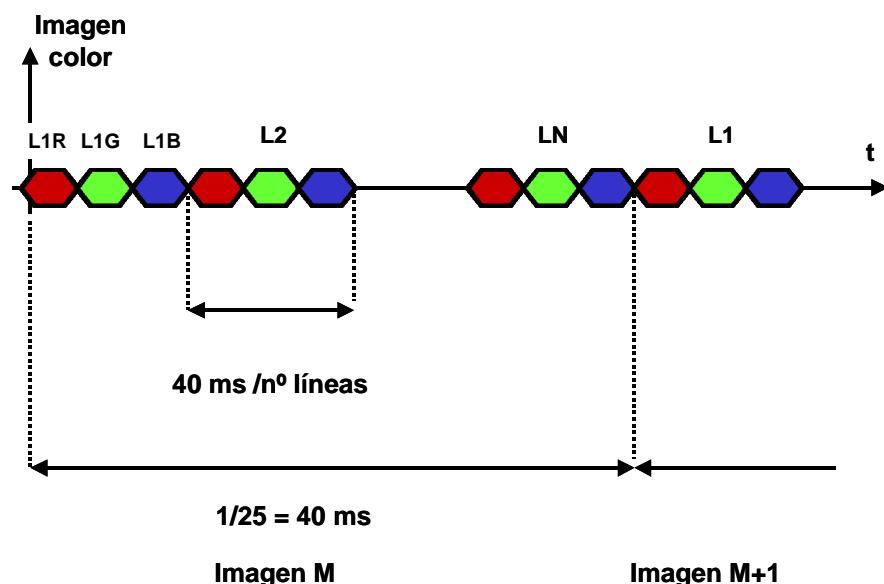


Fig. 1.76 Multiplexación de las componentes de color

1.8.16 Sistemas de vídeo en componentes

No obstante, la generalización de esta idea, sustituyendo las componentes de color R,G,B por la luminancia y las señales diferencia de color, constituye la base para la codificación del color en los denominados sistemas de componentes de color. Pertenece a esta categoría distintos sistemas de televisión y de registro de señal de vídeo en soporte magnético. Entre estos sistemas debe destacarse el MUSE, sistema de alta definición analógico japonés operativo desde finales de los 80; el D2-MAC, tentativa europea para la definición de un formato analógico con definición mejorada que finalmente fue abandonado; la norma digital 4:2:2, sistema digital para el registro, producción e intercambio de material entre estudios; sistemas de registro de señal sobre soporte magnético analógicos como el Betacam o digitales como los sistemas D1, D2, DVCPal, DVCPro; el sistema MPEG-2, estándar para la compresión de imágenes digitales; etc.

En general, la tendencia de todos los sistemas de televisión y vídeo es la de realizar una codificación en componentes de las señales de color. Evidentemente, todos estos sistemas son incompatibles con los sistemas de televisión analógicos convencionales como el NTSC o el PAL.

La codificación del color en componentes admite también diversas variantes dependiendo del sistema considerado. La característica común de todos estos sistemas es que las señales de luminancia y las de diferencia de color pueden ser fácilmente separadas sin la aparición de cruces o interferencias entre ellas. Normalmente, en los sistemas analógicos, la luminancia y las señales diferencia de color están multiplexadas en el tiempo. En la figura 1.77 se muestra una configuración típica de estas señales que se utiliza en el sistema MUSE (también se usaba en el D2-MAC). La señal de luminancia ocupa 2/3 del tiempo de línea activa, mientras que las señales diferencia de color se transmiten en líneas alternas durante el 1/3 del tiempo restante. Nótese que esta codificación aprovecha la menor resolución del sistema visual a las componentes de color mediante un doble artilugio. Por una parte, las señales diferencia de color correspondientes a una línea se comprimen en el tiempo y tienen una duración igual a la mitad que la señal de luminancia. Esto significa, que para un ancho de banda determinado de la señal, las componentes de color tendrán una resolución horizontal igual a la mitad que la luminancia. Por otra parte, la resolución vertical de las señales diferencia de color también se reduce a la mitad, ya que la señal (R-Y) sólo se transmite en las líneas impares y la (B-Y) en las pares. El receptor, para decodificar el color asociado a la línea k, utilizará sistemáticamente la información de las componentes de color correspondientes a la línea k-1 y a la línea k.

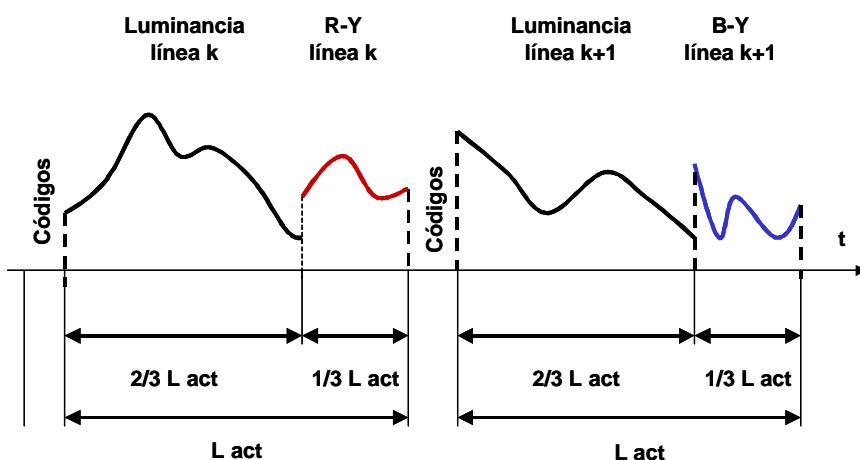


Fig. 1.77 Multiplexación temporal de componentes de color

La codificación utilizada en un sistema de registro de señal de vídeo analógico sobre soporte magnético tiene unas características parecidas. En el sistema Betacam la señal de luminancia y las señales diferencia de color se registran en pistas adyacentes, por lo que su separación también es simple y no existen cruces entre las señales. Las dos componentes de color se comprimen en un factor 2 respecto la luminancia y se registran en la misma pista. El ancho de banda asignado a la luminancia es nuevamente el doble que el de las componentes de color.

Otro sistema de codificación de vídeo que puede considerarse en componentes es el S-Vídeo. En este caso, la componente de luminancia y las componentes de croma se proporcionan mediante cables separados en paralelo. El S-Vídeo se utiliza como un formato doméstico para la interconexión entre equipos de vídeo analógico. Prácticamente todos los receptores modernos incorporan una entrada en S-Vídeo que permite conectarlos directamente con una cámara de vídeo o un magnetoscopio. La señal en S-Vídeo, al tener las componentes de luminancia y color separadas, proporciona una mayor calidad de imagen que la señal de vídeo compuesto. Otra posibilidad es interconectar usando directamente las tres componentes de color R, G y B por separado. En los equipos profesionales es habitual utilizar este tipo de interconexión. En los modelos domésticos el Euroconector permite la interconexión entre distintos equipos de vídeo mediante la transmisión de las tres componentes de color por separado y en paralelo.

La codificación de las componentes de luminancia y diferencia de color en sistemas digitales es algo más compleja y dejaremos su análisis detallado para capítulos posteriores.

1.8.17 Sistemas de vídeo compuesto

Cuando se introdujeron los primeros sistemas de televisión en color era muy importante mantener la compatibilidad con los sistemas de blanco y negro ya operativos. Ello exigía, como hemos visto, transmitir la señal de luminancia con el mismo formato que esperaban los receptores en blanco y negro y superponer la información de croma de manera que pudiera ser decodificada por un receptor en color, pero que tuviera una incidencia mínima sobre la calidad de la señal que se obtendría al usar un receptor en blanco y negro. Estas condiciones imponen que la señal de color de un sistema compatible debe tener la forma:

$$x_c(t) = Y(t) + c_r(t) \quad (1.80)$$

donde $Y(t)$ representa la señal en blanco y negro y $c_r(t)$ la información de croma. Es evidente que como la señal $Y(t)$ se transmite durante todo el tiempo no es posible una multiplexación temporal de las dos señales, como ocurre en el caso de los sistemas de componentes de color. Por otra parte, tampoco se consideró apropiado multiplexar en frecuencia la señal de luminancia y las diferencias de color, por cuanto el ancho de banda de las señales de televisión habría aumentado considerablemente y además, habría dificultado notablemente la retrocompatibilidad del sistema. En efecto, en este supuesto, si una emisora transmitiese en blanco y negro, los receptores en color hubieran esperado la señal de crominancia fuera de la banda de la señal de luminancia, por lo que la señal que finalmente fuera demodulada sería imprevisible, a no ser que se asignaran guardas de transmisión considerables, lo que hace inviable esta alternativa.

En consecuencia, las señales $Y(t)$ y $c_r(t)$ deben compartir tiempo y banda de frecuencia, por lo que deberán imbricarse de manera que su interferencia mutua sea mínima y que, a su vez, la decodificación de la información de color pueda realizarse con sistemas electrónicamente simples. La solución adoptada por los sistemas NTSC y PAL consiste en modular en amplitud las componentes diferencia de color utilizando dos portadoras ortogonales (modulación en cuadratura). El sistema SECAM utiliza una modulación en frecuencia de las señales diferencia de color. En todos los casos,

las frecuencias portadoras de las componentes de color están dentro de la banda útil de la señal de luminancia.

Tomemos como ejemplo el sistema NTSC, la señal de croma se obtiene como:

$$c_r(t) = I(t) \cdot \cos(\omega_0 t) + Q(t) \cdot \sin(\omega_0 t) \quad (1.81)$$

Las señales $Q(t)$ e $I(t)$ están relacionadas con las componentes diferencia de color mediante una relación matricial que ya hemos establecido en el apartado 1.4.6 y que repetimos para mayor claridad:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.783 & -0.269 \\ 0 & 0.478 & 0.414 \end{bmatrix} \cdot \begin{bmatrix} Y \\ (R-Y) \\ (B-Y) \end{bmatrix} \quad (1.82)$$

La estructura de $Q(t)$ e $I(t)$ es idéntica a la de la señal de luminancia, es decir, cada una de estas señales proporciona la información sobre el valor de las componentes de color en un punto de la imagen, en una secuencia de líneas y campos, con las mismas características que la empleada para la señal de luminancia. Las señales $Y(t)$, $Q(t)$ e $I(t)$ están pues perfectamente sincronizadas, de modo que, al aplicarlas a un decodificador NTSC, obtendremos las señales $R(t)$, $G(t)$, $B(t)$ que nos indicarán las componentes de señal que debemos aplicar al tubo de color para obtener las imágenes en color.

Sin embargo, las señales $Q(t)$ e $I(t)$ son moduladas en cuadratura de acuerdo con la ecuación 1.81, para que no interfieran con la señal de luminancia. Un receptor en color debe demodular estas señales antes de aplicarlas a la matriz de decodificación 1.82. La modulación en cuadratura de ambas componentes permite recuperar ambas componentes siempre que el receptor disponga de una referencia de fase de la portadora utilizada en el transmisor. Veremos los detalles del demodulador en el capítulo dedicado a los sistemas de color compatibles, pero ya podemos adelantar que será necesario transmitir, junto con las señales $I(t)$ y $Q(t)$ una señal de referencia que permita sincronizar la fase del receptor con la del transmisor.

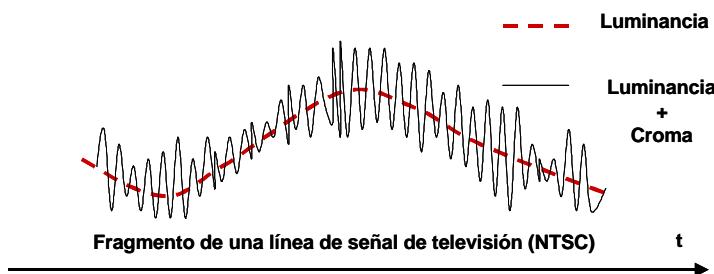


Fig. 1.78 Superposición de las señales de luminancia y croma en el tiempo

Si un receptor en blanco y negro recibe la señal NTSC, sólo será capaz de interpretar la información de luminancia, por lo que las componentes de color serán consideradas como una interferencia. En la figura 1.78 se representa esquemáticamente la forma de onda de la señal NTSC recibida en un fragmento de línea activa. De acuerdo con este esquema, a la señal de luminancia se superpone una

componente de alta frecuencia que contiene la información sobre las componentes de color en su amplitud y fase. Sin embargo, para el receptor en blanco y negro esta componente de alta frecuencia se interpreta como la propia señal de luminancia, por lo que será representada tal cual en la pantalla del receptor.

Es importante, por tanto, elegir la frecuencia de la portadora de la señal de croma de manera que su efecto visual en un monitor de blanco y negro sea mínimo. Por ello, la portadora de color es de alta frecuencia (aunque dentro de la banda), para que sea el propio sistema visual humano el que integre las rápidas variaciones de nivel que se superponen con la señal de luminancia. Además, la frecuencia portadora se elige de modo que en dos líneas sucesivas los efectos de la interferencia se cancelen. Así, si en una determinada posición de la pantalla, la interferencia ha provocado un aumento del nivel de la luminancia, en la posición inmediatamente superior e inferior (líneas adyacentes) la señal de croma provocará una disminución de la misma. Además, el signo de la interferencia también varía de cuadro a cuadro, por lo que se obtiene un complejo patrón interferente que varía tanto en el espacio como en el tiempo y que es apenas perceptible por el ojo que integra toda esta información.

Una posible frecuencia portadora, que permite obtener estos patrones de interferencia poco perceptibles, es utilizar un múltiplo impar de la mitad de la frecuencia de línea. En concreto, en el sistema NTSC se toma como frecuencia portadora el valor:

$$\omega_0 = 2\pi \cdot \frac{f_{Línea}}{2} \cdot 455 = 2\pi \cdot (3,578 \text{ MHz}) \quad (1.83)$$

Desde el punto de vista espectral, este valor de frecuencia portadora supone que las señales de luminancia y croma quedan espectralmente entrelazadas con una interferencia mínima entre ellas. No obstante, es evidente que esta codificación de las componentes de color, compartiendo tiempo y frecuencia con la señal de luminancia, introducirá algunas limitaciones de calidad inherentes a casi todos los sistemas de vídeo compuesto. Todas las hipótesis sobre la cancelación e integración del patrón interferencias suponen que la señal de croma no presenta rápidas variaciones espaciales o temporales. Justificaremos todos estos resultados y veremos un análisis detallado de estos efectos en capítulos posteriores.

1.9 Muestreo espacial de las imágenes: filas y columnas

Las diferentes etapas de muestreo de la escena han permitido representarla como una señal analógica que proporciona información sobre las componentes de luminancia y color en cada punto del espacio y del tiempo. Si ahora aplicamos un muestreo temporal de esta señal y codificamos cada una de las muestras con una palabra binaria, obtendremos una versión digital de la escena que permitirá un fácil tratamiento de la información con procesadores digitales o su transmisión mediante sistemas de comunicación digitales.

1.9.1 Conceptos básicos

Consideremos en primer lugar el caso de una señal de televisión en blanco y negro. La propia estructura de la señal indica que el procedimiento más simple para su digitalización consiste en tomar

muestras equiespaciadas en cada una de sus líneas. Si las muestras se sincronizan con el inicio de las líneas, se obtendrá un muestreo de la imagen con estructura rectangular, en la que cada muestra representa los valores de luminancia obtenidos sobre una retícula rectangular como la que se muestra en la figura 1.79. De este modo, la imagen puede tratarse como una matriz $u[n,m]$, donde el primer índice indica la fila y el segundo la columna a la que pertenece cada muestra. Estas muestras suelen también denominarse elementos de imagen o *píxeles* (que se corresponde con la abreviación anglosajona de *picture elements*).

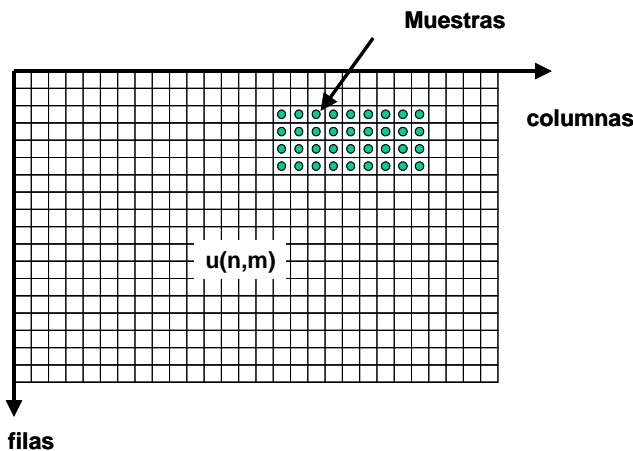


Fig. 1.79 Retícula de muestreo rectangular para la digitalización de las imágenes

El número de muestras que se toman en cada línea de la imagen debe estar en consonancia con el número de líneas con objeto de que la resolución en ambas direcciones tenga valores parecidos. Así pues, en primera aproximación (veremos los detalles más adelante), en un sistema como el europeo que utiliza 575 líneas activas, el número de muestras para tomadas en cada línea debería ser de unas $(4/3) \times 575 = 766$ muestras, donde hemos multiplicado por la relación de aspecto para tener en cuenta que la imagen no es cuadrada. De acuerdo con esta propuesta, la distancia física entre los elementos de imagen en la retícula de muestreo es idéntica en ambas direcciones, por lo que a menudo se dice que a esta estructura de muestreo le corresponde un píxel cuadrado. No obstante, a menudo nos encontramos con sistemas que muestrean la señal de vídeo a una velocidad distinta, obteniendo un número de muestras por línea que puede ser mayor o menor que el propuesto anteriormente. En estos casos se dice que se trabaja con sistemas con elementos de imagen rectangulares.

En la figura 1.80 se muestran distintas versiones de la misma imagen muestreada con distintas resoluciones. En todos los casos, los elementos de imagen son cuadrados.

Por otra parte, es necesario establecer el número de bits con el que debe codificarse cada muestra, o lo que es equivalente, el número de posibles niveles de luminancia que deben asignarse a la señal para que el sistema visual humano sea incapaz de distinguir entre los niveles de gris originales y los cuantificados.



Fig. 1.80 Versiones de una misma imagen con distintas resoluciones

Para determinar este parámetro debemos previamente caracterizar la capacidad del sistema visual para diferenciar entre dos niveles de luminancia distintos. Se considera que para distinguir un círculo de luz con una luminancia L_a sobre un fondo con luminancia L_b , es necesario que se verifique la siguiente relación numérica:

$$\frac{\Delta L}{L_b} \geq \sigma \quad (1.84)$$

donde ΔL representa la diferencia entre los dos niveles de luminancia y σ es una constante, denominada fracción de Weber, cuyo valor depende del observador y del nivel de iluminación total al que está sometida en la retina. En condiciones de luz diurna, el valor de esta constante está comprendido entre 0.02 y 0.05.

Podemos utilizar la relación anterior para calcular el número de niveles de gris que el ojo será capaz de distinguir en un sistema de televisión. Para ello supongamos que el sistema trabajará con unos niveles de luminancia que están situados entre L_{\max} y L_{\min} . La relación entre estos dos niveles se define como el contraste (K) de la imagen:

$$K = \frac{L_{\max}}{L_{\min}} \quad (1.85)$$

Los valores de contraste máximos que puede admitir la retina humana es del orden de 100 por lo que no tiene sentido trabajar con valores de luminancia que vayan más allá de estos límites.

Con estas hipótesis, el valor de luminancia que puede distinguirse inmediatamente después del mínimo será, de acuerdo con la ecuación 1.84:

$$L_1 = (1 + \sigma) \cdot L_{\min} \quad (1.86)$$

Análogamente, el segundo nivel de luminancia que podremos distinguir será:

$$L_2 = (1 + \sigma) \cdot L_1 = (1 + \sigma)^2 \cdot L_{\min} \quad (1.87)$$

Con lo que se infiere que el número máximo de niveles que podrán distinguirse puede obtenerse de la ecuación:

$$L_{\max} = (1 + \sigma)^m \cdot L_{\min} \quad (1.88)$$

Tomando como valor máximo de contraste $K=100$ y un valor típico de $\sigma=0.05$ obtenemos que el número máximo de niveles de gris que pueden distinguirse es:

$$m = \frac{\log(L_{\max} / L_{\min})}{\log(1 + \sigma)} \approx 92 \quad (1.89)$$

Una alternativa, de carácter más experimental, para determinar el número de niveles necesarios para codificar las muestras consiste en presentar distintos tipos de imágenes a espectadores que deben valorar cuál es el número mínimo de niveles a partir del que no se aprecia ninguna mejora en la calidad de las imágenes. Esta experiencia suele dar que el número de niveles de gris se sitúa entre los 45 y 60, lo que indica que la aproximación realizada con el modelo anterior proporcionaba un resultado algo sobreestimado. En la figura 1.81 se muestra una misma imagen representada con distintos números de niveles.



Fig. 1.81 Imágenes obtenidas utilizando distintos números de niveles de gris

De los resultados anteriores se deduce que para codificar correctamente las imágenes se requiere un número de niveles que requiere un mínimo de 6 o 7 bits. Debido a que toda la estructura de las memorias están basadas en palabras de 8 bits, se decidió cuantificar los niveles de luminancia con 8 bits, con lo que se obtiene un total de 256 niveles posibles, que cubren perfectamente las necesidades del sistema visual.

En el caso en que se trate de imágenes en color, puede optarse por muestrear las 3 componentes de color RGB o bien la luminancia y las señales diferencia de color. En el primer caso, la resolución de las tres imágenes debe ser la misma y cada muestra de las componentes de color se suele cuantificar con 8 bits. En el caso de tomar la luminancia y las señales diferencia de color, es habitual muestrear estas últimas con una retícula de muestreo que incluya un menor número de muestras espaciales debido a la menor resolución del sistema visual al color. Normalmente la reducción es en un factor 4, que supone que se utilizan la mitad de filas y columnas. Tanto las muestras correspondientes a la luminancia como a las señales diferencia de color suelen muestrearse con 8 bits.

1.9.2 Sistemas de adquisición

Existen multitud de sistemas para la digitalización de imágenes, que en una primera aproximación podrían clasificarse en sistemas que obtienen un único fotograma de la escena o sistemas que digitalizan ininterrumpidamente la secuencia de imágenes asociada a una señal de vídeo.

Entre los sistemas para la digitalización de un único fotograma se incluyen los *scanners*, las cámaras de fotografía electrónica y las tarjetas digitalizadoras de fotogramas a partir de una señal de vídeo. En todos estos sistemas el formato digital de la imagen proporciona indudables ventajas tanto para su almacenamiento masivo sobre soportes magnéticos u ópticos, como para su tratamiento y análisis mediante ordenador. Existe un gran número de programas tanto de ámbito doméstico como profesional para el retoque y mejora imágenes en formato digital cuyo uso se ha popularizado.

Las dimensiones de la imagen que proporcionan estos sistemas dependen de la tecnología utilizada. En el caso de *scanners* y cámaras fotográficas, suelen utilizarse sensores del tipo CCD que proporcionan una excelente resolución de la imagen, comparable, como mínimo, a la calidad de los sistemas de televisión de radiodifusión. En los *scanners* se utilizan CCD que exploran las líneas de la imagen de manera secuencial (CCD's de línea), obteniéndose resoluciones típicas superiores a los 600 elementos de imagen por pulgada (d.p.i: dots per inch). En el caso de las cámaras de fotografía electrónica se utilizan sensores CCD matriciales con características parecidas a los utilizados en las cámaras de vídeo, obteniendo una calidad parecida. Una ventaja inherente al uso de sensores CCD es, como veremos en el capítulo dedicado a cámaras, que estos sistemas exploran la imagen sobre una retícula de muestreo rectangular, por lo que la conversión a señal digital puede realizarse directamente, direccionando cada uno de los elementos de imagen y convirtiendo el valor de carga adquirido a una palabra binaria.

En las tarjetas digitalizadoras de señal de vídeo, la señal de partida suele ser de vídeo compuesto (obtenido generalmente a partir de una cámara o de un reproductor de vídeo) o bien las componentes de color RGB. El sistema identifica los impulsos de sincronismo de línea y de campo propios de la señal para determinar los instantes en los que debe realizar el muestreo de las líneas de la imagen. El muestreo de un fotograma se realiza cuando el usuario activa una determinada función, que normalmente puede controlar por *software*. El muestreo se inicia en el cuadro siguiente al instante marcado por el usuario y el resultado se almacena generalmente en una memoria propia de la tarjeta que en algunos casos permite la digitalización de varios fotogramas consecutivos. Una vez realizada la adquisición de la imagen, los resultados pueden transferirse a la memoria propia del ordenador.

Aunque generalmente permiten que el usuario defina la resolución con la que se desea digitalizar la imagen, la calidad final depende siempre del tipo de formato de entrada, que excepto para aplicaciones profesionales específicas, suele ser PAL o NTSC. Así, es posible encontrar en el mercado tarjetas digitalizadoras que proporcionan resoluciones de 800 o 1000 líneas a partir de una señal PAL que originalmente sólo contiene 575 líneas activas. El procedimiento utilizado para obtener las líneas restantes es, evidentemente, la interpolación, que en la mayoría de productos es, a lo sumo, lineal.

Los sistemas que digitalizan ininterrumpidamente la secuencia de imágenes suelen utilizarse en equipos de televisión y vídeo digital. La secuencia de bits obtenida puede ser almacenada directamente sobre cinta magnética, como es el caso de los magnetoscopios digitales profesionales, o ser tratada mediante un potente procesador que se encarga de comprimir la información para poder transmitirla en un ancho de banda razonable o almacenarla en soportes de acceso aleatorio (disco duro, CD-ROM).

En general, el proceso de digitalización, almacenamiento y transmisión se realiza mediante sistemas profesionales que sólo se encuentran en estudios de televisión o productoras de vídeo en formato digital. Recientemente han aparecido equipos domésticos que permiten realizar el proceso de digitalización y edición de la señal de vídeo con el uso de tarjetas específicas para ordenador. Estos sistemas proporcionan actualmente una calidad de vídeo comparable al VHS, aunque se espera que pronto puedan obtenerse mejores calidades.

Los formatos digitales más extendidos actualmente en estudios de televisión, y que pueden considerarse como el punto de partida de los distintos equipos para el registro o la transmisión digital, son las normas 4:2:2 y 4:2:0, que definen la conversión a señales digitales tanto para los sistemas de 525 líneas/60 Hz como para los sistemas 625 líneas/50 Hz, manteniendo un gran número de parámetros comunes entre ambos formatos.

1.9.3 Formatos de televisión digital

En 1982, el ITU-R (anteriormente CCIR), en su recomendación 601, desarrolló un conjunto de especificaciones para señales de televisión digital en estudios de televisión o de producción de vídeo. El objetivo de esta recomendación era facilitar el intercambio de programas a escala internacional. Las recomendaciones definen muchos parámetros comunes entre el formato americano y europeo con el objeto de que los fabricantes puedan incluir varios módulos comunes en equipos para 525 y 625 líneas. El uso de este formato permite la interconexión entre distintos equipos digitales.

Existen básicamente dos variantes básicas conocidas como 4:2:2 y 4:4:4, donde los dígitos indican la proporción entre las muestras dedicadas a la luminancia y las dedicadas a las componentes de color. El sistema más utilizado es el 4:2:2, en el que están codificadas la luminancia y las señales diferencia de color. El formato 4:4:4 puede utilizarse también con componentes RGB. En las tablas adjuntas se detallan las características más significativas de estos dos formatos.

Las componentes diferencia de color utilizadas en estos dos formatos han sido definidas previamente en el apartado 1.4.19, en la ecuación 1.51. Las componentes se registran con una corrección de gamma para adaptarse a las características de los *displays* de representación.

Tabla 1.3 Características del formato 4:2:2

ITU-601	Parámetros de codificación en componentes 4:2:2	
Parámetros	Sistemas 525/60	Sistemas 625/50
Señales codificadas: Y, C _R ,C _B	Las coordenadas utilizadas se denotan E _Y , E _{R-Y} y E _{B-Y} . Su relación con las coordenadas del sistema NTSC se discute en el texto.	
Número de muestras por línea completa: • Luminancia • Señales diferencia de color	858 429	864 432
Estructura de muestreo	La retícula de muestreo es ortogonal y se repite según un patrón de línea, campo y cuadro. Las muestras de las componentes diferencia de color se corresponden con las muestras impares de la luminancia (1,3,5,...)	
Frecuencias de muestreo: • Luminancia • Señales diferencia de color	13.5 MHz 6.75 MHz La tolerancia para la frecuencia de muestreo debe coincidir con la tolerancia para la frecuencia de línea del standard de televisión en color analógico de partida. Las frecuencias de muestreo de 13.5 MHz y 6.75 MHz son múltiplos enteros de 2.25 MHz, que es el mínimo común múltiplo de las frecuencias de línea de los sistemas de 525 y 625 líneas.	
Codificación de las muestras	Se utiliza un cuantificador uniforme de 8 bits por muestra del tipo PCM, tanto para la señal de luminancia como para las señales diferencia de color.	
Número de muestras digitales por línea activa: • Luminancia • Señales diferencia de color	720 360	
Relación de tiempo horizontal señales analógica y digital • Desde el fin de la línea activa a O _H	16 períodos de reloj de luminancia	12 períodos de reloj de luminancia

Correspondencia entre los niveles de la señal de vídeo y los niveles de cuantificación: • Escala • Señal de luminancia • Señales diferencia de color	0 a 255 220 niveles de cuantificación donde el nivel de negro se corresponde con el nivel 16 y el nivel de blanco con el 235. El nivel de la señal puede superar, ocasionalmente, el nivel 235. 225 niveles de cuantificación situados en la parte central de la escala del cuantificador con el nivel de señal cero asociado al valor 128.
Uso de palabras código	Las palabras código se corresponden con los niveles 0 y 255 y se utilizan únicamente con propósitos de sincronización. Los niveles 1 a 254 están disponibles para la señal de vídeo.

Tabla 1.4 Características del formato digital 4:4:4

ITU R-601	Parámetros de codificación en componentes 4:4:4	
Parámetros	Sistemas 525/60	Sistemas 625/50
Señales codificadas: Y, C _R , C _B o R, G, B	Las señales se denotan como E _Y , E _R -E _Y , E _B -E _B o como E _R , E _G , E _B . Su relación con las componentes RGB del sistema NTSC se comenta en el texto.	
Número de muestras por línea completa de cada señal	858	864
Estructura de muestreo	Retícula de muestreo ortogonal con repetición en base a la línea, el campo y el cuadro. La estructura de muestreo hace que las tres componentes coincidan y que también coincidan con las posiciones de las muestras de la luminancia en el sistema 4:2:2.	
Frecuencia de muestreo para cada señal	13.5 MHz	
Tipo de codificación	Cuantificador uniforme de un mínimo de 8 bits tipo PCM. (8 ó 12)	
Duración de la línea activa expresada en número de muestras	Mínimo de 720	
Correspondencia entre los niveles de señal de vídeo con los 8 bits más significativos • Escala • R,G,B o señal de luminancia • Señales diferencia de color	0-255 220 niveles de cuantificación donde el nivel de negro se corresponde con el nivel 16 y el nivel de blanco con el 235. El nivel de la señal puede superar, ocasionalmente, el nivel 235. 225 niveles de cuantificación situados en la parte central de la escala del cuantificador con el nivel de señal cero asociado al valor 128.	

Para tener una idea general del volumen de datos binarios que supone la digitalización de una señal de vídeo, consideremos como ejemplo el caso del estándar 4:2:2. La luminancia se muestrea a una frecuencia de 13.5 MHz tomando 8 bits por muestra, lo que da un total de 108 Mbps. Además, cada una de las señales de diferencia de color se muestran a 6.75 MHz, nuevamente con 8 bits por muestra, lo que produce un nuevo flujo de 108 Mbps. En total tenemos un flujo de 216 Mbps asignados exclusivamente a la señal de vídeo digital, sin tener en cuenta las componentes de audio o de información adicional que pueden desear incluirse en la señal de televisión. Este flujo de datos, si bien puede ser soportado en la transferencia de datos entre equipos de un estudio, es excesivo para su transmisión directa al usuario, ya que requeriría utilizar un considerable ancho de banda.

Por ello, en televisión digital es fundamental aplicar procedimientos y tratamientos específicos de los datos que permitan comprimir la cantidad de información que debe transmitirse al canal.

Las dimensiones de la componente de luminancia en los formatos 4:4:4 y 4:2:2 son de 720x576 en los sistemas de 625/50 y de 720x480 en los de 525/30. La diferencia en el número de filas entre ambos sistemas se debe a los estándares de partida analógicos. En el sistema de 625 líneas sólo son activas 575 que se reparten entre los dos campos. Esto da lugar a un total de 287,5 líneas activas por campo. Al digitalizar cada uno de los campos, las medias líneas deben tomarse como una fila completa de la imagen digital, por lo que obtenemos un total de 288 líneas por campo, es decir, 576 (288x2) líneas por imagen. Para el sistema de 525 líneas puede aplicarse un razonamiento similar.

En el formato 4:4:4 se toma una muestra de las dos componentes de color por cada muestra de luminancia. Esto significa que los tamaños de las matrices diferencia de color coinciden con los de las matrices de luminancia. En cambio, en el sistema 4:2:2, las componentes de croma son submuestreadas en un factor 2 dentro de cada fila. Esto significa que los tamaños de las matrices de croma serán de 360x576 en los sistemas 625/50 y de 360x480 en los de 525/60. El submuestreo de la información de croma resulta lógico, puesto que la resolución del sistema visual es menor en la información de croma que en la de luminancia. No obstante, en este formato, sólo se reduce la frecuencia de muestreo en el sentido horizontal. El número de muestras de croma en el eje vertical sigue coincidiendo con el de la componente de luminancia, de modo que la frecuencia de muestreo de la información de croma se mantiene constante a 6,75 MHz. En la figura 1.82 se representa cómo se distribuyen las muestras de luminancia y croma en el formato 4:2:2.

Para reconstruir la información en un *display* es necesario recuperar todas las componentes R, G y B en cada elemento de imagen. En el formato 4:4:4 no existe ningún problema, pues disponemos de la información de croma y luminancia de cada píxel. En el formato 4:2:2 suele realizarse una interpolación lineal para obtener los niveles de croma que no han sido transmitidos o almacenados. Una alternativa a la interpolación lineal consiste en utilizar las mismas componentes de croma que el píxel de la izquierda.

La recomendación ITU-R 656 establece el estándar de interconexión entre dispositivos que utilizan el formato 4:2:2 (básicamente magnetoscopios digitales, editores y mesas profesionales). En la versión en paralelo de la interficie se utiliza un conector del tipo DB25 por el que se van transmitiendo las muestras de 8 bits de las componentes de luminancia y croma, multiplexadas en el tiempo. El orden de transmisión, de acuerdo con la notación de la figura 1.82, es C_{r1}, Y₁, C_{b1}, Y₂, C_{r3}, Y₃, C_{b3}, Y₄, C_{r5}, Y₅, C_{b5}, Y₆, etc. Junto con los bits de cada una de las muestras se transmite un reloj de 27 MHz (un

periodo por muestra). También está definida una versión en serie del interfaz que utiliza un cable coaxial de 75 ohmios con conectores del tipo BNC. En este caso se utiliza una tasa de 243 Mbps, debido a que, en este modo, se transmiten 9 bits por muestra.

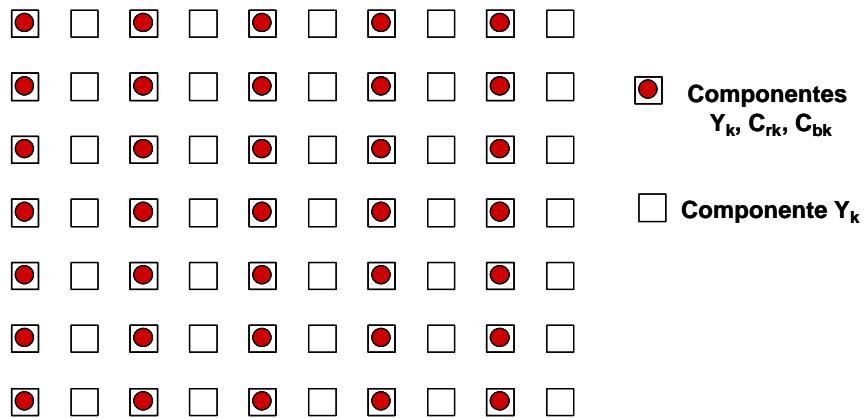


Fig. 1.82 Distribución espacial de las muestras de luminancia y croma en el formato 4:2:2

Formato 4:2:0

El formato 4:2:0 es una simplificación del 4:2:2 que se utiliza en un gran número de aplicaciones para el registro de vídeo en soporte magnético o en disco duro. Es el formato de entrada utilizado en la versión de MPEG-2 que se utiliza en la transmisión de televisión digital (Main Level, Main Profile, MP@ML). Se obtiene reduciendo a la mitad la frecuencia de muestreo de las componentes de croma en el sentido vertical. Con ello, se iguala la densidad de muestras de croma en las dos direcciones. Las muestras de croma se obtienen a partir de las muestras del formato 4:2:2, promediando dos filas consecutivas. En la figura 1.83 se muestra la posición de las muestras de croma respecto a las de luminancia, indicando explícitamente que los valores de croma pueden asociarse a la posición intermedia entre dos líneas (promedio de los valores en cada una de las líneas).

En definitiva, en el formato 4:2:0 el tamaño de las matrices de croma se reduce a 360x240 para el estándar 525/60 y a 360x288 para el estándar 625/50, mientras que las matrices de luminancia se mantienen con el mismo tamaño que en los formatos 4:4:4 o 4:2:2.

Otras relaciones de submuestreo

Además del formato 4:2:2 y el 4:2:0, también están definidos otros formatos de submuestreo de la señal de vídeo como el 4:1:1 y el 4:1:0, que a menudo se confunden con los primeros. En la tabla 1.5 se proporcionan los tamaños de las imágenes utilizados en cada uno de ellos y los factores de diezmado de la señal de croma que se utilizan en cada caso. El formato 4:1:0 fue utilizado durante bastante tiempo por Intel en el DVI (*Digital Video Interactive*). Actualmente, tanto el 4:1:1 como el 4:1:0 están prácticamente abandonados. También existe normativa definida para los formatos 3:2:2, 3:1:1 y 3:1:0, pero prácticamente no han sido utilizados.

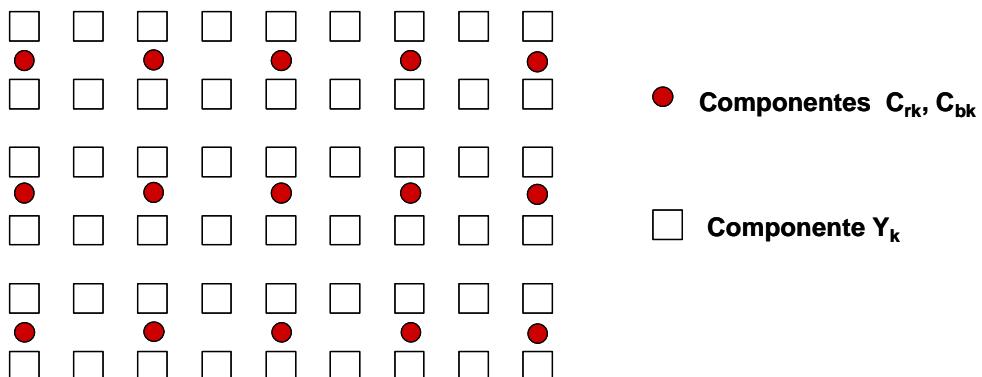


Fig. 1.83 Distribución de las componentes de luminancia y croma en el formato 4:2:0

Tabla 1.5 Formatos de submuestreo de las señales de crominancia

Formato	Muestras Y línea	Muestras Y columna	Muestras C línea	Muestras C columna	Submuestreo horizontal	Submuestreo vertical
4:4:4	720	576/480	720	576/480	---	---
4:2:2	720	576/480	360	576/480	2:1	---
4:2:0	720	576/480	360	288/240	2:1	2:1
4:1:1	720	576/480	180	576/480	4:1	---
4:1:0	720	576/480	180	144/120	4:1	4:1

Formatos reducidos

En algunas aplicaciones donde no resulta necesaria una excesiva calidad de la imagen de vídeo digital suelen emplearse reducciones sobre el tamaño de la imagen. Con ello, se consigue una importante reducción respecto a la tasa de bits original, que puede hacer factible la codificación de la señal en soportes de baja densidad o canales de reducido ancho de banda. Un ejemplo típico de aplicaciones que utilizan formatos reducidos son los ficheros de vídeo para Windows AVI. Otro ejemplo es el MPEG-1, diseñado para codificar señal de vídeo sobre un soporte CD-ROM, que en principio fue diseñado sólo para almacenar información de audio. Todas estas aplicaciones suelen partir de una reducción del tamaño de las imágenes y, además, suelen aplicar codificadores que comprimen la información de vídeo.

Uno de los formatos de vídeo reducido más populares es el SIF (*Source Intermediate Format*) que, esencialmente, consiste en un submuestreo de un factor 2 del formato 4:2:0 que se aplica tanto a las componentes de luminancia como de croma. Los tamaños de las matrices de luminancia para el formato SIF son de 360x288 para el estándar de 625 líneas y 360x240 para el 525 líneas. Las matrices de croma también se submuestrean en un factor 2 en cada dirección respecto las matrices de croma del 4:2:0 (180x144 para 625 líneas y 180x120 para 525). También se realiza una reducción de la

frecuencia de imagen a 25 Hz para el sistema europeo y a 30 Hz para el americano. Con ello, las imágenes resultantes no son entrelazadas. En estas condiciones, se obtiene una calidad equivalente al formato de vídeo analógico VHS.

Las muestras espaciales del formato SIF se obtiene aplicando filtros espaciales y temporales sobre el formato 4:2:0, para eliminar las componentes de frecuencia elevada y posteriormente submuestreando las componentes de luminancia y de croma resultantes. La posición relativa entre las componentes de luminancia y croma, junto con las componentes descartadas se representan en la figura 1.84. La posición de las muestras de croma respecto a las de luminancia se debe a que se utilizan los puntos centrales de las señales filtradas. Con ello se obtiene una muestra de croma centrada en cada cuatro muestras de luminancia, que se utilizará para recomponer todas las componentes de color (RGB) de la imagen.

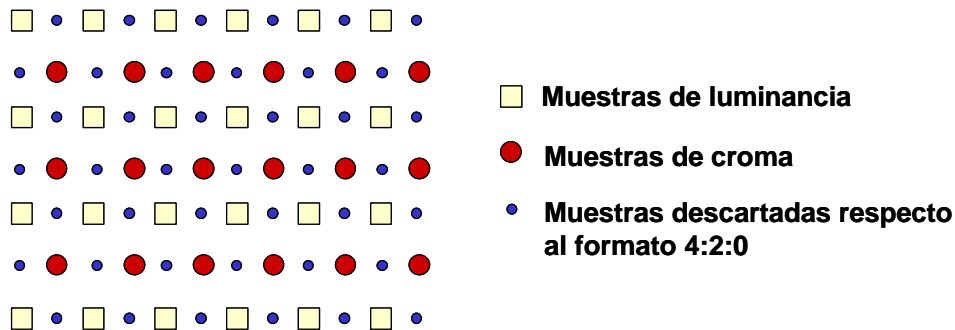


Fig. 1.84 Disposición de las muestras en el formato SIF respecto al 4:2:0

El formato CIF (*Common Intermediate Format*) es un compromiso entre el formato SIF para 625 y 525 líneas. Utiliza 360x288 muestras de resolución de luminancia (europeo) y una frecuencia de refresco de 30 Hz (americano).

Los formatos QSIF y QCIF (Quater) se obtienen reduciendo de nuevo la resolución espacial en un factor 4 (factor 2 en cada dirección) y la resolución temporal en un factor de 2 o 4. Estos formatos suelen utilizarse para la transmisión de señales de vídeo telefonía con el estándar de compresión H261 o para la transmisión de vídeo en directo por internet. Los submuestreos espaciales se realizan filtrando las señales SIF (o directamente la 4:2:0). Las posiciones espaciales de las componentes de luminancia y croma resultantes son parecidas a las del formato SIF.

Formatos de píxel cuadrado

Todos los formatos digitales descritos corresponden a relaciones de aspecto 4:3 y se obtienen submuestreando las componentes de croma y/o luminancia del formato 4:4:4. Recordemos que la recomendación ITU-R 601 establecía una única frecuencia de muestreo para las señales procedentes del NTSC o del PAL, por lo que el número de elementos en cada fila es idéntico para ambos estándares. El problema es que como el número de filas es distinto, el área asociada a cada elemento de imagen no es cuadrada, sino rectangular. En el sistema europeo, para que en una pantalla con una

relación de aspecto 4:3 puedan presentarse 720 columnas y 576 filas, es necesario que los píxeles sean algo más anchos que altos ($4/720 > 3/576$). En cambio, en el sistema americano, los píxeles deben ser algo más altos que anchos.

En aplicaciones informáticas es conveniente que los píxeles tengan una relación de aspecto cuadrada, ya que las tarjetas de visualización y los programas de tratamiento de imagen asumen esta propiedad. Por ello, es conveniente redefinir unos formatos alternativos para poder gestionar las señales de vídeo digital mediante ordenador. Los formatos utilizados en entornos informáticos se suelen denominar formatos de píxel cuadrado y son equivalentes a los formatos que hemos considerado hasta ahora, con la salvedad de que se modifica el número de elementos por fila para forzar que la relación de aspecto del píxel sea cuadrada.

Así, en el estándar de 625 líneas, el tamaño de una imagen 4:2:0 será de 768x576 píxeles ($768 = (4/3) \times 576$) mientras que en el estándar de 525 líneas tendremos un tamaño de imagen de 640x480 píxeles ($640 = (4/3) \times 480$). Este formato es conocido como VGA y, como vemos, proviene del estándar americano. El resto de formatos SIF, CIF y QCIF de píxel cuadrado se obtienen dividiendo las resoluciones espaciales por factores de 2. En la tabla 1.6 se comparan los tamaños de cada formato en las versiones de 625 líneas, 525 líneas y formatos cuadrados.

Tabla 1.6 Comparación entre formatos digitales

	Europeo TV	Americano TV	Europeo Cuadrado	Americano Cuad.
4:2:0	720x576	720x480	768x576	640x480
SIF	360x288	360x240	384x288	320x240
CIF	360x288	360x288	384x288	384x288
QSIF	180x144	180x120	192x144	160x120

Un formato adicional que se utiliza con bastante frecuencia es el SIF de 352x288 píxeles en el sistema europeo y el de 352x240 píxeles en el sistema americano. Este formato se emplea en aplicaciones de compresión de vídeo en MPEG-1 y es una reducción del formato SIF convencional adaptada para poder descomponer las imágenes en bloques de 16x16 píxeles, a los que se aplican técnicas de compensación de movimiento entre fotogramas consecutivos. En este caso, 360 no es un número divisible por 16, por lo que debe reducirse hasta 352 para que puedan fraccionarse las imágenes en un número entero de bloques. Obsérvese que esta modificación no es necesaria en el MPEG-2, que usualmente trabaja en el formato 4:2:0, debido a que 720 sí que es divisible por 16.

Formatos digitales de alta definición

También están definidos los formatos de alta definición digitales, que básicamente consisten en doblar el número de píxeles en cada dirección. En el estándar europeo el tamaño de imagen de alta definición está definido en 1440x1152 para relaciones de aspecto de 4:3 y en 1920x1152 para relaciones de aspecto panorámicas.

2 La señal de televisión en blanco y negro

2.1 Introducción

En este capítulo se exponen los detalles relativos a la forma de onda de la señal de televisión en blanco y negro, tomando como base el sistema analógico de 625 líneas/50 Hz utilizado en Europa. Aunque hoy en día prácticamente se ha abandonado la transmisión de señales en blanco y negro y el parque de receptores monocromos es muy reducido, sigue siendo necesario comprender el formato de esta señal, pues constituye la base sobre la que se definieron los sistemas de televisión en color compatibles. Por otra parte, las componentes de la señal de televisión están directamente relacionadas con las características del tubo de imagen, por lo que será necesario revisar sus principios de funcionamiento para poder justificar los valores y niveles de algunos parámetros de la señal. Así pues, las formas de onda de la señal de vídeo compuesto pueden considerarse una herencia de las restricciones tecnológicas de los primeros receptores y tubos de imagen, por lo que resulta esencial comprender sus características para poder interpretar y entender las señales de televisión analógicas. Aprovechando este breve análisis del tubo en blanco y negro, también veremos los principios básicos de los tubos de imagen de color.

Nuestro objetivo no es, ni mucho menos, realizar un análisis detallado del receptor de televisión, sino simplemente sentar las bases de la televisión analógica para posteriormente poder abordar los modernos sistemas de televisión digital. Por ello, obviaremos, tanto en este capítulo como en el siguiente, cualquier comentario que haga referencia a aspectos circuitales del receptor o del transmisor.

En la primera parte se examinan brevemente los elementos básicos de un tubo de imagen, estableciendo la función de cada uno de ellos. La relación no lineal entre la tensión de control del tubo y la intensidad del haz de electrones nos llevará a la necesidad de realizar una corrección de los niveles de luminancia (también necesaria en los sistemas de color) denominada corrección gamma. Posteriormente se analizarán las formas de onda que permiten controlar la posición del haz en la pantalla, considerando en primer lugar una respuesta ideal de los circuitos de generación de las señales y de las bobinas de deflexión del tubo. Una vez analizado el caso ideal, se evalúa la incidencia sobre el sistema de televisión de los circuitos reales.

El cuerpo principal de este capítulo se dedica a la definición de las señales de sincronismo de línea y de campo que, por otra parte, son idénticas a las que se utilizan en los sistemas de color. Se ha dedicado especial atención a la señal de sincronismo de campo y a toda la secuencia de borrado de

campo, ya que debido a su complejidad, se ha creído conveniente justificar la presencia de cada una de sus componentes en vez de, simplemente, indicar cuál era su forma de onda.

2.2 Principios de funcionamiento del tubo de rayos catódicos en blanco y negro

El dispositivo más utilizado para la representación de imágenes de televisión es el tubo de rayos catódicos. Recientemente han aparecido otros dispositivos para la representación de imágenes como son las pantallas de cristal líquido, de plasma, sistemas de retroproyección por tubos de imagen o láser, etc., que aunque se prevé que en un futuro cercano puedan sustituir al tubo de rayos catódicos, por el momento su aplicación en receptores de televisión domésticos es prácticamente testimonial.

En este apartado pretendemos proporcionar una perspectiva muy general de los elementos que constituyen un tubo de rayos catódicos en blanco y negro y sus principios de funcionamiento. La comprensión de este dispositivo de representación de imágenes resulta fundamental para entender las distintas componentes que se insertan en la estructura de la señal de vídeo. Téngase en cuenta que en el momento de definir la forma de onda de la señal de vídeo en blanco y negro se tuvieron en cuenta, sobre todo, los diversos aspectos tecnológicos relacionados con el receptor de televisión y con el sistema de representación de imágenes. El objetivo fundamental era que la señal de vídeo proporcionara información suficiente al receptor para mantener constantemente un correcto sincronismo con el transmisor, y que todo ello pudiera implementarse con circuitos electrónicos relativamente simples y económicos. Veremos que los sincronismos se insertan dentro de la información de vídeo de manera que sea muy simple su extracción y separación respecto a la componente de luminancia. Asimismo, sus formas de onda están diseñadas para que sean fácilmente separables los dos tipos de información de sincronismo. Algunas especificaciones y tiempos de guarda en las señales de sincronismo pueden parecer excesivamente restrictivos desde el punto de vista de la tecnología actual, por lo que siempre deberemos tener presente el estado de la electrónica en el momento en que fueron definidas.

En la figura 2.1 se representa esquemáticamente una sección transversal de un tubo de imagen monocromo indicando sus componentes básicos. A continuación describimos brevemente la función y características principales de cada uno de estos elementos.

2.2.1 Cátodo, filamento y rejilla

La función principal del cátodo es la de generar los electrones que formarán el haz. Normalmente, está formado por un tubo de níquel recubierto por un material rico en bario y que es calentado mediante un filamento de tungsteno. La temperatura alcanzada y el tipo de material utilizado para el recubrimiento son los factores clave que determinan la temperatura del tubo de rayos catódicos. Algunos sistemas proporcionan una reserva con un compuesto de bario que durante la vida del tubo va refrescando constantemente el área activa. Con ello se consigue aumentar la vida del tubo aproximadamente en un factor tres. Sin embargo, la temperatura de este tipo de cátodos es superior a la de los sistemas convencionales por lo que aumenta la probabilidad de fallo y se reduce la fiabilidad del sistema. Recientemente, se están introduciendo algunos sistemas con cátodos de estado sólido que aprovechan el efecto de inyección de electrones desde el silicio a un dióxido de silicio. Este tipo de sistemas se conocen con el nombre de cátodos fríos y tienen la ventaja de que requieren niveles de tensión mucho

menores para controlar el haz, con lo que se reduce notablemente la complejidad del amplificador de vídeo.

Los electrones generados por el cátodo son acelerados por la tensión positiva aplicada a la rejilla g2 y pasan a través de un pequeño orificio situado en la rejilla g1. La tensión de la rejilla g1 controla la intensidad del haz. Cuanto más negativa sea la polarización de esta rejilla menor será la intensidad del haz. La modulación de la intensidad del haz siguiendo la señal de luminancia puede realizarse modificando la tensión de la rejilla g1, o bien, como es habitual en los receptores modernos, actuando directamente sobre la tensión del cátodo. En este último caso, la tensión de la rejilla g1 determina la intensidad global del haz actuando directamente sobre el brillo de la pantalla.

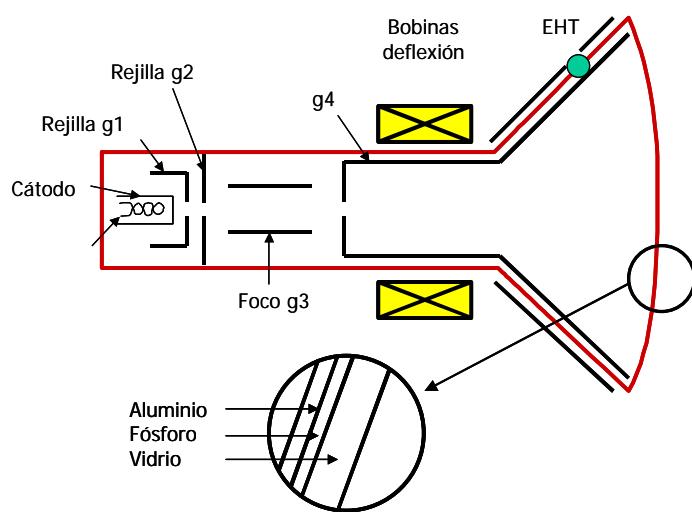


Fig. 2.1 Estructura y componentes básicos de un tubo de imagen en blanco y negro

2.2.2 Rejillas de aceleración y enfoque

Una vez los electrones que constituyen el haz han pasado a través de la rejilla g1, son acelerados por la segunda rejilla g2 (a veces denominada primer ánodo) y penetran en la rejilla g3, cuya función es tratar de enfocar el haz concentrando toda la energía en una pequeña sección de forma aproximadamente circular. Posteriormente, los electrones son nuevamente acelerados hacia la pantalla por la rejilla g4 y el ánodo.

2.2.3 Deflexión del haz

La deflexión del haz puede realizarse mediante sistemas electrostáticos o magnéticos. La deflexión electrostática sólo puede usarse para ángulos de deflexión pequeños, por lo que la mayoría de los

tubos domésticos utilizan una deflexión magnética. Para ello se usan dos bobinas que controlan independientemente la desviación del haz en sentido horizontal (líneas) y vertical (campos).

2.2.4 Pantalla

El haz de electrones impacta contra la pantalla, donde activa los fósforos que convierten la corriente de electrones en luz. Justo de antes de impactar contra los fósforos, el haz pasa a través de una fina capa de aluminio situada entre los fósforos y el cristal. La función de esta película de aluminio es doble. Por una parte protege a los fósforos de los posibles iones que hayan quedado remanentes en el tubo después de su sellado y que al ser bombardeados por el haz de electrones podrían alcanzar altas temperaturas y quemar los fósforos. Esta película de aluminio es suficientemente fina para permitir el paso de los electrones para disipar la energía de los iones. Además, garantiza que toda la luz generada por los fósforos es reflejada hacia el espectador y no hacia el interior del tubo.

En los tubos de color existe una máscara adicional que se encarga de asegurar que cada haz de color impacta únicamente sobre el tipo de fósforos al que está destinado. El inconveniente principal de esta máscara es que intercepta gran parte de la energía del haz, reduciendo por tanto la intensidad de luz. Toda la superficie interior de la cavidad del tubo está recubierta con un material de grafito conductor, que se interconecta con la película de aluminio y el electrodo g4. Este sistema permite garantizar que toda la corriente generada en el cátodo se retorna a la fuente de alimentación.

2.2.5 Forma del raster

En televisión se denomina *raster* al trazado que realiza el haz de electrones cuando explora la pantalla. Idealmente este trazado debería estar formado por líneas rectas (con una ligera inclinación debida al barrido vertical de la pantalla) que están uniformemente separadas. Sin embargo, si el haz sólo se controla mediante sistemas de deflexión lineales, la forma del *raster* en la pantalla no sería lineal. Este efecto, que se ilustra en la figura 2.2, produce una distorsión con forma de cojín en la pantalla y es tanto más importante cuanto más plana es la pantalla.

Existen varios procedimientos para corregir esta distorsión. Una primera alternativa es controlando el devanado de las bobinas de deflexión para que produzcan una distorsión de barril (inversa a la distorsión de cojín, ver figura 2.3) que cancele la primera distorsión. Otra solución consiste en deformar las señales de corriente de deflexión electrónicamente para cancelar la distorsión. Este sistema se utiliza sobre todo en tubos de color de tres cátodos. Otra posibilidad consiste en disponer un conjunto de pequeños imanes permanentes, colocados sobre la periferia de las bobinas o en el exterior de la cavidad de la pantalla, que distorsionan el campo de deflexión del haz y que cancelan la distorsión de cojín. En los tubos actuales de blanco y negro o color suele utilizarse una combinación de todas estas alternativas.

2.2.6 Modulación del haz mediante la señal de luminancia

La intensidad del haz de electrones se controla regulando la tensión existente entre el cátodo y la primera rejilla g1. En la práctica, la modulación de la intensidad del haz mediante la señal de

luminancia puede aplicarse a cualquiera de los dos terminales o incluso a ambos. Actualmente, la técnica más utilizada es la modulación de la tensión aplicada al cátodo, manteniendo fija la tensión de la rejilla g1.

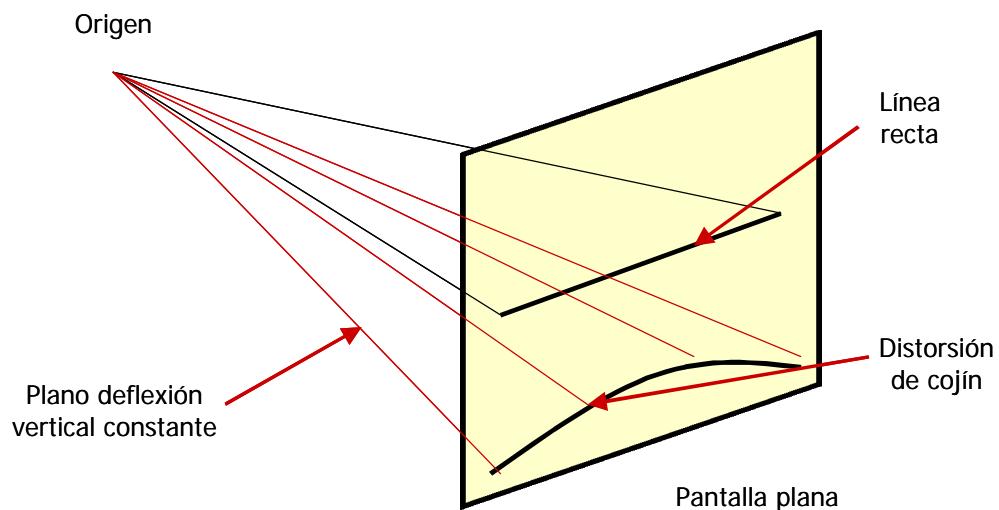


Fig. 2.2 Forma del raster en una pantalla plana: distorsión de cojín

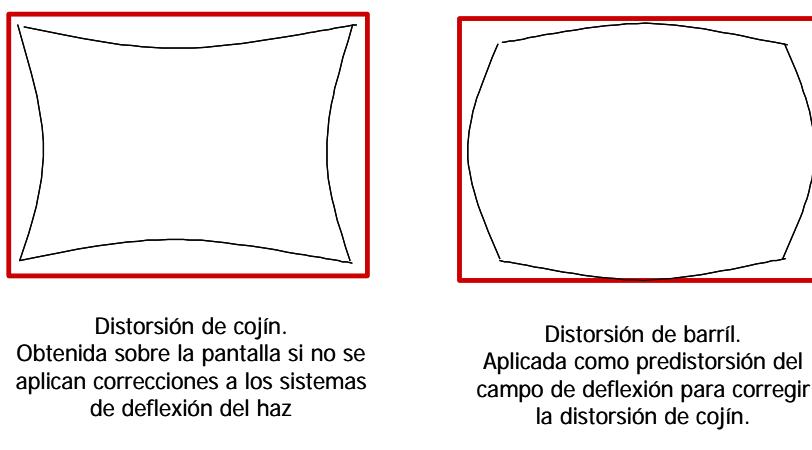


Fig. 2.3 Distorsión de cojín y de barril

La relación entre la tensión aplicada al terminal de control de intensidad del haz y la iluminación es, sin embargo, fuertemente no lineal. Esta relación se representa gráficamente en la figura 2.4 y se corresponde con la ecuación:

$$I_{haz} = \frac{K(1+D)^3 \cdot V_{dr}^3}{(1+D \cdot V_{dr}/V_{co})^{3/2} V_{co} 3/2} \quad (2.1)$$

para el caso en que la tensión de control de intensidad se aplique en el cátodo y:

$$I_{haz} = K(V_{dr}^3 / V_{co} 3/2) \quad (2.2)$$

cuando se aplica sobre la rejilla. D representa el factor de penetración en el fósforo, V_{dr} la tensión aplicada al terminal de control y V_{co} la tensión de corte del haz. Esta no linealidad se conoce con el nombre de gamma y desempeña un papel muy importante en la configuración de las señales a transmitir que analizaremos con cierto detalle en el apartado siguiente.

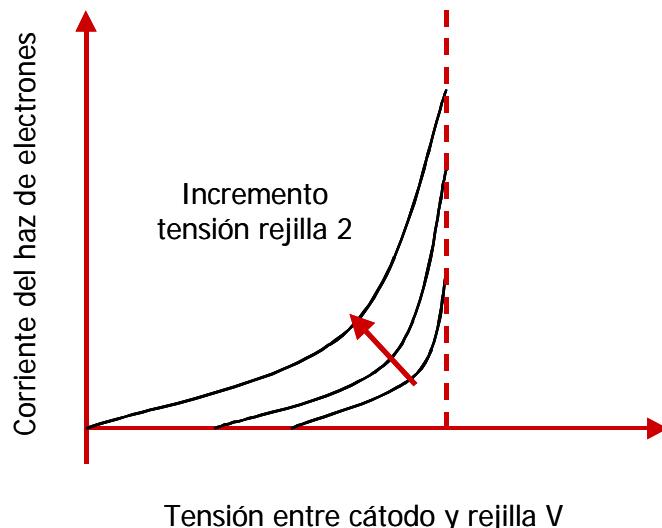


Fig. 2.4 Relación no lineal entre la intensidad del haz y la tensión aplicada al tubo de imagen

2.2.7 Principios de funcionamiento de los tubos de color

El principio de funcionamiento de un tubo de color es parecido al de un tubo de blanco y negro, con la salvedad de que en vez de un único haz que representa la información de luminancia, disponemos de un cañón electrónico para cada componente de color. La formación de imágenes en color se debe a que cada uno de los haces incide sobre materiales electroluminiscentes (fósforos) con componentes de radiación en cada uno de los colores básicos. Existen un gran número de alternativas tecnológicas para la formación de imágenes en color, cuya descripción detallada está más allá de los objetivos de este texto. Los primeros tubos en color fueron propuestos por RCA y estaban basados en tres cañones de electrones dispuestos en una configuración delta cuyos haces impactaban sobre los fósforos de color a

través de una máscara que disponía de una rejilla de puntos. La disposición de los cañones en delta, la máscara y la pantalla con los fósforos se representan en la figura 2.5. Si las señales que se aplican a cada cañón corresponden a las componentes de señal R, G y B y los distintos haces se exploran a través de la pantalla de forma que cada haz, al pasar por la máscara incide sobre el fósforo de color que le corresponde, obtendremos las imágenes en color. A pesar de que el principio de funcionamiento de este tubo de color es muy simple, los requisitos y ajustes necesarios para que se produzcan imágenes con la suficiente definición y pureza son muy complejos. Actualmente, esta tecnología es totalmente obsoleta, aunque sus principios de funcionamiento no han variado excesivamente. La máscara estaba situada unos 2,3 cm por detrás de la pantalla y era una placa metálica con tantos agujeros como tríadas de fósforos existen en la pantalla. La configuración geométrica de los cañones es tal que el haz de electrones sólo puede impactar contra el tipo de fósforo que le corresponde al pasar a través de la máscara. Para que esto sea posible es preciso disponer de diversos sistemas para controlar los sistemas de direccionamiento de los haces de forma separada. Cada haz debe pasar a través del centro de deflexión del tubo, para lo cual deben introducirse dos magnetos, denominados anillos de pureza, que se montan aproximadamente sobre los primeros ánodos del tubo y mediante los que se corrige la posición axial de los haces antes de iniciar su deflexión. Las bobinas de deflexión pueden moverse hacia delante y hacia atrás del tubo para ajustarse al centro de deflexión. El procedimiento de ajuste es complejo y requiere el ajuste individual de cada componente de color hasta obtener una buena pureza. Además, los haces deben converger en todos los puntos de la pantalla, lo que sólo se consigue mediante una combinación de campos magnéticos estáticos y dinámicos que se aplican de forma individual a cada haz.

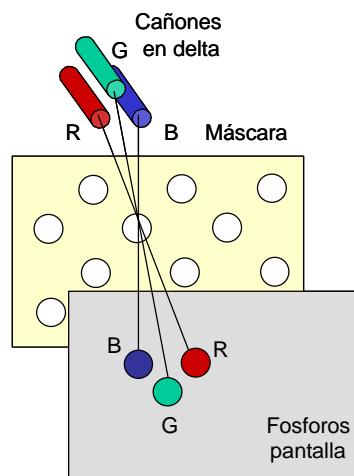


Fig. 2.5 Principios del tubo de color con cañones en delta

La disposición de los cañones en delta se sustituyó por la disposición en línea, que simplificaba considerablemente los problemas de convergencia, debido a la disposición simétrica de los dos cañones que estaban fuera del centro del tubo. Además, las máscaras evolucionaron hacia perforaciones del tipo rectangular, lo que permitía un mayor rendimiento lumínico del tubo (más brillo). La deposición de los fósforos en la pantalla se realizó mediante bandas verticales continuas

que reducían la sensibilidad al polo magnético terrestre, aumentando la pureza y permitiendo un procedimiento de desmagnetización más simple. En la figura 2.6 se muestra un esquema con tres cañones en línea que utiliza fósforos rectangulares.

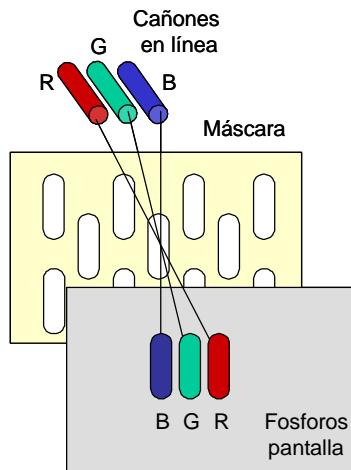


Fig. 2.6 Principios del tubo con cañones en línea y máscaras rectangulares

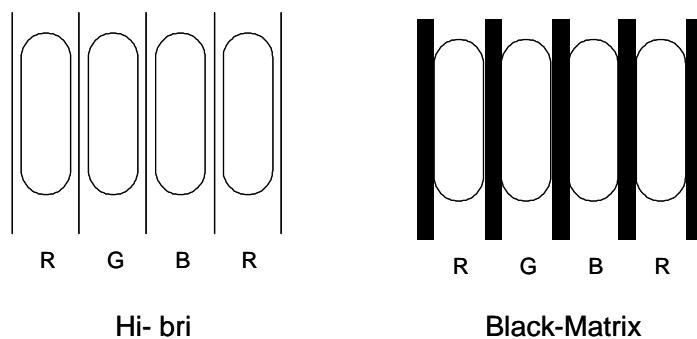


Fig. 2.7 Comparación entre un tubo convencional y un tubo Black-Matrix

Otra mejora introducida en los tubos es el denominado sistema de *black-matrix*, que permite obtener pantallas con menor índice de reflexión de la luz. En una pantalla convencional, cuando está apagada se observa un tono grisáceo, que es el color de los fósforos y que además refleja bastante la luz ambiente. Cuando la pantalla está encendida, el nivel de negro de la imagen queda representado por este color grisáceo y las reflexiones del ambiente que producen los fósforos. Para obtener un elevado contraste es necesario que las partes de la imagen que están activas produzcan una cantidad de luz muy superior a la que reflejan los fósforos. El sistema *black-matrix*, que se obtiene insertando bandas negras entre los fósforos, tal y como se ilustra en la figura 2.7, permite mejorar la relación de contraste reduciendo la luz reflejada por la pantalla. En una pantalla que tenga una distancia entre tríadas de 0.6

mm, las bandas negras son de 0.08 mm, lo que proporciona un total de banda negra por triplete de 0.24 mm, es decir, un 40 % del ancho de la triplete. Esto significa que con esta técnica se absorbe un 40% de la luz incidente, proporcionando una menor intensidad de luz reflejada y por tanto un aumento considerable del contraste que puede obtenerse con el tubo. Además, la inserción de las bandas negras proporciona una distancia de guarda entre los fósforos que aumenta la pureza y reduce la sensibilidad del sistema al campo magnético terrestre.

La introducción del tubo Triniton supuso un avance muy significativo en la producción de tubos en color, ya que se trataba de un tubo muy estable y que requería muchos menos ajustes. Esta tecnología fue desarrollada por Sony en 1968 y sigue siendo la base de los tubos Black-Triniton (Triniton con Black-Matrix) y Flat-Display actuales. La idea básica se representa en la figura 2.8 y consiste en que en vez de tres cañones electrónicos, existe uno solo con tres cátodos, los cuales se encuentran alineados horizontalmente a una distancia inferior a la que se obtiene mediante la configuración de tres cañones en línea. A continuación existe un sistema de electrodos que hacen de lente electrónica mediante la que se realiza el enfoque de los haces que se cruzan en el centro del tubo. El haz verde sigue una trayectoria recta, mientras que los otros dos son deflectados por la placas a4 de modo que cuando atraviesan la máscara van a parar sobre su fósforo correspondiente. La máscara en estos tubos está formada por tiras verticales metálicas continuas, curvada por una sola dirección (curvatura cilíndrica) y sujetada mediante un bastidor mediante el que se consigue mantenerla en tensión. En la pantalla se depositan tiras de fósforos verticales y las bandas negras que las separan y absorben la luz ambiental. La disposición cilíndrica de la pantalla reduce las deformaciones de la imagen. La única convergencia necesaria es la dinámica horizontal, que además se corrige fácilmente actuando de forma electrónica sobre las bobinas deflectoras.

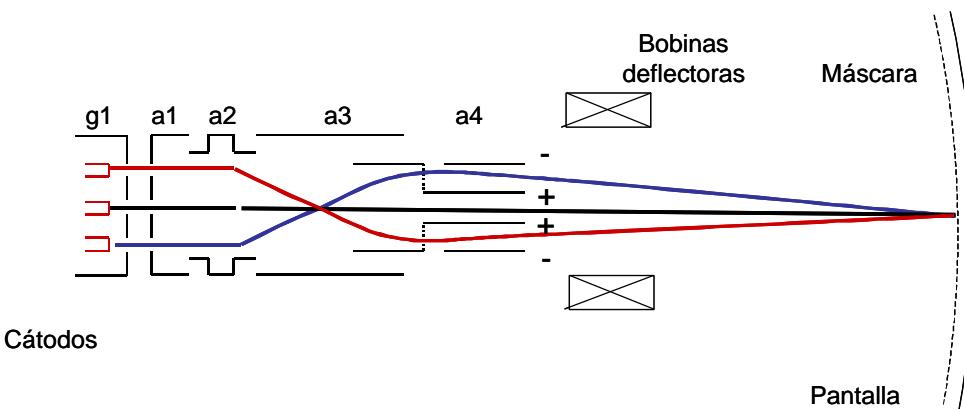


Fig. 2.8 Principios de funcionamiento del tubo triniton

2.3 Corrección gamma

La relación no lineal entre la tensión aplicada al cátodo y la iluminación de la pantalla es un problema propio del receptor que de algún modo debe ser corregido, pues en caso contrario la iluminación que observará el espectador no se corresponderá con la de la escena. La razón de que esta relación no

lineal reciba el nombre de gamma se debe a que, en general, las expresiones anteriores pueden simplificarse y expresarse como:

$$I_{\text{haz}} = K(V_{dr})^{\gamma} \quad (2.3)$$

donde en general el valor del parámetro gamma depende del tipo de control de haz que se implemente y del tipo de fósforos utilizados. Los valores más típicos de gamma están situados entre 2 y 3.

La solución más obvia parece, en principio, corregirla en el propio receptor, pues se trata de un problema originado en éste y que no se produce en las cámaras (que por lo general presentan una excelente linealidad entre la tensión sensada y la cantidad de luz incidente). Sin embargo, ya estamos acostumbrados a que, a veces, en televisión no suelen primar las soluciones más obvias, sino las más económicas. Por ello, para no encarecer el receptor mediante la introducción de circuitos no lineales, se decidió realizar esta corrección en el transmisor, pasando las señales de luminancia o de croma a través de unos sistemas no lineales que realizaran una transformación inversa, de modo que cuando esta señal se aplicara directamente al tubo de imagen, se obtuviera una iluminación proporcional a la que originalmente había en la escena.

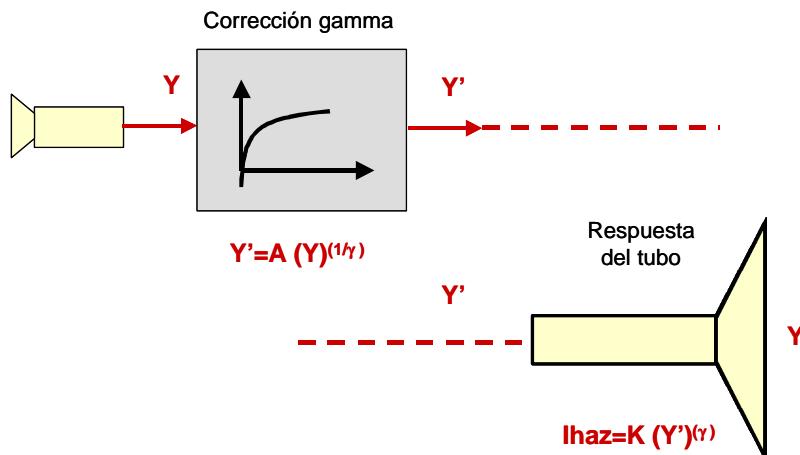


Fig. 2.9 Principio de la corrección gamma en el transmisor

La idea general se representa en la figura 2.9 donde se indica que la señal de luminancia originada en la cámara se pasa a través de un dispositivo no lineal (implementado mediante diodos y transistores que, por supuesto, trabajarán en zona no lineal) que modifica su amplitud de acuerdo con la relación:

$$Y' = A(Y)^{1/\gamma} \quad (2.4)$$

donde A representa un factor de ganancia constante.

Al aplicar esta señal de luminancia corregida directamente al tubo de imagen se obtiene una corriente del haz que tiene una relación de proporcionalidad con la luminancia original:

$$I_{haz} = K(Y')^\gamma = cte \cdot Y \quad (2.5)$$

De este modo, podemos garantizar que la imagen que finalmente será visualizada por el espectador se corresponderá con la luminancia original de la escena.

El valor de la constante gamma que debe utilizarse está normalizado por el ITU a un valor de 2.2 para los sistemas de televisión en blanco y negro y a un valor de 2.8 para los sistemas en color. En este último caso, las señales a las que se aplica la corrección de gamma son las componentes R, G y B obtenidas en la cámara. Con ello, se obtienen las componentes:

$$\begin{aligned} R' &= A(R^{1/\gamma}); \\ G' &= A(G^{1/\gamma}); \\ B' &= A(B^{1/\gamma}); \end{aligned} \quad (2.6)$$

que posteriormente se combinan, de acuerdo con las ecuaciones presentadas en el capítulo anterior, para obtener la señal de luminancia y las componentes diferencia de color que deben transmitirse en el sistema de color. Así, en el caso del sistema PAL, las señales que finalmente se transmiten vienen determinadas por las expresiones:

$$\begin{aligned} Y' &= 0.3 \cdot R' + 0.59 \cdot G' + 0.11 \cdot B'; \\ (R - Y)' &= R' - Y'; \\ (B - Y)' &= B' - Y'; \end{aligned} \quad (2.7)$$

Sin embargo, como probablemente el lector ya habrá observado, el uso de una constante gamma distinta para los sistemas en blanco y negro y los de color supone una cierta pérdida de compatibilidad entre ambos. En efecto, cuando se transmite la señal de color y se visualiza en un receptor en blanco y negro, la luminancia que se aplica al tubo de imagen se ha obtenido corrigiendo las componentes R, G, B mediante una gamma de valor 2.8, mientras que la no linealidad del tubo monocromo tiene un valor de gamma de 2.2. Así pues, la señal de luminancia que se aplica al receptor en blanco y negro ha sido sobrecorregida, con lo que la información que recibe el espectador no se corresponde con los valores originales de la cámara. El efecto global es que la imagen obtenida es algo más clara que la original. No obstante, esta no linealidad es tolerable.

Otro problema derivado de la corrección de gamma aparece en sistemas para la representación de imágenes basados en pantallas de cristal líquido (L.C.D.). En estos sistemas la linealidad entre la tensión de control y la iluminación de la pantalla es excelente, por lo que no requerirían ninguna corrección. No obstante, al haberse aplicado una corrección de gamma en la transmisión, es necesario introducir en estos receptores un sistema no lineal que elimine la primera transformación y restaure las señales de luminancia o de color originalmente obtenidas en la cámara. Vemos pues que, desde el punto de vista de estos monitores, la corrección de gamma realizada en el receptor no parece una decisión muy acertada. Por todo ello, es muy posible que los futuros sistemas de televisión no realicen la corrección de gamma y dejen esta tarea, si es necesaria, al receptor.

2.4 Señales de deflexión del haz

2.4.1 Caso ideal

Las señales que se aplican a las bobinas de deflexión del haz tienen, en el caso ideal, una forma de diente de sierra. El periodo de las señales coincide con el periodo de línea para la corriente de deflexión horizontal y con el de campo para la vertical. Así pues, para el estándar europeo la frecuencia de la señal de barrido horizontal es de 15625 Hz, mientras que la de barrido vertical es de 50 Hz. Su amplitud oscila entre unos valores máximo y mínimo que se corresponden con las corrientes que deben aplicarse a las bobinas de deflexión para posicionar el haz en los extremos de la pantalla. Las formas de onda se representan esquemáticamente en la figura 2.10, donde hemos supuesto que la corriente horizontal mínima se corresponde con la posición del haz en el extremo de izquierdo de la pantalla y el mínimo de la corriente vertical con la posición del haz en el extremo superior. Este convenio a veces se intercambia, por lo que en algunos textos las formas de onda pueden aparecer invertidas. Es evidente que en la práctica estas formas de onda sólo podrán ser aproximadas, pues la corriente en una bobina no puede cambiar bruscamente. Sin embargo, supongamos por el momento que estas señales son ideales y que la desviación del haz es lineal con la corriente aplicada.

En este supuesto, al aplicar estas corrientes a las bobinas de deflexión, el haz realiza una exploración de la pantalla de acuerdo con el patrón indicado en la figura 2.11. Suponiendo que el primer campo se iniciase en el extremo superior derecho de la pantalla, el haz iría explorando cada línea hasta el extremo derecho, tras lo cual retornaría inmediatamente hasta el extremo izquierdo. Nótese que la línea no es perfectamente horizontal, sino que tiene una pequeña inclinación debido a que la corriente de deflexión vertical disminuye ligeramente durante la exploración de la línea. Esta inclinación ha sido exagerada en la representación de la figura 2.11 con objeto de poder observar las transiciones entre líneas y campos con mayor claridad. El desplazamiento vertical entre los dos extremos de una línea debería de ser de $2H/N_L$ ya que la señal de barrido vertical debe realizar la exploración completa de la pantalla (es decir, toda su altura: H) en el tiempo asignado a $N_L/2$ líneas.

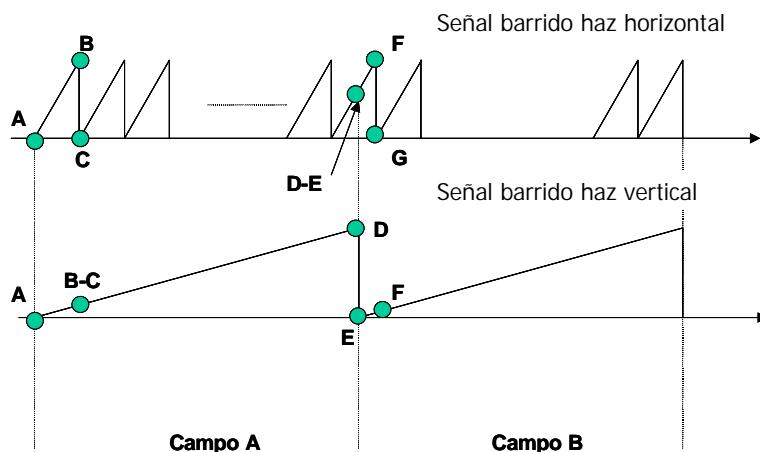


Fig. 2.10 Señales ideales de deflexión del haz en un tubo de imagen en blanco y negro

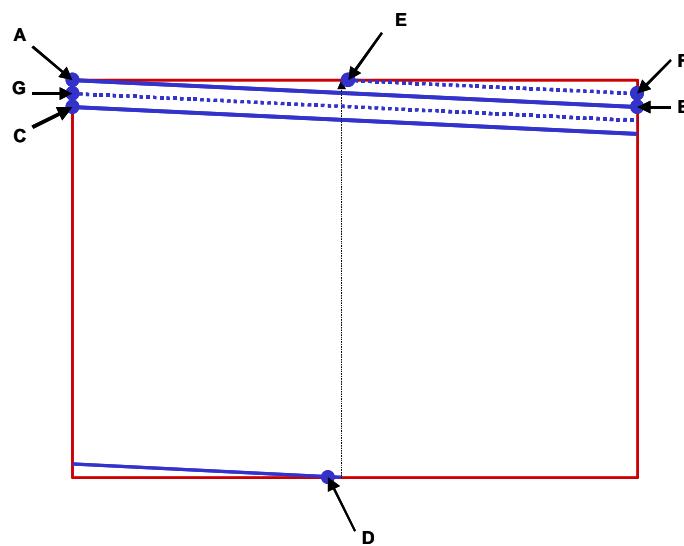


Fig. 2.11 Exploración del haz en la pantalla con señales de barrido ideales

Este ligero desplazamiento vertical permite que la exploración de la siguiente línea se inicie en una posición vertical situada algo por debajo de la línea precedente. En las figuras 2.10 y 2.11 se representan las posiciones sobre la pantalla y las tensiones asociadas a estos puntos característicos con los símbolos A (punto de inicio de la exploración), B (punto final de la primera línea) y C (punto inicial de la siguiente línea, que está situado en la misma posición vertical que el punto B).

Por otra parte, como el número de líneas de los sistemas de televisión analógicos es, como ya se había argumentado en el capítulo anterior, impar, el final del primer campo se produce cuando la señal de barrido horizontal se encuentra en la mitad de la pantalla (posición correspondiente a media línea). Esto significa que el haz retorna al extremo superior de la pantalla con una posición horizontal centrada, tal y como se ha representado en la figura 2.11. Los puntos y las tensiones asociadas al final del primer campo y el inicio del segundo se representan respectivamente con las letras D y E. La posición en la pantalla de esta primera media línea se trazaría justo por encima de la primera línea del campo anterior. Además, esta nueva línea alcanza el extremo derecho de la pantalla en una posición vertical (punto F) de valor mitad al que le correspondía en el punto B, de modo que el inicio de la siguiente línea (punto G, que se encuentra situado en la misma posición vertical que el F) se representaría entre las dos primeras líneas del primer campo. De esta forma, las líneas correspondientes a cada uno de los dos campos quedan perfectamente imbricadas, obteniendo un correcto entrelazado de las imágenes.

2.4.2. Señales de deflexión del haz: aproximación al caso real

En principio, si las señales de deflexión del haz fueran ideales podríamos transmitir la información de luminancia en todo instante de tiempo, sin necesidad de proporcionar los tiempos de guarda definidos por las partes activa y no activa de las líneas y los campos. Sin embargo, la desviación de los

generadores de diente de sierra respecto al supuesto ideal impone que no pueda transmitirse la señal de luminancia durante el periodo de tiempo en el que se produce la conmutación del generador. Sin embargo, estos tiempos de guarda se utilizan, tal y como hemos visto en capítulo anterior, para transmitir la información de sincronismo, por lo que no son totalmente desaprovechados.

La forma exacta de las señales de deflexión del haz es difícil de predecir, ya que dependen del circuito electrónico utilizado para su síntesis, que, como es obvio, presenta diferencias significativas dependiendo tanto del modelo de receptor como del fabricante. Además, los circuitos de deflexión horizontal y vertical suelen ser muy distintos, ya que las frecuencias que deben generar son muy dispares (50 Hz y 15625 Hz). Téngase en cuenta que el modelo circuital de las bobinas de deflexión depende de la frecuencia, por lo que las cargas de los circuitos de barrido horizontal y vertical son considerablemente diferentes, lo que exige el empleo de filosofías de generación de la señal distintas. En nuestro caso tan sólo nos interesa tener una idea general de cómo las formas de onda reales se apartan del caso ideal y de cómo ésta desviación influye en la definición de los parámetros de la señal de televisión. Por ello, obviamos el problema circuital de la generación de las señales y nos referiremos a una hipotética forma de onda no ideal, que se representa en la figura 2.12 y que aplicaremos tanto para el barrido vertical como el horizontal.

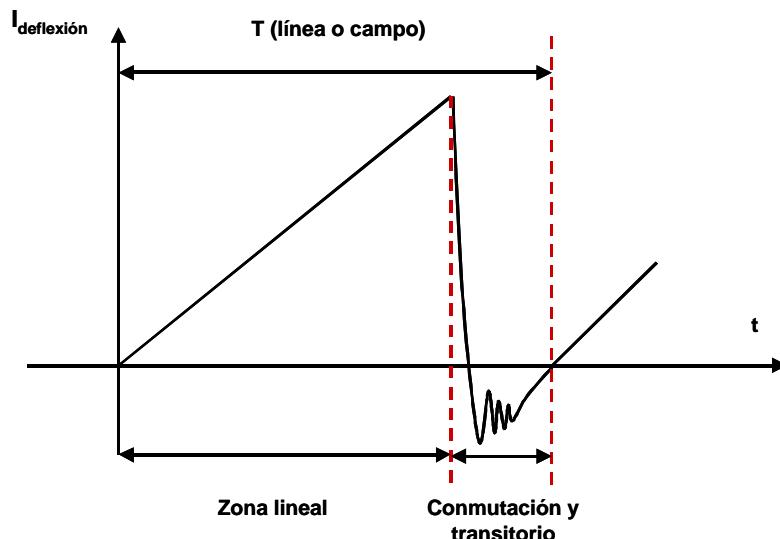


Fig. 2.12 Forma de onda no ideal de un generador en diente de sierra

Esta señal genérica tiene una zona lineal que se corresponde con la exploración uniforme del haz sobre la pantalla que se obtenía con el diente de sierra ideal. Sin embargo, la discontinuidad del diente de sierra, que permitía el retorno inmediato del haz para iniciar la exploración de una nueva línea o campo, desaparecerá y, aunque es posible obtener un tiempo de conmutación reducido, es importante tener en cuenta este aspecto y, como mínimo, definir una señal de televisión que permita que los fabricantes puedan utilizar generadores de diente de sierra simples y económicos. Obsérvese que las bobinas de deflexión representan una carga inductiva importante que limita la capacidad del generador

de commutar rápidamente. Esta zona de transición suele ser no lineal y varía notablemente de un generador a otro. Su efecto sobre la exploración del haz en la pantalla es que el retorno al inicio de la línea o del campo no se realiza inmediatamente, sino que se invierte un cierto tiempo finito. Durante este periodo de tiempo es importante garantizar que la energía del haz sea prácticamente nula, puesto que de lo contrario el retorno del haz sería visible en la pantalla. Nótese que debido a la variabilidad que presenta esta región para distintos generadores, es absurdo intentar aprovecharla para volcar parte de la información de luminancia.

Otro aspecto que se ha indicado en la figura 2.12 es la aparición de un régimen transitorio, con oscilaciones amortiguadas, que aparecen antes iniciar la zona lineal y que se conocen con el nombre de oscilaciones de *Barkaunsen*. Estas oscilaciones, cuyo nivel ha sido exagerado en la gráfica, son comunes a todos los generadores de diente de sierra y suele considerarse que han desaparecido completamente después de un tiempo de aproximadamente un 10 % del periodo de la señal. Como en el caso anterior, es importante garantizar que no se transmite información de luminancia durante el tiempo en que éste se mantiene transitorio. Por tanto, la señal de televisión deberá proporcionar un tiempo de guarda, durante el que no se transmite información de luminancia y que permite que el haz retrone a su posición de origen. Este tiempo de guarda será distinto para el retorno de línea y el de campo, ya que los dos generadores tienen características diferentes. En el caso de la señal de desviación horizontal, el tiempo de guarda es de unos 12 μs , dejando por tanto un total de 52 μs para la transmisión de la información de luminancia. Los 12 μs de guarda se aprovechan para enviar la señal de sincronismo de línea y, en el caso de televisión en color, para transmitir una referencia de fase de la portadora de la información de croma.

Durante el cambio de campo, la señal de luminancia deja de transmitirse durante un tiempo de 1,6 ms (25 líneas). Nuevamente, durante este periodo de tiempo se transmiten las señales de sincronismo de campo y en este caso, se utilizan algunas líneas para la transmisión de la señal de teletexto y para la transmisión de líneas de prueba (test), cuya posición exacta indicaremos más adelante. Es importante observar que sólo una parte de estos tiempos de guarda se corresponden con el retorno horizontal o vertical del haz. Los tiempos de guarda o, en otras palabras, las partes no activas de la señal de televisión, están sobredimensionados para permitir acomodar distintos tipos de circuitos de barrido y proporcionar tiempo suficiente para garantizar que el receptor está en condiciones de detectar el envío de las señales de sincronismo o iniciar una nueva exploración de una línea o campo. Veremos los detalles de las distintas componentes que se transmiten durante estos periodos en las secciones siguientes.

2.5 Sincronismo de línea

El sincronismo de línea se transmite con el objeto de informar al receptor de la ocurrencia del final de una línea e inicio de otra. Esta señal debe poder ser separada con facilidad de la señal de luminancia, por lo que se transmite con una polaridad invertida respecto a la primera. En la figura 2.13 se representa la forma de onda de la componente de sincronismo de línea, junto a una hipotética información de luminancia asociada a las líneas entre las que está situado el pulso de sincronismo.

La amplitud del impulso de sincronismo representa aproximadamente un 30 % de la excursión total de la señal de televisión y se sitúa por debajo del nivel de referencia o de cero voltios. Es decir, si

normalizamos la excusión de esta señal a 1 voltio pico a pico, obtendremos que las tensiones de la señal de luminancia van desde 0 voltios (nivel de negro) a 0.7 voltios (nivel de blanco), mientras que los impulsos de sincronismo se situarían a un nivel de -0.3 voltios. Debemos comentar, no obstante, que los niveles indicados se corresponden con una señal de televisión normalizada a 1 voltio y que en la práctica es habitual disponer de esta señal atenuada o amplificada. Por otra parte, también es frecuente encontrar representaciones de la señal de televisión en las que los niveles de los pulsos de sincronismo y la señal de luminancia están invertidos respecto a las formas de onda de la figura 2.13. La diferencia entre las dos representaciones es una inversión de la señal que puede producirse en cualquier punto de la cadena de transmisión, por lo que no debe ser significativa. En este texto usaremos indistintamente las dos representaciones, por lo que es importante que el lector identifique por el contexto cuál de las dos estamos utilizando. En general, cuando hagamos referencia a las señales de sincronismo, éstas se representarán como positivas, mientras que si el aspecto central de la discusión es la señal de información de luminancia o color de la escena, se tomarán los sincronismos como negativos.

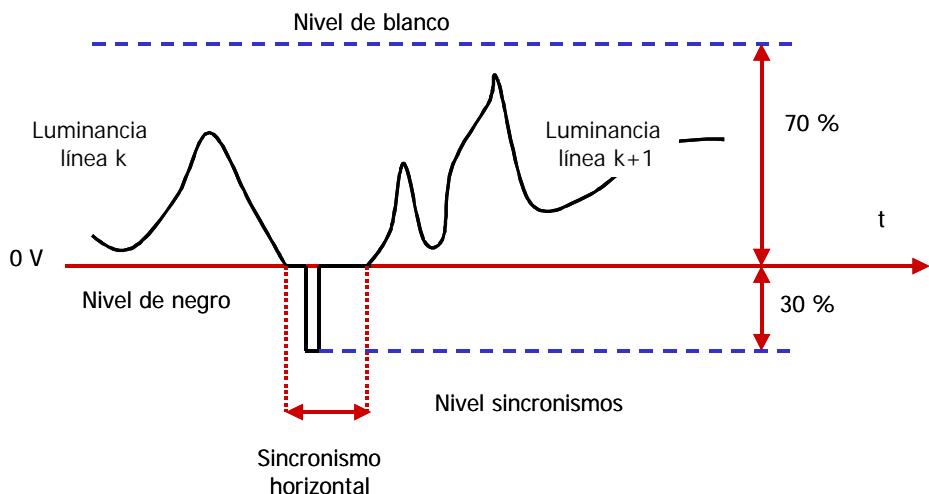


Fig. 2.13 Inserción del sincronismo horizontal entre las líneas de la señal de TV

Aunque, como veremos, la forma de onda de la señal de sincronismo vertical es bastante más compleja que la de sincronismo horizontal, los niveles utilizados para transmitir este tipo de sincronismo son también de signo opuesto al de la luminancia y del mismo nivel que los impulsos de sincronismo horizontal. Por ello, la separación de las componentes de sincronismo y de luminancia puede realizarse de manera muy simple, ya que basta con comparar la señal recibida con el nivel de negro (0 voltios).

Un posible diagrama de bloques genérico para el control de las señales que se envían al monitor se representa en la figura 2.14, donde vemos que la señal de luminancia se aplica al tubo de imagen para modular la intensidad del haz, mientras que las señales de sincronismo se envían como referencia a los

circuitos de generación de las señales de deflexión del haz. La perfecta coordinación entre estas tres señales permite visualizar la imagen en el monitor.

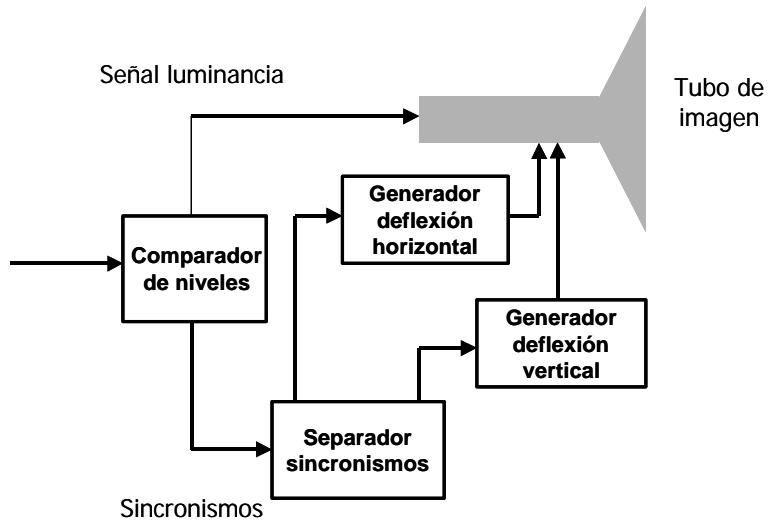


Fig. 2.14 Diagrama de bloques del sistema de separación de sincronismo y luminancia aplicado al tubo

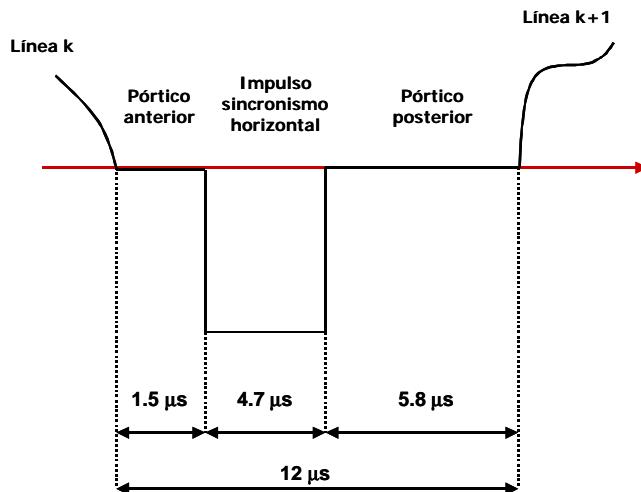


Fig. 2.15 Detalle del impulso de sincronismo horizontal

En la figura 2.15 se representa un detalle del fragmento de señal correspondiente al impulso de sincronismo horizontal donde se indican tres partes diferenciadas, cumpliendo cada una de ellas una misión específica. El impulso de sincronismo tiene una duración de 4.7 µs y está precedido por una zona denominada pórtico anterior, que tiene el mismo nivel que el negro y una duración de 1.7 µs. La

misión del pórtico anterior es garantizar que el impulso de sincronismo está precedido por una zona con nivel de negro constante para eliminar cualquier dependencia del inicio del pulso de sincronismo con el nivel de luminancia con que finaliza la línea. Esta utilidad se ilustra en la figura 2.16, donde se indica que si una línea termina con un nivel de gris próximo al blanco, se requiere de un cierto tiempo para que el nivel de la señal retorne al negro. Si se recibiera el pulso de sincronismo durante esta transición, podría ocurrir que no fuera detectado o que fuera detectado con retraso. La función del pórtico anterior es por tanto la de garantizar que el receptor se encuentra en óptimas condiciones, independientemente del nivel de gris con el que finalizan las distintas líneas, para detectar el impulso de sincronismo de línea, que será utilizado como referencia temporal para reinicializar el generador de barrido horizontal.

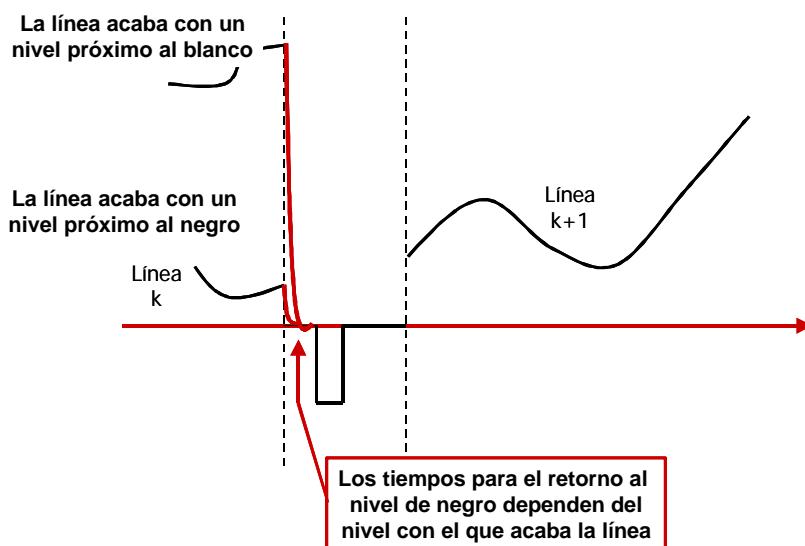


Fig. 2.16 Ilustración de la misión del pórtico anterior

En principio, la detección del sincronismo suele realizarse mediante un filtro paso alto, detectando la primera transición del impulso. Este momento es, en principio, el punto en que el generador de diente de sierra conmuta para reinicializar la exploración del haz. El tiempo restante desde este momento y hasta que se recibe la información asociada corresponde al propio impulso de sincronismo y al pórtico posterior. Todo este periodo de tiempo, que se extiende hasta 10.5 μ s, se utiliza como guarda para garantizar que el generador de la señal de barrido ha tenido tiempo suficiente para situarse en la zona lineal y puede aplicarse la señal de luminancia.

Consideraremos ahora, con algún detalle, cómo afectaría a la imagen que se visualiza en la pantalla el uso de dos generadores de barrido horizontal que tuvieran unos tiempos de transición distintos. Para ello supongamos que se dispone de dos generadores que entregan unas señales como las que se representan en la figura 2.17. Es importante observar que los dos generadores cumplen con las condiciones exigidas por la señal de televisión, ya que aunque uno de ellos conmute mucho más

rápidamente que el otro, los dos están dentro de los límites esperados. Si no fuera así, ya hemos comentado que las no linealidades y transitorios del generador serían visibles en la pantalla deteriorando gravemente la imagen.

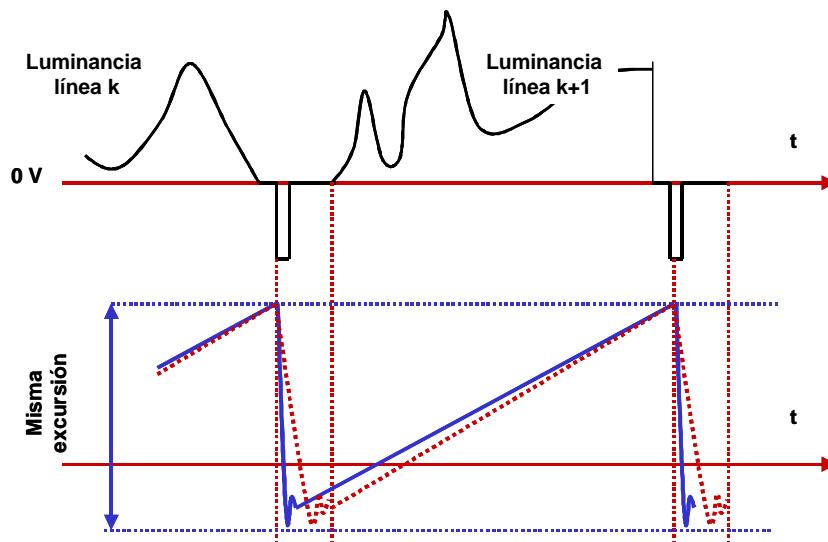
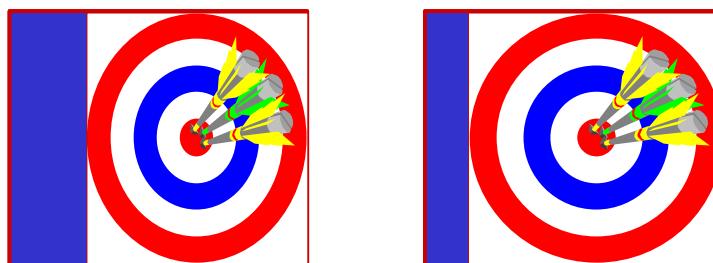


Fig. 2.17 Efectos de las diferencias entre generadores de barrido de pantalla



La imagen correspondiente al generador que comuta más rápidamente se genera más hacia la derecha de la pantalla.

El segundo generador tiene un nivel de tensión más bajo cuando se recibe la luminancia por lo que la exploración se inicia más hacia la izquierda.

Fig. 2.18 Imágenes obtenidas con los dos generadores de barrido considerados

En la figura 2.18 se representan las imágenes que se obtendrían en el supuesto de que la excusión de señal de los dos generadores fuera idéntica y que se aplicaran a un mismo tubo de rayos catódicos. Nótese que cuando se recibe la señal de luminancia, el generador más rápido tiene un nivel de tensión superior al del otro generador, por lo que la presentación de la imagen se empezará en una posición horizontal más avanzada, tal y como se ha indicado en el gráfico. Sin embargo, es evidente que este problema no es, en absoluto, grave, por lo que puede solventarse controlando el nivel de continua

(*offset*) y la amplificación de las señales de barrido. En efecto, en la figura 2.19 se han representado las mismas señales de barrido que las consideradas en la figura 2.18, pero con unos niveles de continua y excursiones adaptadas para que ambas señales permitan un barrido del haz desde un extremo a otro de la pantalla.

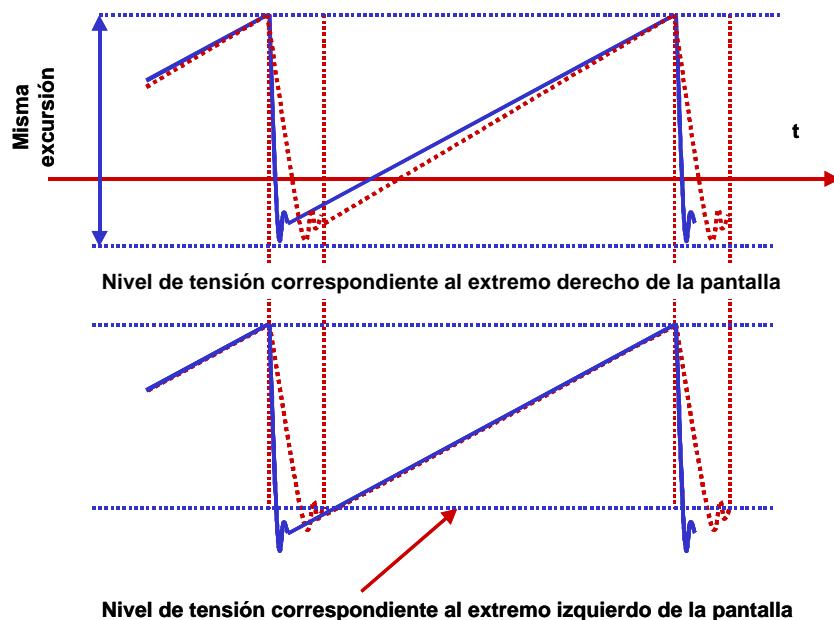


Fig. 2.19 Ajuste de las señales de barrido para explorar todo el ancho de la pantalla

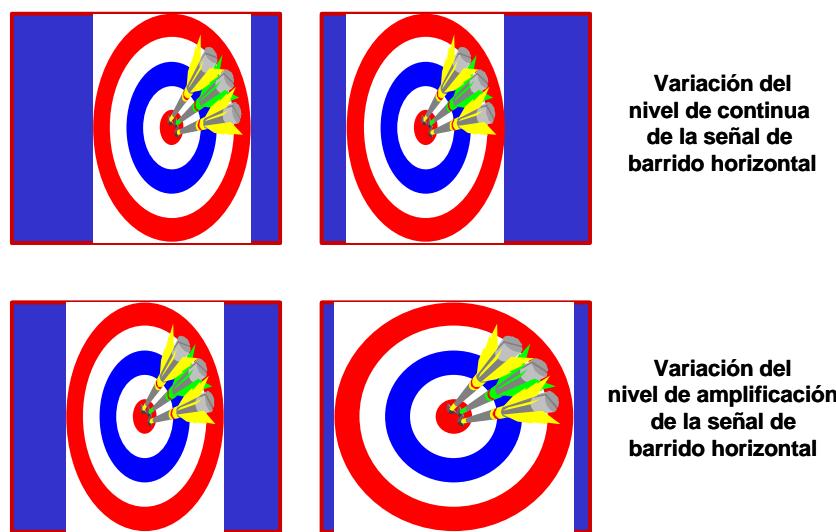


Fig. 2.20 Efectos del control de continua y amplificación sobre la imagen

El cambio de nivel de continua de la señal de barrido produce un desplazamiento horizontal de la zona de la pantalla en la que se presenta la imagen, mientras que la amplificación o atenuación de los niveles representa una ampliación o reducción del tamaño horizontal de la imagen. Ambos efectos se ilustran en la figura 2.20. Los efectos que sobre la imagen introduce la amplificación de la señal de barrido o el *offset* son, evidentemente, generalizables al subsistema de barrido vertical, por lo que no repetiremos estos argumentos cuando consideremos los sincronismos verticales.

Los receptores antiguos solían proporcionar al usuario controles para el ajuste de los niveles de continua y amplificación de las señales de barrido. Estos ajustes se han suprimido en la mayor parte de receptores actuales que, ya vienen preajustados de fábrica o, incluso, son capaces de realizar un ajuste automático de estas señales cada vez que se conectan. No obstante, en los monitores para ordenador es habitual que el usuario tenga acceso a estos controles.

La sincronización de los generadores de barrido puede realizarse en principio de dos maneras:

- a) Disparo directo.
- b) Disparo controlado (*Line Flywheel*)

El primer método consiste en disparar el generador de barrido cada vez que se detecta la llegada de un impulso de sincronismo de línea. Fue un sistema que se utilizó sólo en los primeros receptores y que está en desuso, por cuanto es muy sensible al ruido (disparos no deseados) y es posible que algunos pulsos no sean correctamente detectados, con lo que se perdería la información correspondiente a esa línea.

El sistema utilizado en prácticamente todos los receptores se basa en un sistema PLL (*Phase Locked Loop*) simple o dual, de modo que la información de sincronismo se utiliza únicamente para ajustar la frecuencia interna de la señal de barrido que se está generando. En el caso de que aparezcan ruidos espurios que activan el sistema de detección de sincronismos, éstos sólo modifican ligeramente la frecuencia del sistema sin que se produzcan efectos visibles. Análogamente, si se pierde alguno de los impulsos, el sistema interno sigue generando la señal de barrido sin que el error sea observado por el espectador.

En la figura 2.21 se representa el diagrama de bloques de un sistema de generación de señal de barrido horizontal que utiliza esta técnica. La señal de sincronismo se pasa a través de un bloque encargado de detectar los pulsos de sincronismo de línea y que normalmente es un filtro paso alto o derivador que recorta una de las polaridades de la derivada. Esta señal se introduce, junto con una señal parecida, pero generada internamente, a un comparador de fase cuya tensión de salida controla la frecuencia de un oscilador (oscilador controlado por tensión – *V.C.O.*). En la salida del VCO existe un conformador de onda que obtiene la señal en diente de sierra a partir de la señal proporcionada por el oscilador. Esta señal, que se amplifica y a la que se añade un nivel de continua adecuado, se aplica a las bobinas de deflexión horizontal. La referencia de sincronismo interna se obtiene derivando la señal obtenida en la salida de este conformador.

La polaridad de señal en la salida del comparador de fase depende de cuál de los dos impulsos se recibe antes, y su nivel es proporcional a la diferencia de tiempos. En el supuesto de que el impulso procedente del sincronismo externo se recibiera antes que el interno, indicaría que la frecuencia de las

señales de barrido internas es más baja que el ritmo con la que recibimos las líneas, por lo que el comparador de fase generará una tensión positiva que provocará un aumento, proporcional a la diferencia de tiempos, de la frecuencia de la señal interna. Esta filosofía de control de la frecuencia del generador de barrido converge, siempre que los parámetros del bucle de realimentación estén correctamente diseñados, en una situación en la que la señal de barrido coincide, en frecuencia y fase, con la cadencia con la que se transmiten las líneas de la señal de televisión.

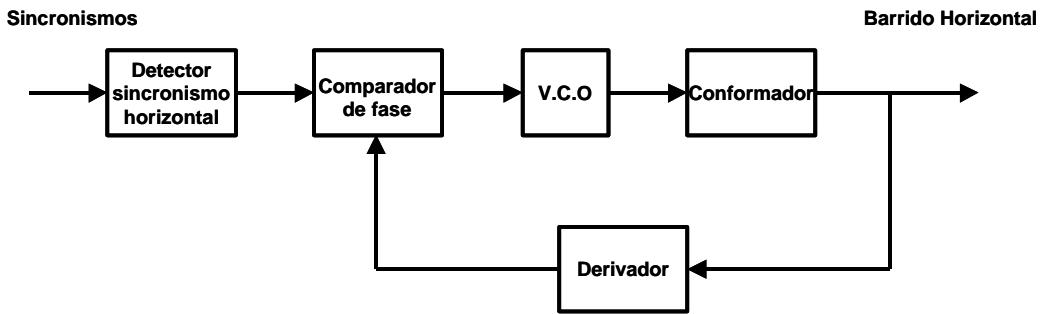


Fig. 2.21 Sistema de control de frecuencia del generador de barrido horizontal

2.6 Sincronismo de campo

La señal de sincronismo de campo es considerablemente más compleja que la de sincronismo de línea, pues la señal de luminancia deja de transmitirse durante un total de 25 líneas en cada campo, tiempo que se utiliza no sólo para insertar el pulso de inicio de campo, sino todo un conjunto de componentes que se encargarán de asegurar que los circuitos que deben detectar esta señal están en las mismas condiciones iniciales y finales en los dos campos. Además, algunas de estas 25 líneas, que se denotan con el nombre de líneas de borrado de campo, se aprovechan para transmitir señales de prueba y de teletexto. Las formas de onda correspondientes al periodo de borrado de campo se representan en las figura 2.22 y figura 2.23, donde se indica la secuencia exacta de impulsos de sincronismo y las señales que se producen en cada campo. En estas gráficas los impulsos de sincronismo se han indicado con polaridad positiva, mientras que la señal de luminancia se representa con valores de tensión negativos. Esta representación es justamente la inversa que la utilizada en la figura 2.13. Sin embargo, ya allí habíamos comentado que, a menudo dependiendo del contexto, las polaridades de la luminancia y los sincronismos pueden aparecer invertidas. En este caso, estamos centrando la atención sobre los impulsos de sincronismo, por lo que nos resultará algo más cómodo suponer que estas señales son positivas.

La estructura de la señal de borrado de campo no resulta aparentemente trivial. Examinaremos con detalle y justificaremos cada una de sus componentes dentro de poco, pero por el momento describamos las distintas partes en las que se descompone.

Consideremos inicialmente el campo impar (ver figura 2.22). Aunque el inicio del campo se produce en la línea 1, la señal de luminancia deja de transmitirse un poco antes, en la mitad de la línea 623, con el objeto de disponer de tiempo suficiente para que los circuitos de detección de sincronismo de campo estén preparados para su correcta recepción. En la mitad de la línea 623 se inicia una secuencia

de impulsos con los mismos niveles y duración que los impulsos de línea, pero que tienen una cadencia doble. Estos impulsos se conocen con el nombre de impulsos de pre-igualación y su función es que las señales que reciben los circuitos de detección de sincronismo de campo sean iguales, tanto para el caso de que se trate del campo par como del impar, sin que ello suponga una pérdida de los impulsos de sincronismo horizontal. Hay un total de cinco impulsos de pre-igualación que se extienden durante dos líneas y media.

El impulso de sincronismo de campo propiamente dicho se inicia en la línea 1 y se extiende nuevamente durante 2 líneas y media (hasta la mitad de la línea 3). Este impulso está caracterizado por tener, durante la mayor parte del tiempo, una tensión positiva. Sin embargo, durante este periodo se siguen enviando los pulsos de sincronismo de línea, aunque ahora se señalizan invertidos y ligeramente adelantados, utilizando, como en el caso de la pre-igualación, una cadencia doble de la de la frecuencia de línea.

Una vez finalizado el impulso de sincronismo de campo, se envían cinco impulsos de post-igualación, que se extienden desde la mitad de la línea 3 hasta el final de la 5, cuya forma de onda es idéntica a la de los impulsos de pre-igualación. La función de estos impulsos no es tan importante como la de los de pre-igualación, y se encargan de que, una vez detectado el sincronismo de campo, las señales aplicadas a los circuitos de detección sean iguales en los dos campos, por lo que ambos se restauran a unas condiciones iniciales parecidas. Además, estos impulsos garantizan que la forma de onda sea simétrica respecto al centro del impulso de sincronismo de campo.

Después de los impulsos de post-igualación se incluyen un total de 17,5 líneas en negro. Su función es la de garantizar que los circuitos de barrido han tenido tiempo para conmutar la señal de diente de sierra y que se genera una señal estable que permite aplicar la señal de luminancia al tubo. Estas 17,5 líneas se extienden desde el inicio de la línea 6 hasta la mitad de la línea 23 y contienen los impulsos de sincronismo horizontal a la cadencia propia de la frecuencia de línea. Algunas de estas líneas se aprovechan para transmitir otro tipo de señales que no corresponden a información de vídeo. Entre estas señales se incluye la información de teletexto (entre las líneas 11 a 15), códigos identificativos de la emisora y señales para el control de la calidad de recepción de la señal en distintos puntos de la red de distribución (líneas 16 a 19).

Estableciendo un paralelismo entre la señal de sincronismo de línea y la de sincronismo de campo, podríamos decir que los impulsos de pre-igualación cumplen una función similar a la del pórtico anterior, mientras que el conjunto formado por los impulsos de post-igualación y las 17,5 líneas en blanco tienen una función análoga a la del pórtico posterior.

La forma de onda de la señal de borrado del campo par tiene unas características parecidas a las descritas para el impar. En este caso el impulso de sincronismo de campo se inicia en la mitad de la línea 313 y se extiende hasta el final de la 315. Está precedido por cinco impulsos de pre-igualación por lo que la señal de luminancia correspondiente al campo impar deja de transmitirse al final de la línea 310. Entre la línea 316 y la mitad de la 318 están situados los impulsos de post-igualación a los que siguen las 17,5 líneas en negro. La señal de luminancia vuelve a transmitirse al principio de la línea 336, utilizando las líneas 324 a 328 para la información de teletexto y las 329 a 332 se usan para transmitir información de control.

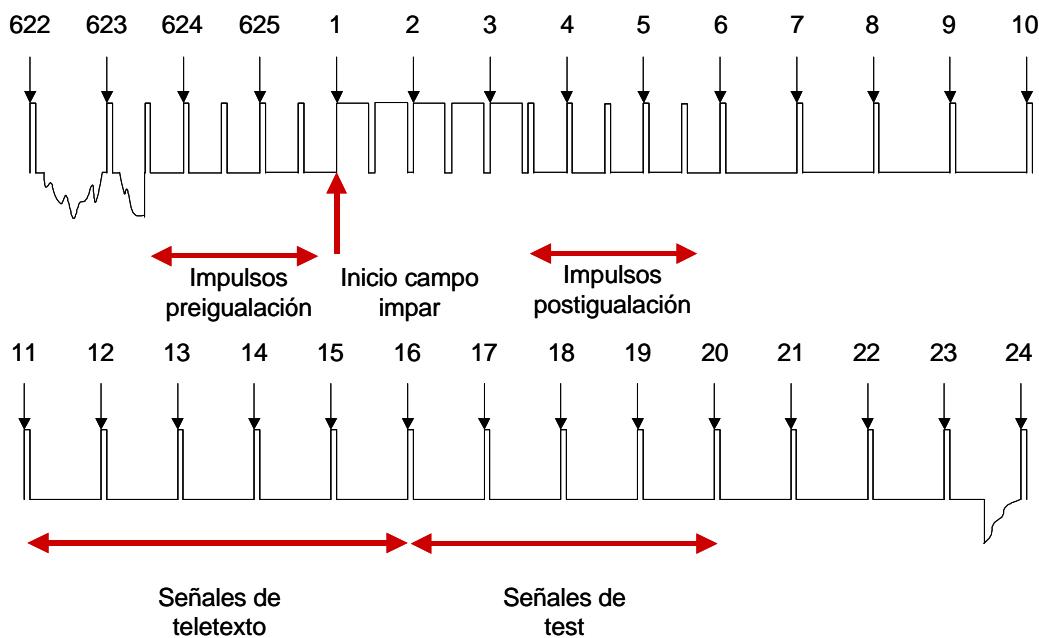


Fig. 2.22 Detalle de las señales de sincronismo de campo impar

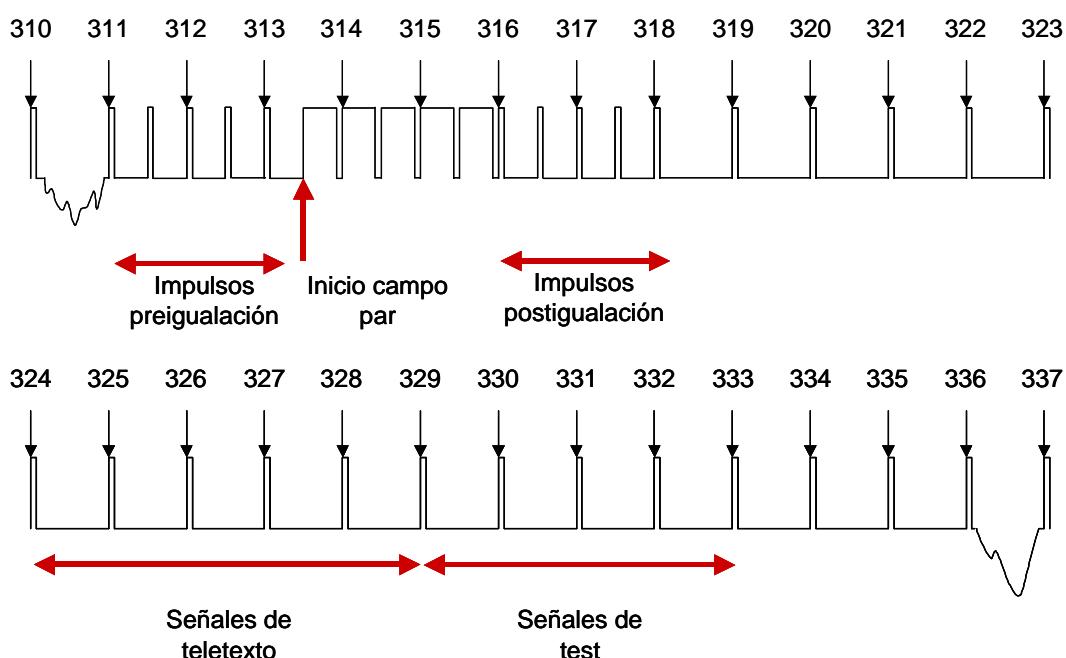


Fig. 2.23 Detalle de la señal de sincronismo de campo par

Las formas de onda de las señales de borrado correspondientes a los campos par e impar son por tanto muy parecidas, resultando incluso difíciles de identificar en la pantalla de un osciloscopio o monitor de forma de onda, en los que, por lo general, no se dispone de las etiquetas que identifican los números de línea de la señal de televisión. El procedimiento más fácil para determinar si se trata del campo par o del impar es ver si el campo anterior acaba en una línea completa o en media línea de luminancia (campo par e impar, respectivamente).

Una vez descritas las distintas componentes de la señal de barrido de campo, procederemos a justificar el por qué estas señales adoptan estas formas de onda que en un principio pueden parecer un tanto caprichosas. Para ello seguiremos un razonamiento natural y progresivo, intentando introducir el pulso de sincronismo de campo de una forma muy sencilla y viendo las medidas que deben tomarse para garantizar que los circuitos de detección de sincronismo de línea y de campo sean lo más simples posibles. En general, podríamos decir que la complejidad de la señal de borrado de campo está condicionada por la restricción de que los circuitos de detección deben ser simples y económicos.

En la figura 2.24 se han representado los impulsos de sincronismo de línea que se obtendrían en la salida del separador de sincronismos y los puntos en los que corresponde insertar la señal de sincronismo de campo. Como estamos considerando un sistema de 625 líneas, cada uno de los campos debe tener una duración de 312,5 líneas. Esto significa que si suponemos que el campo impar se inicia en la línea 1, el par debe iniciarse en la mitad de la línea 313, es decir, cuando han transcurrido 312 líneas completas más media línea adicional.

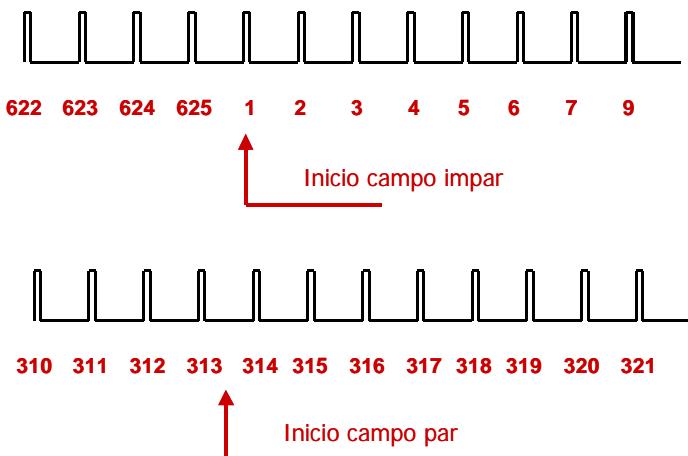


Fig. 2.24 Puntos de inserción de la señal de sincronismo de campo par e impar

Por otra parte, hemos visto que para simplificar el circuito separador de sincronismos, los impulsos de inicio de línea tenían una polaridad distinta a la de la señal de luminancia, por lo que parece lógico que la señal de sincronismo de campo también deba seguir manteniendo esta misma polaridad. Con ello, todos los sincronismos podrán aislar la señal de luminancia mediante un único circuito. Sin embargo, esto plantea el problema de cómo separar las componentes correspondientes al sincronismo de línea y las asociadas al sincronismo de campo. La forma que en principio parece más evidente es utilizar un filtrado de la señal de sincronismo, tal y como se simboliza en la figura 2.25. Los impulsos

de sincronismo de línea tienen una duración corta, por lo que podrán detectarse sin problemas si se utiliza un filtro paso alto. El impulso de sincronismo de campo podría ser de mayor duración, de hecho la cadencia con la que transmiten los campos es muy lenta si se compara con la de las líneas, con lo que podrían ser detectados mediante un filtrado paso bajo o integrador. Utilizando esta sencilla filosofía, podríamos pensar en introducir un impulso de larga duración (en concreto se define una duración de dos líneas y media) como señalización para el impulso de sincronismo de campo.

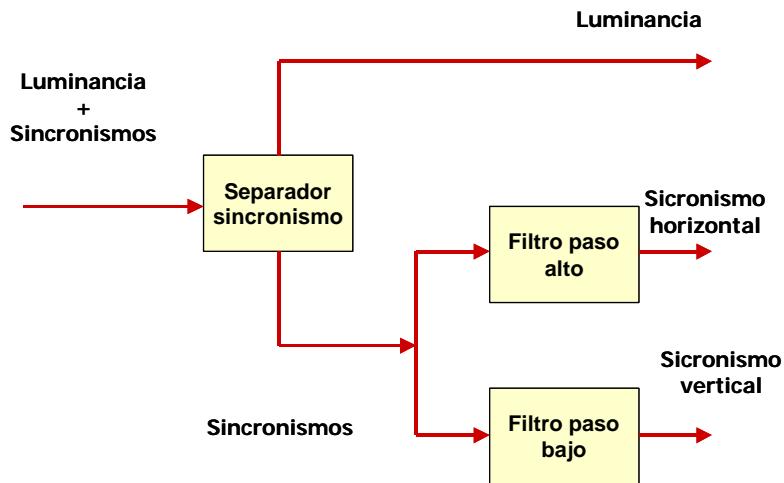


Fig. 2.25 Diagrama de bloques del subsistema de separación de sincronismos

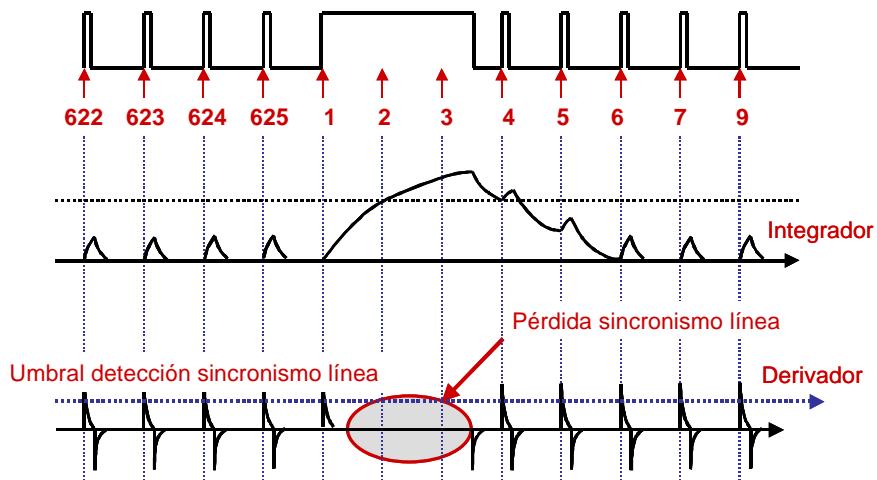


Fig. 2.26 Señales en la salida de los detectores de sincronismo para un impulso de campo directo

Esta sencilla propuesta de señalización se ilustra, para el campo impar, en la figura 2.26, donde se observa la respuesta que se obtendría al aplicar un filtro paso bajo y uno paso alto a la señal de sincronismos. En principio, la respuesta obtenida parece satisfactoria y sin duda nos indica que

estamos en el buen camino. La integración de la señal produce un resultado que permitirá detectar la presencia del impulso de sincronismo de campo utilizando un simple comparador de nivel. En efecto, el nivel de la señal en la salida del filtro paso bajo sólo excederá el umbral prefijado cuando estemos en presencia de un impulso de baja frecuencia, ya que los impulsos de sincronismo de línea no tienen la duración suficiente como para obtener un nivel elevado de señal en la salida.

Por otra parte, la señal en la salida del filtro paso alto mantiene un nivel elevado cada vez que se recibe un impulso de línea, por lo que, mediante otro comparador de nivel, puede mantenerse la sincronización con los circuitos de barrido horizontal. Sin embargo, la introducción directa del pulso de sincronismo de campo impide la transmisión del sincronismo de línea en toda la extensión de esta señal. En principio, este problema no es grave, por cuanto hemos comentado que la mayoría de receptores sólo utilizan los impulsos de sincronismo de línea para corregir las posibles desviaciones del circuito de generación de barrido horizontal. Por ello, el generador de barrido interno podría mantenerse sin ningún problema durante estas 2,5 líneas sin que se produjeran desajustes importantes y, en todo caso, utilizar los impulsos de sincronismo de línea posteriores para cancelar cualquier desalineamiento que se haya producido.

No obstante, la señal de televisión debe contemplar una amplia gama de filosofías de recepción de la señal y es posible, sobre todo en los primeros receptores de televisión, que existan fabricantes que prefieran utilizar circuitos de generación de la señal de barrido horizontal que se activen directamente mediante la señal de televisión. En estos sistemas, es necesario que la salida del filtro paso alto siga manteniendo los impulsos de sincronismo horizontal durante la recepción del sincronismo de campo, pues en caso contrario podrían producirse desviaciones importantes y sobretensiones en los circuitos de generación de barrido.

Detalle del sincronismo de campo

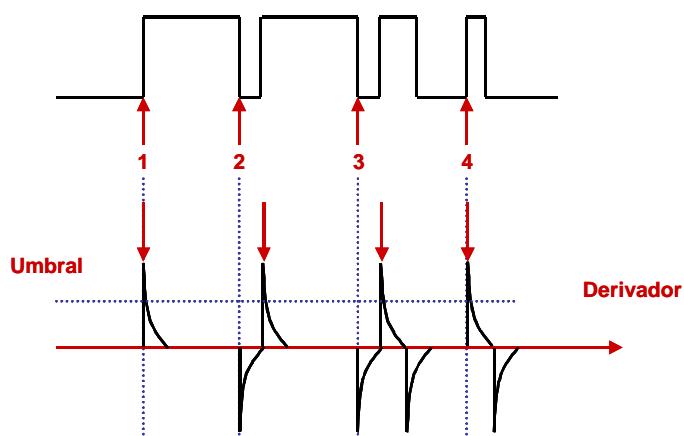


Fig. 2.27 Introducción de los pulsos de sincronismo de línea dentro del impulso de campo

Una primera propuesta para mantener la detección de los impulsos de sincronismo de línea durante la transmisión del sincronismo de campo se ilustra en la figura 2.27. La señalización del sincronismo de línea se realiza desactivando el pulso de sincronismo de campo durante el tiempo correspondiente a un

pulso de línea. La salida del filtro paso alto detectará esta nueva señalización, produciendo un tren de pulsos como el ilustrado en esta misma figura. La alternativa parece buena en primera instancia aunque se observa que se introduce un retardo en la detección del sincronismo de línea debido a la propia inversión de la señal. No obstante, esta propuesta sugiere claramente la solución definitiva: si se introduce el retorno a cero un poco antes del inicio de la línea, adelantando ligeramente el pulso de modo que su final coincida con el inicio de una línea, podremos mantener perfectamente el sincronismo horizontal. Esta última idea se ilustra en la figura 2.28, donde ahora puede comprobarse la correcta detección del inicio de las líneas.

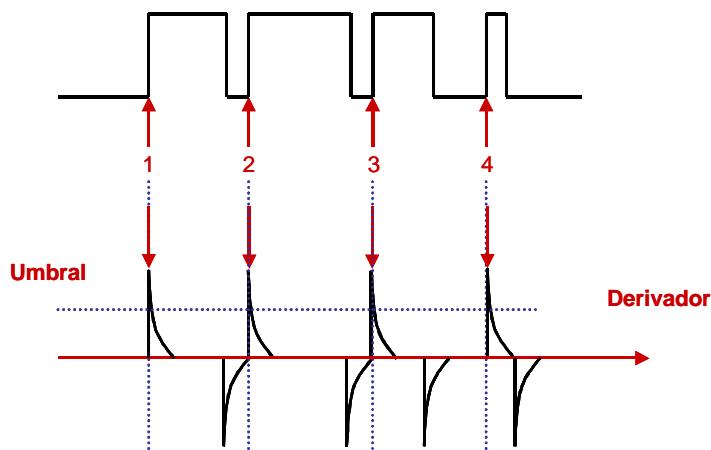


Fig. 2.28 Inserción del sincronismo de línea en el de campo

Sin embargo, si examinamos más detenidamente las formas de onda actuales, veremos que aún existen problemas serios con la detección del sincronismo de campo. En efecto, en la figura 2.29 se representan simultáneamente las formas de onda de los impulsos de sincronismo de campo par e impar. Los inicios de los dos pulsos de sincronismo se han representado sobre la misma vertical, de modo que se observa un desfase entre la posición relativa de los sincronismos de línea en cada gráfica. Esto es así debido a que el campo par empieza en la mitad de una línea, mientras que el impar lo hace al principio de línea, por lo que aparece una cierta simetría entre las dos señales.

El resultado de esta simetría es que, cuando las señales se aplican al circuito integrador encargado de detectar el inicio de campo, la evolución de la forma de onda para el campo par y el impar es distinta, alcanzando el nivel fijado por el umbral antes en un campo que en el otro. Nótese que el campo impar empieza con una línea completa, por lo que el impulso de sincronismo de línea tarda más en aparecer que en el campo par. Por ello la tensión de umbral se alcanza antes para este primer campo.

Este problema es realmente grave, por cuanto si uno de los campos se detecta antes, supondrá que éste tendrá mayor duración que el otro, generando una señal de barrido distinta para cada campo, con lo que el entrelazado dejará de ser correcto. La solución al problema exige que las dos señales sean idénticas, con lo que su integración también producirá idénticos resultados. Para ello, es necesario incorporar los impulsos de sincronismo de línea en las mismas posiciones relativas dentro de la señal,

es decir, añadiremos un pulso de sincronismo en el campo par allí donde estén situados los impulsos correspondientes al campo impar y viceversa. Esta solución supone que los impulsos de línea tendrán una frecuencia de ocurrencia doble en toda la región donde queramos que las dos señales sean idénticas. La forma de esta solución se representa en la figura 2.30.

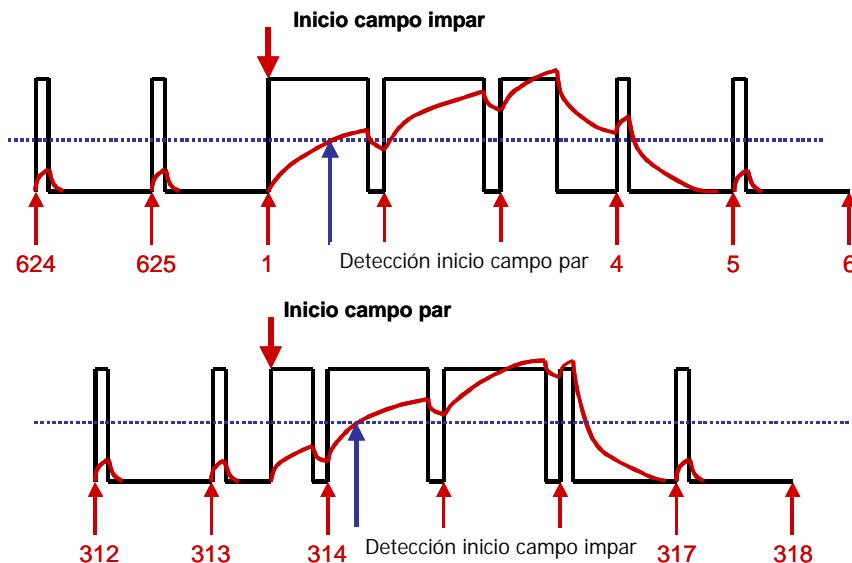


Fig. 2.29 Problemas en la detección del inicio de campo

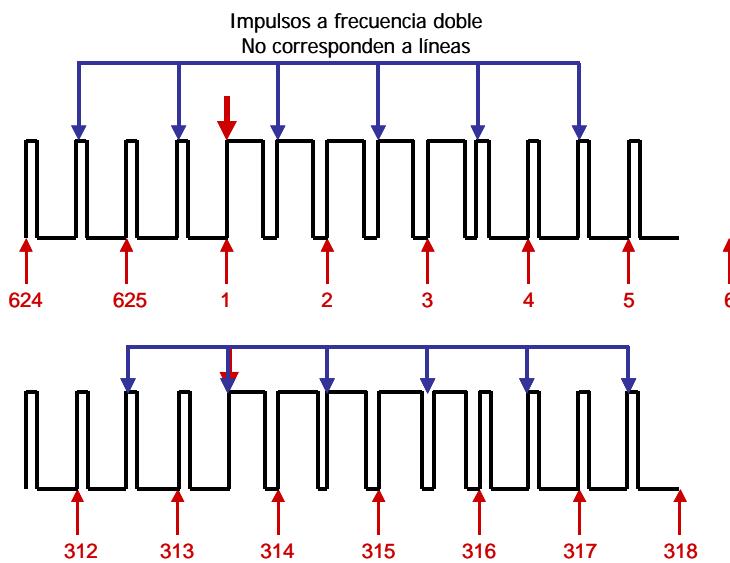


Fig. 2.30 Forma de onda definitiva del segmento próximo al sincronismo de campo

El lector podría objetar que utilizando esta forma de onda, los circuitos de detección del sincronismo de línea se dispararán a una frecuencia doble durante este tramo de señal. En efecto, este es el precio que tenemos que pagar para garantizar que las señales en la salida del filtro paso bajo se mantienen idénticas, produciéndose la detección del inicio de campo en la misma posición relativa, independientemente de que se trate del campo par o el impar. No obstante, este problema no es en absoluto grave, por cuanto, conociéndolo, resulta muy fácil inhibir el disparo de los circuitos de barrido horizontal hasta que no tengan un nivel de tensión próximo al que debe producirse la conmutación. Así pues, en general, los circuitos de detección del impulso de sincronismo de línea se mantendrán inhibidos mientras la tensión de barrido no supere un determinado nivel que indique que ya estamos explorando el extremo derecho de la pantalla, por lo que debemos estar preparados para la recepción de un nuevo pulso de sincronismo horizontal.

La inserción de estos impulsos a doble frecuencia de la de línea se realiza no sólo durante el pulso de sincronismo de campo, sino que se extiende a un total de 7,5 líneas, entre las que se incluyen las 2,5 anteriores al pulso de sincronismo, las 2,5 propias del pulso y las 2,5 posteriores. La función de los impulsos de pre-igualación (2,5 líneas precedentes) es la de garantizar que la señal aplicada al integrador cuando se inicia el pulso de sincronismo es la misma, independientemente del campo que estemos considerando. En el supuesto de que no existieran estos impulsos de pre-igualación, el estado inicial del integrador podría ser diferente en cada uno de los campos, ya que en el campo par, el último impulso de sincronismo de línea se hubiera recibido sólo media línea antes, mientras que en el impar se habría recibido una línea completa antes. Con la presencia de los impulsos de pre-igualación puede asegurarse que las posibles asimetrías de señal entre los dos campos provienen como mínimo de 2,5 líneas antes de la llegada del impulso de campo, lo que suele considerarse como un tiempo suficiente para que la memoria del integrador ya no tenga relevancia. La función de los impulsos de post-igualación es, como ya hemos mencionado, menos importante, y se encarga de que la descarga del integrador sea parecida, independientemente de si el campo es par o impar.

2.6.1 Sincronización de los generadores de barrido vertical

La sincronización de los circuitos de barrido vertical puede realizarse mediante un método directo, es decir, reinicializando el generador en diente de sierra cada vez que se detecta la presencia de un impulso de sincronismo de campo, o mediante métodos indirectos basados en las referencias de los impulsos de sincronismo horizontal.

La razón de que en este caso pueda usarse un método directo, mientras que en el sincronismo de línea, casi invariablemente, se use un sistema basado en un PLL, se debe por una parte a que la frecuencia de trabajo es mucho más lenta (tardaría mucho tiempo en controlarse el VCO por lo que las derivas podrían ser considerables), y por otra a que, al tratarse de un pulso detectado mediante un integrador, suele ser algo menos sensible al nivel de ruido.

En el caso de utilizar un sincronismo directo, el generador de barrido vertical suele estar diseñado para que commute automáticamente a una frecuencia interna si se produce un error en la detección del pulso de sincronismo de campo. Con esta solución pueden tolerarse algunos errores, aunque si se produce una ráfaga de fallos en la detección del sincronismo, se observará que la pantalla se desplaza hacia arriba o hacia abajo a una velocidad que dependerá de la diferencia entre la frecuencia del generador

interno y la de la señal de televisión. Este efecto puede comprobarse en algunos receptores antiguos que permiten que el usuario pueda controlar el umbral de detección de la señal de sincronismo de campo. Si se ajusta este umbral a un nivel excesivamente elevado, podrá forzarse que los impulsos de sincronismo no se detecten correctamente y que el barrido vertical empiece a realizarse a la frecuencia interna, observándose cómo en la pantalla la imagen se desplaza en sentido vertical.

No obstante, los receptores más modernos suelen utilizar un dispositivo de sincronismo indirecto que conjuga la señal de sincronismo horizontal obtenida mediante la técnica del *fly-wheel* (una señal muy estable y exenta de ruido) con la señal de sincronismo vertical directa. Las características de este sistema de sincronización indirecta, cuyo diagrama de bloques se representa en la figura 2.31, son excelentes.

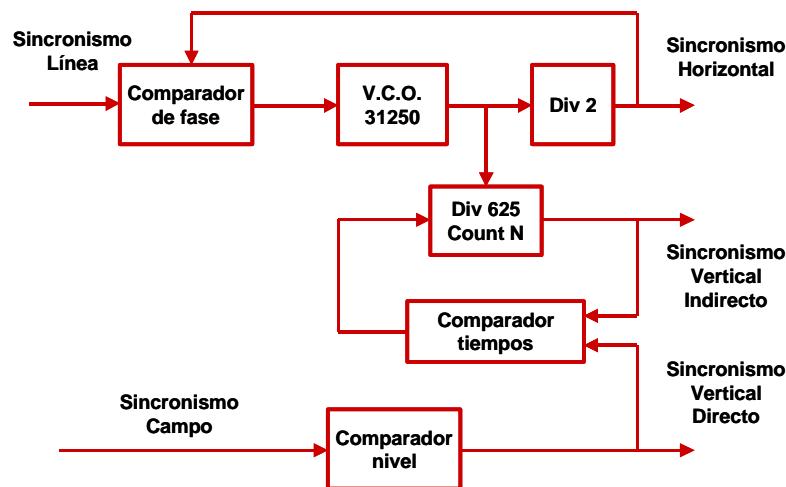


Fig. 2.31 Diagrama para la obtención indirecta del sincronismo de campo

La idea básica consiste en utilizar el mismo sistema que suministra la referencia de fase de la señal de sincronismo horizontal para obtener la señal de sincronismo vertical. El bucle del PLL debe modificarse ligeramente, provocando que el oscilador controlado por tensión trabaje a una frecuencia doble de la de línea. Esto se consigue insertando un divisor de frecuencia por 2 en la salida del VCO, tal y como se indica en el diagrama de bloques. Esta modificación del circuito no representa una pérdida de la fase de la señal de sincronismo horizontal, por cuanto el comparador de fase sigue trabajando con la señal de sincronismo externa y la salida del divisor de frecuencia.

Al doblar la frecuencia del oscilador interno podemos obtener una señal cuya frecuencia coincide exactamente con la frecuencia de campo utilizada en el transmisor. Para ello basta con dividir la frecuencia del VCO por 625 (para lo que normalmente se utiliza un contador de pulsos), ya que al haber doblado la frecuencia del oscilador cada cuenta de 625 corresponderá con el tiempo de campo. La frecuencia obtenida coincide exactamente con la del transmisor, ya que el procedimiento utilizado para su obtención es esencialmente el mismo. Recuérdese que al seleccionar el número de líneas en el sistema de televisión ya habíamos tenido en cuenta que admitiera una descomposición en números primos sencilla con el objeto de poder dividir fácilmente su frecuencia.

Sin embargo, aunque disponemos de la frecuencia de campo exacta, la fase de la señal en la salida del divisor por 625 dependerá de las condiciones en que se haya inicializado el circuito.

Para adquirir la fase exacta, se utilizará la señal de sincronismo de campo obtenida por un método directo. El procedimiento es sumamente sencillo. Un comparador de fase determina si las señales de sincronismo de campo indirecto y directo están en fase. Si no es así, se reduce el número de pulsos que debe contar el divisor por 625, de manera que la frecuencia de la señal generada internamente aumente ligeramente y consiga ponerse en fase con la señal de sincronismo externa. Este método de sincronización es muy parecido al utilizado en el PLL. Durante el tiempo de adquisición de la referencia de fase, el receptor utiliza la señal de sincronismo directo para controlar los generadores barrido vertical, pero una vez se ha adquirido correctamente la referencia de fase, lo cual se detecta por la coincidencia de los pulsos indirectos con los directos, se utiliza preferentemente la señal indirecta para el control de los circuitos de barrido vertical.

La señal indirecta es mucho más estable que la directa y no requiere de ningún control por parte del usuario. Esta señal se mantiene como base para la sincronización del barrido vertical mientras no se detecte una discrepancia con la señal directa durante un número elevado de campos, situación que, en condiciones de recepción normales, sólo se produce cuando el usuario cambia de emisora. En este caso, se vuelve a utilizar la señal de sincronismo directa hasta que se determine la fase exacta del campo.

Las ventajas del método indirecto respecto al directo son evidentes, puesto que se obtiene un perfecto sincronismo con la emisora, aun en el caso de que el nivel de ruido provoque la pérdida de un elevado porcentaje de impulsos de sincronismo vertical. Además, no requiere ningún oscilador interno a la frecuencia de campo ni el ajuste del nivel de disparo. El entrelazado entre los dos campos es, por otra parte, perfecto, ya que se eliminan completamente las posibles imprecisiones temporales en la detección del pulso de sincronismo. Además, el sistema puede modificarse para que pueda conmutar directamente entre los estándares de 50 Hz y 60 Hz.

2.7 Espectro de la señal de televisión en blanco y negro

En el capítulo anterior se realizaron diversas consideraciones relativas al ancho de banda de la señal de televisión, llegando a la conclusión de que un ancho de banda de unos 5 MHz proporciona una imagen con calidad más que satisfactoria para la mayoría de espectadores. Sin embargo, no hemos realizado ningún análisis para determinar las características espectrales de la señal de televisión. Evidentemente, la forma y características de la transformada de Fourier de la señal dependerán de la escena que se transmita, pero podemos realizar ciertas hipótesis genéricas que nos permitan tener una idea aproximada de los detalles espectrales de esta señal.

La hipótesis más simple, por lo que respecta al cálculo de la transformada de Fourier, es la de suponer que se transmite una escena estática que tiene simetría vertical. Bajo esta suposición, el cálculo de la transformada es relativamente sencillo, ya que se trata de determinar el espectro de una señal periódica. En efecto, al tener simetría vertical todas las líneas de la imagen son iguales y al ser estática los cuadros se repiten indefinidamente. Aparte de que esta hipótesis es la única que podemos tratar analíticamente con cierta simplicidad, produce unos resultados bastante aproximados al espectro real

de la señal. Esto es debido a que en la mayoría de escenas reales, si bien existe movimiento y no todas las líneas son iguales, es verdad que entre una imagen y la siguiente o una línea y la siguiente suelen producirse pocos cambios, de modo que la señal de televisión tendrá una estructura casi-periódica. Evidentemente, la aproximación es tanto más precisa cuanto más parecidas sean las peculiaridades de la escena a las hipótesis que realizamos.

Para determinar el espectro de una señal de televisión con las características mencionadas descompondremos su estructura en las componentes siguientes:

$$x(t) = p(t) \cdot w(t) + c(t) \quad (2.8)$$

donde $x(t)$ representa la señal de televisión, con los sincronismos de línea y de campo incorporados y $p(t)$, $w(t)$ y $c(t)$ son señales artificiales que nos permiten descomponer la señal $x(t)$ como una combinación analítica de señales periódicas, cuya transformada de Fourier resulta muy sencilla de estimar.

La señal $p(t)$ representa, tal y como se indica en la figura 2.32, las líneas de la señal de televisión con los sincronismos de línea incorporados, pero sin la presencia de la señal de sincronismo de campo. Por ello, la señal $p(t)$ es perfectamente periódica y sus características espetrales podrán ser calculadas con facilidad. La señal $w(t)$ es una ventana que toma el valor 0 allí donde la señal $p(t)$ debería incorporar el sincronismo de campo y valor 1 en el resto de valores. Es evidente por tanto que esta señal también es periódica con un periodo de $1/50 \text{ Hz} = 20 \text{ ms}$. El producto entre la señal $p(t)$ y $w(t)$ da como resultado una señal que es idéntica a la señal $x(t)$ en todo instante de tiempo exceptuando las 25 de borrado de campo en las que toma el valor cero. Finalmente, la señal $c(t)$ es una señal que representa las componentes de borrado de campo que no están incluidas en el producto $p(t)$ y $w(t)$. Como las formas de onda de las zonas de borrado de campo par e impar, aun en el supuesto de que no se transmita teletexto ni información de control, no son exactamente iguales, esta componente también tendrá un periodo de 40 ms. Nótese, sin embargo que la diferencia entre la forma de onda del campo impar y del campo par es muy sutil, ya que la única diferencia es que la distancia entre el último impulso de postigualación y la siguiente señal de sincronismo de línea es de media línea en el campo impar y de una línea completa en el par. Por ello, muchas veces se supone que esta señal tiene un periodo fundamental de 20 ms. Mediante esta descomposición de la señal podemos expresar la transformada de Fourier como:

$$X(f) = P(f) * W(f) + C(f) \quad (2.9)$$

donde $P(f)$, $W(f)$ y $C(f)$ representan las transformadas de Fourier de las señales periódicas $p(t)$, $w(t)$ y $c(t)$ respectivamente y el símbolo $*$ representa el producto de convolución entre las dos transformadas. La transformada de Fourier de una señal periódica puede calcularse directamente a partir del desarrollo en una serie de Fourier en exponentiales complejas de la señal. Para la señal $p(t)$ el desarrollo en serie de Fourier viene dado por la siguiente expresión:

$$p(t) = \sum_{k=-\infty}^{\infty} a_k \cdot \exp\left(\frac{2\pi}{T_{Línea}} \cdot k \cdot t\right) \quad (2.10)$$

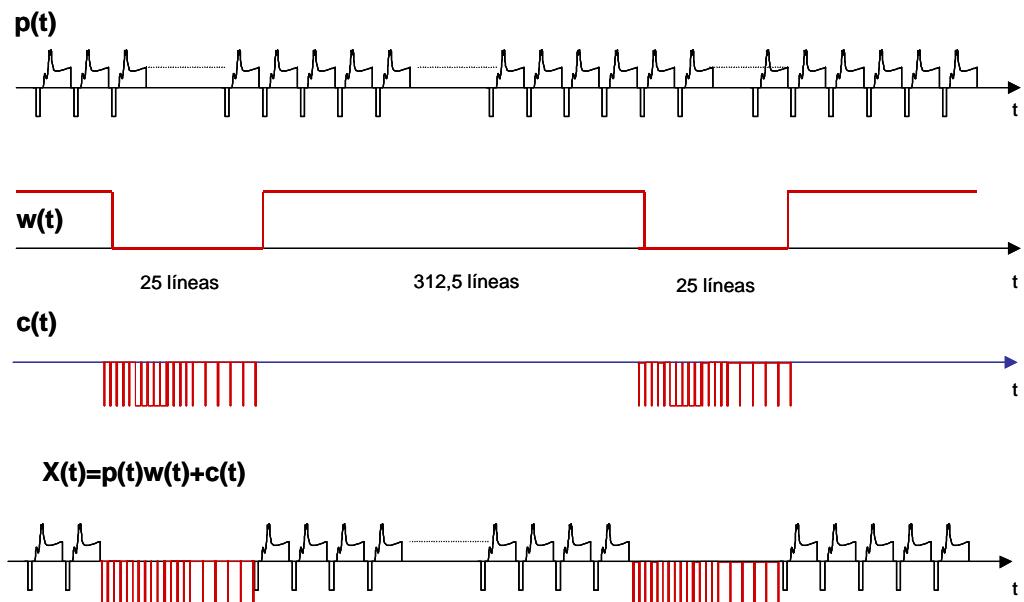


Fig. 2.32 Descomposición de la señal $x(t)$ en tres componentes periódicas

donde los coeficientes a_k dependen de la forma de onda de la señal en un periodo y se determinan a partir de la relación:

$$a_k = \frac{1}{T_{Línea}} \cdot \int_{T_{Línea}} p(t) \cdot \exp\left(-\frac{2\pi}{T_{Línea}} \cdot k \cdot t\right) dt \quad (2.11)$$

donde el símbolo bajo la integral indica que los límites de integración pueden tomarse en cualquier parte de la señal mientras se considere un periodo completo de la misma.

Si aplicamos la transformada de Fourier a la ecuación 2.10 obtenemos:

$$P(f) = F(p(t)) = \sum_{k=-\infty}^{\infty} a_k \cdot F\left(\exp\left(\frac{2\pi}{T_{Línea}} \cdot k \cdot t\right)\right) = \frac{1}{T_{Línea}} \sum_{k=-\infty}^{\infty} a_k \cdot \delta(f - k \cdot f_{Línea}) \quad (2.11)$$

donde $\delta(t)$ representa la función delta de Dirac. Este resultado nos indica que la transformada de Fourier de la señal está formada por un tren de deltas situadas en múltiplos de la frecuencia de línea y cuyas amplitudes relativas dependen de la forma de onda concreta que adopte la señal. Teniendo en cuenta que el ancho de banda de la señal de televisión se extiende hasta unos 5MHz, deberemos considerar la presencia de aproximadamente 320 armónicos ($320 \times f_{Línea}$ es aproximadamente 5 MHz). Este resultado se representa esquemáticamente en la figura 2.33.

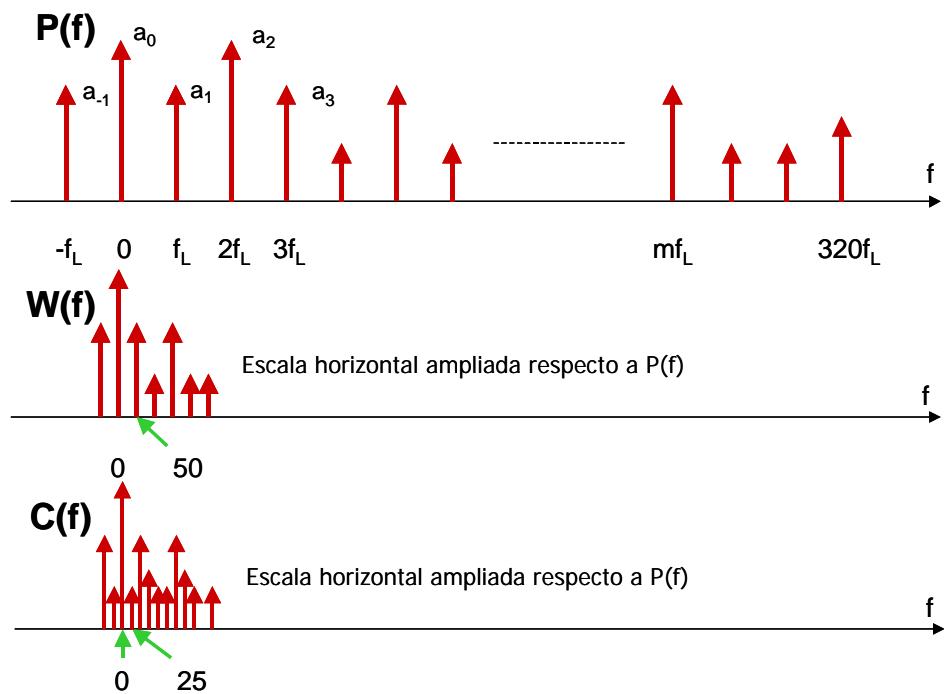


Fig. 2.33 Representación de los espectro $P(f)$, $W(f)$ y $C(f)$

Por otra parte, como las señales $w(t)$ y $c(t)$ también son periódicas, sus transformadas de Fourier, representadas en la figura 2.33, darán un resultado parecido al obtenido en (2.12), aunque en este caso las funciones delta estarán separadas 50 Hz para el caso de $w(t)$ y 25 Hz para $c(t)$:

$$W(f) = F(w(t)) = \frac{1}{T_{campo}} \sum_{k=-\infty}^{\infty} b_k \cdot \delta(f - k \cdot f_{campo}) \quad (2.13)$$

$$C(f) = F(c(t)) = \frac{1}{T_{cuadro}} \sum_{k=-\infty}^{\infty} c_k \cdot \delta(f - k \cdot f_{cuadro});$$

Para obtener el espectro final de la señal basta con convolucionar el tren de deltas $W(f)$ con $P(f)$ y sumar $C(f)$. La convolución de un espectro genérico con un tren de deltas da como resultado el desplazamiento del espectro a las frecuencias a las que están desplazadas las deltas junto con un factor de escala directamente relacionado con la amplitud de las funciones delta. En efecto:

$$W(f) * P(f) = \int_{-\infty}^{\infty} W(v) \cdot P(f - v) \cdot dv = \\ \frac{1}{T_{Línea}} \cdot \int_{-\infty}^{\infty} W(v) \cdot \left(\sum_{k=-\infty}^{\infty} a_k \cdot \delta(f - kf_{Línea} - v) \right) \cdot dv = \frac{1}{T_{Línea}} \cdot \sum_{k=-\infty}^{\infty} a_k \cdot W(f - kf_{Línea}) \quad (2.14)$$

En la figura 2.34 se representan los sucesivos desplazamientos espectrales de $W(f)$ a múltiplos de la frecuencia de línea. Nótese que como la frecuencia de línea es muy superior a la de campo, las deltas correspondientes al espectro $W(f)$, que están separadas por sólo 50 Hz, permanecen muy agrupadas entorno a los múltiplos de $f_{Línea}$. Esto implica que aparecen amplias regiones espectrales, situadas entre los múltiplos de $f_{Línea}$, en las que la señal no tiene contenidopectral. La contribución de $C(f)$ a las característicasspectrales de la señal no es excesivamente importante, ya que únicamente afecta a una pequeña región centrada en el origen.

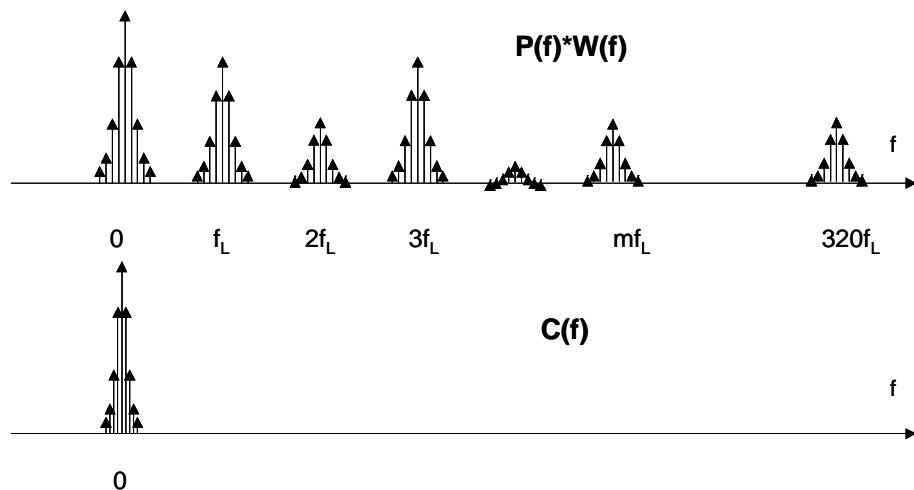


Fig. 2.34 Espectro aproximado de la señal de televisión

En resumen, podríamos decir que el espectro de la señal de televisión concentra su energía en múltiplos de la frecuencia de línea donde se ubican los paquetes de funciones delta separadas 50 Hz, que se corresponden con el espectro de la señal $W(f)$. En la región centrada en el origen, la contribución de $C(f)$ al espectro global representa la aparición de nuevas componentes separadas 25 Hz. Sin embargo, los valores de estas componentes son de pequeña amplitud y no suelen ser consideradas casi nunca en la práctica.

Aunque consideremos un número elevado de armónicos en la señal $w(t)$, prácticamente toda la energía de la señal de televisión queda concentrada alrededor de los múltiplos de la frecuencia de línea, ya que la separación entre las componentes de $W(f)$ es mucho menor que la frecuencia de línea. Con ello, una importante zona del espectro queda libre para la inserción de otras señales sin que se produzcan interferencias significativas con la luminancia. Estos espacios libres en el espectro se aprovecharán para insertar la información de color.

Hasta ahora hemos supuesto que la señal proporcionara información sobre una escena estacionaria y con todas sus líneas iguales. Esta situación no se produce casi nunca en la práctica, por lo que el espectro de la señal de televisión sólo se corresponde con el que hemos obtenido de manera aproximada. En general, las líneas en una señal de televisión suelen presentar una elevada correlación entre sí y los cambios que se producen entre dos cuadros consecutivos casi nunca son excesivamente

importantes, con lo que la hipótesis que hemos realizado suele ser bastante plausible. No obstante, siempre podríamos pensar en algún contraejemplo, con abundantes cambios de plano y con poca simetría vertical y cuya transformada de Fourier no tuviera nada en común con la obtenida en este apartado. Sin embargo, estas imágenes son, desde un punto de vista estadístico, muy poco probables, por lo que, en la práctica, el modelo propuesto resulta simple pero eficaz. Basta considerar, por tanto, que para imágenes reales las señales no serán perfectamente periódicas, por lo que las teóricas funciones delta de Dirac se convertirán en impulsos de amplitud finita y adquirirán una cierta anchurapectral, aunque mantendrán la energía concentrada alrededor de los múltiplos de la frecuencia de línea.

Para que el lector compruebe que el modelo anterior es bastante robusto, podemos considerar cómo afectaría al espectro de la señal si la imagen siguiera siendo estacionaria pero la pantalla estuviera ahora dividida en dos mitades de modo que las formas de onda de las líneas correspondientes a la mitad superior y la inferior son distintas. En este ejemplo es fácil de ver que la señal $p(t)$ puede descomponerse en dos partes tal y como se indica en la figura 2.35.

$$p(t) = p_1(t) \cdot w_1(t) + p_2(t) \cdot w_2(t) \quad (2.15)$$

donde tanto $p_1(t)$ como $p_2(t)$ son señales periódicas con las mismas características que la señal $p(t)$ que habíamos considerado anteriormente.

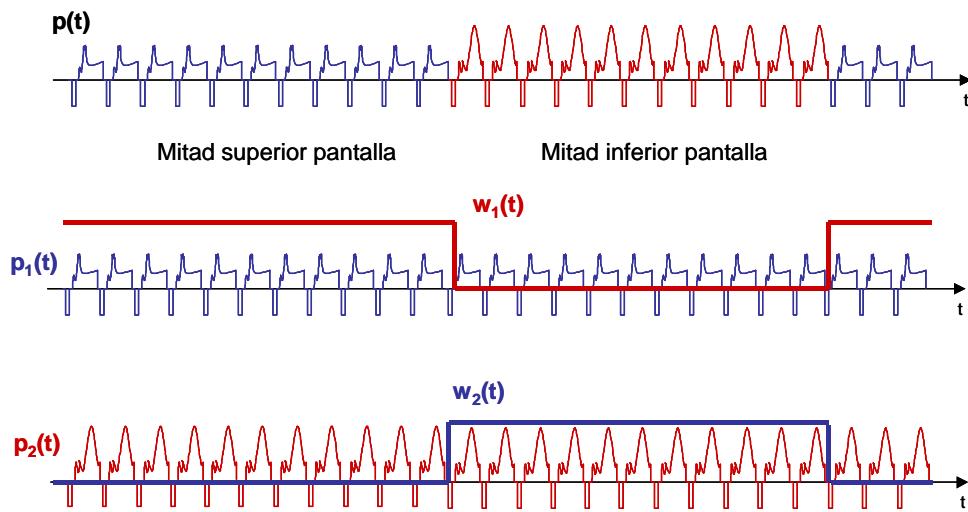


Fig. 2.35 Descomposición de la señal $p(t)$ en señales periódicas

Las señales $w_1(t)$ y $w_2(t)$ son también periódicas y actúan como ventanas para combinar las dos imágenes en las dos mitades de la pantalla. Su frecuencia fundamental es de 50 Hz, ya que las dos formas de onda deben comutarse en cada campo de la señal de televisión. Téngase en cuenta que en la figura 2.35 únicamente se ha representado un campo de la señal, aunque la generalización del dibujo es evidente.

La transformada de Fourier de la señal $p(t)$ vendrá dada en este caso por:

$$P(f) = P_1(f)*W_1(f) + P_2(f)*W_2(f) \quad (2.16)$$

En esta expresión vemos que $P(f)$ es la suma de las dos componentes que se obtienen como el producto de convolución de una señal periódica con frecuencia fundamental $f_{\text{Línea}}$ y otra con frecuencia fundamental f_{campo} . El resultado es, pues, un conjunto de deltas con una distribución espectral idéntica a la que hemos obtenido anteriormente. Obsérvese que sin pérdida de generalidad hubiera sido posible incluir en las funciones $w_1(t)$ y $w_2(t)$ regiones que anularan la señal allí donde debe incorporarse el sincronismo de campo, con lo que para obtener el espectro completo de la señal únicamente faltaría sumar la componente $C(f)$.

En este último ejemplo se observa que no es necesario que todas las líneas sean idénticas para que el espectro de la señal se concentre en los múltiplos de la frecuencia de línea. De hecho, podríamos seguir dividiendo la imagen en regiones verticales manteniendo su estructura espectral. Este resultado confirma que podemos suponer que, en general, las características spectrales que hemos obtenido se mantienen siempre que la imagen presente ciertos patrones de repetición en las líneas y cierto grado de estacionariedad temporal, lo que en general ocurre para la mayoría de aplicaciones prácticas.

Las particularidades del espectro de la señal de luminancia pueden generalizarse también para las señales diferencia de color o R, G, B en aquellos sistemas que utilicen una descomposición de la señal en líneas y campos. Así, las señales R, G, B que se aplican al tubo de imagen, o las componentes diferencia de color, evidentemente antes de su modulación, que se transmiten en sistemas analógicos como el NTSC, PAL o SECAM, o las señales RGB presentes en un Euroconector, tienen todas ellas una estructura espectral parecida.

3 La señal de televisión en color: sistemas compatibles

3.1 Introducción

Los sistemas de televisión en color compatibles, codifican la información de crominancia mediante las señales diferencia de color o, como hemos visto en el capítulo 1, mediante una combinación lineal de éstas. Con ello se obtiene no sólo la compatibilidad directa (que un televisor monocromo pueda visualizar una señal de color), sino también la retrocompatibilidad o compatibilidad indirecta (que un televisor de color pueda visualizar una señal en blanco y negro).

Además, es necesario que la información de crominancia se superponga tanto en el tiempo como en banda de frecuencia con la información de luminancia, de modo que un receptor en color pueda interpretar correctamente la superposición de ambas señales, mientras que un receptor monocromo represente la información de croma como si se tratara de una interferencia. Es pues importante definir una codificación de las señales diferencia de color que pueda ser fácilmente interpretada por los receptores de color (mediante soluciones circuitales no excesivamente complejas), pero que represente un reducido nivel de interferencia visible para la señal de luminancia en los sistemas en blanco y negro.

Las soluciones de los sistemas NTSC, PAL y SECAM para abordar este problema son, a pesar de sus diferencias, conceptualmente muy parecidas. En todos ellos se utiliza alguna combinación lineal de las señales de color que son moduladas mediante una portadora cuya frecuencia se sitúa en la parte alta de la banda de luminancia. En los sistemas NTSC y PAL se utiliza una modulación de amplitud donde las señales diferencia de color están en cuadratura, mientras que en el sistema SECAM se realiza una modulación en frecuencia de las componentes de croma. En todos los casos, la frecuencia de la portadora de la información de color se selecciona para que la interferencia sobre la señal de luminancia sea mínima.

En este capítulo describiremos los principios generales de los sistemas NTSC y PAL. Este último puede considerarse como una mejora del NTSC, por lo que es fundamental comprender los principios de éste para posteriormente poder entender las mejoras que introduce el PAL. Por ello, dedicaremos la mayor parte de este capítulo a describir las características y particularidades del sistema NTSC, preparando el terreno para la posterior descripción de las modificaciones que introduce el PAL. Como en el caso de los sistemas de televisión monocromos, las descripciones del funcionamiento de los sistemas de color serán al nivel de diagramas de bloques sin entrar en los detalles de los circuitos electrónicos empleados para realizar estas funciones. Con ello, pretendemos únicamente que el lector

tenga una idea general sobre los principios de funcionamiento de cada sistema. Aunque la transmisión de señales de televisión analógicas se irá sustituyendo de forma paulatina por sistemas digitales, probablemente, los sistemas de color compatibles seguirán todavía vigentes como formatos de intercambio doméstico de vídeo durante bastantes años. Actualmente, existe un gran parque de receptores con este tipo de decodificadores y un elevado número de reproductores de vídeo, cámaras, etc., que siguen utilizando los sistemas NTSC y PAL. Probablemente, en un futuro próximo, los receptores de televisión y cámaras de vídeo serán compatibles con distintos formatos de señal de vídeo tanto analógicos como digitales.

La parte final del capítulo se dedicará a exponer brevemente la modulación que se aplica en las señales de televisión en banda base para su radiodifusión terrena. El objetivo es que el lector pueda situar la posición espectral de los distintos canales de televisión y la situación de las componentes de audio y vídeo en el espectro. Los sistemas de radiodifusión por satélite o cable utilizan, evidentemente, bandas de frecuencias distintas a las que nosotros haremos referencia y que están adaptadas a las características de estos canales de transmisión. Sin embargo, los decodificadores de satélite o cable que el usuario tiene instalados en el punto de recepción reconvierten las señales recibidas a la banda de radiodifusión terrena o directamente a señal de vídeo compuesto en banda base (conexión del decodificador de satélite a través de Euroconector) para que puedan ser correctamente decodificadas por el receptor.

3.2 Las señales de color en los sistemas NTSC y PAL

Las siglas del sistema NTSC corresponden a la *National Television System Committee*, que fue la comisión encargada de definir los parámetros del sistema de televisión en color compatible que se adoptó en los Estados Unidos en 1953. Se trata de un sistema de 525 líneas con una frecuencia de 60 campos por segundo (30 imágenes por segundo) en el que se superponen la señal de luminancia y la de croma de acuerdo con la siguiente expresión:

$$x_{NTSC}(t) = Y'(t) + I'(t) \cdot \cos(2\pi f_0 t + 33^\circ) + Q'(t) \cdot \sin(2\pi f_0 t + 33^\circ) \quad (3.1)$$

donde $Y'(t)$ corresponde a la información de luminancia e $I'(t)$ y $Q'(t)$ representan una combinación lineal de las señales diferencia de color $(R-Y)'$ y $(B-Y)'$ que proporciona al receptor la información de color. El símbolo ‘ \cdot ’ indica que se realiza una corrección gamma sobre las señales, que al tratarse de un sistema de color, adoptará un valor de 2,8.

Antes de adentrarnos en los secretos y particularidades de este sistema de televisión, empezaremos comparando algunos de los parámetros que utiliza con los del sistema PAL. Esto nos permitirá obtener una mejor perspectiva de las peculiaridades de cada uno de los sistemas.

Prescindiendo de la información de luminancia, la información de color proporcionada por el sistema NTSC es:

$$croma_{NTSC}(t) = Q'(t) \cdot \sin(2\pi f_{NTSC} t + 33^\circ) + I'(t) \cdot \cos(2\pi f_{NTSC} t + 33^\circ) \quad (3.2)$$

Mientras que en el sistema PAL viene dada por:

$$croma_{PAL}(t) = U'(t) \cdot \operatorname{sen}(2\pi f_{PAL} t) \pm V'(t) \cdot \cos(2\pi f_{PAL} t) \quad (3.3)$$

donde en este caso las señales U' y V' son una versión ponderada por constantes de las señales $(B-Y)'$ y $(R-Y)'$.

Una primera inspección a estas ecuaciones revela alguna semejanza y bastantes diferencias entre los dos sistemas, que conviene comentar con cierto detalle. En primer lugar, la similitud es que ambos utilizan una modulación de amplitud con portadoras en cuadratura de las componentes de color. En efecto, cada una de las señales de diferencia de color se multiplican por unas portadoras que están en cuadratura (desfasadas 90° entre ellas) y que, como veremos, permitirán recuperar las señales de color originales, siempre que el receptor pueda disponer de un oscilador en perfecta sincronía con el del transmisor.

La diferencia principal entre los dos sistemas consiste en el cambio de signo que experimenta la componente V' en el sistema PAL. Aunque en esta versión de la ecuación del PAL no se indica explícitamente, el signo de esta componente varía de una línea a otra, permitiendo una codificación del color mucho más robusta e insensible al ruido que en el sistema NTSC. Dejaremos para más adelante la justificación de que esta alternancia en el signo de la componente V' proporcione mayor fiabilidad al sistema en la codificación del color, digamos por el momento que constituye la esencia del sistema PAL, que debe su nombre (*Phase Alternating Line*) a esta particular manera de codificar el color.

El resto de diferencias no son tan significativas desde un punto de vista conceptual. Las frecuencias portadoras, aunque distintas, están pensadas en los dos casos para que la interferencia visual que introduce la croma en un receptor en monocromo sea lo más reducida posible.

3.2.1 Señales diferencia de color en el NTSC y el PAL

Hemos visto en el primer capítulo que cualquier combinación lineal de las señales diferencia de color $(R-Y)$, $(G-Y)$ y $(B-Y)$ permitía obtener sistemas compatibles y retrocompatibles. En el caso del sistema PAL se toman como señales diferencia de color directamente las $(R-Y)$ y $(B-Y)$, aunque ligeramente atenuadas por unos coeficientes diseñados para que el margen dinámico de la señal de crominancia sea el adecuado. Las componentes U y V pueden expresarse como:

$$\begin{aligned} U' &= 0.493 \cdot (B - Y)' \\ V' &= 0.877 \cdot (R - Y)' \end{aligned} \quad (3.4)$$

Los coeficientes que atenúan a las señales diferencia de color han sido calculados para que los niveles de la señal de crominancia PAL se mantengan en un margen de valores adecuados para la señal de vídeo compuesto (es decir, la suma de la señal de luminancia y de crominancia). En efecto, podría ocurrir que al sumar la señal de croma a la de luminancia se superara en exceso el nivel de blanco, o incluso que se redujera por debajo de los niveles de sincronismo. Este último supuesto podría ser realmente grave pues la señal de croma podría pasar a través del circuito separador de sincronismos y activar los circuitos de barrido. Además, teniendo en cuenta que la señal de vídeo compuesto deberá

ser finalmente modulada en amplitud para su transmisión, es importante asegurar que los niveles se mantienen dentro de unos límites en los que la posible sobremodulación de la señal no sea perjudicial para su recepción. En consecuencia, estos coeficientes están determinados para que los niveles de amplitud de la señal de croma se mantengan, independientemente de los colores transmitidos, dentro de los márgenes en los que se puede garantizar una correcta decodificación de la señal.

Las señales de color transmitidas en el sistema PAL suelen representarse mediante vectores en el plano U', V' que se indica esquemáticamente en la figura 3.1. Dado un vector en este plano y la componente de luminancia asociada pueden calcularse, utilizando las relaciones ya mencionadas, las componentes R, G, B. Así por ejemplo, el vector (U,V)=(0.3,0.2) con una luminancia de 0.4 se correspondería con las coordenadas R,G,B siguientes:

$$\begin{aligned} R &= \frac{V}{0.877} + Y = 0.63 \\ B &= \frac{U}{0.493} + Y = 1 \\ G &= \frac{Y - 0.3 \cdot R - 0.11 \cdot B}{0.59} = 0.17 \end{aligned} \quad (3.5)$$

Análogamente, dadas unas componentes R, G, B asociadas a un determinado color, es relativamente simple determinar la componente de luminancia y las de croma U y V. El procedimiento utilizado para el color R=0.3, G=0.5, B=0.7 se indica en las siguientes ecuaciones:

$$\begin{aligned} Y &= 0.3 \cdot R + 0.59 \cdot G + 0.11 \cdot B = 0.46; \\ U &= 0.493 \cdot (B - Y) = 0.12 \\ V &= 0.877 \cdot (R - Y) = -0.14 \end{aligned} \quad (3.6)$$

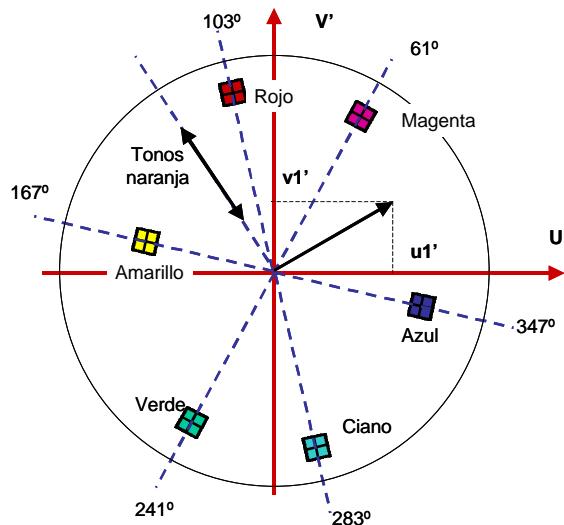


Fig. 3.1 Diagrama de componentes diferencia de color U', V' y colores representativos

Para introducir el efecto de la corrección gamma en estas ecuaciones, basta con considerar que se ha aplicado este factor de corrección sobre las componentes R, G, B y mantener las mismas relaciones algebraicas. Así, el color R=0.0343, G=0.143, B=0.368 sin corrección gamma se corresponde con el $R'=(R)^{1/2.8}=0.3$, $G'=0.5$ y $B=0.7$ por lo que sus componentes Y', U' y V' tendrán los mismos valores que los obtenidos en la ecuación 3.6. Del mismo modo, si aplicamos las ecuaciones 3.5 con los valores de Y', U' y V' corregidos por la gamma, obtendremos las componentes R', G' y B' que deberemos aplicar directamente al tubo de imagen.

Con los procedimientos esbozados arriba podríamos representar los distintos matices de color sobre el plano U',V'. En la figura 3.1 se han representado aproximadamente las posiciones de los colores más habituales. En este diagrama los cambios en el módulo del vector (U',V') se corresponden directamente con cambios en la saturación o intensidad del color (más o menos cantidad de blanco); en otras palabras, al variar el módulo del vector el matiz de color se conserva. En la Figura se indican los ángulos de las líneas en las que encontraremos los matices de los colores primarios, sus complementarios y los naranjas. Al desplazarnos a través de estas rectas obtendremos versiones del mismo color con distintas saturaciones e intensidades. El color blanco y toda la gama de grises están representados en el punto central, ya que este caso las componentes diferencia de color son nulas.

Forma de onda de una línea PAL con la carta de barras de color

Usando las relaciones 3.6 entre las componentes RGB y las variables U y V podemos determinar la forma de onda de una línea de señal PAL si conocemos los colores que deben representarse en cada posición de la línea. Para ello, basta con determinar los niveles de luminancia y de las señales diferencia de color para cualquier combinación RGB y representar la señal de acuerdo con las ecuaciones 3.1 y 3.3. Una señal que habitualmente se utiliza para el ajuste de receptores es la carta con barras de colores. Esta carta contiene todos los colores primarios, los complementarios, el blanco y el negro. Existen versiones de esta tabla con distintos niveles de saturación de los colores. En este ejemplo supondremos que los colores son totalmente puros.

En la tabla 3.1 se proporcionan los niveles RGB asociados a cada uno de los colores. Como suponemos que los colores son puros y tienen el nivel máximo los valores RGB originales y los corregidos en gamma coinciden ya que $1^y=1$. También se determinan los valores de las variables Y, U y V. A partir de estos valores puede determinarse el módulo y la fase de la señal de croma que se suma a la luminancia. En efecto, podemos expresar la ecuación 3.3 como

$$\text{croma}_{PAL}(t) = \sqrt{U'^2(t) + V'^2(t)} \sin\left(2 \cdot \pi \cdot f_{PAL} t \pm \arctan\left(\frac{V'(t)}{U'(t)}\right)\right) \quad (3.7)$$

donde la fase tomará el signo positivo o negativo en función de si la componente V'(t) está invertida o no. A efectos de representación de la forma de onda de la señal, por el momento, únicamente nos interesa el módulo de la componente de croma.

La forma de onda de la señal resultante se representa en la figura 3.2. Los niveles de tensión representados en esta figura se corresponden con una normalización de la señal entre -0,3 Voltios (impulsos de sincronismo) y 0,7 Voltios (nivel máximo de la luminancia para el blanco, Y=1). Por lo tanto, los valores obtenidos en la tabla deben previamente normalizarse a estos niveles de tensión. El procedimiento de normalización es bastante simple. Así, para el color amarillo el nivel medio de tensión vendrá dado por $0,7 \times Y_{\text{amarillo}} = 0,623$ Voltios. La amplitud de la portadora de color es de $0,7 \times P_{\text{amarillo}} = 0,7 \times 0,4493 = 0,3145$ Voltios. En consecuencia, la excursión de tensión de esta barra de color se situará entre $0,623 - 0,3145 = 0,3085$ Voltios y $0,623 + 0,3145 = 0,9375$ Voltios. El resto de valores de tensión asociados a los otros colores pueden determinarse de forma parecida.

Tabla 3.1 Valores normalizados de Y, U y V para la barra de colores puros

	R	G	B	Y	U	V	mod.crom
Blanco	1	1	1	1,00	0,0000	0,0000	0,0000
Amarillo	1	1	0	0,89	-0,4388	+0,0965	0,4493
Ciano	0	1	1	0,70	+0,1479	-0,6139	0,6315
Verde	0	1	0	0,59	-0,2909	-0,5174	0,5936
Magenta	1	0	1	0,41	+0,2909	+0,5174	0,5936
Rojo	1	0	0	0,30	-0,1479	+0,6139	0,6315
Azul	0	0	1	0,11	+0,4388	-0,0965	0,4493
Negro	0	0	0	0,00	0,0000	0,0000	0,0000

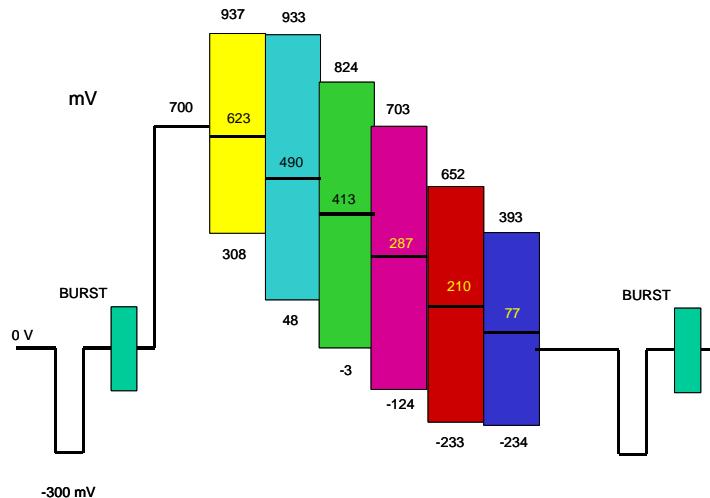


Fig. 3.2 Forma de onda de una línea de la señal de color PAL para una barra de colores

En la figura 3.2 sólo hemos representado los límites de amplitud de señal para cada una de las barras de color. La codificación de la información de color está tanto en los niveles de amplitud como en la fase de modo que al pasar de un color a otro, no sólo se producen los cambios de nivel sino que también existen cambios de fase. La señal de burst que aparece en la figura permite que el receptor pueda obtener una referencia de fase para decodificar correctamente los colores. Veremos los detalles de esta señal más adelante. Obsérvese también que los niveles de amplitud que se producen para los colores primarios y los complementarios puros no superan, en ningún caso, el nivel de sincronismo, por lo que todavía resulta posible identificar la posición de estos impulsos en la forma de onda de la señal sin necesidad de emplear circuitos complejos. De hecho, los factores de escala por los que hemos multiplicado las componentes (B-Y) y (R-Y) se han diseñado específicamente para que los niveles máximos de las señales de color cumplan con estos requisitos. Si los colores de la barra no son puros, sino que tienen algún índice de saturación, las excursiones de la señal respecto al valor de luminancia disminuyen (hay más blanco), por lo que tampoco se produce la sobremodulación de la señal de color.

Componentes de color para la señal NTSC

En el sistema NTSC se eligieron las componentes de crominancia teniendo en cuenta las características de resolución de colores del ojo humano. Para ello se utilizaron los resultados de un sencillo experimento que fue llevado a cabo por König en 1894. Este experimento consiste en situar un tablero de colores a una cierta distancia de los espectadores e ir alejándolo progresivamente para determinar si todos los colores dejan de percibirse a una misma distancia o existen colores que dejan de percibirse antes que otros.

El tablero está dividido en zonas del mismo tamaño, pero con distintos matices de color. El número de zonas en que se divide es muy elevado, de modo que están presentes prácticamente todas las tonalidades. El resultado del experimento indica que la resolución espacial del ojo no es igual para todos los colores. En efecto, cuando el tablero está situado en una posición cercana a los espectadores es posible distinguir todos los colores. A medida que el tablero se aleja de los espectadores algunos colores dejan de percibirse, mientras que otros siguen distinguiéndose correctamente. Este fenómeno empieza a producirse cuando el ángulo subtendido por los rectángulos desciende por debajo de los 20° de grado y afecta principalmente a las tonalidades situadas entre el azul, el magenta y el verde. El sistema visual distingue que se trata de colores distintos basándose en la sensación de brillo que le producen, pero no en su color. A partir de este momento, si el tablero continua alejándose, cada vez son más los colores que dejan de distinguirse. Las tonalidades que tardan más en desaparecer son las correspondientes al amarillo y ciano, que dejan de percibirse como colores cuando el ángulo subtendido por los rectángulos del tablero se sitúa en torno a los 10° de grado. Si el tablero sigue alejándose, ya no podremos apreciar ningún matiz de color aunque seguiremos distinguiendo las distintas zonas en las que está dividido por la sensación de brillo que producen mientras el ángulo con el que observemos cada fragmento sea superior al minuto de grado. Éste es el punto a partir del cual integraremos toda la información observando el tablero como un todo y que, como es lógico, coincide con el límite de la agudeza visual.

La explicación de este fenómeno se debe a que la densidad de los tres tipos de conos en la retina no es idéntica, por lo que la resolución cromática es más baja para unas tonalidades que otras. Cuando las

zonas del tablero son lo suficientemente grandes como para activar un número considerable de los tres tipos de sensores, podemos apreciar todas las tonalidades, pero si el área de excitación se reduce sólo se excitarán mayoritariamente dos tipos de conos, con lo que la visión se volverá dicromática.

Al definir el NTSC se aprovechó esta característica del sistema visual para reducir, en lo posible, el ancho de banda de una las componentes diferencia de color. La idea básica es muy simple y consiste en que una de las señales diferencia de color coincida con las tonalidades en las que el sistema visual presenta menor resolución. Esta señal puede transmitirse con un ancho de banda menor, debido a que estos colores sólo podrán ser distinguidos si ocupan una región que ocupe un área mínima en pantalla. En caso contrario, no es necesario transmitir los cambios que se producen en esta componente de señal puesto que se trata de colores que tampoco serán percibidos por el espectador. El límite del ancho de banda de esta componente de color puede determinarse teniendo en cuenta la resolución de estos colores que, de acuerdo con el experimento de König, se corresponde con un ángulo de visión de $20'$ de grados. La otra componente de color deberá transmitirse con un ancho de banda que esté relacionado con la resolución de $10'$ de grado a la que dejan de percibirse todos los colores. Este mismo argumento puede usarse para justificar el mayor ancho de banda asignado a la señal de luminancia y que se debe a la mejor resolución del sistema visual ($1'$ de grado) para las sensaciones de brillo que las de color.

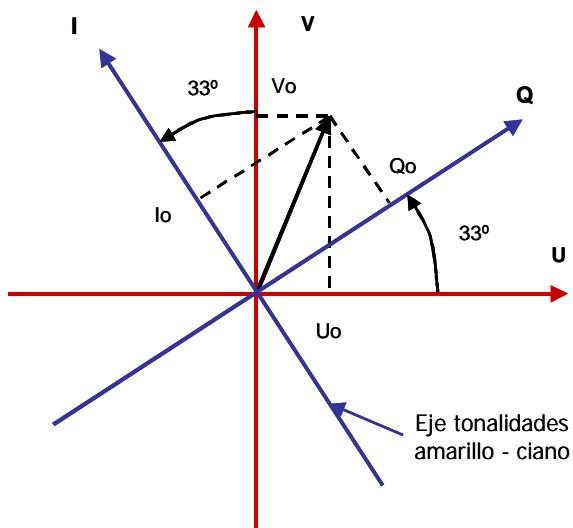


Fig. 3.3 Posición relativa entre los ejes U y V e I y Q

Utilizando esta idea, el sistema NTSC definió las dos señales diferencia de color como una rotación sobre las componentes U y V, de modo que las nuevas coordenadas se correspondieran aproximadamente con las tonalidades de mayor y menor resolución de color del sistema visual humano. Los nuevos ejes de coordenadas para las señales diferencia de color se representan en la figura 3.3, donde se indica un giro de 33° respecto las componentes U y V del PAL. Este giro centra uno de los ejes sobre la tonalidad amarilla (la de mayor resolución) y el otro, ortogonal al primero,

sobre una tonalidad intermedia entre el magenta y el azul (la región con menor resolución al color). Las nuevas coordenadas recibieron los nombres I y Q atendiendo a la fase de la portadora con la que se modula la crominancia en el sistema NTSC. La componente I (In-phase) corresponde a la de mayor ancho de banda y está en fase con la portadora cosenoidal, mientras que la componente Q (Quadrature) está en cuadratura con esta portadora.

La relación analítica entre las componentes del sistema NTSC y el PAL es por tanto:

$$\begin{bmatrix} I \\ Q \end{bmatrix} = \begin{bmatrix} \cos 33^\circ & -\sin 33^\circ \\ \sin 33^\circ & \cos 33^\circ \end{bmatrix} \cdot \begin{bmatrix} V \\ U \end{bmatrix} \quad (3.8)$$

Evidentemente, teniendo en cuenta la relación entre las componentes U y V con las señales diferencia de color, también podemos expresar:

$$\begin{bmatrix} I \\ Q \end{bmatrix} = \begin{bmatrix} 0.783 & -0.269 \\ 0.478 & 0.414 \end{bmatrix} \cdot \begin{bmatrix} R - Y \\ B - Y \end{bmatrix} \quad (3.9)$$

O bien, obtener la relación matricial entre las componentes Y, I y Q con las señales R, G, B que proporcionará la cámara

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3.10)$$

Esta matriz se aplica a las señales R,G,B procedentes de la cámara después de haber realizado la corrección gamma siguiendo un esquema como el indicado en la figura 3.4. Las plantillas de los filtros de las señales I' y Q' se representan en la figura 3.5. El filtro correspondiente a la señal I tiene un ancho de banda de 1.3 MHz y utiliza una curva suave con una banda de transición relativamente amplia, mientras que el filtro de la señal Q, con un ancho de banda de 0.5 MHz, tiene una banda de transición mucho más abrupta, de modo que todas las componentes que superan la frecuencia de corte son fuertemente atenuadas.

En la figura 3.4 también se representan unas células de retardo que se aplican al canal de luminancia y a la señal diferencia de color I. La presencia de estos retardos se debe a que los filtros que se aplican a las tres componentes tienen anchos de banda distintos, por lo que las señales también experimentan retardos distintos. Las células de retardo intentan compensar estas diferencias para que las tres señales estén en fase y por lo tanto proporcionen, en cada instante de tiempo, información sobre un mismo elemento de imagen. La componente que experimenta un mayor retardo es la Q, ya que es la que se pasa a través del filtro con menor ancho de banda. Por ello, sólo es necesario compensar los retardos de las otras dos componentes. En el supuesto de que no se realizara esta corrección, las informaciones correspondientes a la luminancia y a cada una de las señales diferencia de color no estarían correctamente sincronizadas, lo que se observaría en la pantalla como un ligero desplazamiento horizontal entre las tres componentes.

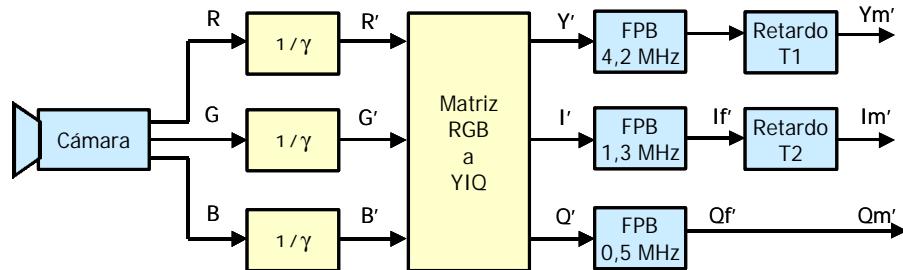


Fig. 3.4 Obtención de la señal de luminancia y componentes I, Q en el NTSC

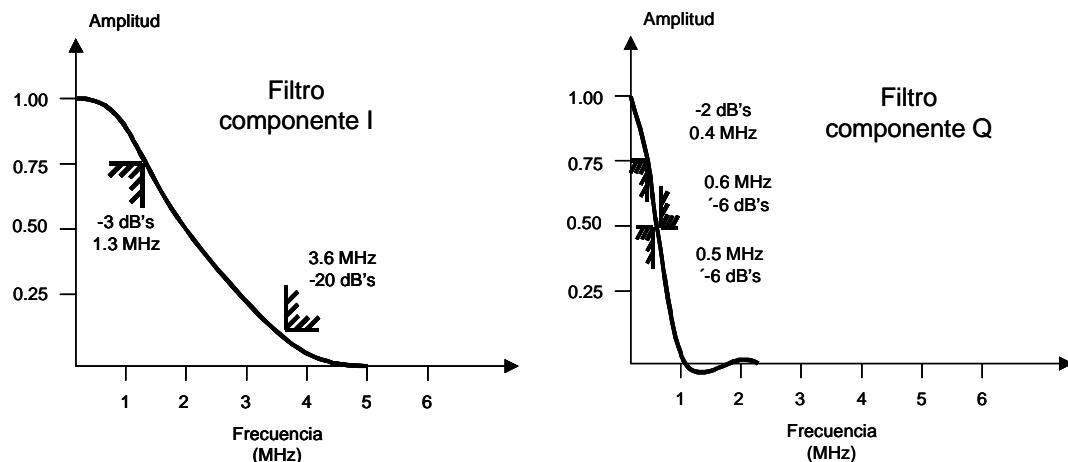


Fig. 3.5 Plantillas de los filtros de señal diferencia de color I, Q en el sistema NTSC

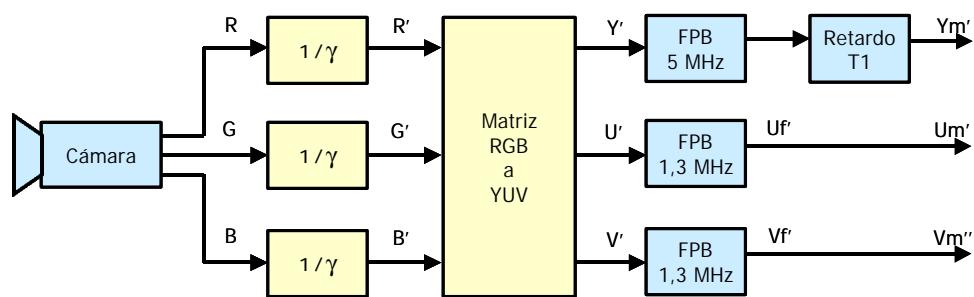


Fig. 3.6 Obtención de la señal de luminancia y componentes U y V en el sistema PAL

Así, en una región de la imagen en la que apareciera un contorno uniforme, veríamos que el cambio de color se produciría en una posición situada ligeramente más hacia la derecha de la zona en la que se produce el cambio de brillo. Además, como primero se recibiría la componente I y después la Q, observaríamos cambios de matices en las zonas de los contornos que no se corresponderían con la imagen real. En la figura 3.4 las señales en la salida de los filtros y en la salida de las células de retardo se indican respectivamente como I_f' , Q_f' y I_m' , Q_m' . No obstante, en general suele prescindirse de estos subíndices para distinguir las distintas versiones de las señales, entendiéndose que siempre se realiza este filtrado y retardo de señales y que, por contexto, dependiendo del punto del sistema al que hagamos referencia, sabremos siempre de qué señales se trata.

En el sistema PAL la obtención de las señales Y' , U' y V' sigue un esquema parecido al utilizado en el NTSC (ver figura 3.6), aunque en este caso los dos filtros de las señales diferencia de color son idénticos, por lo que únicamente deberemos retardar la señal de luminancia. Las plantillas utilizadas para los filtros de las componentes U' y V' se representan en la figura 3.7.

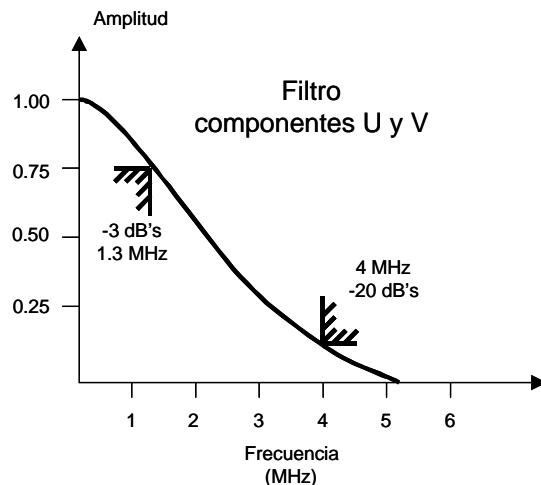


Fig. 3.7 Plantillas de los filtros de las componentes U y V en el sistema PAL

3.2.2 Modulación de las señales diferencia de color en los sistemas NTSC y PAL

Una vez establecida la relación entre las señales diferencia de color que se utilizan en los sistemas PAL y NTSC, podremos comprobar que las señales transmitidas al canal, aunque con diferencias ciertamente sutiles, tienen muchas características en común.

En efecto, desarrollando los términos en seno y coseno en la expresión de la señal de croma para el sistema NTSC, obtenemos la siguiente expresión:

$$\begin{aligned}
 croma_{NTSC}(t) &= Q'(t) \cdot \sin(2\pi f_N t + 33^\circ) + I'(t) \cdot \cos(2\pi f_N t + 33^\circ) = \\
 &= Q'(t) \cdot (\sin(2\pi f_N t) \cdot \cos(33^\circ) + \cos(2\pi f_N t) \cdot \sin(33^\circ)) + \\
 &\quad + I'(t) \cdot (\cos(2\pi f_N t) \cdot \cos(33^\circ) - \sin(2\pi f_N t) \cdot \sin(33^\circ)) = \\
 &= B'(t) \cdot \sin(2\pi f_N t) + A'(t) \cdot \cos(2\pi f_N t)
 \end{aligned} \tag{3.11}$$

donde hemos definido las señales $A'(t)$ y $B'(t)$ como las componentes en fase y cuadratura respecto a la portadora cosenoideal. Estas señales pueden expresarse en función de las señales $Q'(t)$ e $I'(t)$ como

$$\begin{aligned}
 A'(t) &= Q'(t) \cdot \sin(33^\circ) + I'(t) \cdot \cos(33^\circ); \\
 B'(t) &= Q'(t) \cdot \cos(33^\circ) - I'(t) \cdot \sin(33^\circ);
 \end{aligned} \tag{3.12}$$

relación que puede expresarse matricialmente como

$$\begin{bmatrix} A'(t) \\ B'(t) \end{bmatrix} = \begin{bmatrix} \cos 33^\circ & \sin 33^\circ \\ -\sin 33^\circ & \cos 33^\circ \end{bmatrix} \cdot \begin{bmatrix} I'(t) \\ Q'(t) \end{bmatrix} \tag{3.13}$$

donde se observa claramente que las componentes $A'(t)$ y $B'(t)$ se corresponden con un giro de 33° en el sentido horario de las componentes $I'(t)$ y $Q'(t)$.

Llegados a este resultado, cabe cuestionarnos qué sentido tiene que el sistema NTSC realice un primer giro de 33° de las componentes V' y U' en el sentido antihorario, si posteriormente estas señales serán transmitidas mediante unas portadoras desfasadas que equivalen a realizar un giro de 33° en el sentido horario de las componentes de color. Una primera inspección a este resultado parece indicar que el sistema NTSC está transmitiendo las mismas señales que se envían en el sistema PAL (exceptuando, claro está, el cambio de signo en la componente V'). En efecto, si sustituimos directamente la ecuación 3.8 en la ecuación 3.13 obtendremos:

$$\begin{bmatrix} A'(t) \\ B'(t) \end{bmatrix} = \begin{bmatrix} \cos 33^\circ & \sin 33^\circ \\ -\sin 33^\circ & \cos 33^\circ \end{bmatrix} \cdot \begin{bmatrix} \cos 33^\circ & -\sin 33^\circ \\ \sin 33^\circ & \cos 33^\circ \end{bmatrix} \cdot \begin{bmatrix} V'(t) \\ U'(t) \end{bmatrix} = \begin{bmatrix} V'(t) \\ U'(t) \end{bmatrix} \tag{3.14}$$

lo que nos indica que las señales $A'(t)$ y $B'(t)$ coinciden con las que se transmiten en el sistema PAL.

Sin embargo, hemos de tener en cuenta que las señales I' , Q' que se aplican a los mezcladores han sido previamente filtradas, por lo que no en todos los casos las componentes $A'(t)$ y $B'(t)$ coincidirán con las señales $U'(t)$ y $V'(t)$ que se transmitirían en el sistema PAL. Consideraremos algunos casos particulares para aclarar estas diferencias.

Supongamos que la escena que desea transmitirse produce una señal $Q'(t)$, con todas sus componentes espectrales situadas dentro del ancho de banda de 0.5 MHz que se utiliza en el filtro de la componente en cuadratura. En este caso, es evidente que la señal que se aplica al modulador, llamémosla Q'_f , coincide exactamente con la componente $Q'(t)$, por lo que las señales $A'(t)$ y $B'(t)$ serán prácticamente iguales a las $U'(t)$ y $V'(t)$ que se transmitirían en el PAL.

No obstante, en el caso en que la señal $Q'(t)$ tenga componentes espectrales de frecuencia superior al ancho de banda del filtro, la señal $Q'_{\text{f}}(t)$ no coincidirá con la original. Esto significa que se transmitirán matices de color distintos a los que realmente existen en la escena, ya que no se envía la componente completa, sino únicamente la parte de baja frecuencia. El problema, como hemos visto, no es grave, ya que el sistema visual será incapaz de distinguir entre los matices reales y los recibidos. Sin embargo, ya no es posible sustituir la ecuación 3.8 en la 3.13, puesto que la componente que realmente se transmite no es exactamente la $Q'(t)$. En consecuencia, las señales $A'(t)$ y $B'(t)$ ya no coinciden con las $U'(t)$ y $V'(t)$.

En resumen, la filosofía empleada en el sistema NTSC para la construcción de las señales diferencia de color permite ahorrar algo de ancho de banda en una de las componentes, sin que suponga una pérdida en la calidad de la señal. Si las componentes de color sólo tienen contenido en baja frecuencia, la información de crominancia transmitida en el sistema NTSC y el PAL coinciden. En cambio, si existen componentes de alta frecuencia, las señales de croma en ambos sistemas no coinciden, aunque el espectador no es capaz de apreciar las diferencias entre los matices de las informaciones recibidas. Veremos en secciones posteriores que uno de los principales problemas del sistema NTSC es que pueden producirse degradaciones en el matiz de color obtenido en el reproductor respecto al matiz original de la escena. Sin embargo, estos problemas son debidos a motivos completamente ajenos al tipo de señales diferencia de color que se transmiten.

En la siguiente sección se describirá con detalle la señal en banda base para el sistema NTSC y se analizarán los diagramas de bloques del codificador y el decodificador NTSC. Asimismo, discutiremos el problema al que nos hemos referido en el párrafo anterior y cuya resolución llevó a la introducción del sistema PAL.

3.3 El sistema NTSC

3.3.1 Selección de la frecuencia portadora de la señal de croma

La superposición de la señal de croma con la información de luminancia es, como ya se discutió en el capítulo 1, la única forma de conseguir que los sistemas de televisión en color sean compatibles con los de blanco y negro. Sin embargo, esta estrategia de transmisión de la información de color hace inevitable que aparezcan interferencias entre las componentes de luminancia y las de color. Estas interferencias afectarán especialmente a los receptores en blanco y negro, ya que en su diseño no se ha tenido en cuenta la presencia de la señal de croma.

Una vez decidido el sistema de modulación y las señales diferencia de color que se utilizarán para la transmisión de la croma, sólo queda seleccionar la frecuencia portadora de la señal. Esta frecuencia portadora deberá elegirse, utilizando como criterio principal que la interferencia que introduce la señal de color en los receptores de blanco y negro sea mínima. Si expresamos la ecuación 3.1 en función de una única portadora cosenoide obtenemos:

$$\begin{aligned} x_{\text{NTSC}}(t) &= Y'(t) + I'(t) \cdot \cos(2\pi f_0 t + 33^\circ) + Q'(t) \cdot \sin(2\pi f_0 t + 33^\circ) = \\ &= Y'(t) + \sqrt{Q'^2(t) + I'^2(t)} \cdot \cos(2\pi f_0 t + 33^\circ - \arctg(Q'(t)/I'(t))) \end{aligned} \quad (3.15)$$

donde se observa que a la señal de luminancia se superpone una señal cosenoidal cuya amplitud y fase proporciona la información relativa a las señales diferencia de color. En la figura 3.8 se ha representado una línea típica de la señal de televisión en color donde se indica explícitamente la superposición entre las dos componentes. En principio, un receptor en color deberá estar preparado para separar correctamente las componentes de luminancia y color de la señal recibida. En cambio, un receptor en blanco y negro interpretará la componente cosenoidal como una interferencia que se superpone a la señal de luminancia y que, al no poder separarla de la primera, será representada como si fuera la propia luminancia en la pantalla del receptor.

Así pues, deberemos elegir la frecuencia portadora de la información de crominancia de acuerdo con un triple criterio. Por una parte, los receptores en color deberán poder separar con facilidad ambas componentes utilizando filtros u otros sistemas de fácil implementación. Además, deberemos intentar elegir la frecuencia portadora para minimizar las intermodulaciones entre ambas señales, que podrían provocar que parte de la información de crominancia fuera interpretada como luminancia y viceversa. Finalmente, admitiendo que es inevitable que los receptores en blanco y negro representen en la pantalla toda la información recibida, sin distinguir si se trata de croma o luminancia, la portadora de color deberá tener una frecuencia que introduzca un patrón de interferencia poco visible para el espectador.

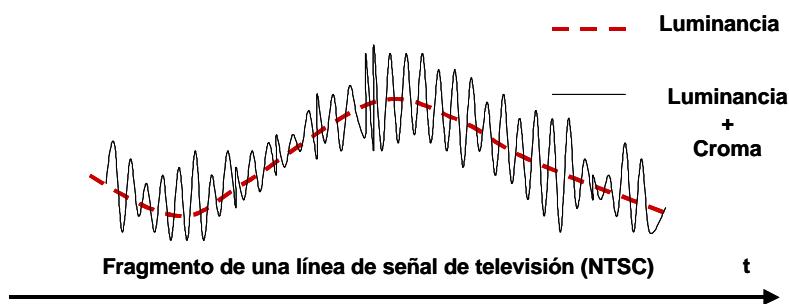


Fig. 3.8 Superposición de la luminancia y la croma en el sistema NTSC

Debemos comentar en este punto que en todo momento trabajaremos con la hipótesis de que los receptores monocromos suponen que se está transmitiendo una señal en blanco y negro puro. En la actualidad, muchos receptores en blanco y negro realizan algún tipo de filtrado previo sobre la señal recibida presuponiendo que únicamente recibirán señales en color. Con ello se consigue reducir apreciablemente la interferencia que produce la croma sobre la luminancia. Sin embargo, cuando se diseñaron los primeros sistemas en color, debía considerarse el efecto de estas señales sobre receptores preparados exclusivamente para recibir señales en blanco y negro. Por tanto, este es el problema que debemos considerar como criterio de selección de la frecuencia portadora.

Imbricación espectral entre la luminancia y las componentes de color

Todas las consideraciones previas nos llevan a elegir una frecuencia portadora que sea un múltiplo impar de la mitad de la frecuencia de línea.

$$f_{NTSC} = (2n+1) \frac{f_{Línea}}{2} \quad (3.16)$$

La razón fundamental de esta selección es que el contenido espectral de la señal de luminancia está concentrado en los múltiplos de la frecuencia de línea, por lo que al modular las señales diferencia de color nos interesarán que su energía se concentre en aquellas regiones espectrales donde la señal de luminancia presente un contenido energético mínimo. De este modo, las componentes de color y luminancia quedan imbricadas espectralmente, reduciéndose la interferencia entre ambas y facilitando la separación de estas componentes en el receptor.

La imbricación espectral entre las señales se ilustra en la figura 3.9. La señal de luminancia y las dos componentes diferencia de color tienen, como hemos visto en el capítulo anterior, un espectro en el que la mayor parte de la energía se concentra en los múltiplos de la frecuencia de línea. Si elegimos una frecuencia portadora que sea un múltiplo impar de la mitad de la frecuencia de línea, es decir, que esté situada justo entre dos múltiplos de la frecuencia de línea, desplazaremos el contenido espectral de las señales I y Q a frecuencias en las que la señal de luminancia tiene un contenido prácticamente nulo. De este modo, los dos espectros quedan imbricados, compartiendo la misma banda de frecuencias, pero con una interferencia mutua mínima. Además, veremos que es posible obtener una muy buena separación entre las dos componentes utilizando un tipo especial de filtros periódicos (*Comb Filters*).

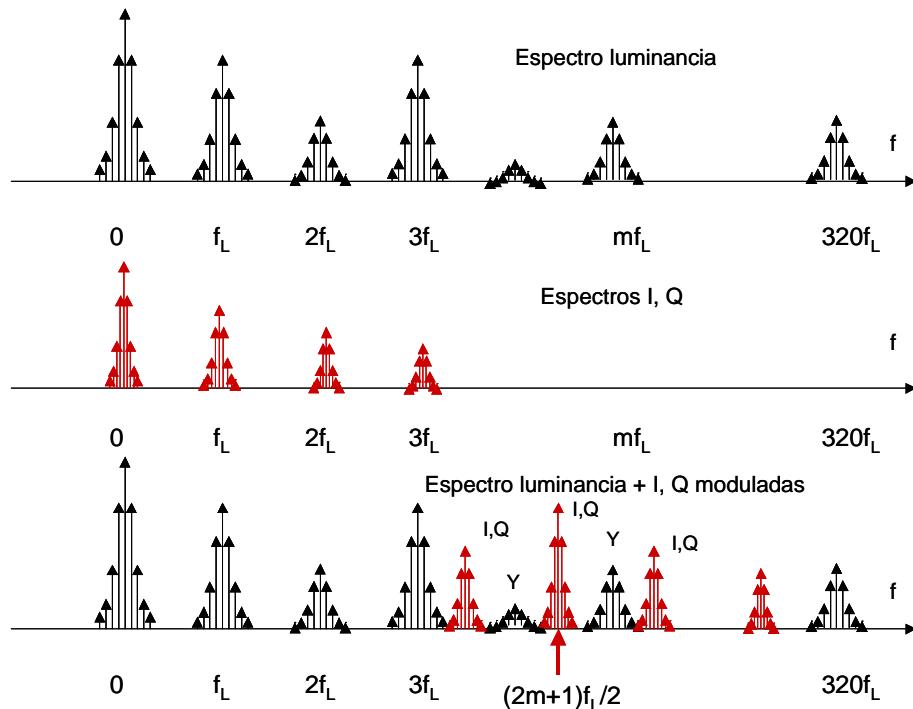


Fig. 3.9 Imbricación entre los espectros de la luminancia y la croma

En la figura 3.9 hemos querido mantener la presencia de las componentes, separadas 50 Hz y que son debidas a la repetición de la señal en cada campo. Debería notarse que esta representación no se ha realizado a escala, en cuyo caso hubiéramos obtenido una mayor concentración de los espectros en los múltiplos de la mitad de la frecuencia de línea, dejándose un considerable espacio intersticial entre las componentes de luminancia y de croma.

El valor de la frecuencia portadora de la señal de croma debe ser suficientemente elevado para que el patrón de interferencias que se produce en la pantalla de un receptor monocromo sea lo menos visible posible. Si la frecuencia es elevada podemos intuir, a partir de la figura 3.8, que la imagen representada en el receptor tendrá cambios de luminancia muy rápidos que, siempre que produzcan un patrón poco definido, serán integrados por el espectador, resultando poco visibles. Sin embargo, debemos notar que la frecuencia portadora no puede aumentarse excesivamente, ya que tendremos que mantener las señales de croma dentro del ancho de banda asignado a la componente de vídeo en un canal de televisión, que en el sistema NTSC es de aproximadamente 4,2 MHz. Téngase en cuenta que este ancho de banda es una restricción impuesta por los sistemas en blanco y negro que ya estaban operativos en el momento de definir el NTSC. Los sistemas en color compatibles deben siempre acomodarse en los canales previamente utilizados por los sistemas en blanco y negro, puesto que si no fuera así, podrían introducirse interferencias en sistemas ya operativos. Además, debe tenerse presente que la portadora de sonido está situada en 4,5 MHz y debe mantenerse en esta posición si pretendemos que el sistema sea compatible.

La frecuencia que finalmente se eligió para el sistema NTSC es:

$$f_{NTSC} = \frac{455}{2} \cdot f_{Línea} = 227,5 \cdot f_{Línea} \quad (3.17)$$

Si sustituimos la frecuencia de línea por su valor numérico obtenemos:

$$f_{NTSC} = 227,5 \times (525 \times 30 \text{ Hz}) = 3,583125 \text{ MHz} \quad (3.18)$$

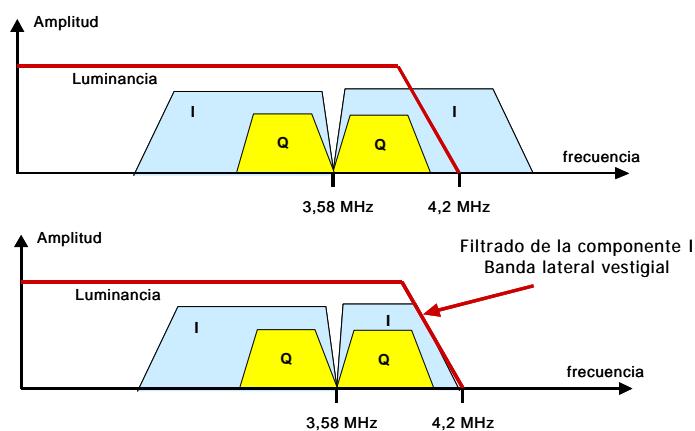


Fig. 3.10 Ubicaciones espectrales de las componentes de luminancia y croma

En la figura 3.10 se representa la posición relativa de los espectros de la señal de luminancia y de cada una de las componentes de color. Aunque sigue manteniéndose la imbricación espectral entre las señales, los espectros se han representado como continuos para facilitar la localización de los límites del ancho de banda de cada componente.

La portadora elegida desplaza las señales I y Q a la parte alta del espectro. La señal Q, con un ancho de banda de 1 MHz, se extiende desde los 3 MHz hasta los 4 MHz, por lo que tiene cabida en el ancho de banda asignado a la componente de vídeo de un canal del sistema NTSC. No obstante, la señal I, con un ancho de banda total de 2,6 MHz, se situará entre los 2,2 MHz hasta los 4,8 MHz con lo cual excedería el límite asignado a la señal de vídeo. La solución adoptada consiste en eliminar, mediante filtrado, la parte alta del espectro de la señal I, por lo que realmente esta señal se transmite con su banda lateral inferior completa más una banda lateral vestigial superior de 0,5 MHz. Veremos que este formato de transmisión de la señal no supone ningún problema para su demodulación.

3.3.2 Patrones de interferencia en el sistema NTSC

Se conocen con el nombre de patrones de interferencia las componentes de la imagen que aparecen en un receptor monocromo como consecuencia de la existencia de la información de croma en la señal de televisión. La señal recibida puede descomponerse en la componente de luminancia más la de croma

$$x_{NTSC}(t) = Y'(t) + croma_{NTSC}(t) \quad (3.19)$$

Desde el punto de vista de un receptor monocromo, la señal $Y'(t)$ es la componente deseada mientras que la croma deberá ser considerada como una interferencia. La imagen que veremos en la pantalla del receptor será, pues, una superposición de estas dos señales, interesándonos que la componente de croma sea lo menos molesta posible. La elección de una frecuencia portadora de valor elevado condiciona que las componentes de croma varíen muy rápidamente en el tiempo, lo que significa que en la pantalla se representarán cambios de luminancia muy bruscos que el sistema visual, en lo posible, integrará. Decimos que, *en lo posible, integrará* debido a que la portadora de la señal de croma tiene una frecuencia de 3,58 MHz que, en principio, debería resultar visible, pues se encuentra dentro de los límites en los que el ojo tiene aún una buena resolución. De hecho, si la interferencia permaneciera estacionaria en la pantalla, sería muy molesta y su presencia sería intolerable para los usuarios de receptores monocromos.

Sin embargo, comprobaremos que al haber elegido como frecuencia portadora un múltiplo impar de la mitad de la frecuencia de línea, la interferencia generada en la pantalla cambia completamente su fase de una imagen a otra. De este modo, los puntos de la pantalla que en la imagen actual aparecen con una luminancia superior a la real aparecerán, en la imagen siguiente, con una luminancia inferior a la real. La baja visibilidad de la interferencia se debe, por tanto, a la acción conjunta de integración del sistema visual tanto en el espacio como en el tiempo. El fenómeno que se produce es parecido al que se discutió en el capítulo 1 con motivo del efecto Kell. En aquel caso, no era posible que el ojo observara una imagen en la que todas las líneas de un campo eran blancas y las del otro campo negras, debido al efecto conjunto de rápida variación espacial y temporal de la imagen, que provocaba que se integraran ambas informaciones en una imagen gris. No obstante, en este caso debe tenerse en cuenta

que la frecuencia portadora de la croma no está en el límite de resolución del ojo, por lo que la interferencia, aunque poco molesta y por lo tanto tolerable, será perceptible.

Evidentemente, el patrón de interferencia generado por la señal de croma depende de la propia información que contiene la señal. No obstante, con el objeto de esclarecer los conceptos expuestos en los párrafos anteriores, representaremos un ejemplo concreto de imagen o patrón interferente. Para ello, supondremos que se transmite una señal con luminancia constante a la que se superpone una señal de croma cuyas componentes de color tampoco varían en el tiempo. Este ejemplo se correspondería con una imagen estática de color uniforme en toda la pantalla. La forma de onda asociada a una de las líneas se representa esquemáticamente en la figura 3.11, donde el color de la imagen dependería del valor medio de la señal (luminancia) y de la amplitud y fase de la portadora de color (componentes I y Q).

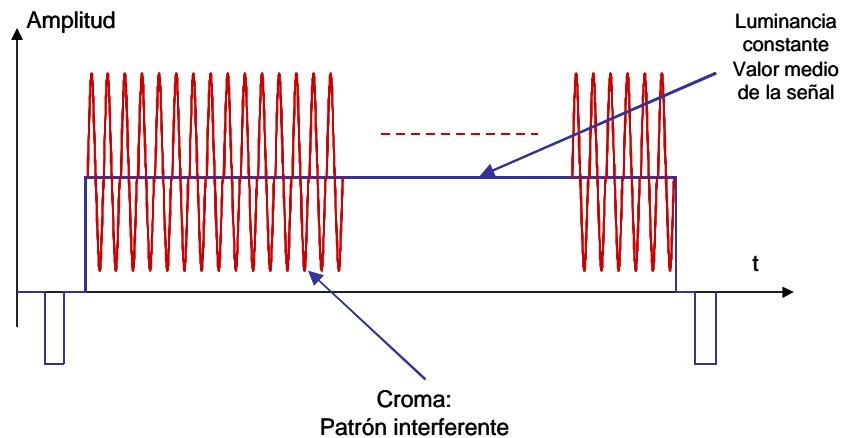


Fig. 3.11 Señal genérica para la que se determina la imagen del patrón de interferencia

La señal que desearíamos que se representara en un receptor en blanco y negro sería únicamente la de luminancia, que en este caso concreto se correspondería con el valor medio de la señal. Sin embargo, la croma se superpondrá a la luminancia y, por lo tanto, la imagen que veremos en la pantalla se corresponderá con la forma de onda de la señal de la figura 3.11, es decir, obtendremos rápidas variaciones de la luminancia a lo largo de la pantalla. En resumen, el receptor representa la imagen como si se tratara de una señal en blanco y negro, ya que en principio no está diseñado para descifrar la información de color.

La imagen que se observará en la pantalla puede descomponerse en la suma de la señal deseada, que en este caso es un nivel de gris constante, más la interferencia sinusoidal debida al color. Esta interferencia se corresponde con una imagen cuyos niveles de brillo a lo largo de una línea dependen de la amplitud de la componente sinusoidal y varían de acuerdo con la frecuencia portadora de la señal de croma. Por lo tanto, como la frecuencia portadora del color es 227,5 veces la frecuencia de línea, se representarán 227,5 ciclos de esta interferencia en cada línea de la imagen (en realidad sólo serán visibles los ciclos que estén dentro de la línea activa).

En la figura 3.12 se ilustra el procedimiento general para ver la forma gráfica del patrón interferente generado por la portadora de color. Las partes correspondientes al semiperíodo negativo y positivo de la señal sinusoidal se representan respectivamente como pequeños cuadrados de color oscuro y claro. En realidad la luminancia asociada a la señal tiene transiciones sinusoidales entre los niveles máximo y mínimo de brillo, pero para simplificar la visualización del patrón interferente, supondremos que estas transiciones pueden aproximarse por estos pequeños cuadrados. Así, el brillo máximo de la interferencia estará situado en el centro de un cuadrado claro, mientras que el mínimo nivel de brillo se corresponderá con el centro de un cuadrado oscuro.

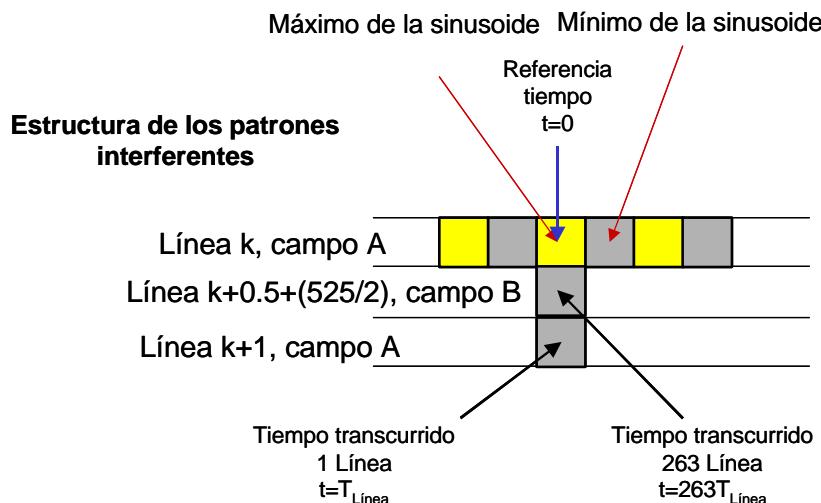


Fig. 3.12 Cálculo del patrón interferente producido en la pantalla por la portadora de color

Consideremos un punto de la pantalla en el que se produce un máximo nivel de la interferencia. Supondremos que este punto, tal y como se indica en la figura 3.12, pertenece a la línea k del campo A y en él tomaremos la referencia de origen de tiempos. El punto situado en la misma vertical, pero dos líneas más hacia abajo, corresponde a la línea $k+1$ del campo A y se representará en la pantalla justo una línea después del punto que hemos tomado como referencia. El tiempo que transcurre entre la representación de los dos elementos es, por tanto, igual al tiempo de línea ($T_{\text{Línea}}$), por lo que la portadora de color habrá completado un total de 227,5 ciclos. En efecto:

$$n^{\circ} \text{ ciclos} = \frac{T_{\text{Línea}}}{T_{\text{portadora de color}}} = \frac{227,5 \cdot f_{\text{Línea}}}{f_{\text{Portadora}}} = 227,5 \text{ ciclos} \quad (3.20)$$

De acuerdo con este resultado, la señal habrá experimentado un cambio de fase de 180° , por lo que si en el instante de referencia la interferencia tenía un nivel de brillo máximo, ahora será mínimo.

Determinemos ahora el nivel de la portadora en la línea de la imagen situada justo por debajo del elemento de referencia. Esta línea corresponde al siguiente campo de la imagen y se representará

cuando hayan transcurrido un total de $((525/2)+1/2)=263$ líneas desde que se ha representado el primer elemento. El término de $525/2$ líneas se debe a que ha transcurrido un campo completo entre los dos elementos y la media línea adicional se añade debido a que el número de líneas en cada campo no es entero. La figura 3.13 puede ayudar a comprender mejor la presencia de estos dos términos. En este ejemplo se considera un sistema de televisión de 5 líneas y dos campos. El elemento de referencia ha sido indicado mediante un pequeño cuadrado, mientras que los puntos obtenidos después de cada periodo de línea completa se representan mediante círculos. Nótese que en este ejemplo el punto situado inmediatamente debajo del elemento de referencia se traza después de $(\text{número total líneas}/2)+0.5$ líneas, es decir, después de 3 líneas completas.

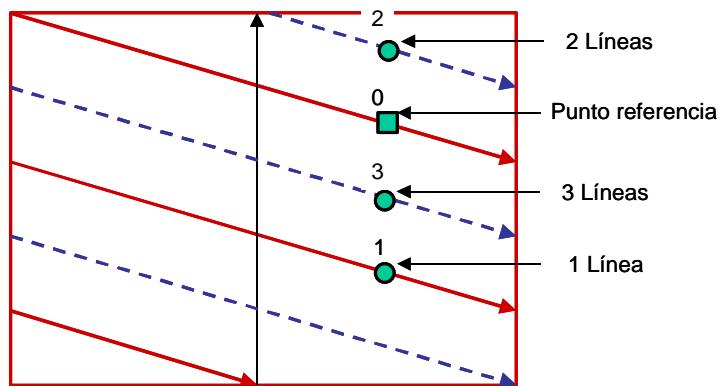


Fig. 3.13 Cálculo del número de líneas entre un elemento de imagen y el inmediatamente inferior en un sistema simplificado de 5 líneas y 2 campos

Volviendo a nuestro sistema de 525 líneas, el número de ciclos de la portadora de color que se habrán completado durante estas 263 líneas puede calcularse como

$$\frac{263 \cdot T_{\text{Línea}}}{T_{\text{portadoracolor}}} = \frac{263 \times 227,5 \cdot f_{\text{Línea}}}{f_{\text{Línea}}} = 59832,5 \text{ ciclos} \quad (3.21)$$

Nuevamente este resultado nos indica que la portadora de color, después de estas 263 líneas, volverá a estar en contrafase respecto al valor que tomaba en el elemento de imagen de referencia. Los niveles de brillo que toma la portadora de color en los tres elementos considerados se han representado en la figura 3.12.

Si el elemento de referencia coincide con un máximo de brillo de la interferencia, los dos elementos situados en las líneas inmediatamente inferiores tomarán valores mínimos. Podríamos repetir este razonamiento para cualquier elemento situado sobre la línea k, o sobre cualquiera de las líneas del campo A, obteniendo una imagen aproximada de cómo se visualizará la interferencia en la pantalla. El resultado obtenido se ha representado en la figura 3.14, donde se observa que los máximos y mínimos de la interferencia quedan superpuestos en un patrón complejo en los que los niveles límite se alternan tanto en el sentido vertical como en el horizontal.

Si la frecuencia de la portadora de color hubiera sido un múltiplo de la frecuencia de línea, se hubiera obtenido un patrón interferente formado por líneas verticales que alternan los valores máximos y mínimos. Así pues, aunque en el sentido horizontal este patrón interferente presentaría cambios a una frecuencia razonablemente alta, en el sentido vertical los niveles permanecerían constantes, siendo por tanto muy visibles y molestos.

Hasta ahora se ha considerado la forma de patrón interferente considerando una única imagen. La principal ventaja de la frecuencia portadora que se ha elegido es que este patrón se invierte completamente de una imagen a la siguiente, permitiendo, como ya se ha comentado, que la imagen que se presenta al espectador presente rápidas variaciones no sólo en las direcciones vertical y horizontal, sino también en el tiempo.

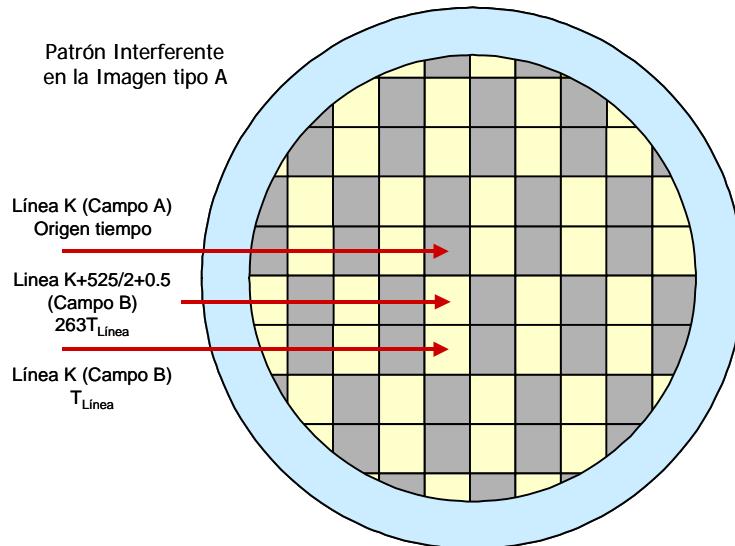


Fig. 3.14 Forma del patrón interferente correspondiente a una imagen completa

En efecto, refiriéndonos nuevamente a la figura 3.12, está claro que el elemento de referencia volverá a representarse en la siguiente imagen una vez que hayan transcurrido las 525 líneas propias del sistema NTSC. Podemos calcular el número de ciclos de la portadora de color que transcurren durante este intervalo de tiempo, obteniendo, de nuevo, que la portadora estará en contrafase respecto al valor que tomaba en este mismo punto de la pantalla en la imagen anterior. En efecto:

$$\frac{525 \cdot T_{\text{Línea}}}{T_{\text{portadora de color}}} = \frac{525 \times 227,5 \cdot f_{\text{Línea}}}{f_{\text{Línea}}} = 119437,5 \text{ ciclos} \quad (3.22)$$

Así, la imagen asociada a la interferencia varía de imagen a imagen, de manera que, en cada punto de la pantalla, la interferencia que se superpone a la señal deseada tomará valores de signo cambiado oscilando a una frecuencia de 12,5 Hz (equivalente a dos imágenes). En la figura 3.15 se representan

los patrones obtenidos en esta secuencia de dos imágenes que se han denominado imagen tipo A e imagen tipo B.

Evidentemente, la secuencia temporal del patrón interferente abarca únicamente dos imágenes, ya que la portadora de color estará nuevamente en fase con el punto de referencia cuando hayan transcurrido 2×525 líneas, que se corresponden con $2 \times 119437,5 = 238875$ ciclos completos de la portadora.

Es importante recordar que todo el análisis del patrón interferente se ha basado en que la señal que se transmitía era estacionaria y uniforme en toda la pantalla. En un caso real se producirán variaciones tanto en la señal de luminancia como en la amplitud y la fase de la portadora de color, por lo que la imagen de la croma no será tan estable como hemos postulado. Las variaciones en la amplitud de la señal de croma supondrán que en determinadas regiones la amplitud de la imagen interferente será mayor que en otras. Por otra parte, los cambios de fase pueden significar que en las zonas en las que se produzcan cambios de tonalidades el patrón de interferencias deje de tener la estructura periódica que hemos discutido. Sin embargo, estas zonas están ligadas a los contornos de la imagen en los que, debido a que varían tanto la luminancia como la tonalidad de la imagen, la interferencia será poco visible. En zonas con brillo y color uniformes, que por otra parte son las más abundantes y en donde la interferencia es más fácil de percibir, se mantendrá la estructura espacial y temporal del patrón interferente.

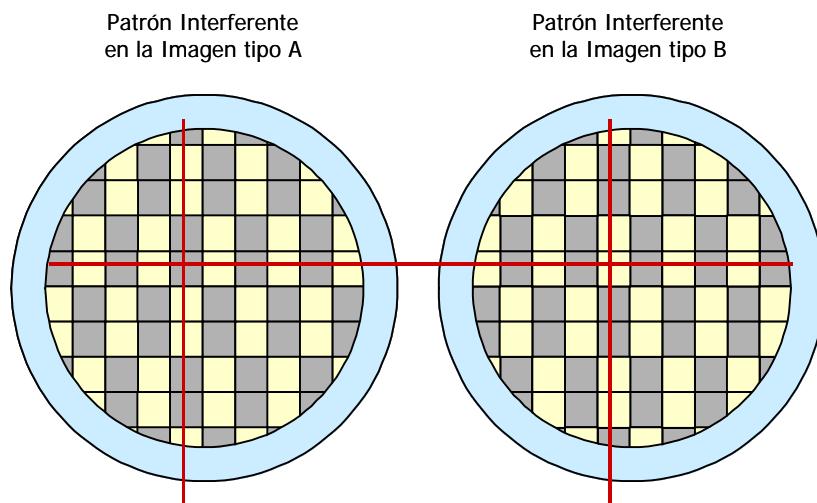


Fig. 3.15 Variación temporal del patrón interferente entre imágenes consecutivas

Nuevamente, podríamos plantearnos qué hubiera ocurrido con la evolución temporal del patrón de interferencias en el supuesto de que la frecuencia portadora hubiera sido un múltiplo de la frecuencia de línea. El resultado es evidente, ya que el número de ciclos de la portadora de color en una línea hubiera sido un entero perfecto, de manera que el patrón permanecería estacionario en el tiempo aumentando considerablemente su visibilidad.

Otro aspecto que se deriva directamente de esta discusión es que el hecho de que cualquier frecuencia que no sea un múltiplo impar de la mitad de la frecuencia de línea generará un patrón interferente que resultará más molesto para el espectador. En efecto, la frecuencia seleccionada permite que tanto variaciones en la dirección vertical como en la temporal muestren la mayor tasa de cambios posible con lo que facilitan la integración de la imagen interferente y minimizan su visibilidad. Cualquier otro valor de la frecuencia podría provocar la aparición de patrones estacionarios o que se desplazaran muy lentamente sobre la pantalla, con lo que su visibilidad sería mayor. De hecho, en el siguiente apartado veremos que en el sistema NTSC tuvo que modificarse ligeramente la frecuencia de imagen respecto a la utilizada en los sistemas en blanco y negro para provocar que la posible intermodulación entre la portadora de color y de sonido tuviera una frecuencia que resultara poco molesta cuando se visualiza en la pantalla.

3.3.3 Modificación de la frecuencia de imagen

Al diseñar el sistema NTSC tuvieron que tenerse en cuenta los diversos factores que podrían generar interferencias en la imagen por el hecho de añadir una nueva componente que soportaba la información de color. Uno de estos factores es que, en el sistema en blanco y negro sobre el que se basó el NTSC, la señal de sonido se transmitía en frecuencia modulada sobre una portadora situada en los 4.5 MHz. Los parámetros de esta modulación del sonido no podían modificarse en lo más mínimo, puesto que cualquier cambio sobre la frecuencia portadora hubiera significado importantes degradaciones sobre las señales demoduladas por los receptores en blanco y negro ya existentes.

En principio, la portadora de sonido está fuera de la banda de vídeo, por lo que parece que si la separación entre las dos señales que se realiza en el receptor es correcta, no debería aparecer ningún problema. Sin embargo, al diseñar el sistema debe tenerse en cuenta que no todos los receptores serán ideales y que es posible que algunos sistemas de separación de vídeo y audio muestren algún tipo de comportamiento no lineal.

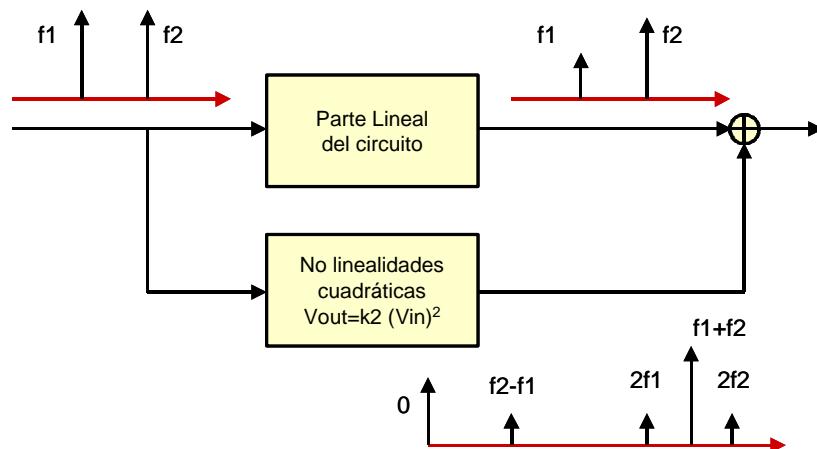


Fig. 3.16 Aparición de componentes de intermodulación debidas a las no linealidades del circuito

La presencia de no linealidades en alguno de los circuitos puede provocar la aparición de nuevas componentes de señal, que aparecen en el mismo receptor y que pueden producir interferencias sobre las componentes deseadas. La situación se ilustra en la figura 3.16, donde se representa un sistema que incorpora una no linealidad de tipo cuadrático en su salida. Si en la entrada están presentes dos componentes sinusoidales de frecuencias distintas, en la salida de la parte lineal del circuito siguen manteniéndose estas dos componentes, probablemente con amplitudes y fases distintas, pero con la misma frecuencia. Sin embargo, en la salida de la parte cuadrática del circuito aparecen señales que no se corresponden directamente con las frecuencias de entrada, sino con las distintas combinaciones suma y diferencia entre las dos frecuencias.

Para el ejemplo que hemos considerado, puede entenderse fácilmente este resultado si desarrollamos la salida que se obtiene cuando en la entrada aparecen dos componentes sinusoidales:

$$\begin{aligned} k_2(A_1 \cdot \sin(\omega_1 t) + A_2 \cdot \sin(\omega_2 t))^2 = \\ k_2 \left(\frac{1}{2} \cdot (A_1^2 + A_2^2) - \frac{1}{2} \cdot (A_1^2 \cdot \cos(2\omega_1 t) + A_2^2 \cdot \cos(2\omega_2 t)) \right) + \\ + k_2 (A_1 \cdot A_2 (\cos((\omega_1 - \omega_2)t) - \cos((\omega_1 + \omega_2)t))) \end{aligned} \quad (3.23)$$

La señal de salida está, pues, formada por la superposición de componentes continuas, componentes de frecuencia doble y componentes cuya frecuencia se obtiene como la suma y la diferencia de las frecuencias de las señales de entrada. El nivel de estas señales depende del grado de no linealidad de los circuitos que queda representado por la amplitud de la constante k_2 . Si se consideraran no linealidades de orden superior, los resultados obtenidos serían parecidos.

Es evidente que estas no linealidades serán más pronunciadas en algunos receptores que en otros y que en principio no se trata de un problema del sistema, sino del receptor. No obstante, al diseñar el sistema de televisión deben tenerse en cuenta todas las posibilidades para evitar, en lo posible, la presencia de componentes indeseadas o en el caso en que sea imposible evitar su presencia, intentar que su efecto sobre la calidad de la imagen sea lo menos perceptible posible.

De todas las componentes obtenidas en la salida de la no linealidad de segundo orden, la más perjudicial para el sistema de televisión es la que se obtiene como la diferencia entre la frecuencia portadora del sonido y la de la croma, ya que su valor cae dentro del ancho de banda de la señal de vídeo y por lo tanto podría hacerse visible en la pantalla.

Esta diferencia de frecuencias, si mantenemos las 30 imágenes por segundo que se utilizaban en el precursor en blanco y negro del sistema NTSC, tomaría el valor:

$$f_{dif} = f_{audio} - f_{color} = 4.5MHz - 227,5 \times 30 \times 525 = 916875Hz \quad (3.24)$$

Si expresamos esta frecuencia en función de la frecuencia de línea del sistema obtenemos:

$$f_{dif} = \frac{f_{dif}}{f_{Línea}} \cdot f_{Línea} = 58,2143 \cdot f_{Línea} \quad (3.25)$$

valor que en principio debería ser corregido si se desea evitar que esta interferencia sea visible en la pantalla. Para ello, es necesario modificar ligeramente alguno de los parámetros del sistema de televisión para conseguir que la interferencia sea lo menos molesta posible, es decir, que pueda expresarse como un múltiplo impar de la mitad de la frecuencia de línea. Teniendo en cuenta el resultado de la ecuación 3.25, el objetivo es que la diferencia entre la frecuencia de la portadora de audio y la de color sea 58,5 veces la frecuencia de línea.

Imponiendo esta condición sobre la ecuación 3.24 obtenemos:

$$f_{dif} = f_{audio} - 227,5 \times f_l \times N_v = 58,5 \cdot f_l \times N_v \quad (3.26)$$

donde f_l representa la frecuencia de imagen y N_v el número de líneas del sistema. De todos estos parámetros, el único que en principio puede ser modificado sin alterar la compatibilidad es la frecuencia de imagen. Recuérdese que la frecuencia de imagen se había elegido de 30 Hz para poder usar la red eléctrica como señal patrón para sincronizar las distintas cámaras de un estudio de televisión en los albores de la televisión. En el momento en que se definió la televisión en color, este procedimiento de sincronismo ya se había abandonado completamente. Además, los receptores regeneran siempre la señal de barrido vertical a partir de los sincronismos de campo, por lo que la frecuencia de imagen puede modificarse sin que afecte el correcto sincronismo con la señal recibida.

El valor de la nueva frecuencia de imagen es:

$$f_l = \frac{f_{audio}}{(58,5 + 227,5) \times N_v} = 29,97 \text{ imágenes/s} \quad (3.27)$$

valor que evidentemente no supone ninguna variación observable por el espectador.

Con esta nueva frecuencia de imagen se modifican alguno de los parámetros básicos del sistema de televisión. La nueva frecuencia y periodo de línea toman los valores:

$$f_{Línea} = f_l \times N_v = 15734,26 \text{ Hz}; \quad T_{Línea} = 63,556 \mu\text{s} \quad (3.28)$$

y la frecuencia de la portadora de color será:

$$f_{color} = 227,5 \times f_{Línea} = 3,579545 \text{ MHz} \quad (3.29)$$

Aunque el sistema NTSC admite sólo una desviación de ± 10 Hz sobre esta frecuencia es habitual en la literatura referirse a este valor como 3,58 MHz, sin dar cuenta de todos los decimales. También, al referirnos a la frecuencia de imagen o de campo, es habitual seguir utilizando, por simplicidad, los guarismos de 30 y 60 Hz.

Es fundamental que independientemente de la precisión con la que se consiga obtener la frecuencia portadora de la información de color, su valor coincida exactamente con el valor de la frecuencia de línea multiplicada por el factor 227,5. Para ello, en el transmisor se utiliza un único generador, del que

se obtienen, por división de frecuencias, las dos señales. El esquema utilizado habitualmente se representa en la figura 3.17, donde sólo se indica cómo obtener una señal con una frecuencia doble a la de línea (que indica la cadencia de los sincronismos en las primeras líneas de la parte de borrado de campo). El procedimiento esbozado consiste en dividir la portadora de color por $5 \times 7 \times 13 = 455$, con lo que obtenemos una señal cuya frecuencia es la mitad que la teórica frecuencia de línea. Nótese que no es posible obtener directamente la frecuencia de línea, ya que sólo pueden implementarse divisores de frecuencia en factores enteros. Esta señal se introduce a un comparador de fase cuya salida actúa sobre un VCO. La señal de salida de este oscilador se divide por un factor 4 y se introduce en el comparador de fase. Cuando el circuito se estabiliza, la frecuencia de las señales en la entrada del comparador de fase será la misma por lo que la salida del VCO tendrá una frecuencia doble a la de línea. A partir de esta señal, es posible obtener el resto de señales que habrán de incorporarse como sincronismos a la señal de vídeo. Así, la propia frecuencia de línea se obtendría dividiendo por 2 la frecuencia de esta señal y la frecuencia de campo dividiéndola por $525 = 3 \times 7 \times 5 \times 5$.

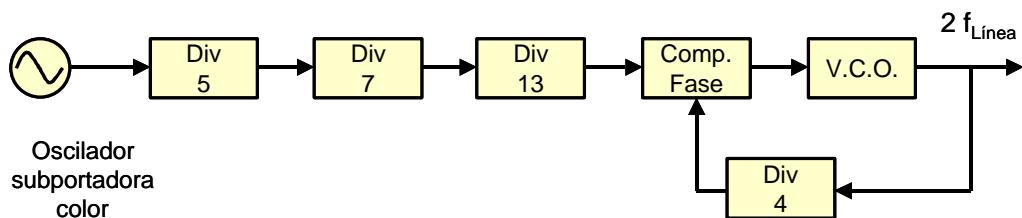


Fig. 3.17 Obtención de la frecuencia de línea a partir de la portadora de color en el transmisor

3.3.4 Demodulación de las componentes I y Q: caso ideal

En este y los siguientes apartados presentaremos los sistemas y principios utilizados para la recuperación de las señales diferencia de color I y Q a partir de la señal de croma. Estos demoduladores constituyen la parte esencial de la decodificación de la información en el receptor, por lo que comprender su principio de funcionamiento es crucial tanto para tener una visión general del sistema como para entender las distintas componentes que conforman la señal de un sistema de color compatible.

En una primera aproximación al problema supondremos que la señal de croma se corresponde exactamente con la ecuación 3.2, es decir, por el momento ignoraremos que la parte alta del espectro de la señal I es filtrada y que esta componente se transmite con banda lateral superior vestigial. Una vez desarrollada esta aproximación al problema, se tratará el caso real, comprobando que puede aplicarse el mismo esquema de demodulación. Además, veremos que una correcta demodulación de las señales exige que el receptor disponga de un oscilador en fase con la portadora de color utilizada en el transmisor. Esto exigirá transmitir una referencia de frecuencia y fase de la portadora de color que se incorporará a la señal de vídeo y cuyos detalles expondremos también en este apartado. Los principios de demodulación de las señales diferencia de color expuestos en esta y las siguientes secciones son comunes tanto para el decodificador de NTSC como en el de PAL. Por ello, los resultados obtenidos se utilizarán en la discusión del PAL sin necesidad de justificarlos.

Considérese el esquema de la figura 3.18 que se aplica sobre la señal de croma del sistema NTSC, cuya expresión repetimos aquí por claridad:

$$\text{croma}_{\text{NTSC}}(t) = Q'(t) \cdot \sin(2\pi f_c t + 33^\circ) + I'(t) \cdot \cos(2\pi f_c t + 33^\circ) \quad (3.30)$$

donde usamos el símbolo f_c para referirnos de forma abreviada a la frecuencia portadora de la información de croma en el NTSC. Los dos canales del demodulador son parecidos. La señal de croma se mezcla con portadoras generadas en el propio receptor, que, para un correcto funcionamiento del sistema, deberán estar en fase con la portadora de la señal recibida, y el resultado se pasa a través de filtros adaptados al ancho de banda de las señales diferencia de color.

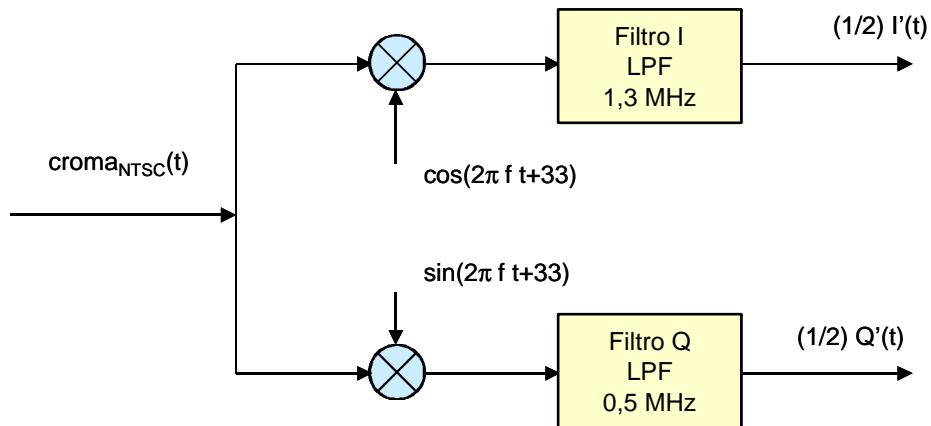


Fig. 3.18 Esquema básico de un demodulador de croma

El resultado del producto de la señal de croma por la portadora en coseno de la rama superior del demodulador será:

$$\begin{aligned} & \text{croma}_{\text{NTSC}}(t) \times \cos(2\pi f_c t + 33^\circ) = \\ & = \frac{1}{2} \cdot I'(t) + \frac{1}{2} \cdot Q'(t) \cdot \sin(4\pi f_c t + 66^\circ) + \frac{1}{2} \cdot I'(t) \cdot \cos(4\pi f_c t + 66^\circ) \end{aligned} \quad (3.31)$$

lo que establece que la señal $I'(t)$ ha sido trasladada nuevamente al origen, por lo que podremos recuperarla después del filtrado paso bajo, adaptado al ancho de banda de esta señal, que se indica en la figura. Nótese que el resto de componentes que aparecen en la salida del mezclador están centradas a una frecuencia doble de la portadora de vídeo, por lo que serán eliminadas por el filtro paso bajo.

Análogamente, el resultado del producto de la señal de croma por la portadora en seno da como resultado:

$$\begin{aligned} & \text{croma}_{\text{NTSC}}(t) * \sin(2\pi f_c t + 33^\circ) = \\ & = \frac{1}{2} \cdot Q'(t) - \frac{1}{2} \cdot Q'(t) \cdot \cos(4\pi f_c t + 66^\circ) + \frac{1}{2} \cdot I'(t) \cdot \sin(4\pi f_c t + 66^\circ) \end{aligned} \quad (3.32)$$

que también indica que en la salida de la rama inferior del demodulador dispondremos de la señal diferencia de color $Q'(t)$. Así pues, en este caso ideal el demodulador recupera perfectamente las dos señales diferencia de color.

3.3.5 Demodulación con errores de fase

Consideremos ahora el efecto de un error de fase α en las señales generadas en el propio receptor. Para el caso de la componente en fase obtenemos:

$$\begin{aligned} croma_{NTSC}(t) * \cos(2\pi f_c t + 33^\circ + \alpha) = \\ = \frac{1}{2} \cdot I'(t) \cdot \cos \alpha - \frac{1}{2} \cdot Q'(t) \cdot \sin \alpha + \\ + \frac{1}{2} \cdot Q'(t) \cdot \sin(4\pi f_c t + 66^\circ + \alpha) + \frac{1}{2} \cdot I'(t) \cdot \cos(4\pi f_c t + 66^\circ + \alpha) \end{aligned} \quad (3.33)$$

De acuerdo con este resultado, existen dos componentes que siguen manteniéndose centradas en una frecuencia doble a la de portadora y que por lo tanto serán rechazadas por el filtro paso bajo. Sin embargo, en la región de baja frecuencia aparece tanto la componente $I'(t)$ (la deseada) como la $Q'(t)$ (interferencia), por lo que ambas señales pasarán a través del filtro. Se dice en este caso que existe cruce de color en el decodificador, ya que la señal que en principio debía corresponder únicamente a la componente $I'(t)$ contiene parte de información de la componente en cuadratura. Nótese que estas señales están ponderadas por el coseno y el seno del error de fase, lo que indica que el cruce de color es tanto más importante cuánto mayor sea este error.

El resultado obtenido para el canal en cuadratura es parecido:

$$\begin{aligned} croma_{NTSC}(t) \times \sin(2\pi f_c t + 33^\circ + \alpha) = \\ = \frac{1}{2} \cdot Q'(t) \cdot \cos \alpha + \frac{1}{2} \cdot I'(t) \cdot \sin \alpha + \\ + \frac{1}{2} \cdot Q'(t) \cdot \cos(4\pi f_c t + 66^\circ + \alpha) + \frac{1}{2} \cdot I'(t) \cdot \sin(4\pi f_c t + 66^\circ + \alpha) \end{aligned} \quad (3.34)$$

Ahora bien, sólo una fracción de la energía, correspondiente a la región de baja frecuencia de la señal $I'(t)$, pasará a través del filtro paso bajo, ya que el ancho de banda de esta componente es superior a la del filtro. No obstante, sigue apareciendo cruce entre las dos componentes de color.

La degradación de la señal debida a estos errores de fase es muy grave, ya que podría modificar todos los matices de color de la escena. Por ello, debe controlarse con sumo cuidado que los generadores de las señales utilizadas para la demodulación estén perfectamente sincronizados con la portadora de croma utilizada en el transmisor. Con este objetivo, se decidió incorporar una pequeña réplica de la portadora de color utilizada en el transmisor para que el receptor pudiera usarla como referencia y sincronizarse perfectamente con ella. Esta réplica de la portadora de color se denomina *salva* y se transmite en cada línea de la señal, ubicándose en una parte libre del pórtico posterior del sincronismo de línea.

3.3.6 Inserción de la salva de la portadora de color

La señal de salva se introduce en la parte correspondiente al pórtico posterior de la señal de televisión y se utiliza para informar al receptor sobre la frecuencia exacta y la fase que se está utilizando en el transmisor como portadora de la información de color. Tal y como se indica en la figura 3.19, está situada a una distancia de $5.3 \mu s$ del impulso anterior del sincronismo de línea y tiene una duración aproximada de $2.5 \mu s$, durante los que se transmiten aproximadamente unos 9 ciclos de la portadora de color. La amplitud es de un 40 % respecto al nivel de blanco, que, si éste se normaliza a 0.7 voltios, corresponderá a una tensión de unos 0.28 voltios. Este nivel de amplitud es aproximadamente igual al nivel del impulso de sincronismo y se utiliza en el receptor para ajustar los sistemas de control automático de ganancia. Con ello es posible obtener un correcto equilibrado de las ganancias del canal de luminancia y del de croma. Las tolerancias respecto a estos valores teóricos se muestran en la misma figura y son, como puede comprobarse, bastante elevados.

La introducción de la salva se realiza mediante puertas de transmisión que se abren durante el intervalo de tiempo correspondiente y permiten introducir una copia, con la amplitud y fase adecuada, de la señal del oscilador de croma en la señal de vídeo. Veremos algunos esquemas a nivel de diagramas de bloques cuando analicemos el codificador NTSC.

La parte de la señal de vídeo en la que se introduce la salva de color corresponde a la parte de línea no activa y por lo tanto es una región en la que no se transmite información ni de luminancia ni de color y durante la que los circuitos de barrido horizontal están realizando el retorno hacia el extremo izquierdo de la pantalla. La mayoría de receptores actuales inhiben las señales que se aplican al tubo de imagen durante este intervalo de tiempo para garantizar que el retorno del haz se realiza sin activar ningún fósforo de la pantalla y que por lo tanto es invisible para el espectador.

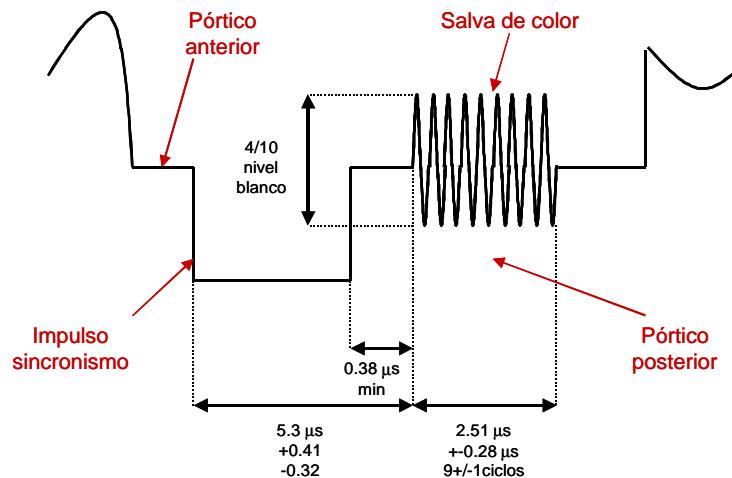


Fig. 3.19 Salva de color para recuperar la información de fase de la portadora

No obstante, en la señal de televisión está prevista cualquier eventualidad y, como ya debe parecer obvio a estas alturas, también se contempló la posibilidad de que algunos fabricantes no inhibieran las señales que se aplican al tubo de imagen durante el tiempo de línea no activa. En este supuesto, como los demoduladores continúan operando, descodificarán la salva y obtendrán las señales diferencia de color asociadas al nivel de amplitud y la fase de esta componente. Si estas señales diferencia de color se aplican junto con la luminancia a la matriz de decodificación RGB, obtendremos las señales que posteriormente se aplicarán al tubo de imagen y que, dependiendo de la amplitud y la fase de la señal de salva, podrán ser más o menos visibles.

Hemos visto que la amplitud de la señal de salva se ha elegido del mismo valor que los impulsos de sincronismo para poder equilibrar fácilmente las ganancias de los canales de luminancia y de croma. Por tanto, el único parámetro que puede ser elegido libremente es el desfasaje relativo entre la portadora de color y la señal de salva. Este desfasaje puede tomar el valor que se considere más oportuno, siempre que tanto el transmisor como el receptor se pongan de acuerdo en cuál es el valor con el que se va a transmitir. La situación se ilustra en la figura 3.20, donde se muestra que si el desfasaje es conocido, el receptor siempre puede regenerar la salva con la misma fase que en el transmisor.

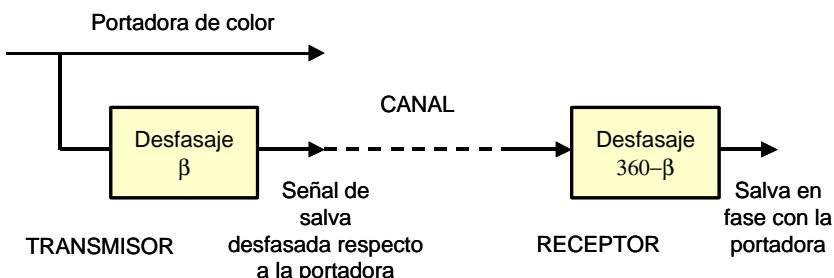


Fig. 3.20 Desfase de la salva de color respecto a la portadora

Para determinar la fase más adecuada deberíamos calcular las señales RGB que se obtendrían si se demodulara la señal de salva. El procedimiento para realizar este cálculo es relativamente sencillo. En efecto, supongamos que la salva tiene una amplitud genérica A y un desfasaje β respecto a la señal patrón $\cos(2\pi f_c t + 33^\circ)$. Las señales que obtendremos en la salida del demodulador de color serán:

$$\begin{aligned} I'(t) &= LPF_{1.3MHz} \left(A \cos(2\pi f_c t + 33^\circ + \beta) \times \cos(2\pi f_c t + 33^\circ) \right) = \frac{A}{2} \cdot \cos(\beta); \\ Q'(t) &= LPF_{0.5MHz} \left(A \cos(2\pi f_c t + 33^\circ + \beta) \times \sin(2\pi f_c t + 33^\circ) \right) = -\frac{A}{2} \cdot \sin(\beta); \end{aligned} \quad (3.35)$$

Evidentemente, la luminancia asociada a la señal de salva será cero, ya que la señal está situada sobre el pótico posterior, que tiene un nivel de negro. Si introducimos los valores de $Y'(t)$, $I'(t)$ y $Q'(t)$ en la matriz de decodificación R,G,B (es decir, la relación inversa a la ecuación 3.10), obtendremos:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 0 \\ \frac{A}{2} \cdot \cos\beta \\ -\frac{A}{2} \cdot \sin\beta \end{bmatrix} \quad (3.36)$$

$$= A \cdot \begin{bmatrix} 0.478\cos\beta - 0.310\sin\beta \\ -0.136\cos\beta + 0.323\sin\beta \\ -0.552\cos\beta - 0.851\sin\beta \end{bmatrix}$$

resultado que muestra la dependencia de las componentes RGB respecto a la amplitud y la fase de la señal de salva. El objetivo sería elegir un valor de fase cuyas componentes RGB fueran lo menos visibles posible. Como la amplitud está prefijada, bastará con ver los valores que toman las tres variables en función del ángulo de desfase.

En la figura 3.21 se han representado las tres componentes R,G,B para valores de la variable β entre 0 y 2π tomando la amplitud A como unitaria. Los valores negativos significan que la señal que se aplicará al tubo de imagen no producirá ninguna energía en el haz resultante, por lo que equivalen a tener una componente de color nula. El ángulo de desfasaje que finalmente se eligió fue el de 147° , para el cual sólo la componente G es positiva. El valor de intensidad del haz G asociado a la señal de salva se sitúa sobre el 16% del valor máximo que puede tomar esta señal. La elección del valor 147° se debe a un doble motivo. Por una parte se trata de un ángulo en el que dos de las componentes se anulan y la otra mantiene un valor reducido, pero, además, es un ángulo muy simple de generar, si se considera que, sumado a los 33° de desfase de la portadora en coseno da un total de 180° , que se podrá implementar con una simple inversión de signo. Adicionalmente, este ángulo de desfasaje coincide con la inversión de la portadora de la señal diferencia de color U, lo que, como veremos, se utiliza en algunos receptores para obtener directamente las componentes U y V a partir de la señal de croma del sistema NTSC.

En los apartados dedicados al codificador y decodificador NTSC veremos algunas de las alternativas más usadas para la inserción de la salva en la señal de vídeo y para la sincronización del oscilador de color del receptor basándose en la información proporcionada por esta señal.

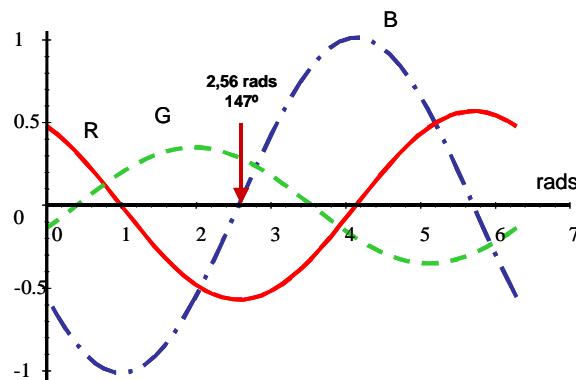


Fig. 3.21 Componentes RGB en función de la fase de la salva

3.3.7 Demodulación de las componentes de color: banda lateral superior vestigial

Hasta ahora hemos analizado el esquema del demodulador de las señales diferencia de color en el supuesto de que las dos componentes se transmitieran en doble banda lateral, comprobando la necesidad de disponer de una referencia de la portadora de color para poder obtener un correcto sincronismo con la portadora de color. En este apartado comprobaremos que el mismo esquema de demodulación puede aplicarse al sistema real, en el que la banda superior de la señal I ha sido recortada a 4.2 MHz para acomodar la señal de vídeo dentro de los límites establecidos por la normativa. El análisis demodulador con esta hipótesis requiere la introducción de la transformada de Hilbert para analizar los efectos que produce la supresión de una de las bandas superiores. El lector puede omitir todo este apartado sin pérdida de continuidad con el resto del texto. El análisis permitirá comprobar que el hecho de recortar la banda superior de la señal I introduce una atenuación de 3 dB's en la parte de alta frecuencia de esta señal y puede provocar, si los filtros del canal Q no están perfectamente ajustados, la aparición de un pequeño cruce de componentes de color.

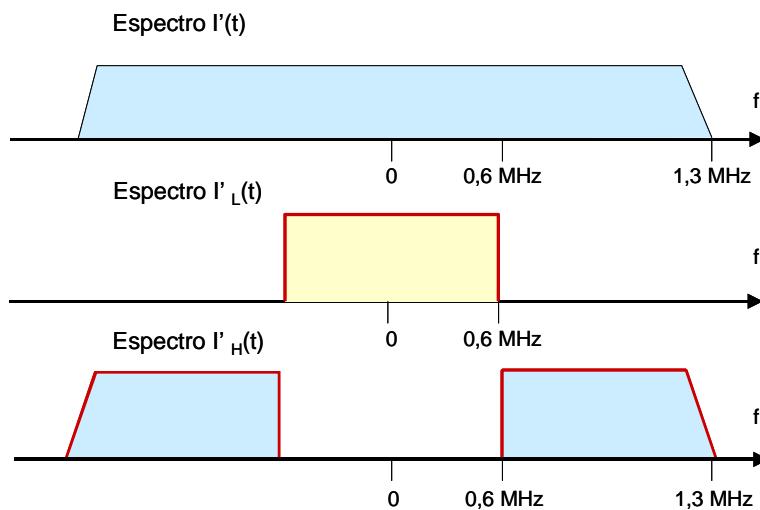


Fig. 3.22 Descomposición de la señal $I'(t)$ en componentes de baja y alta frecuencia

El primer objetivo, para poder evaluar el tratamiento de señal que realiza el demodulador, es obtener una expresión analítica de la señal $I'(t)\cos(2\pi f_c t + 33^\circ)$ filtrada. Para ello es conveniente descomponer la señal en banda base $I'(t)$ en dos partes de alta y baja frecuencia, tal y como se esboza en la figura 3.22. De esta forma $I'(t)$ puede expresarse como la suma de estas dos componentes

$$I'(t) = I'_L(t) + I'_H(t) \quad (3.37)$$

Los anchos de banda de estas dos componentes han sido ajustados para que, al modular la señal y eliminar la banda superior de $I'_H(t)$, obtengamos la versión en banda lateral superior vestigial que se utiliza en el NTSC. El resultado final al que queremos llegar se representa en la figura 3.23, donde suponemos que los filtros utilizados para recortar la banda superior son ideales. Los factores de

amplitud $\frac{1}{2}\exp(\pm j(\pi/180)33)$ existentes en cada una de las bandas se deben a la modulación por el coseno. Téngase en cuenta que la transformada de Fourier de $I'(t)\cos(2\pi f_c t + 33^\circ)$ viene dada por:

$$F\{I'(t) \cdot \cos(2\pi f_c t + 33^\circ)\} = \frac{1}{2} \left\{ I'(f - f_c) \cdot e^{j\frac{\pi}{180} \cdot 33} + I'(f + f_c) \cdot e^{-j\frac{\pi}{180} \cdot 33} \right\} \quad (3.38)$$

donde $I'(f)$ es la transformada de Fourier de la señal $I'(t)$ en banda base. Sobre este espectro se realiza el filtrado de la banda lateral que supera los 4,2 MHz.

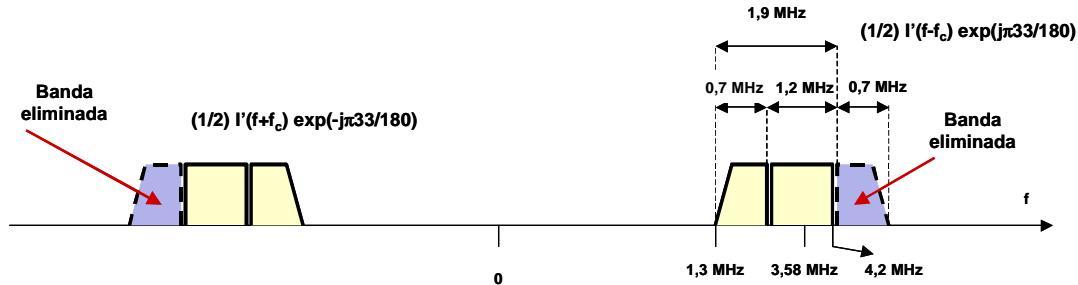


Fig. 3.23 Transformada de Fourier de la señal $I'(t)$ modulada y filtrada

Para obtener la expresión matemática en el dominio temporal de la señal modulada y filtrada, es necesario recurrir a la transformada de Hilbert de la componente $I'_H(t)$. Esta transformada se define, en el dominio frecuencial, como el resultado de filtrar la señal mediante un filtro $H_{Hilbert}(f)$ cuya respuesta en frecuencia se representa en la figura 3.24 y que responde a la expresión:

$$H_{Hilbert}(f) = \begin{cases} -j & \text{si } f > 0 \\ j & \text{si } f < 0 \end{cases} \quad (3.39)$$

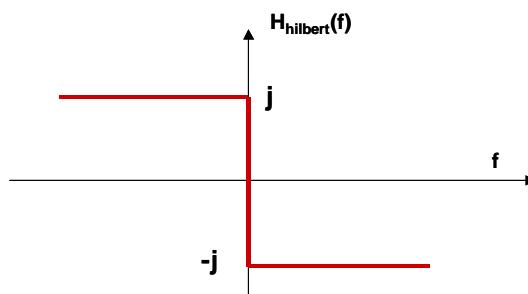


Fig. 3.24 Respuesta en frecuencia de un transformador de Hilbert

El resultado pasar la señal $I'_H(t)$ a través de este filtro puede expresarse como:

$$\hat{I}_H(t) = F^{-1}(I_H(f) \cdot H_{Hilbert}(f)) \quad (3.40)$$

donde F^{-1} representa la transformada inversa de Fourier e $I_H(f)$ la transformada de la componente $I_H(t)$.

Alternativamente, la transformada de Hilbert $\hat{I}_H(t)$ puede expresarse en el dominio temporal como la convolución entre la señal $I'_H(t)$ con un sistema cuya respuesta impulsional es $1/(\pi t)$. Así:

$$\hat{I}_H(t) = I'_H(t) * \left(\frac{1}{\pi t} \right) = \int_{-\infty}^{\infty} I'_H(\tau) \cdot \frac{1}{\pi(t-\tau)} d\tau \quad (3.41)$$

Esta expresión deja claro que la transformada de Hilbert de una señal real será también una señal real, ya que todas las funciones que aparecen dentro del símbolo integral son reales. La transformada de Hilbert de una señal sólo puede ser implementada de forma aproximada, ya que el tipo de filtrado involucrado es ideal y por lo tanto irrealizable. Sin embargo, en nuestro caso únicamente nos interesa la representación matemática de la señal y no los valores que toma.

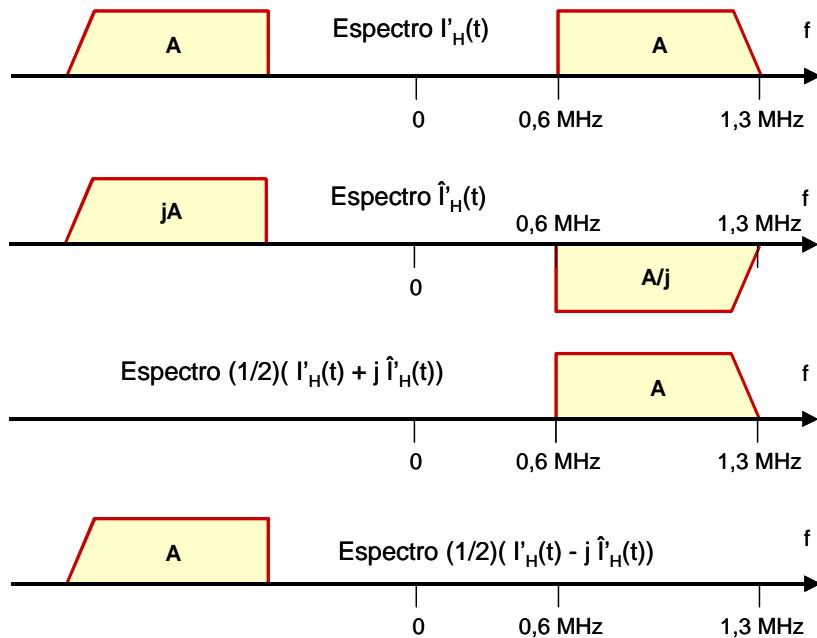


Fig. 3.25 Obtención de las bandas superiores de la señal en función de la transformada de Hilbert

En la figura 3.25 se representan las transformadas de Fourier de las señales $I'_H(t)$, $\hat{I}_H(t)$ y de las combinaciones $\frac{1}{2}I'_H(t)+j\hat{I}_H(t)$ y $\frac{1}{2}I'_H(t)-j\hat{I}_H(t)$, donde j representa el número complejo $j^2=-1$. Los espectros de estas dos últimas señales son particularmente interesantes, ya que, como puede

comprobarse gráficamente, se corresponden con las partes de la banda lateral de la señal $I'(t)$ que serán eliminadas durante el posterior filtrado de la señal modulada. Restando estas señales de $I'(t)$ podemos construir las funciones $L_1(t)$ y $L_2(t)$:

$$\begin{aligned} L_1(t) &= I'(t) - \frac{1}{2}(I'_H(t) + j\hat{I}_H(t)), \\ L_2(t) &= I'(t) - \frac{1}{2}(I'_H(t) - j\hat{I}_H(t)), \end{aligned} \quad (3.42)$$

cuyas transformadas de Fourier se representan en la figura 3.26 y no son más que la versión en banda base de las partes positiva y negativa del espectro representado en la figura 3.23. Por lo tanto, para obtener la expresión matemática de la versión modulada y filtrada de la componente de color $I'(t)$, bastará con desplazar el espectro de $L_1(t)$ a la frecuencia f_c y el de $L_2(t)$ a la frecuencia $-f_c$, escalando las amplitudes de ambos adecuadamente. Si denominamos $L(t)$ a la componente $I'(t)$ una vez modulada y filtrada, obtenemos la siguiente relación:

$$L(t) = L_1(t) \cdot \frac{1}{2} \cdot \exp\left(j2\pi f_c t + \frac{\pi}{180} 33^\circ\right) + L_2(t) \cdot \frac{1}{2} \cdot \exp\left(-j2\pi f_c t - \frac{\pi}{180} 33^\circ\right) \quad (3.43)$$

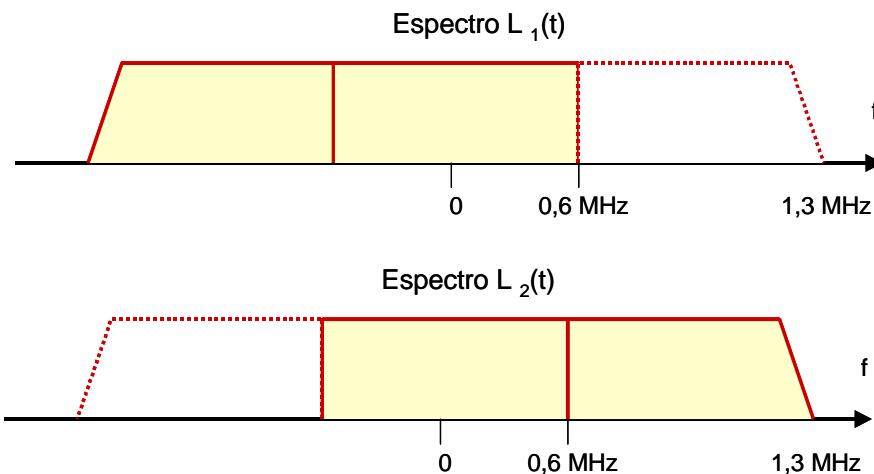


Fig. 3.26 Versiones en banda base de la parte positiva y negativa del espectro de la señal

Si sustituimos los valores de $L_1(t)$ y $L_2(t)$ en función de $I'(t)$, $I'_H(t)$ y la transformada de Hilbert de esta última, obtenemos

$$L(t) = \left(I'(t) - \frac{1}{2} \cdot I'_H(t)\right) \cdot \cos(2\pi f_c t + 33^\circ) + \frac{1}{2} \cdot \hat{I}'_H(t) \cdot \sin(2\pi f_c t + 33^\circ) \quad (3.44)$$

de modo que la señal de croma completa que se utiliza en el sistema NTSC podrá expresarse como la suma de la componente en fase obtenida en la ecuación anterior más la componente en cuadratura:

$$\begin{aligned} croma_{NTSC}(t) = & \left(I'(t) - \frac{1}{2} \cdot I'_H(t) \right) \cdot \cos(2\pi f_c t + 33^\circ) + \\ & + \frac{1}{2} \cdot \hat{I}'_H(t) \cdot \sin(2\pi f_c t + 33^\circ) + Q'(t) \cdot \sin(2\pi f_c t + 33^\circ) \end{aligned} \quad (3.45)$$

Si multiplicamos esta señal por la componente en fase de la parte superior del demodulador de color obtendremos:

$$\begin{aligned} croma_{NTSC}(t) \cdot \cos(2\pi f_c t + 33^\circ) = & \\ = & \frac{1}{2} \cdot \left(I'(t) - \frac{1}{2} \cdot I'_H(t) \right) + \frac{1}{2} \cdot \left(Q'(t) + \frac{1}{2} \cdot \hat{I}'_H(t) \right) \cdot \sin(4\pi f_c t + 66^\circ) + \\ & + \frac{1}{2} \cdot \left(I'(t) - \frac{1}{2} \cdot I'_H(t) \right) \cdot \cos(4\pi f_c t + 66^\circ) \end{aligned} \quad (3.46)$$

Finalmente, la señal en la salida del filtro paso bajo será:

$$\frac{1}{2} \cdot \left(I'(t) - \frac{1}{2} \cdot I'_H(t) \right) = \frac{1}{2} \cdot \left(I'_L(t) + \frac{1}{2} \cdot I'_H(t) \right) \quad (3.47)$$

Es decir, el hecho de filtrar la banda lateral superior de la componente $I'(t)$ introduce una atenuación en las componentes de alta frecuencia de la versión de $I'(t)$ que recuperamos en el decodificador. Evidentemente, esta atenuación es conocida a priori, por lo que siempre puede corregirse en el propio receptor amplificando la región del espectro que está atenuada. En la figura 3.27 se representa aproximadamente el espectro de la señal demodulada y se sugiere la forma que debería tener el filtro paso bajo del decodificador para ecualizarla completamente. En la práctica, sin embargo, los receptores que decodifican las señales $I'(t)$ y $Q'(t)$ no suelen utilizar esta compensación de frecuencia, aplicando directamente la señal de la ecuación 3.47 a la matriz de decodificación RGB. Con ello, se produce un pequeño cambio sobre el color real en estas componentes de alta frecuencia. La variación es poco perceptible, ya que, al tratarse de componentes de alta frecuencia, suelen coincidir con los contornos de la imagen, donde el cambio de matiz es mucho menos importante que el cambio en la señal de luminancia. Las regiones uniformes corresponden a componentes de baja frecuencia en las que el color quedará correctamente decodificado.

Espectro componente en fase decodificada

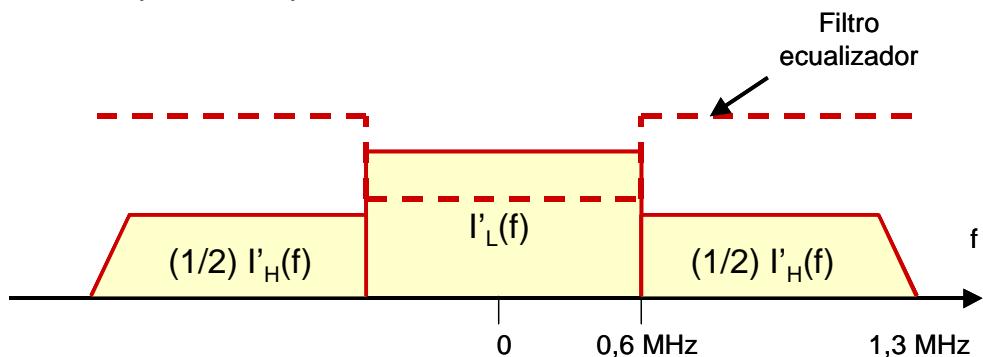


Fig. 3.27 Espectro de la componente en fase y filtro ecualizador

Por otra parte, el resultado que obtenemos al multiplicar la señal de croma por la componente en cuadratura es:

$$\begin{aligned} & \text{croma}_{NTSC}(t) \times \sin(2\pi f_c t + 33^\circ) = \\ & = \frac{1}{2} \cdot (Q'(t) + \frac{1}{2} \cdot \hat{I}_H(t)) - \frac{1}{2} \cdot (Q'(t) + \frac{1}{2} \cdot \hat{I}'_H(t)) \cdot \cos(4\pi f_c t + 66^\circ) + \\ & + \frac{1}{2} \cdot (I'(t) - \frac{1}{2} \cdot I'_H(t)) \cdot \sin(4\pi f_c t + 66^\circ) \end{aligned} \quad (3.48)$$

En la parte de banda base recuperamos la señal $Q'(t)$ más la transformada de Hilbert de $I'_H(t)$. No obstante, esta última señal es la parte de alta frecuencia de la señal $I'(t)$ (componentes superiores a 0.6 MHz), por lo que será eliminada por el filtro paso bajo del decodificador de la componente $Q'(t)$. Así pues, este filtro eliminará la presencia de componentes interferentes debidas al filtrado de la banda superior de $I'(t)$ en el canal en cuadratura. Sin embargo, debe tenerse en cuenta que, en un caso real, si la banda de transición del filtro no es suficientemente abrupta, pueden aparecer componentes de la señal $\hat{I}'_H(t)$ en el canal $Q'(t)$, deteriorando parte de la información de color. Este problema se conoce con el nombre de cruce de color (*cross-color*) ya que su efecto es la aparición, en el canal en cuadratura, de una pequeña interferencia altamente correlada con la componente en fase.

3.3.8 Demodulación directa de las componentes U y V

La demodulación de las señales diferencia de color sobre los ejes $I'(t)$ y $Q'(t)$ que se ha descrito en el apartado anterior sólo se utilizó en los primeros receptores NTSC. La ventaja de este tipo de demodulación es que permite recuperar la señal $I'(t)$ con el mismo ancho de banda con que ha sido transmitida y que el cruce entre las dos componentes de color es prácticamente nulo. Sin embargo, el problema de este sistema de demodulación es que los filtros de las componentes $I'(t)$ y $Q'(t)$ tienen anchos de banda muy distintos y, en consecuencia, los retardos que se introducen en las dos señales, son también muy diferentes. Eliminar esta descompensación de los dos canales requiere introducir una unidad de retardo sobre la señal $I'(t)$ para sincronizarla con la señal $Q'(t)$, lo que, juntamente con la unidad de retardo, que también debe aplicarse al canal de luminancia, encarece y complica notablemente el diseño del receptor. Además la matriz de decodificación de las componentes R', G' y B' a partir de las señales Y' , I' y Q' es algo más compleja que la matriz asociada a las componentes Y', U' y V' .

Por todo ello, la mayoría de los fabricantes prefieren realizar una demodulación directa de las señales sobre los ejes U' y V' . Este tipo de demodulación supone una reducción del ancho de banda global con el que se recuperan las señales diferencia de color. No obstante, la simplificación y reducción de costes en el receptor justifican esta pérdida de resolución en la señal de croma.

En la figura 3.28 se ha representado el diagrama de bloques de un demodulador de las componentes $U'(t)$ y $V'(t)$, observándose que la diferencia más notable respecto al demodulador convencional es que las señales que se mezclan con la croma están ahora desfasadas 33° respecto a la portadora de color. Este desfase entre los osciladores del receptor y la portadora de color introduce un giro sobre

los ejes de demodulación de 33° (en el sentido horario), por lo que las señales que recuperaremos son directamente las componentes $U'(t)$ y $V'(t)$.

El análisis del caso ideal en el que no se considera el filtrado de la banda superior de la componente $I'(t)$ resulta trivial. En efecto, la señal obtenida después del mezclador en la rama superior del demodulador será:

$$\begin{aligned} croma_{NTSC}(t) \times \cos(2\pi f_c t) = \\ = \frac{1}{2} \cdot I'(t) \cdot \cos(33^\circ) + \frac{1}{2} \cdot Q'(t) \cdot \sin(33^\circ) + \\ + \frac{1}{2} \cdot Q'(t) \cdot \sin(4\pi f_c t + 33^\circ) + \frac{1}{2} \cdot I'(t) \cdot \cos(4\pi f_c t + 33^\circ) \end{aligned} \quad (3.49)$$

donde las componentes en banda base se corresponden, de acuerdo con las ecuaciones 3.13 y 3.14, con la componente $\frac{1}{2}V'(t)$. Debe tenerse en cuenta, como ya se ha observado anteriormente, que esta relación sólo es estrictamente cierta cuando la escena original contiene todas las componentes de la señal $Q'(t)$ dentro del ancho de banda de transmisión. En caso contrario, al haber filtrado esta señal diferencia de color, las componentes obtenidas no se corresponderán directamente con la señal $V'(t)$, aunque las diferencias de matices no serán apreciadas por el observador.

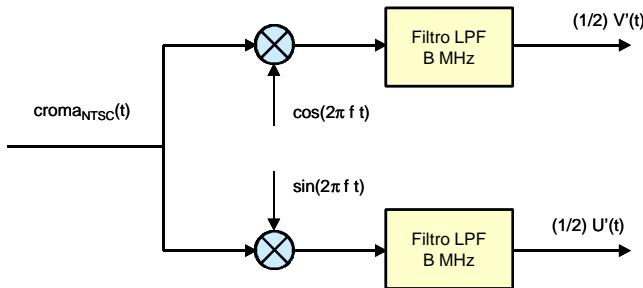


Fig. 3.28 Diagrama de bloques de un demodulador de componentes U' y V'

Análogamente, podríamos calcular la señal obtenida como resultado del producto de la croma por la portadora en seno. La parte de baja frecuencia de este resultado se corresponde con una combinación de las señales $I'(t)$ y $Q'(t)$, que, siempre que apliquen las condiciones comentadas en el párrafo anterior, coincide con la señal $\frac{1}{2}U'(t)$.

El análisis del demodulador, cuando se considera la señal de croma NTSC que incluye el modelo de la señal $I'(t)$ en banda lateral superior vestigial, es algo más complejo. En la rama superior del demodulador, el producto entre la croma y la componente en coseno tiene las siguientes componentes:

$$\begin{aligned} croma_{NTSC}(t) \times \cos(2\pi f_c t) = \\ = \frac{1}{2} \cdot I'(t) \cdot \cos(33^\circ) + \frac{1}{2} \cdot Q'(t) \cdot \sin(33^\circ) + \frac{1}{4} \hat{I}'_H(t) \cdot \sin(33^\circ) - \frac{1}{4} \cdot I'_H(t) \cdot \cos(33^\circ) + \\ + \frac{1}{2} \cdot Q'(t) \cdot \sin(4\pi f_c t + 33^\circ) + \frac{1}{2} \cdot I'(t) \cdot \cos(4\pi f_c t + 33^\circ) + \\ + \frac{1}{4} \cdot \hat{I}'_H(t) \cdot \sin(4\pi f_c t + 33^\circ) - \frac{1}{4} \cdot I'_H(t) \cdot \cos(4\pi f_c t + 33^\circ) \end{aligned} \quad (3.50)$$

En la rama inferior, el resultado del producto entre la portadora en seno y la señal de croma es:

$$\begin{aligned} & \text{croma}_{\text{NTSC}}(t) \times \text{sen}(2\pi f_c t) = \\ & = \frac{1}{2} \cdot Q'(t) \cdot \cos(33^\circ) - \frac{1}{2} \cdot I'(t) \cdot \text{sen}(33^\circ) + \frac{1}{4} \cdot \hat{I}'_H(t) \cdot \cos(33^\circ) + \frac{1}{4} \cdot I'_H(t) \cdot \text{sen}(33^\circ) + \quad (3.51) \\ & + \frac{1}{2} \cdot I'(t) \cdot \text{sen}(4\pi f_c t + 33^\circ) - \frac{1}{2} \cdot Q'(t) \cdot \cos(4\pi f_c t + 33^\circ) - \\ & - \frac{1}{4} \cdot \hat{I}'_H(t) \cdot \cos(4\pi f_c t + 33^\circ) - \frac{1}{4} \cdot I'_H(t) \cdot \text{sen}(4\pi f_c t + 33^\circ) \end{aligned}$$

Si tenemos en cuenta que, para las dos señales, las componentes centradas al doble de la frecuencia portadora serán eliminadas por el filtrado paso bajo, es evidente comprobar que las señales obtenidas se corresponden con las del caso ideal más unas componentes interferentes que dependen de $I'_H(t)$ y su transformada de Hilbert.

Así, para la componente en fase, la señal no deseada viene dada por:

$$\text{Componente no deseada canal fase} = \frac{1}{4} \cdot \hat{I}'_H(t) \cdot \text{sen}(33^\circ) - \frac{1}{4} \cdot I'_H(t) \cdot \cos(33^\circ) \quad (3.52)$$

mientras que para la componente en cuadratura es:

$$\text{Componente no deseada canal cuadratura} = \frac{1}{4} \cdot \hat{I}'_H(t) \cdot \cos(33^\circ) + \frac{1}{4} \cdot I'_H(t) \cdot \text{sen}(33^\circ) \quad (3.53)$$

Como en ambos casos las interferencias dependen de la señal $I'_H(t)$ o de su transformada de Hilbert, pueden ser eliminadas siempre que los filtros paso bajo del demodulador tengan un ancho de banda del orden de los 0,6 MHz. Nótese también que, como recuperamos las señales $U'(t)$ y $V'(t)$, los dos filtros deben tener el mismo ancho de banda, por lo que ya no existen diferencias en el tiempo de retardo que sufre cada uno de los canales.

Evidentemente, la reducción del ancho de banda de las dos señales diferencia de color a 0,6 MHz afecta a la resolución de la señal de color, que queda considerablemente reducida. La solución adoptada en los receptores reales consiste en utilizar filtros paso bajo con un ancho de banda algo superior a los teóricos 0,6 MHz, intentando buscar una solución de compromiso entre la pérdida de resolución en la información de croma y el cruce de componentes de color y nivel de las interferencias. El ancho de banda típico para estos filtros suele situarse en torno a los 0.9 MHz.

3.3.9 Diagrama de bloques de un codificador NTSC

El objeto de este apartado es dar una visión general del esquema de un codificador NTSC, combinando los distintos elementos que constituyen la señal y que ya hemos presentado en secciones previas. El diagrama de bloques genérico del codificador se ha representado en la figura 3.29. La parte correspondiente al tratamiento de las señales R,G,B procedentes de la cámara hasta la obtención de las componentes $I'(t)$ y $Q'(t)$ en banda base ya ha sido presentada anteriormente y se corresponde directamente con el esquema de la figura 3.3. Cabe remarcar que aunque hemos indicado que en el receptor suele realizarse la demodulación directa de las señales $U'(t)$ y $V'(t)$ con objeto de reducir el número de unidades de retardo que deben aplicarse a las señales de luminancia y diferencia de color,

en el codificador, por la propia definición del sistema NTSC, deben mantenerse las dos unidades de retraso para garantizar que todas estas componentes están sincronizadas. Téngase en cuenta que el coste global de un equipo transmisor es muy superior al de un receptor, por lo que este aspecto tiene una repercusión prácticamente nula sobre el sistema.

En este diagrama de bloques se supone que la señal que se genera en el bloque de subportadora de color es directamente la componente en fase $\cos(2\pi f_c t + 33^\circ)$, de modo que esta señal se multiplica directamente por la señal $I'(t)$ en banda base. Para obtener la versión en cuadratura es necesario retardar la subportadora un ángulo $\pi/2$. En la figura se utiliza un signo negativo para indicar explícitamente que se corresponde con un retraso y no con un adelanto de la fase.

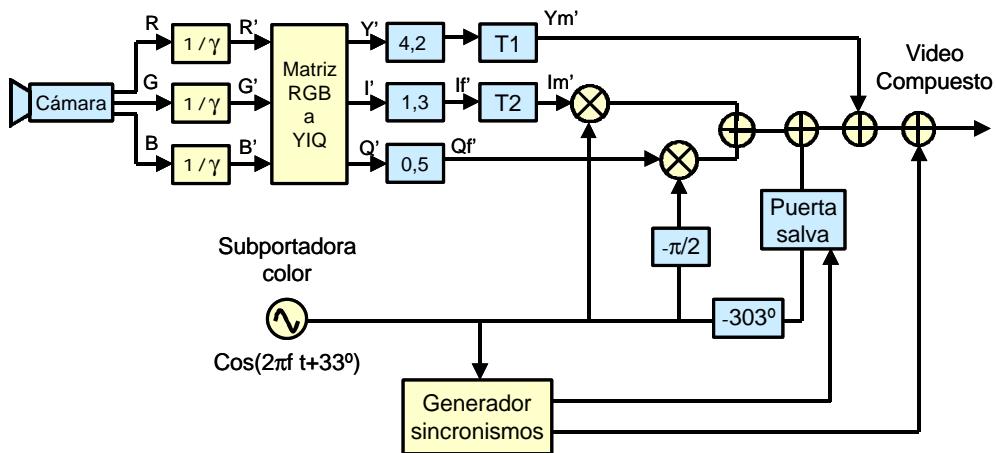


Fig. 3.29 Diagrama de bloques de un codificador NTSC

La señal de salvamento puede obtenerse mediante un retraso de 303° sobre la misma subportadora. Téngase en cuenta que la salvamento debe tener un desfase de 180° respecto al eje correspondiente a la señal $U'(t)$. El desfase de 303° se obtiene como consecuencia de sumar los 90° grados de desfase entre la portadora en coseno y seno, los 33° debidos al giro de las componentes $I'(t)$ y $Q'(t)$ respecto a las componentes $U'(t)$ y $V'(t)$, más los 180° correspondientes a la inversión de la fase de la componente V' . En algunos textos suele aparecer un adelanto de la fase de 57° respecto a la portadora de la componente en fase o un adelanto de 147° respecto a la componente en cuadratura. En este texto hemos preferido indicar los desfases como negativos, debido a que en la práctica sólo pueden realizarse retardos, y no adelantos, de señales.

La subportadora de color se utiliza como señal de referencia para la obtención de los sincronismos. Para ello se utiliza un circuito como el que hemos discutido en la figura 3.17, que obtiene una señal con una frecuencia doble a la de línea y que, evidentemente, está perfectamente sincronizada con la portadora de color. Esta señal es la referencia de fase para determinar las posiciones de todos los pulsos de sincronismo. Así, los impulsos de pre-igualación y post-igualación deben generarse a la misma cadencia que esta referencia, los pulsos de sincronismo de línea tendrán una frecuencia mitad a la de referencia, y los de sincronismo de campo tendrán una cadencia 525 veces menor, por lo que

bastará con dividir la frecuencia de la señal de referencia por 525 para obtener una señal que proporcione la referencia de fase del sincronismo de campo. Una vez obtenidas las señales de referencia para cada tipo de sincronismo, éstas se utilizan para activar los circuitos de conformación de pulsos para cada tipo de sincronismo y se insertan a la señal de luminancia y croma tal y como se indica en el diagrama de bloques. La salva de color se inserta a la señal de croma mediante una puerta que está controlada por el propio generador de sincronismos y que sólo se mantiene activa durante una fracción del tiempo correspondiente al pórtico posterior del sincronismo de línea.

A la señal de vídeo compuesto que se obtiene en la salida del codificador debe aún incorporarse la señal de audio y aplicarla a un modulador que adapte esta señal en banda base para su transmisión a distancia.

El procedimiento utilizado para generar la salva en un codificador real no se corresponde con el que se ha representado en la figura 3.29, ya que en la práctica resultaría complejo ajustar un retraso a 303° con precisión suficiente. La filosofía que normalmente se utiliza es la de aprovechar la propia estructura del modulador de croma mediante un esquema como el de la figura 3.30, de modo que el generador de sincronismos controla cuándo los mezcladores están conectados a las señales diferencia de color $I'(t)$ y $Q'(t)$ o a las constantes para la generación de la salva de color. La estrategia de conmutación entre las dos posibles entradas al modulador de croma es idéntica a la que se realizaba en la puerta de salva.

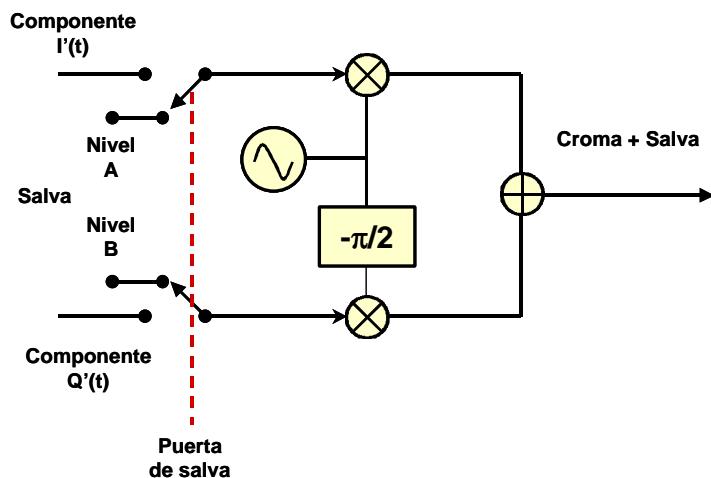


Fig. 3.30 Diagrama de bloques del circuito de inserción de la salva de color

Los niveles A y B están ajustados para que la señal obtenida en la salida del modulador de croma tenga la amplitud y fase de la salva. Suponiendo que trabajamos con excursiones de señal normalizadas a 1 voltio y que la amplitud de la salva es aproximadamente de 0,3 voltios, los valores de A y B pueden calcularse resolviendo la siguiente ecuación:

$$A \cdot \cos(2\pi f_c t + 33^\circ) + B \cdot \sin(2\pi f_c t + 33^\circ) = -0.3 \cdot \sin(2\pi f_c t) \quad (3.54)$$

Si desarrollamos las funciones trigonométricas obtenemos:

$$(A \cdot \cos 33^\circ + B \cdot \sin 33^\circ) \cdot \cos(2\pi f_c t) - (A \cdot \sin 33^\circ - B \cdot \cos 33^\circ) \cdot \sin(2\pi f_c t) = \\ = -0.3 \cdot \sin(2\pi f_c t) \quad (3.55)$$

que resulta un sistema de ecuaciones cuya solución viene dada por

$$A = 0,1633; \quad B = -0,2516 \quad (3.56)$$

Así pues, con estos niveles de continua en cada uno de los dos canales de diferencia de color obtendremos la señal de salva directamente de los moduladores de croma. La conmutación entre las señales diferencia de color y los valores de continua para la salva permite también garantizar que la señal de croma permanece inactiva mientras se transmiten las señales de sincronismo.

3.3.10 Diagrama de bloques del decodificador NTSC

En principio, el proceso de decodificación de la señal sigue un proceso inverso al del codificador. En éste, las componentes de luminancia, croma, sincronismos y salva de color se generan en módulos independientes y, finalmente, todas las componentes se superponen en la señal de vídeo compuesto. El decodificador debe obtener las señales a aplicar al tubo de imagen, es decir, las componentes RGB y las señales de deflexión del haz a partir de la señal de vídeo compuesto, por lo que es esencial separar todas las componentes de la señal y mantener un correcto sincronismo entre ellas. En la figura 3.31 se representa un posible diagrama de bloques de un decodificador NTSC genérico. Cabe decir que en la práctica pueden encontrarse diversas variantes de este esquema, sobre todo en lo respecta a la tecnología utilizada en cada uno de los bloques. No obstante, los conceptos presentados pueden considerarse de validez general para la mayoría de receptores.

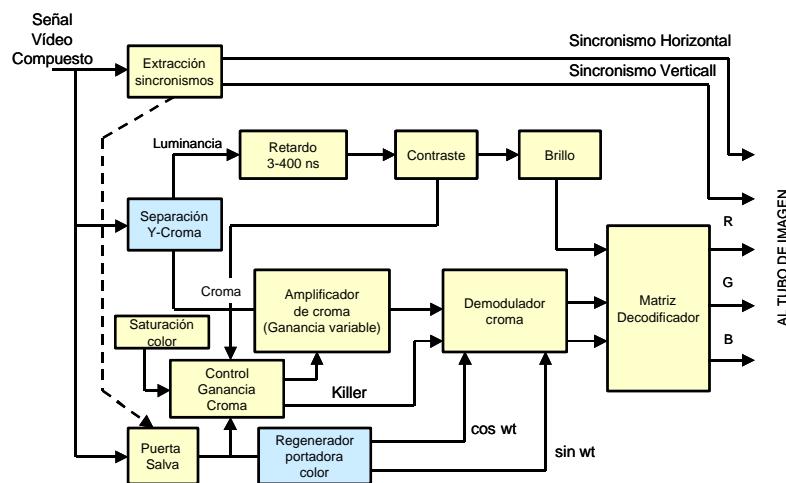


Fig. 3.31 Diagrama de bloques de un decodificador NTSC

La separación de los sincronismos se realiza determinando el nivel de negro de la señal recibida y aislando los impulsos de sincronismo de la información de vídeo basándonos en la polaridad de la señal. La detección de los sincronismos horizontales y verticales y la generación de las señales de barrido del haz requiere, como es obvio, un procedimiento análogo al utilizado en el caso de los receptores en blanco y negro. En los receptores de color, la detección de los sincronismos proporciona una referencia sobre la posición de la señal de salva. Esta señal se utiliza como referencia de fase en el sistema regenerador de la portadora de color, cuyos detalles consideraremos más adelante.

Las componentes de luminancia y croma se discriminan mediante filtrado y se procesan en canales independientes. El ajuste de la ganancia del canal de luminancia y el nivel de continua permiten ajustar respectivamente el contraste y el brillo general de la imagen. La señal de croma debe amplificarse antes de ser aplicada al demodulador para equilibrar la ganancia con el canal de luminancia. El control de la ganancia se realiza teniendo en cuenta el nivel de amplitud de la señal de salva recibida y los valores de los controles de saturación de color y contraste a los que el usuario tenga ajustado su receptor. En este control de ganancia suele tenerse en cuenta el nivel de ruido con el que se recibe la señal de croma. De este modo, cuando la amplitud de la salva decrece, se reduce automáticamente la saturación de color en el monitor, consiguiendo que el ruido sea menos molesto para el espectador. También suele incorporarse un sistema de inhibición del canal de croma, denominado supresor de color (*Killer*), que desactiva la demodulación de las componentes de color cuando se detecta la ausencia de la salva debido a que la señal se transmite en blanco y negro (siempre que el transmisor anule la salva de color en este tipo de transmisiones).

En el esquema de la figura 3.31 hemos supuesto que el receptor emplea una demodulación directa de las componentes $U'(t)$ y $V'(t)$ de modo que sólo debe introducirse una célula de retardo de señal en el canal de luminancia. El valor de este retardo depende, como ya hemos comentado, de los filtros del demodulador y oscila entre los 3 y 400 ns. No obstante, la demodulación de las componentes $U'(t)$ y $V'(t)$ o las originales $I'(t)$ y $Q'(t)$ sólo afectará a la propia estructura del demodulador (que deberá demodular con la referencia de fase de 33° respecto a la salva e incorporar un retardo para el canal $I'(t)$) y a la matriz de decodificación de las componentes R', G', B' .

Una vez identificada la función de los distintos bloques que intervienen en el proceso de decodificación de la señal, es conveniente profundizar en los detalles de dos elementos que juegan un papel fundamental en las características del receptor: el filtro de separación de croma-luminancia y el sistema de regeneración de la portadora de color.

3.3.11 Separación de las señales de luminancia y croma

Al seleccionar la frecuencia portadora de la señal de croma, se ha tenido en cuenta que la interferencia que introduce en el canal de luminancia sea mínima. Hemos visto que para que el sistema siga manteniendo la compatibilidad con los sistemas en blanco y negro es necesario imbricar espectralmente la información de croma y la de luminancia. Sin embargo, aunque desde un punto de vista teórico siempre es posible separar las dos señales, es evidente que no parece trivial, ni por supuesto económico, el diseño de un sistema que recomponga las dos señales originales en el receptor, eliminando cualquier tipo de cruce entre ellas.

Básicamente coexisten dos filosofías para separar estas señales que se conocen con el nombre de filtro muesca (*notch filter*) y filtro peine (*comb filter*). El uso de una u otra alternativa depende del coste y calidad del receptor, utilizándose la primera para receptores normales y la segunda en receptores de calidad. En los apartados siguientes examinamos los detalles de estas dos estrategias.

Filtro muesca (*notch filter*)

La idea de este tipo de filtrado de la señal de vídeo compuesto es utilizar un filtro paso banda, centrado en la frecuencia de 3,58 MHz, para obtener la señal de croma y un filtro de banda eliminada para obtener la luminancia. El filtrado de la señal de vídeo se ilustra en la figura 3.32.

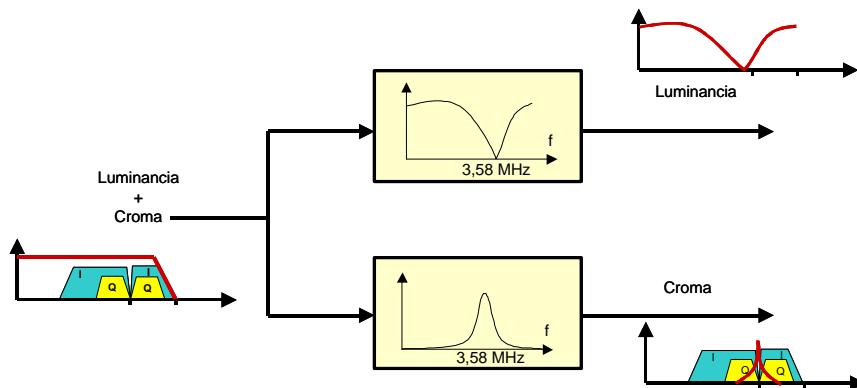


Fig. 3.32 Separación de las componentes de croma y luminancia mediante filtro muesca

El filtro utilizado para la componente de luminancia recibe el nombre de filtro muesca, ya que consiste en situar un cero de transmisión en la frecuencia portadora de la señal de croma. El resultado global obtenido con este filtro es parecido al que se obtendría con un filtro paso bajo aplicado a la señal de luminancia. De hecho, el filtro empieza a atenuar la señal de luminancia en frecuencias algo inferiores a los 3,58 MHz (normalmente del orden de 3 MHz) y aunque tiene cierta ganancia para frecuencias superiores a los 4 MHz, el contenido de la señal de luminancia a estas frecuencias puede considerarse prácticamente nulo.

Este procedimiento de separación de las componentes de la señal de vídeo puede realizarse mediante filtros convencionales de bajo coste. En principio, si las imágenes no tienen excesivo contenido en alta frecuencia, el sistema funciona correctamente. Es evidente que si pudiéramos garantizar que la señal de luminancia no contiene información más allá de los 3 MHz, la separación entre las dos componentes mediante este sencillo filtrado sería perfecta. Sin embargo, en regiones de la imagen con mucho detalle, la señal de luminancia suele presentar cierto contenido en alta frecuencia, por lo que la separación entre las dos componentes no será como en un principio esperaríamos. Analicemos cómo afecta este filtrado al contenido en alta frecuencia en cada uno de los canales.

La señal de luminancia perderá las componentes de alta frecuencia situadas en torno a la frecuencia portadora del color, con lo que la imagen que se observará en la pantalla será una versión de la escena original con menor detalle. En principio, esta pérdida de detalle sería tolerable por el sistema visual humano, ya que, como hemos visto, un sistema con 3 MHz de ancho de banda proporciona una calidad de imagen bastante aceptable. Indudablemente se apreciaría una pérdida de resolución respecto a un sistema que mantuviera el ancho de banda nominal de la señal de luminancia, pero, no obstante, es admisible.

El problema es más grave en lo que respecta a la señal de croma. En efecto, las componentes de alta frecuencia de la señal de luminancia estarán ahora presentes en el canal de color. Las componentes próximas a la frecuencia portadora de la croma serán ahora demoduladas como información de color de modo que, en los contornos de la imagen (regiones con contenido espectral de alta frecuencia), serán decodificadas con falsas coloraciones. Este problema suele también conocerse con el nombre de *cross color* debido a que el efecto es parecido al cruce de componentes de color, puesto que en ambos casos los colores decodificados no se corresponden con los originales. Sin embargo, la naturaleza del problema es muy distinta a la que habíamos considerado en el proceso de demodulación de las señales de color, puesto que ahora se debe a que la señal de luminancia introduce errores en las señales diferencia de color, mientras que en el otro caso se trataba de un verdadero cruce entre las dos componentes de color.

Así pues, el uso de un filtro muesca para la separación de las componentes de luminancia y color suele producir resultados aceptables en la mayor parte de las imágenes, aunque, en algunos contornos de las imágenes, pueden aparecer colores falsos que no se corresponden con la escena original.

Filtros peine (*Comb filters*)

Este tipo de filtros requieren utilizar células que retarden la señal de vídeo, por lo que su implementación es bastante más costosa que el filtro muesca. Su respuesta en frecuencia toma valores de ganancia nula en frecuencias uniformemente espaciadas, por lo que su aspecto tiene la forma de un peine. La implementación más simple, y también la más habitual, se representa en la figura 3.33, donde además se indica que la componente de luminancia se obtiene como la suma de señal de vídeo original más la señal de vídeo retardada, mientras que la croma se obtiene como la diferencia entre ambas señales. Analicemos la respuesta en frecuencia del filtro para cada una de las dos componentes.

En el caso del canal suma (luminancia), la salida puede expresarse como:

$$y_L(t) = x(t) + x(t-T) \quad (3.57)$$

donde $x(t)$ representa la señal de vídeo compuesto y T el retardo que se aplica a esta señal. Si aplicamos la transformada de Fourier a los dos términos de la igualdad podemos determinar la respuesta del filtro:

$$Y(f) = X(f)(1 + e^{-j2\pi fT}) \quad (3.58)$$

cuyo módulo viene dado por:

$$|H(f)| = \left| \frac{Y(f)}{X(f)} \right| = \left| 1 + e^{-j2\pi f T} \right| = 2 \cdot |\cos(\pi f T)| \quad (3.59)$$

Este resultado indica que la respuesta del filtro es periódica presentando sus valores máximos en las frecuencias $f_{k\max} = k/T$ y con valores nulos en las frecuencias $f_{knul} = 1/2T + k/T$.

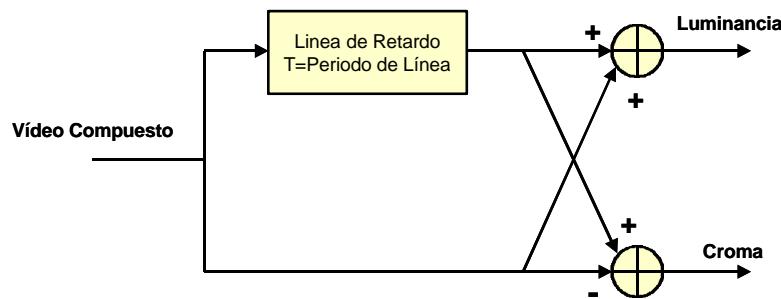


Fig. 3.33 Diagrama de un filtro peine para la separación de croma y luminancia

El módulo de la respuesta en frecuencia del filtro se ha representado en la figura 3.34, donde hemos tomado T igual al periodo de línea. Para este valor de T , el filtro presenta ganancia máxima en los múltiplos de la frecuencia de línea y ganancia nula en los múltiplos impares de la mitad de la frecuencia de línea. Esta disposición de los máximos y los nulos permite que las componentes de luminancia, centradas en los múltiplos de la frecuencia de línea, puedan pasar a través del filtro, mientras que las componentes de croma, que están situadas en los múltiplos impares de la mitad de la frecuencia de línea, sean canceladas en la salida.

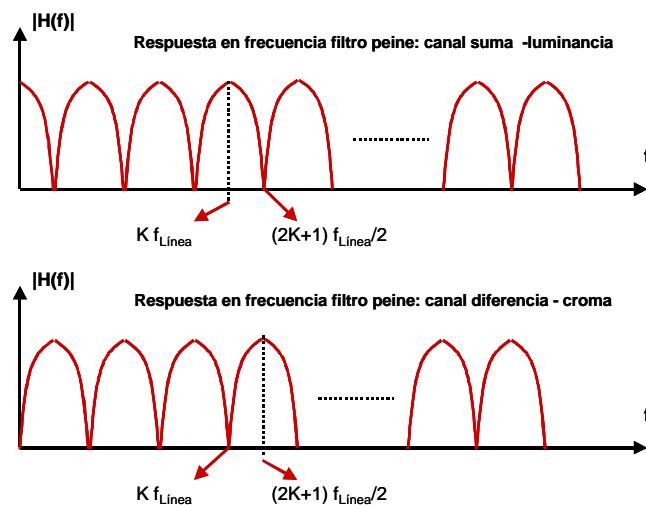


Fig. 3.34 Respuesta en frecuencia de los canales suma y diferencia de un filtro en peine

La respuesta en frecuencia del canal diferencia también se ha representado en figura 3.34, observándose que ahora la disposición de las frecuencias de ganancia nula y máxima se ha invertido respecto al caso anterior. En efecto, la relación entrada salida de este canal es:

$$y_L(t) = x(t) - x(t-T) \quad (3.60)$$

con lo que su respuesta en frecuencia viene dada por:

$$H(f) = \frac{Y(f)}{X(f)} = (1 - e^{-j2\pi fT}) \quad (3.61)$$

cuyo módulo es ahora una función del tipo seno:

$$|H(f)| = 2 \cdot |\sin(\pi fT)| \quad (3.62)$$

que, cuando T es el periodo de línea, presenta ganancia prácticamente nula a la señal de luminancia, debido a que la ganancia del filtro es cero en las frecuencias donde se concentra la mayor parte de la energía de la luminancia. Por el contrario, el filtro deja pasar la componente de croma, debido a que su ganancia es máxima en las regiones donde se sitúa esta señal.

La tecnología utilizada para implementar la línea de retardo que requiere un filtro en peine está basada, en la mayoría de receptores actuales, en sistemas digitales. Por ello, estos sistemas de separación de las componentes de croma y luminancia no se popularizaron hasta mediados de la década de los 80.

La realización de un filtro en peine mediante líneas de retardo analógicas se basa en el uso de transductores de la señal eléctrica a ultrasonidos, debido a que es difícil obtener por medios exclusivamente electrónicos retardos de este orden de magnitud (63,49 μ s en el caso del NTSC). Se trata de sistemas caros que, si bien se utilizan profusamente en los decodificadores de la señal de color en PAL, como parte integrante del filtro separador de croma y luminancia se usan tanto. El principio de funcionamiento de estas líneas de retardo analógicas se ilustra en la figura 3.35. La señal eléctrica se convierte a una onda de presión mediante un transductor electroacústico. Evidentemente, la señal obtenida es ultrasónica, ya que las frecuencias de la señal de vídeo son muy superiores al margen audible. La onda de presión se propaga en el interior de un material cristalino siguiendo el trayecto indicado y sufriendo varias reflexiones que permiten reducir el tamaño físico del retardador manteniendo un trayecto con una longitud aproximada de 17,46 cm (esta longitud debería ser de 17,6 cm para el caso del PAL). La velocidad de propagación de las ondas de presión en este medio es de unos 2750 m/s, de manera que la señal obtenida en la salida del transductor acústico-eléctrico tendrá un retardo de 63,49 μ s respecto la señal de entrada.

Las líneas de retardo digitales pueden realizarse utilizando registros de memorias de desplazamiento del tipo CCD, o bien con sistemas que convierten las señales de entrada directamente a formato digital. En el primer caso no se trata un sistema digital propiamente dicho, ya que, aunque se realiza un muestreo de las señales, estas muestras no se codifican como palabras digitales, sino que se almacenan directamente como una carga eléctrica proporcional a la tensión de entrada. Las cargas se

almacenar en una memoria CCD de línea y se desplazan en el tiempo como si se tratara de un registro de desplazamiento convencional que, en vez de operar con valores binarios, lo hace con elementos de carga eléctrica analógicos. En el capítulo dedicado a cámaras examinaremos con algo más de detalle los principios de funcionamiento de estos registros de desplazamiento. Los sistemas basados en capacidades conmutadas (*switched capacitor*) tienen unas prestaciones y filosofía de funcionamiento parecidas a los sistemas CCD y también pueden usarse para la realización del filtro *comb*.

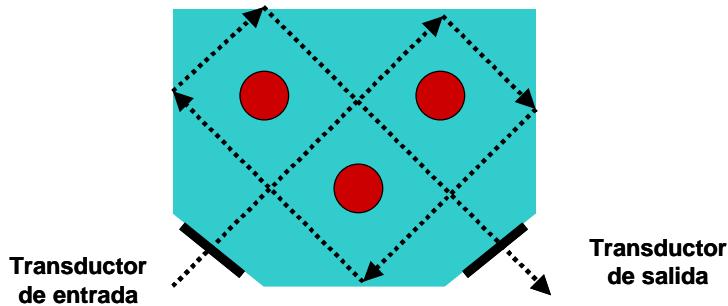


Fig. 3.35 Línea de retardo mediante ultrasonidos con reflexiones múltiples

En los sistemas denominados *digital comb filter* suele realizarse un muestreo de la señal de vídeo compuesto codificándose, mediante un conversor analógico digital, a unos 8 bits por muestra. En este caso la combinación suma y diferencia de las señales se realiza directamente en formato digital utilizando una unidad lógica especialmente diseñada para esta operación. Estos sistemas suelen aprovechar la digitalización de las señales de vídeo para realizar la decodificación de las componentes de color directamente mediante tecnología digital.

La idea de utilizar una línea de retardo para obtener un filtrado de las señales en forma de peine puede extenderse a más de un retardo. En la figura 3.36 se indica un esquema de filtro en peine basado en dos unidades de retardo, cada uno de los cuales es igual al periodo de línea. Esta generalización sólo suele usarse en sistemas basados en el tratamiento digital de las señales y no suele extenderse a más de dos unidades de retardo. La respuesta equivalente que obtendríamos con este sistema puede determinarse con un procedimiento análogo al caso de un único retardo temporal. Para el canal de luminancia, la combinación de las señales en el dominio temporal viene dada por:

$$y_L(t) = x(t) + 2 \cdot x(t-T) + x(t-2 \cdot T) \quad (3.63)$$

que tiene una respuesta en frecuencia:

$$H_L(f) = \frac{Y_L(f)}{X(f)} = \left(1 + 2 \cdot e^{-j2\pi fT} + e^{-j4\pi fT} \right) = (1 + e^{-j2\pi fT})^2 \quad (3.64)$$

cuyo módulo será el cuadrado de la función obtenida en el caso de utilizar un único retardo:

$$|H_L(f)| = 4 \cdot |\cos(\pi fT)|^2 \quad (3.65)$$

Para el caso del canal de croma obtenemos un resultado similar, ya que ahora la respuesta en frecuencia puede expresarse como:

$$H_C(f) = \frac{Y_C(f)}{X(f)} = (1 - 2 \cdot e^{-j2\pi fT} + e^{-j4\pi fT}) = (1 - e^{-j2\pi fT})^2 \quad (3.66)$$

por lo que el módulo también será el cuadrado de la función obtenida para un único retardo:

$$|H_C(f)| = 4 \cdot |\sin(\pi fT)|^2 \quad (3.67)$$

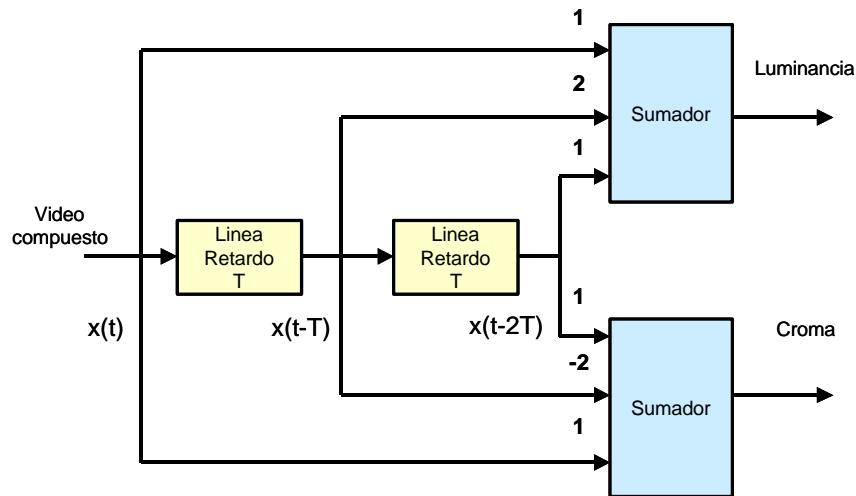


Fig. 3.36 Filtro en peine con dos unidades de retardo

Las magnitudes de las respuestas en frecuencia de estos dos canales se han representado en la figura 3.37. Las dos funciones mantienen las propiedades de ganancia máxima y mínima obtenidas para el filtro *comb* de una sola línea de retardo. Sin embargo, ahora la curva de ganancia es más plana tanto en los alrededores de las frecuencias en las que existe un máximo como en las que presentan ganancia nula. Con ello se consigue distorsionar menos la señal de interés (ya que no se deforman los lóbulos de las componentes frecuenciales) y, lo más importante, eliminar con mayor precisión las componentes indeseadas en cada canal. En efecto, con el filtro *comb* de orden 1 (es decir, con un único retardo), la magnitud de la respuesta en frecuencia alrededor de los puntos con ganancia nula variaba muy rápidamente, por lo que algunas componentes de la señal a cancelar, centradas en las proximidades del cero, podían pasar a través del sistema provocando la aparición de un pequeño cruce entre las componentes de luminancia y crominancia. Evidentemente, el nivel de cruce entre las dos componentes en el caso de un filtro *comb* de orden 1 no es, ni mucho menos, comparable con el que aparece en un sistema basado en el filtro muesca, pero es algo superior al que obtiene con un filtro de orden 2, que ya es prácticamente despreciable.

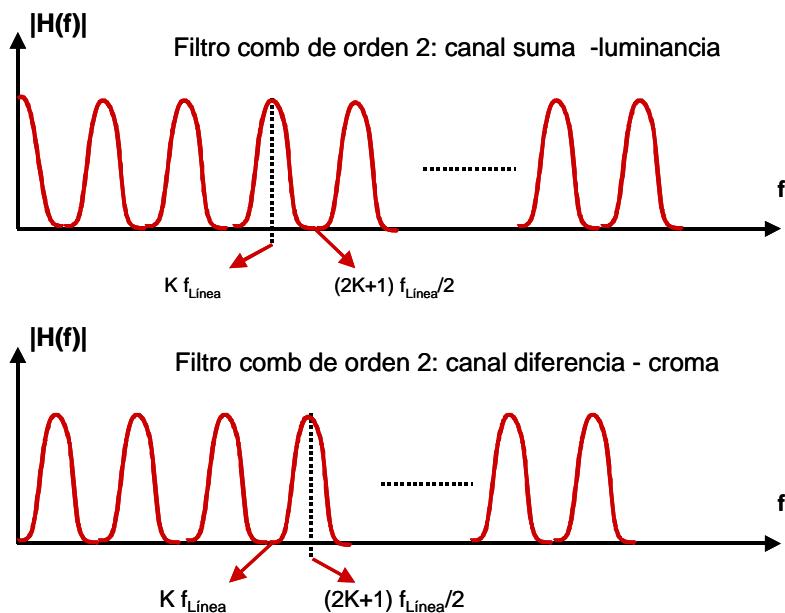


Fig. 3.37 Respuesta en frecuencia de los filtros en peine de orden 2

3.3.12 Regeneración de la portadora de color

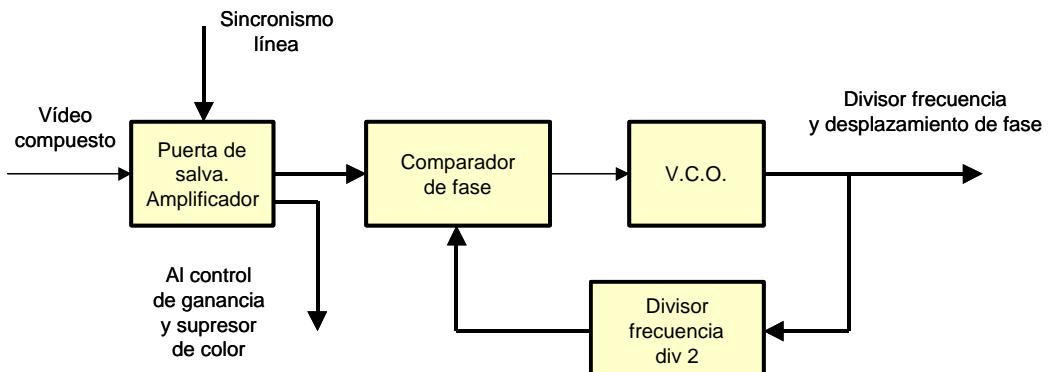


Fig. 3.38 Recuperación de la portadora de color

La regeneración de la portadora de color suele realizarse mediante la estrategia indicada en el diagrama de bloques de la figura 3.38. Básicamente consiste en un oscilador controlado por tensión que genera una señal de frecuencia doble a la de la portadora de color. El interés de que esta señal tenga una frecuencia doble radica en que de este modo puede obtenerse el desplazamiento de 90° de una de las señales del demodulador digitalmente. En efecto, la señal en la salida del regenerador puede

aplicarse a un par de básculas flip-flop tipo D que dividen la frecuencia por 2 y realizan un desplazamiento de un cuarto de periodo de la señal. De este modo el desplazamiento de 90° es muy preciso y no se requiere el uso de sistemas analógicos. El hecho de que las señales que se aplican al demodulador sean cuadradas en vez de sinusoidales no suele ser importante, debido al ancho de banda del propio mezclador de vídeo.

La sincronización entre la referencia interna y la señal de salva se realiza mediante una realimentación de la señal de salida que se compara con la salva externa, como en un PLL convencional. La puerta de entrada al sistema se encarga de garantizar que el oscilador de tensión sólo se controla cuando está presente la salva de color. En la figura 3.38 también se indica que el nivel de la salva se utiliza para actuar sobre la ganancia del amplificador de croma y desactivar los demoduladores de croma cuando la señal que se recibe es en blanco y negro y no contiene la salva.

3.3.13 Problemas del sistema NTSC

El sistema NTSC es particularmente susceptible a la denominada distorsión de fase diferencial. Este fenómeno consiste básicamente en que las no linealidades del canal de transmisión provocan que el retardo que experimentan las señales dependa del nivel de continua sobre el que se transmite la señal. El problema es particularmente importante, ya que supone que el retardo con el que se recibe la portadora de color depende del nivel de luminancia de la señal. Por ello, aparecen pequeños errores de fase entre la salva de color (transmitida con nivel de luminancia nulo) y la propia información de color, cuya portadora tendrá un retardo que depende del nivel de luminancia de cada parte de la escena.

Estas diferencias de fase entre la portadora real y la señal salva suponen que las componentes demoduladas no se corresponden directamente con las originales, sino que aparece un giro del vector de crominancia $U'(t)$, $V'(t)$. En efecto, el error de fase entre la salva de color y la portadora puede considerarse como si existiera un giro en el sistema de coordenadas en el que estamos demodulando la señal o, equivalentemente, si se supone que seguimos demodulando en el mismo sistema de coordenadas, un giro de las componentes del vector de crominancia. La magnitud y dirección del giro depende de los niveles de la señal y, aunque está relacionado con la propia imagen, en primera aproximación podemos considerarlo como aleatorio. La situación se ilustra en la figura 3.39, donde se indica esquemáticamente el carácter aleatorio del giro. Suponemos en este apartado que estamos realizando una demodulación directa de las señales sobre los ejes $U'(t)$ y $V'(t)$, aunque los resultados que derivaremos tienen también validez para el caso en que realicemos la demodulación sobre los ejes $I'(t)$ y $Q'(t)$.

El giro del vector $(U'(t), V'(t))$ supone, como justificaremos seguidamente, un cambio en el matiz del color demodulado que, evidentemente, distorsiona considerablemente la escena, ya que los colores representados no se corresponderán con los reales. El sistema PAL solucionará este problema del NTSC introduciendo una modificación sobre la señal transmitida que intenta convertir los errores de fase del demodulador en errores de amplitud sobre el vector de croma.

Es relativamente fácil comprobar que los errores de amplitud en el vector de croma representan cambios en la saturación y brillo del color representado en la pantalla, pero nunca cambios en el matiz

del color. Es evidente que el sistema visual humano será menos crítico con los errores en la saturación y brillo del color que sobre los cambios en el matiz, por lo que el sistema PAL se considera mucho más robusto que el NTSC.

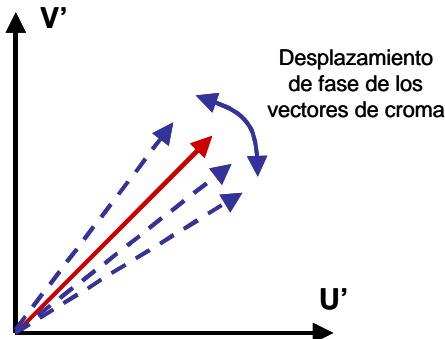


Fig. 3.39 Efectos de la distorsión diferencial de fase sobre el vector de croma

Comprobaremos a continuación que las variaciones en el módulo del vector de croma afectan sólo a la saturación y la intensidad del color, pero no al matiz. Esta misma demostración sirve para indicar que las variaciones en la dirección del vector de croma deben afectar al matiz del color decodificado. Efectivamente, las componentes $U'(t)$ y $V'(t)$ proporcionan información de todos los parámetros del color, por lo que si demostramos que un cambio en el módulo de este vector, manteniendo fija la componente de luminancia, representa sólo un cambio en la saturación y brillo, será evidente que la información de matiz debe depender de la dirección del vector. Es importante dejar claro que, a veces, por abuso del lenguaje, suele decirse que los cambios de dirección del vector de croma se corresponden exclusivamente con cambios de matiz del color, lo cual, no es estrictamente cierto. En efecto, un cambio de dirección lleva implícito una variación del matiz del color, pero también una variación en la saturación y brillo. Desde nuestro punto de vista, lo que nos interesa demostrar es que si logramos reducir el efecto del ruido a un error en el módulo del vector en vez de un error en su dirección, los colores decodificados mantendrán el mismo matiz que los originales. Para ello, bastará con comprobar que el cambio en el módulo del vector de croma no modifica el matiz del color.

Supongamos que las componentes de luminancia y señales diferencia de color que deberíamos demodular en ausencia de ruido y distorsiones son Y_A, U'_A, V'_A . Con estas señales, las componentes R', G', B' que obtendríamos vienen dadas por la relación matricial:

$$\begin{aligned} \begin{bmatrix} R'_A \\ G'_A \\ B'_A \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & -0.509 & -0.194 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} Y'_A \\ (R-Y)_A \\ (B-Y)_A \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & -0.509 & -0.194 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1/0.877 \\ 0 & 1/0.493 & 0 \end{bmatrix} \cdot \begin{bmatrix} Y'_A \\ U'_A \\ V'_A \end{bmatrix} \end{aligned} \quad (3.68)$$

donde la segunda matriz que interviene en el producto se ha obtenido a partir de las relaciones entre las componentes U',V' con las señales diferencia de color (R-Y)', (B-Y)'. Desarrollando el producto obtenemos

$$\begin{bmatrix} R'_A \\ G'_A \\ B'_A \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.1403 \\ 1 & -0.3935 & -0.5804 \\ 1 & 2.0284 & 0 \end{bmatrix} \cdot \begin{bmatrix} Y'_A \\ U'_A \\ V'_A \end{bmatrix} \quad (3.69)$$

Si en vez de utilizar el vector de croma original (U'_A, V'_A), modificamos su módulo en un factor k , las componentes R,G,B que obtendremos serán:

$$\begin{bmatrix} R'_B \\ G'_B \\ B'_B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.1403 \\ 1 & -0.3935 & -0.5804 \\ 1 & 2.0284 & 0 \end{bmatrix} \cdot \begin{bmatrix} Y'_A \\ K \cdot U'_A \\ K \cdot V'_A \end{bmatrix} \quad (3.70)$$

Para relacionar ambos resultados podemos utilizar:

$$\begin{bmatrix} Y'_A \\ K \cdot U'_A \\ K \cdot V'_A \end{bmatrix} = \begin{bmatrix} K \cdot Y'_A \\ K \cdot U'_A \\ K \cdot V'_A \end{bmatrix} + \begin{bmatrix} (1-K) \cdot Y'_A \\ 0 \\ 0 \end{bmatrix} \quad (3.71)$$

que sustituyéndolo en la ecuación 3.70 y desarrollando los productos se obtiene:

$$\begin{bmatrix} R'_B \\ G'_B \\ B'_B \end{bmatrix} = K \cdot \begin{bmatrix} R'_A \\ G'_A \\ B'_A \end{bmatrix} + (K-1) \cdot Y'_A \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (3.72)$$

Es decir, las componentes RGB que decodificamos cuando se modifica el módulo del vector de croma se corresponden con las componentes originales escaladas por el mismo factor (de modo que sólo se modifica el brillo pero no el matiz), más una determinada cantidad de blanco que sólo afectará a la saturación del color. En consecuencia, tal y como pretendíamos demostrar, el color decodificado mantiene el mismo matiz que el original.

En resumen, uno de los principales problemas del sistema NTSC es que los errores de fase en la decodificación del color se traducen directamente en errores sobre el matiz de la imagen representada, lo cual resulta especialmente desagradable en determinadas imágenes, como por ejemplo los rostros de las personas. Para controlar este problema, los receptores NTSC suelen incorporar un mando externo que permite que el espectador pueda ajustar los tonos o matices de color representados. Este control consiste en un sistema que permite desajustar la fase de la portadora de color que se emplea para la demodulación, de modo que el espectador puede actuar sobre los ejes sobre los que se realiza la demodulación compensando el giro debido a los errores de fase. No obstante, este control sólo permite compensar una gama reducida de matices, ya que como hemos comentado el desfase entre la salva y la portadora depende del nivel de luminancia y, por lo tanto, varía en función de la imagen.

Las cartas de ajuste de los países que utilizan el sistema NTSC suelen incorporar una imagen estática con la fotografía de una cara. Ello facilita notablemente el ajuste de matiz en el receptor, debido a que la tonalidad de la piel es una de las que más fácilmente pueden ajustarse pues el sistema visual es muy crítico ante estos errores de color.

Otro de los problemas del sistema NTSC, que ya habíamos comentado anteriormente, es que la asimetría debida al recorte de la banda lateral superior de la componente $I'(t)$ provoca que aparezcan problemas en el matiz de los colores en las regiones donde se producen transiciones de objetos. Por todo ello, el sistema NTSC es mucho menos robusto que el PAL o el SECAM y, en términos irónicos, suele decirse que sus siglas corresponden al acrónimo de *Never Twice The Same Colour*.

3.4 El sistema PAL

El sistema PAL fue propuesto en 1964 por el Dr. Bruch de Telefunken (Alemania) como una alternativa al sistema NTSC que permitía reducir la sensibilidad del receptor a las distorsiones de fase. El sistema fue aceptado por un gran número de países que en la época en que se definió todavía no habían regularizado las transmisiones de televisión en color. Estados Unidos y Japón mantienen el sistema NTSC debido a su considerable implantación en el momento de definir el PAL. Francia y los países de su área de incidencia utilizan un desarrollo propio de H. De France conocido como SECAM y que presenta bastantes aspectos diferenciales con el PAL y el NTSC. El PAL puede considerarse como una evolución natural del NTSC y comparte muchos de los aspectos tecnológicos y de implementación que son comunes a ambos sistemas. Por ello, en esta presentación del PAL nos centraremos principalmente en los aspectos diferenciales entre ambos subsistemas. Suponemos que el lector está familiarizado con las técnicas de modulación y demodulación de las señales de croma, los procedimientos para generar la señal de salva, los filtros para separar las componentes de croma y luminancia y los subsistemas de recuperación de la fase de la portadora de color que, en esencia, son equivalentes en el sistema NTSC y en el PAL con sólo cambiar el número de líneas y las frecuencias de imagen y campo.

En los siguientes apartados se describen las diferencias más significativas entre ambos subsistemas prestando una especial atención a la forma en que se consiguen corregir los errores en la fase de las componentes de color y el proceso de decodificación de la señal PAL.

3.4.1 Corrección de fase mediante la inversión de una de las componentes de croma

En la ecuación 3.3 hemos visto que el sistema PAL realiza una inversión de la componente de color $V'(t)$ en líneas alternadas. La ecuación puede expresarse en función de un índice entero que representa el número de línea:

$$croma_{PAL}^{(k)}(t) = U'(t) \cdot \sin(2 \cdot \pi f_{PAL} t) + (-1)^k \cdot V'(t) \cdot \cos(2 \cdot \pi f_{PAL} t) \quad (3.73)$$

Debe mencionarse que el índice k es un valor entero que se incrementa en cada nueva línea de la secuencia de vídeo y que, aunque está relacionado con la numeración propia de las líneas de la señal de televisión, esta relación, como veremos posteriormente, no es directa.

La ecuación 3.73 define una alternancia en el signo de la componente de color $V'(t)$ que de alguna manera deberá compensarse en el receptor. De hecho, la cancelación de los errores de fase en las componentes de color se produce al corregir esta alternancia de signo en el decodificador. A menudo, las líneas de la señal PAL que no contienen la inversión de signo se denominan *líneas NTSC*, mientras que las líneas invertidas se denominan *líneas PAL*. Es importante que esta notación no resulte confusa, puesto que la esencia del sistema PAL es que aparezcan de forma alternada ambos tipos de líneas. El receptor, para decodificar correctamente las señales, deberá disponer de información precisa sobre el tipo de línea que está recibiendo. Esta información se proporciona al receptor mediante cambios de fase en la señal de salvaje de color que analizaremos posteriormente.

Para ver cómo la alternancia de signo permite reducir los errores de fase en el vector de croma, consideremos el ejemplo que se muestra en la figura 3.40. Por el momento supondremos que deseamos transmitir una imagen homogénea, con un color constante en todas sus líneas. El vector de croma que se desea transmitir se representa en la parte superior de la figura. Debido a la alternancia en la componente $V'(t)$, se transmitirán dos tipos de vectores al canal, una para las líneas no invertidas (líneas NTSC) y otro para las líneas invertidas (líneas PAL). En la figura 3.40 se supone que la línea k corresponde a una línea no invertida y que la línea $k+1$ tiene la componente V' invertida.

Si el canal introduce un desfase constante en los vectores de croma, las señales recibidas experimentarán un giro de un ángulo α en el mismo sentido.

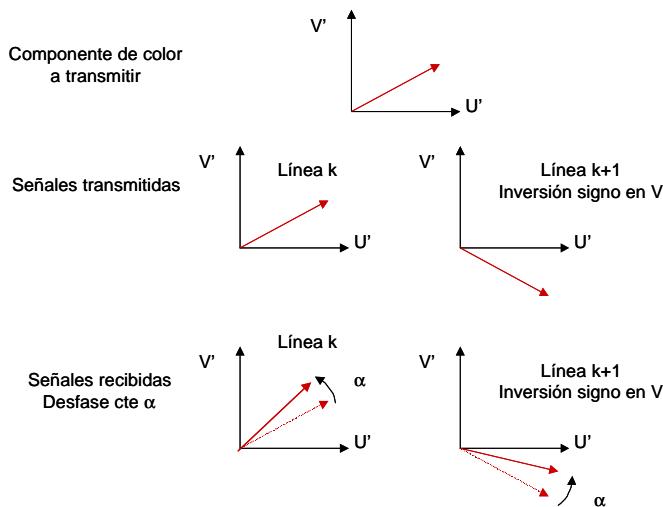


Fig. 3.40 Diagrama de las señales transmitidas y recibidas en el sistema PAL

En el receptor, la rotación de fase que han experimentado las componentes de croma puede cancelarse si se promedian dos líneas consecutivas. El procedimiento de reconstrucción se ilustra en la figura 3.41, donde se indica que la línea k se mantiene igual a como ha sido recibida, mientras que en la línea $k+1$ se realiza la inversión de la componente V' . El resultado de promediar ambas líneas también se muestra en la figura 3.41, observando que, una vez se ha aplicado la corrección de signo, las

rotaciones del canal se traducen en giros opuestos del vector de croma, de modo que al promediar los dos vectores, los errores de fase se cancelan.

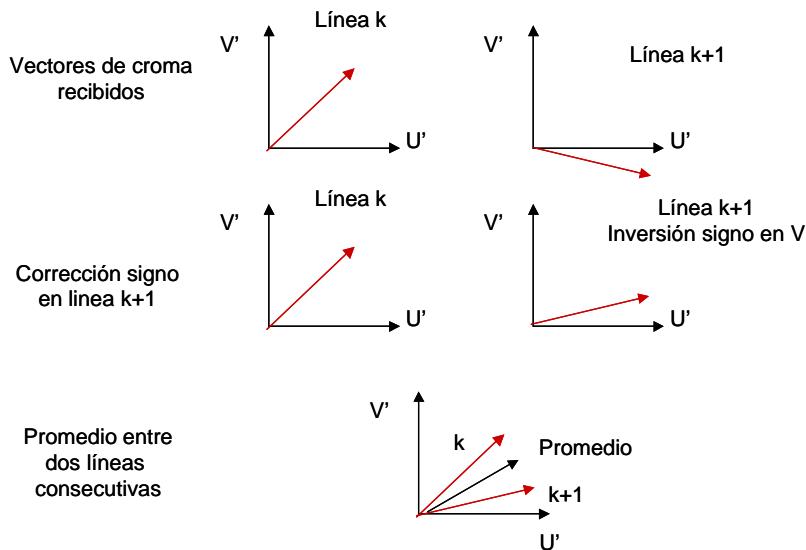


Fig. 3.41 Cancelación de las rotaciones de croma mediante el promedio de dos líneas consecutivas

La inversión de signo de la componente $V'(t)$ y el promedio de las líneas consecutivas debe realizarse en el decodificador. Existen dos alternativas para realizar el promedio que se conocen como el PAL-S (PAL simple) y el PAL D (PAL Delay Line). En el primer caso, las líneas se invierten y se presentan en la pantalla dejando que sea el propio espectador el encargado de realizar el promedio visualmente, corrigiendo los errores de fase. Esta filosofía produce resultados aceptables siempre que la rotación de los vectores de croma sea inferior a unos 5° . En el PAL D el promedio entre dos líneas consecutivas se realiza de forma electrónica, produciendo mejores resultados. En la actualidad, todos los receptores PAL realizan el promedio entre las líneas de forma electrónica. Veremos los diagramas de bloques básicos para cada uno de los procedimientos.

3.4.2 Diagrama de bloques de un decodificador de croma en PAL S

El decodificador de PAL S es prácticamente idéntico al decodificador NTSC, con la salvedad de que debe realizarse la corrección de signo de la componente $V'(t)$. En la figura 3.42 se muestra el esquema simplificado donde se indica que el signo de la componente en coseno debe invertirse de forma sincronizada con los impulsos de sincronismo de línea. La frecuencia del generador de cambio de signo es igual a la mitad de la frecuencia de línea, ya que el ciclo completo (es decir, la línea positiva y la negativa) tiene una duración de dos períodos de línea. Las señales obtenidas en la salida de los mezcladores deben pasarse a través de filtros paso bajo para obtener las componentes $U'(t)$ y $V'(t)$. En el caso del sistema PAL, los dos filtros son en todos los casos idénticos, puesto que las señales U y V tienen el mismo ancho de banda en la transmisión. Estas señales, junto con la componente de

luminancia se aplican a la matriz de decodificación para obtener las componentes RGB que finalmente se aplican al tubo.

El PAL S sólo resulta satisfactorio cuando los errores de fase son muy pequeños (5°). Para errores mayores el sistema visual no puede promediar las líneas consecutivas y la diferencia entre ellas se hace notable. Cuando existen errores de fase grandes, se produce un trazado de bandas por la imagen que tienden a desplazarse en sentido vertical hacia arriba o hacia abajo. Este efecto se conoce con el nombre de efecto Hannover o persiana veneciana.

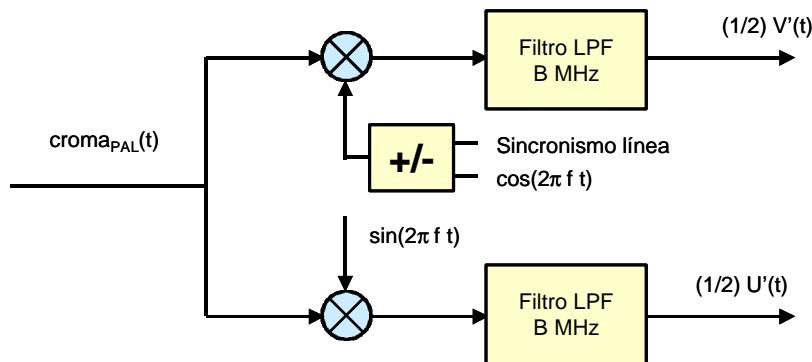


Fig. 3.42 Diagrama de bloques del decodificador de croma en el PAL simple

3.4.3 Diagrama de bloques de un decodificador de croma en PAL D

Hemos comentado que el PAL D realiza un promedio electrónico entre dos líneas consecutivas, con lo que consigue una mejor calidad de las imágenes, eliminando de forma prácticamente completa el efecto Hannover. Para realizar el promedio es necesario introducir una línea de retraso que permita combinar la señal de la línea actual con la de la línea anterior. En los primeros receptores en PAL D, la línea de retraso se construía mediante el uso de transductores de ultrasonidos como los representados en la figura 3.35. Actualmente la mayoría de modelos incorporan líneas de retraso basadas en CCD, en capacidades conmutadas, o directamente mediante el uso de memorias digitales, que producen resultados mucho más estables. En los decodificadores actuales la corrección de signo de la componente $V'(t)$ y el promedio entre líneas consecutivas puede realizarse con las componentes $U'(t)$ y $V'(t)$ demoduladas. En los decodificadores antiguos era necesario realizar el promedio antes de demodular las señales, debido a que las unidades de retraso basadas en ultrasonidos no podían trabajar con componentes de continua. El decodificador de PAL D de la figura 3.43 realiza la corrección de signo de la componente $V'(t)$ y el promedio de líneas con las componentes de croma moduladas.

Para analizar el decodificador de croma supondremos inicialmente que la línea $k+1$ ha sufrido la inversión de la componente V , mientras que la línea k es una línea NTSC convencional. En este caso, las expresiones de las señales en la entrada y en la salida de la línea de retraso vienen dadas por:

$$\begin{aligned} croma_{PAL}^{(k+1)}(t) &= U' \cdot \sin(2 \cdot \pi f_{PAL} t) + V' \cdot \cos(2 \cdot \pi f_{PAL} t) \\ croma_{PAL}^{(k)}(t) &= U' \cdot \sin(2 \cdot \pi f_{PAL} t) - V' \cdot \cos(2 \cdot \pi f_{PAL} t) \end{aligned} \quad (3.74)$$

donde suponemos que las señales U' y V' son constantes para toda la imagen y que el retardo que se aplica a la señal deja inalterada la fase de las portadoras en cuadratura. Analizaremos este problema con detalle más adelante. Por el momento supongamos que al aplicar el retardo de una línea cualquiera de las dos portadoras en la salida, continúan estando en fase con la entrada.

En el canal superior, la suma de ambas componentes da como resultado:

$$S(t) = 2 \cdot U' \cdot \sin(2 \cdot \pi f_{PAL} t) \quad (3.75)$$

mientras que en el canal inferior, la diferencia es

$$D(t) = 2 \cdot V' \cdot \cos(2 \cdot \pi f_{PAL} t) \quad (3.76)$$

Si multiplicamos el canal $S(t)$ por la portadora en fase con el seno y el canal $D(t)$ por la portadora en fase con el coseno y aplicamos los filtros paso bajo, obtendremos las componentes $U'(t)$ y $V'(t)$ deseadas.

En la siguiente línea ($k+2$), los signos de las componentes $V'(t)$ en la entrada y en la salida de la línea de retardo estarán invertidos:

$$\begin{aligned} croma_{PAL}^{(k+2)}(t) &= U' \cdot \sin(2 \cdot \pi f_{PAL} t) - V' \cdot \cos(2 \cdot \pi f_{PAL} t) \\ croma_{PAL}^{(k+1)}(t) &= U' \cdot \sin(2 \cdot \pi f_{PAL} t) + V' \cdot \cos(2 \cdot \pi f_{PAL} t) \end{aligned} \quad (3.77)$$

por lo que la señal en el canal suma seguirá siendo:

$$S(t) = 2 \cdot U' \cdot \sin(2 \cdot \pi f_{PAL} t) \quad (3.78)$$

mientras que en el canal diferencia tendremos una diferencia de signo

$$D(t) = -2 \cdot V' \cdot \cos(2 \cdot \pi f_{PAL} t) \quad (3.79)$$

lo que significa que deberemos invertir esta señal en líneas alternativas para poder obtener la componente de color deseada. La inversión de la señal debe sincronizarse con la alternancia de signo realizada en el transmisor. Para ello, será necesario transmitir la información del signo que se ha utilizado en cada línea. Esta información se introduce en la salva de color y puede ser recuperada en el receptor permitiendo sincronizar la inversión de signo en el canal diferencia con la de las señales recibidas.

En todas las ecuaciones anteriores hemos supuesto que las componentes de color eran iguales en las dos líneas, por lo que en cada uno de los canales se obtenía la componente de interés y se cancelaba la otra componente. Si esto no es así, las señales $S(t)$ y $D(t)$ contienen componentes cruzadas, pero serán eliminadas por el proceso de mezcla y filtrado.

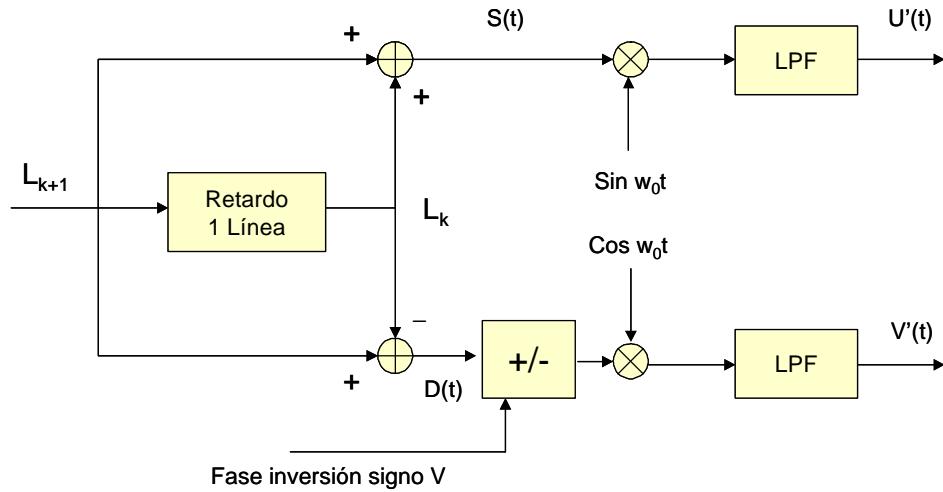


Fig. 3.43 Diagrama del decodificador de croma del sistema PAL D

En efecto, consideremos sólo el caso en que la línea actual es una línea del tipo NTSC y, por tanto, la línea anterior es del tipo PAL. Suponemos que el vector de croma en la línea actual es (U_{k+1}, V_{k+1}) y que el vector de croma en la línea anterior es $(U_k, -V_k)$. Podemos escribir:

$$\begin{aligned} croma_{PAL}^{(k+1)}(t) &= U'_{k+1} \cdot \sin(2 \cdot \pi f_{PAL} t) + V'_{k+1} \cdot \cos(2 \cdot \pi f_{PAL} t) \\ croma_{PAL}^{(k)}(t) &= U'_k \cdot \sin(2 \cdot \pi f_{PAL} t) - V'_k \cdot \cos(2 \cdot \pi f_{PAL} t) \end{aligned} \quad (3.80)$$

La señal en el canal suma vendrá dada por:

$$S(t) = (U'_{k+1} + U'_k) \cdot \sin(2 \cdot \pi f_{PAL} t) + (V'_{k+1} - V'_k) \cdot \cos(2 \cdot \pi f_{PAL} t) \quad (3.81)$$

Al multiplicar esta señal por la componente en seno y filtrar paso bajo obtendremos como componente $U'(t)$ el valor medio entre las dos líneas:

$$U'(t) = \frac{(U'_{k+1} + U'_k)}{2} \quad (3.82)$$

Análogamente, en el canal diferencia obtendremos el promedio entre las dos componentes $V'(t)$

$$V'(t) = \frac{(V'_{k+1} + V'_k)}{2} \quad (3.83)$$

lo cual demuestra que obtenemos como señales de croma los valores promedios entre las dos líneas. Si hubiésemos considerado el caso en que la línea actual es PAL y la anterior NTSC, los resultados obtenidos serían idénticos a los de las ecuaciones 3.82 y 3.83.

El hecho de representar el promedio entre dos líneas consecutivas de la información de croma supone una reducción de la resolución de color. No obstante, esta pérdida de resolución puede ser asumida por el sistema, ya que se trata únicamente de las componentes de color, manteniendo la luminancia distinta para cada línea. Esto significa que las componentes RGB, que se obtienen como la combinación de la luminancia y la croma, son independientes en las distintas filas que constituyen la imagen.

La pérdida de resolución en color se produce sólo en el sentido vertical de la imagen, lo cual tiene incluso cierta lógica. Recordemos que la resolución horizontal de las componentes de croma había sido reducida durante el proceso de transmisión de la señal, aplicando filtros paso bajo, de 1,3 MHz de ancho de banda, a las señales diferencia de color. No obstante, la resolución vertical no ha sido reducida durante la transmisión (se transmiten todas las líneas), por lo que admite una posterior reducción del ancho de banda (mediante el promedio de líneas consecutivas), que en este caso realiza el propio receptor. Con ello, se igualan las resoluciones verticales y horizontales de las componentes de color.

Un aspecto muy importante del demodulador de croma es cómo afecta a las portadoras la unidad de retardo que se aplica a la señal. En el desarrollo anterior hemos comentado que suponíamos que la unidad de retardo no introducía ningún desfase significativo sobre las portadoras en seno y en coseno. Esto nos permitía agrupar los términos de croma en los canales suma y diferencia y demodular correctamente las señales. Si esto no fuera así, y existiera algún desfase entre las portadoras en la entrada y en la salida de la línea de retardo, aparecerían cruces entre las dos componentes diferencia de color que se traducirían en una incorrecta demodulación de la señal. En efecto, si escribimos la ecuación 3.74 teniendo en cuenta un posible desfase de las portadoras entre la entrada y la salida obtenemos:

$$\begin{aligned} croma_{PAL}^{(k+1)}(t) &= U' \cdot \sin(2 \cdot \pi f_{PAL} t) + V' \cdot \cos(2 \cdot \pi f_{PAL} t) \\ croma_{PAL}^{(k)}(t) &= U' \cdot \sin(2 \cdot \pi f_{PAL} t + \varphi) - V' \cdot \cos(2 \cdot \pi f_{PAL} t + \varphi) \end{aligned} \quad (3.84)$$

Si desarrollamos la segunda expresión

$$\begin{aligned} croma_{PAL}^{(k)}(t) &= U' \cdot \cos(\varphi) \cdot \sin(2 \cdot \pi f_{PAL} t) + U' \cdot \sin(\varphi) \cdot \cos(2 \cdot \pi f_{PAL} t) \\ &\quad - V' \cdot \cos(\varphi) \cdot \cos(2 \cdot \pi f_{PAL} t) + V' \cdot \sin(\varphi) \cdot \sin(2 \cdot \pi f_{PAL} t) = \\ &= (U' \cdot \cos(\varphi) + V' \cdot \sin(\varphi)) \cdot \sin(2 \cdot \pi f_{PAL} t) + \\ &\quad + (U' \cdot \sin(\varphi) - V' \cdot \cos(\varphi)) \cdot \cos(2 \cdot \pi f_{PAL} t) \end{aligned} \quad (3.85)$$

y determinamos la señal que obtenemos en el canal suma obtenemos:

$$\begin{aligned} S(t) &= (U'(1 + \cos \varphi) + V' \cdot \sin \varphi) \cdot \sin(2 \cdot \pi f_{PAL} t + \varphi) \\ &\quad + (V'(1 - \cos \varphi) + U' \cdot \sin \varphi) \cdot \cos(2 \cdot \pi f_{PAL} t + \varphi) \end{aligned} \quad (3.86)$$

Finalmente, si aplicamos esta señal al mezclador en seno y la filtramos paso bajo obtenemos la salida de la componente $U'(t)$ que aplicaremos a la matriz de decodificación:

$$U'(t) = \frac{U'(1 + \cos \phi) + V' \cdot \sin \phi}{2} \quad (3.87)$$

Esta expresión coincide con el resultado deseado cuando el desfase es nulo, pero introduce un cruce entre las componentes de color cuando es distinto de cero, resultando tanto más importante cuanto mayor es este desfase. El resultado que obtendríamos si analizamos el canal diferencia en vez del canal suma es parecido.

En consecuencia, es sumamente importante que la línea de retardo mantenga las fases de las dos portadoras en cuadratura, para lo cual el tiempo de retardo deberá ser un múltiplo entero del periodo de la portadora de color. Esta restricción parece imponernos que el tiempo de línea deberá ser un múltiplo entero del periodo de la portadora de color, lo que, como sabemos del análisis del NTSC, supondría la introducción de patrones de interferencia altamente visibles. La solución a este dilema de diseño consiste en dar prioridad a que los patrones de interferencia sean mínimos, seleccionando la frecuencia de la portadora de color para que se cumpla este requisito. Una vez elegida la frecuencia de portadora, la línea de retardo del decodificador PAL se ajusta para que el tiempo de retardo coincida exactamente con un múltiplo entero de periodos de la portadora de color. Esto significa que el tiempo de retardo no coincidirá exactamente con el periodo de línea, sino que será ligeramente inferior. En la práctica, tal y como veremos en el subapartado siguiente, el tiempo de retardo debe ser de 64,056 μs , algo superior al periodo de línea de 64 μs . Esta diferencia significa que el promedio de las componentes de croma no se realiza exactamente entre dos elementos de imagen situados en la misma posición horizontal, sino que el elemento actual ha sido promediado con un elemento de la línea anterior situado ligeramente a su izquierda (desde la posición del espectador). La diferencia entre ambas posiciones es sumamente pequeña (un 0,08 % de la duración de una línea) por lo que no se aprecia ningún efecto secundario.

Debe observarse que el ajuste del retardo de línea del decodificador PAL debe ser muy preciso, puesto que en caso contrario se produciría el cruce entre las componentes de color. Además, la frecuencia de la portadora de color debe ser muy estable y no debe presentar diferencias significativas entre distintas transmisiones. Las especificaciones sobre la frecuencia de portadora de color exigen que la desviación sea inferior a 1 Hz sobre el valor teórico (de un 0,225 unidades por millón respecto al valor real de la portadora).

Una alternativa al decodificador de la figura 3.43 es realizar el promedio entre las componentes de color directamente en banda base. Esta estrategia, que se representa de forma esquemática en la figura 3.44, consiste en utilizar un decodificador PAL S para separar las componentes $U'(t)$ y $V'(t)$ y posteriormente promediarlas de forma independiente. Los retardos con los que se realizan los promedios pueden ser exactamente iguales a un periodo de línea y no se producen cruces entre las componentes de color, aun en el supuesto que el sistema se desajustara. Las líneas de retardo pueden implementarse utilizando circuitos de capacidades conmutadas, dispositivos CCD o mediante la digitalización de las señales croma y la simulación del retardo con memorias digitales. La mayor parte de los decodificadores actuales utilizan esta tecnología, que es más estable y requiere de un menor número de ajustes. No obstante, esta filosofía requiere que los retardos de la señal puedan

realizarse a frecuencias muy bajas (las señales diferencia de color en banda base pueden permanecer en un mismo valor durante tiempos largos), lo cual no es posible implementar mediante líneas de retardo ultrasónicas convencionales. Por ello, este tipo de decodificadores sólo empezaron a utilizarse a partir de la mitad de la década de los 80, cuando la tecnología digital y de CCD resultaron económicamente rentables.

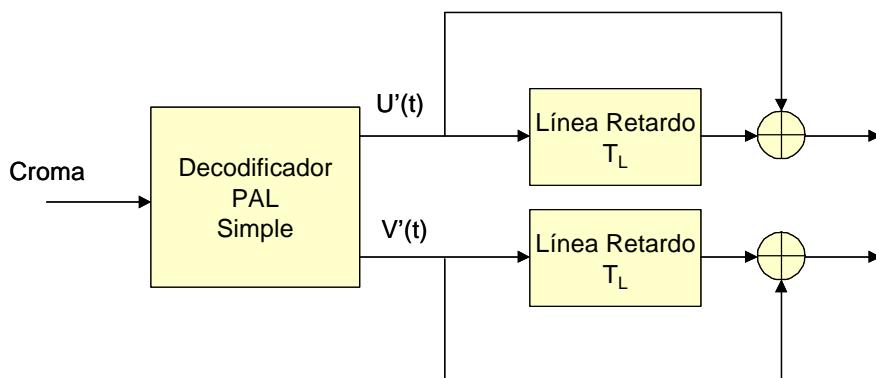


Fig. 3.44 Promedio de las componentes de colores en banda base

3.4.4 Selección de la frecuencia portadora de color

Al analizar el sistema NTSC hemos visto que, si seleccionábamos la frecuencia de la subportadora de color igual a un múltiplo impar de la mitad de la frecuencia de línea, obteníamos unos patrones interferentes que tanto en el dominio espacial como el temporal ofrecían cambios constantes de posición, por lo que la señal resultaba poco visible en los receptores monocromos. En el sistema PAL podríamos intentar utilizar una frecuencia portadora que cumpliera la misma propiedad. No obstante, la inversión de signo en una de las componentes de color introduce cambios significativos en el patrón interferente e imposibilita utilizar una frecuencia portadora que sea un múltiplo impar de la mitad de la frecuencia de línea.

En efecto, el patrón de interferencias de la señal PAL puede considerarse como la superposición de los patrones interferentes asociados a las componentes $U'(t)$ y $V'(t)$. El patrón de la componente $U'(t)$ coincide con el del NTSC, puesto que la señal no sufre ninguna inversión de signo entre líneas consecutivas. En cambio, la componente $V'(t)$ va alternando el signo entre líneas consecutivas, lo que significa que la interferencia también cambiará de signo respecto al patrón del NTSC. Habíamos visto que los elementos de imagen situados en la misma posición horizontal en líneas adyacentes experimentaban un cambio de signo en la amplitud de la interferencia que se superponía con la información de luminancia. Ahora, para la componente $V'(t)$, este cambio de signo se compensa con la inversión de signo propia de la línea, por lo que las interferencias están en fase en líneas adyacentes. Esto significa que aparecerá un patrón interferente formado por líneas verticales que será altamente visible. Además, desde el punto de vista temporal la interferencia también será estacionaria debido a que el número de líneas entre imagen e imagen es impar (625) y por lo tanto la línea actual estará invertida respecto a la misma línea en la imagen anterior, compensando las diferencias de signo que

existían en el patrón interferente del sistema NTSC. La única solución factible a este problema consiste en modificar la frecuencia de la subportadora de color eligiendo un patrón más complejo que no se vea afectado por los cambios de signo de una de las componentes.

La solución que se adoptó fue tomar una frecuencia portadora que cumpliera la siguiente condición:

$$f_{PAL} = (2k+1) \frac{f_L}{2} + \frac{f_L}{4} \quad (3.88)$$

Esta selección resulta bastante lógica si tenemos en cuenta que un múltiplo impar de la mitad de la frecuencia de línea produce un patrón interferente adecuado para la componente $U'(t)$, pero totalmente desfavorable para la componente $V'(t)$. Análogamente, un múltiplo exacto de la frecuencia de línea (múltiplo par de la mitad de la frecuencia de línea) produce un patrón interferente adecuado a la señal $V'(t)$, pero nefasto para la $U'(t)$. La solución más adecuada parece quedarnos con una frecuencia intermedia entre estos dos casos extremos, con lo cual los dos patrones serán complejos y se imbricarán espacialmente en la forma en que se ilustra en la figura 3.45.

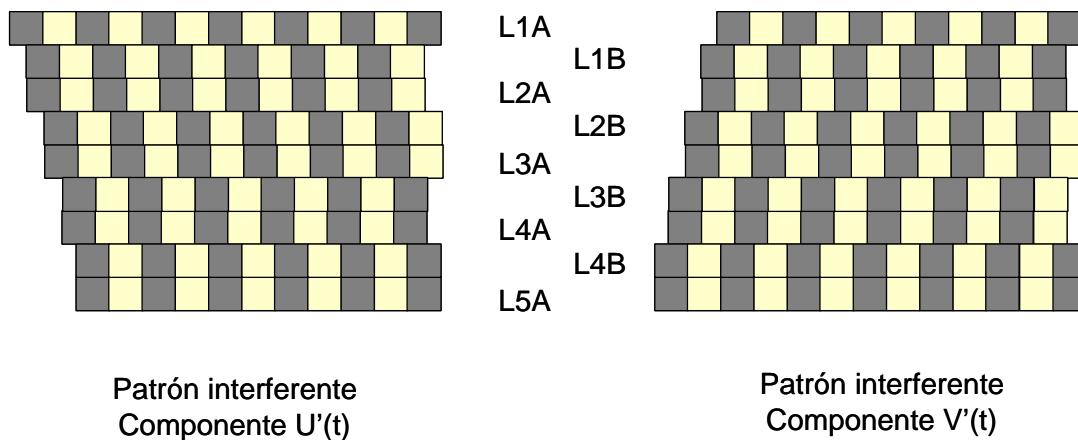


Fig. 3.45 Patrones interferentes espaciales para las componentes $U'(t)$ y $V'(t)$ en el sistema PAL

Con esta frecuencia de portadora, las fases entre las ondas interferentes en líneas consecutivas están desfasadas un ángulo de 90° (retrasadas para la componente $U'(t)$ y adelantadas para la componente $V'(t)$). Los patrones interferentes entre dos líneas consecutivas correspondientes a distintos campos están en fase y las interferencias vuelven a estar en fase cada cuatro líneas consecutivas dentro de un mismo campo. El patrón interferente es algo más visible que en el sistema NTSC debido a que la imbricación de los máximos y mínimos en el sentido vertical es algo más lenta ya que las líneas sucesivas de los dos campos tienen la misma fase. Para mejorar algo esta imbricación, se decidió modificar ligeramente la frecuencia portadora añadiendo al resultado de la ecuación 3.88 la frecuencia de imagen, que en el caso del PAL es de 25 Hz. Con ello, la frecuencia de la subportadora de color pasa a ser:

$$f_{PAL} = (2k + 1) \frac{f_L}{2} + \frac{f_L}{4} + f_V \quad (3.89)$$

Ahora, los patrones interferentes dentro de un mismo campo prácticamente no se han modificado debido a que el incremento de frecuencia es muy pequeño. Sin embargo, el segundo campo quedará con una fase invertida, por lo que la imbricación de las interferencias mejora considerablemente. En la figura 3.46 se muestran los patrones espaciales definitivos del PAL.

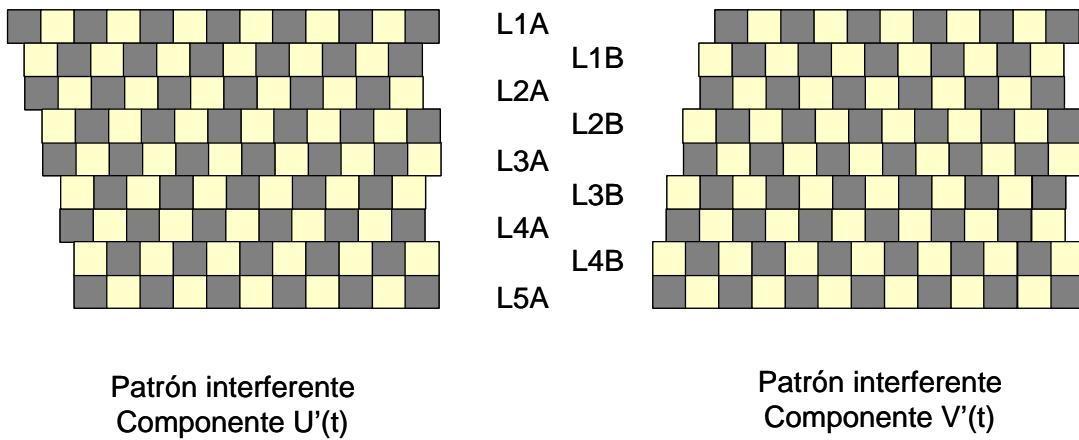


Fig. 3.46 Patrones de interferencia espacial definitivos del sistema PAL

La frecuencia definitiva se selecciona para un valor de k suficientemente elevado para que quede situada sobre la parte alta del espectro de la señal de vídeo. En concreto, el valor seleccionado es $k=283$, con lo que obtenemos una frecuencia definitiva de

$$f_{PAL} = 283,75 \cdot f_L + 25 \text{ Hz} \quad (3.90)$$

que sustituyendo por el valor de la frecuencia de línea de 15625 Hz, obtenemos:

$$f_{PAL} = 4.433.618,75 \text{ Hz} \quad (3.91)$$

valor que debe obtenerse con una precisión de entre $+/- 1$ Hz o $+/- 5$ Hz en función de la variante del sistema PAL que se utilice en la modulación de la señal.

El patrón interferente en el dominio del tiempo también va cambiando de forma parecida a como ocurría en el sistema NTSC. Podemos determinar el número de ciclos de la frecuencia portadora que habrán transcurrido después de una imagen completa multiplicando la frecuencia portadora por el tiempo que transcurre entre dos imágenes. El resultado es de

$$\text{ciclos portadora/imagen} = \frac{f_{PAL}}{25} = 177344,75 \text{ ciclos} \quad (3.92)$$

Lo que significa que el patrón interferente se desplaza un cuarto de periodo en cada imagen. El patrón se repetirá, por tanto, cada cuatro imágenes completas, lo que también significa que la fase relativa entre los sincronismos y la subportadora de color coincide y se repite con una periodicidad de cuatro imágenes, o como suele expresarse habitualmente, con una periodicidad de ocho campos. En efecto, el número de ciclos de la portadora de color después de 8 campos es un entero:

$$\text{ciclos portadora/8 campos} = 4 \cdot \frac{f_{PAL}}{25} = 709379 \text{ ciclos} \quad (3.93)$$

Esto significa que, si deseamos editar para corte la señal PAL, deberemos tener la precaución de realizar los cortes con un número de campos que sea múltiplo de 8. En caso contrario, deberemos regenerar los sincronismos y la portadora de color. La inserción directa de cortes de imagen fragmentando la secuencia de 8 campos producirá la pérdida de sincronismo de la subportadora de croma con la eventual interrupción de la información de color en la pantalla. Este efecto se observa a menudo en videos domésticos cuando se interrumpe la grabación de la escena y se continua posteriormente con otra escena. Durante el inicio de la nueva secuencia se produce una pérdida del sincronismo de la señal de color que se traduce en que, momentáneamente, la imagen se visualiza en blanco y negro.

La secuencia de 8 campos de la señal PAL significa que durante el registro de señales para directo o estudio, las distintas cámaras que intervienen deben estar sincronizadas con la misma portadora de color, ya que, en caso contrario, la conmutación entre distintas cámaras produciría la pérdida de sincronismo de la señal de color. Para ello, es necesario proporcionar una señal de sincronismo global para que todas las cámaras funcionen de forma sincrona. Las cámaras profesionales disponen de una entrada específica para los sincronismos. En aplicaciones semiprofesionales puede utilizarse una cámara maestra que proporciona las señales de sincronismo al resto de cámaras con las que se realiza el registro de la escena. En el NTSC existe un problema parecido, pero, en este caso, la frecuencia de repetición es de 4 campos.

A menudo la frecuencia de la subportadora se expresa en función de la frecuencia de línea por lo que su expresión queda como:

$$f_{PAL} = \left(283 + \frac{3}{4} + \frac{1}{625} \right) \cdot f_L = 283,7516 \cdot f_L \quad (3.92)$$

Otro aspecto fundamental en la selección de la frecuencia portadora de la información de croma es, como ya hemos analizado para el NTSC, las interferencias que pueden producirse como consecuencia de las no linealidades del receptor y la aparición de productos de intermodulación entre la subportadora de color y la subportadora de audio. En este caso, la subportadora de audio se sitúa sobre los 5,5 MHz, con lo que el único producto de intermodulación que entra dentro de la banda de vídeo es la diferencia entre la portadora de audio y de color. La frecuencia de modulación de audio se ha seleccionado para que esta diferencia de frecuencias produzca un patrón interferente con una visibilidad muy reducida. La diferencia entre ambas frecuencias puede expresarse en función de la frecuencia de línea con una expresión muy parecida a la que se obtiene para la frecuencia subportadora de color, por lo que los patrones interferentes son muy parecidos:

$$f_I = 5,5 \text{ MHz} - f_{PAL} = 1,006638125 \text{ MHz} = 68,2484 f_L = 68,25 \cdot f_L - 25 \quad (3.93)$$

3.4.5 Espectro de la señal PAL

Desde el punto de vista espectral, la alternancia de signo de una de las componentes de color tiene una incidencia importante en la imbricación entre las componentes de luminancia y croma. Tal y como hemos visto para el sistema NTSC, las señales de croma en banda base tienen una transformada de Fourier parecida a la de la señal de luminancia y en la que la energía se concentra en los múltiplos de la frecuencia de línea. Las modulaciones en cuadratura de la componente $U'(t)$ y $V'(t)$ pueden analizarse por separado. La componente $U'(t)$ está multiplicada por una función seno que simplemente desplaza el espectro de $U'(t)$ a la frecuencia de la subportadora de color. El análisis es análogo al que hemos realizado para el NTSC y no merece la pena repetirlo. Las componentes espectrales de $U'(t)$ quedan imbricadas con las de luminancia y se centran en las frecuencias

$$f_k = 283,7516 \cdot f_L \pm k \cdot f_L \quad (3.94)$$

En lo que respecta a la componente $V'(t)$, el análisis es algo más complejo, puesto que debe tenerse en cuenta la alternancia de signo de esta componente. Por lo tanto, es necesario separar el análisis del espectro de la señal $V'(t)$ en banda base, con la alternancia de signo incluida, y posteriormente introducir el efecto de la modulación en la componente coseno.

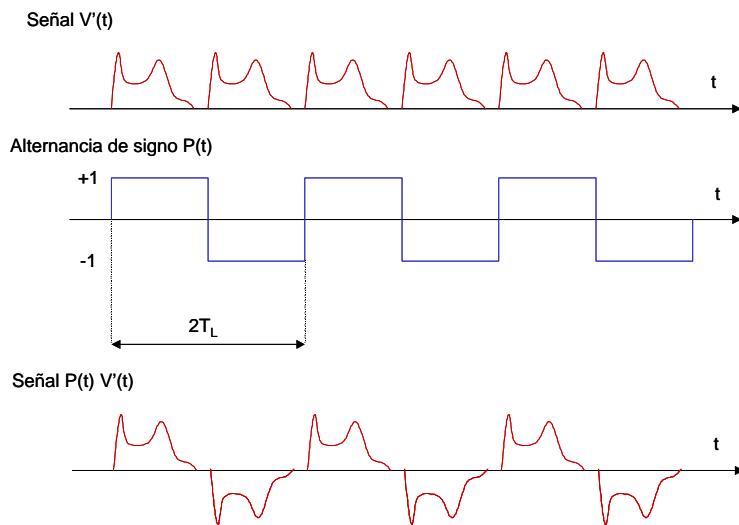


Fig. 3.47 Construcción de la señal $V'(t)$ con la alternancia de signo

En la figura 3.47 se muestran las formas de onda de la componente $V'(t)$ y de la función que realiza la alternancia de signo. Se trata de una función cuadrada con una frecuencia fundamental $f_L/2$, por lo que su espectro estará formado por una serie de rayas espectrales situadas en la frecuencia fundamental de

la señal y en sus múltiplos impares. La transformada de Fourier del producto entre la componente $V'(t)$ y esta señal cuadrada puede determinarse de forma muy simple. La componente $V'(t)$ tiene una serie de rayas espectrales situadas en la componente continua y en los múltiplos de la frecuencia de línea. El resultado de convolucionar este espectro con el de la señal cuadrada es que las nuevas componentes espectrales se sitúan sobre los múltiplos impares de la mitad de la frecuencia de línea. El espectro del producto de las dos señales en banda base se representa en la figura 3.48.

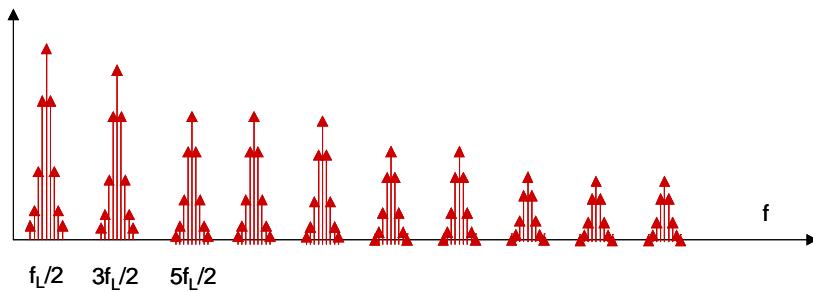


Fig. 3.48 Espectro de la señal $\pm V'(t)$

Ahora falta considerar el efecto de la modulación mediante la portadora en coseno, que no es más que un desplazamiento a la frecuencia subportadora de color de todo el espectro de la señal $\pm V'(t)$. Las componentes espectrales quedan por tanto situadas sobre las frecuencias

$$f_k = 283,7516 \cdot f_L \pm (2k + 1) \cdot \frac{f_L}{2} \quad (3.95)$$

Por tanto, en el sistema PAL, las componentes espectrales de la señal $U'(t)$ y las de la señal $V'(t)$ se sitúan en frecuencias distintas que están imbricadas alrededor de las componentes de luminancia. En la figura 3.49 se representa la situaciónpectral de todas estas componentes en una parte del espectro situado en las proximidades de la frecuencia subportadora de color.

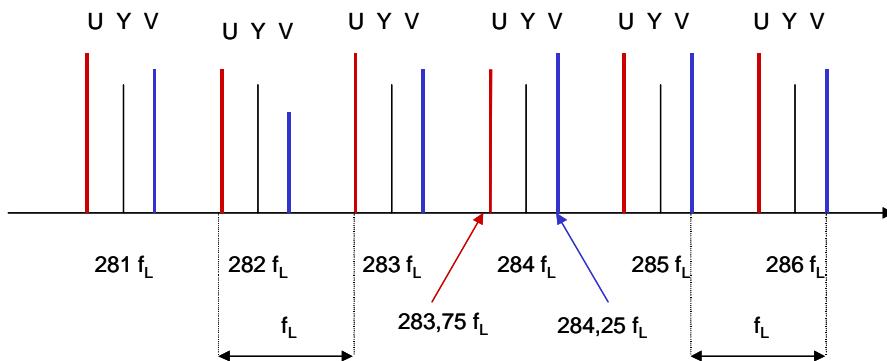


Fig. 3.49 Imbricación espectral entre las componentes de luminancia, $U'(t)$ y $V'(t)$

Recordemos que las componentes de croma en banda base han sido filtradas paso bajo con un ancho de banda de aproximadamente 1,3 MHz. Esto significa que al modularlas en cuadratura se desplazarán a la frecuencia central de la subportadora de color, ocupando, tal y como se muestra en la figura 3.50, un ancho de banda que se sitúa entre los 3,1 MHz y los 5,7 MHz. Este ancho de banda es excesivo para su transmisión en los canales de radiodifusión terrena, por lo que, como en el caso del NTSC, se aplica un filtrado de las componentes de croma que superan los 5 MHz, obteniendo una modulación con banda lateral superior parcialmente suprimida. Ya hemos analizado en el caso del NTSC que, para recuperar las componentes de croma, pueden aplicarse las mismas estructuras de demodulación que se aplicarían cuando no se suprime la banda lateral superior. La diferencia es una reducción de nivel en las componentes de alta frecuencia que resulta poco visible para el usuario, pues no puede comparar con el contenido original de la señal. Además, en el caso del PAL, las dos componentes de croma tienen el mismo ancho de banda, por lo que todos los filtros del demodulador pueden ser idénticos, facilitando el equilibrio en el tratamiento de las dos componentes.

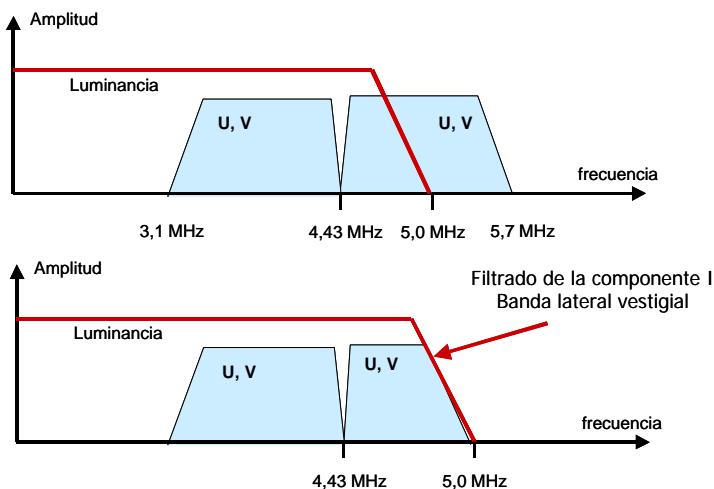


Fig. 3.50 Filtrado de la banda lateral superior para la transmisión de la señal PAL en canales de RF

3.4.6 Generación de sincronismos en el sistema PAL

Hemos visto que es muy importante que la relación entre la frecuencia subportadora de color y la frecuencia de línea sean exactas. Para garantizar que todas las señales de sincronismos de línea, campo e imagen estén perfectamente sincronizadas con la subportadora de color suele emplearse un único oscilador maestro, a partir del cual se obtienen el resto de referencias. El sistema maestro suele ser el oscilador de subportadora de croma, del que se deriva una señal con una frecuencia doble a la frecuencia de línea. A partir de esta última señal pueden obtenerse, como hemos visto en la sección 3.3.3, el resto de sincronismos.

En la figura 3.51 se muestra un diagrama esquemático de cómo puede obtenerse una señal con una frecuencia doble a la de línea a partir de la frecuencia del oscilador de croma. El procedimiento es algo más complejo que en el NTSC debido a que la relación 3.92 entre la subportadora de color y la

frecuencia de línea es también más compleja. El primer mezclador desplaza la frecuencia del oscilador 22 veces la frecuencia de línea, por lo que la frecuencia de la señal en su salida es de $(283,7516-22)f_L$. Al aplicar los sucesivos multiplicadores y divisores de frecuencia indicados en la figura, obtenemos una señal cuya frecuencia es el doble de la frecuencia de línea. A partir de esta señal podemos obtener la frecuencia de línea (dividiendo por 2), la frecuencia de campo (dividiendo por 625) y la frecuencia de imagen (dividiendo por 2×625).

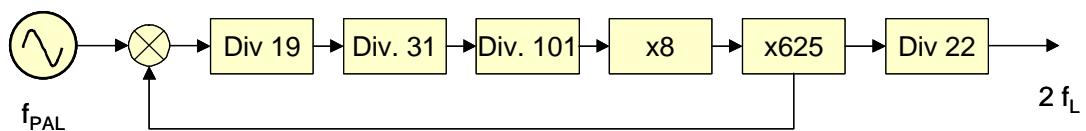


Fig. 3.51 Generación de los sincronismos a partir de la subportadora de color

3.4.7 La señal de salva en el sistema PAL

La señal de salva de color tiene en el sistema PAL una doble función. Por una parte, igual que en el NTSC, permite que el receptor se sincronice con la portadora de color utilizada en el transmisor. La segunda función de la señal de salva es proporcionar información precisa sobre el signo en la componente $V'(t)$ con el que se recibe cada línea. Para ello, la señal de salva incluye unos pocos períodos de la subportadora de color con dos desfasajes distintos en función del tipo de líneas. En el caso de líneas NTSC, la salva se transmite con un desfasaje de 135° respecto a la subportadora utilizada en el transmisor. En el caso de líneas del tipo PAL el desfasaje es de -135° . Las dos posibilidades se muestran en la figura 3.52.

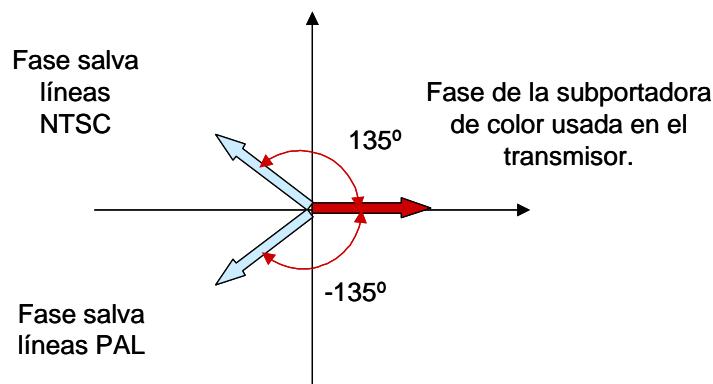


Fig. 3.52 Fase de la salva de color para identificar las líneas NTSC y PAL

Los dos posibles desfasajes se generan en el transmisor utilizando las mismas señales que se aplican a los moduladores de las componentes de croma. Para ello, como ya hemos visto en el sistema NTSC, cada una de las señales moduladoras se ponderan por unos coeficientes y se combinan para obtener los desfasajes deseados de 135° y -135° . Estas señales se insertan en la señal de vídeo compuesto en el

pórtico posterior del sincronismo de línea, de modo análogo a como se realizaba en el sistema NTSC. La inserción de la fase adecuada debe concordar con el signo con el que se modula la componente $V'(t)$.

Las características de amplitud y posición temporal de la salva en el sistema PAL son parecidas a las que hemos descrito para el sistema NTSC, y que han sido representadas en la figura 3.19.

Para sincronizar el oscilador del receptor con la subportadora de color suele utilizarse un circuito que responde al diagrama de bloques de la figura 3.53. La señal de salva y la señal procedente del oscilador local se comparan durante el periodo de tiempo en que se transmite la salva (puerta de salva) y cuya señal de control se obtiene a partir de las señales de sincronismo de línea. El bucle de control del VCO es un PLL convencional en el que la constante de tiempo del integrador es lo suficientemente grande como para que el oscilador sólo pueda adaptarse a la fase de la señal de salva después de varias líneas. El sistema sólo actúa de forma realimentada durante los instantes en los que está presente la salva de color. El resto del tiempo el oscilador funciona de forma autónoma. Como la señal de salva se transmite con cambios de fase de 135° y -135° en líneas sucesivas, el oscilador local converge al valor de fase medio que se está recibiendo, es decir, 180° . Por tanto, el circuito de sincronización con la portadora de color es idéntico al que se utiliza en el sistema NTSC, con la salvedad de que debe garantizarse que las reacciones del sistema sean lo suficientemente lentas como para integrar los cambios sucesivos de fase de la salva.

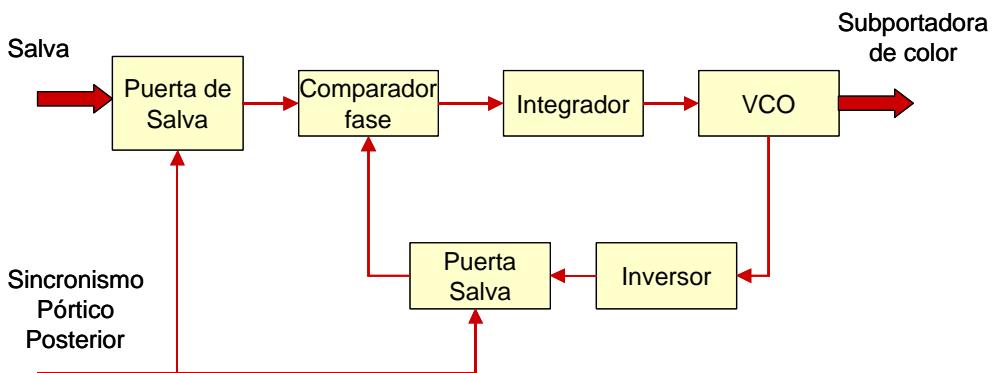


Fig. 3.53 Circuito de sincronización del receptor con la subportadora de color

Una vez obtenida la sincronización del oscilador local con la subportadora de color de la señal recibida, es relativamente simple determinar la fase concreta con la que se recibe la señal de salva para actuar sobre el signo de los demoduladores del decodificador de croma. Para ello, basta con realizar una mezcla entre la señal de salva que se está recibiendo y la subportadora de color que hemos regenerado, aplicando un filtro paso bajo a la señal producto resultante. El signo de la señal en la salida del filtro indicará si se trata de una línea NTSC o una línea PAL. Existen otras variantes tecnológicas para regenerar la subportadora de color y para determinar el signo empleado en cada una de las líneas. Las variantes que hemos descrito son las más simples desde el punto de vista conceptual y se utilizan en un buen número de receptores.

3.4.8 Diagramas de bloques de un codificador y un decodificador PAL

En la figura 3.54 se muestra, a modo de resumen, un diagrama global de todos los elementos que intervienen en la confección de la señal de vídeo en el sistema PAL. La mayor parte de los elementos ya han sido descritos con detalle en apartados anteriores. La señal $P(t)$ que se deriva del generador de sincronismos determina el signo que debe aplicarse a la componente $V'(t)$ y que debe estar sincronizado con la señal que se envía como referencia de fase. Los retardos que se indican en la figura suelen obtenerse, como ya hemos descrito, mediante combinaciones lineales entre las dos portadoras en cuadratura. En la figura no se muestra la inserción de la señal de audio ni se indica el filtrado paso bajo que se aplica a las componentes de croma para obtener la modulación en banda lateral superior vestigial. Normalmente este filtrado se realiza en la etapa previa a los moduladores de radiofrecuencia.

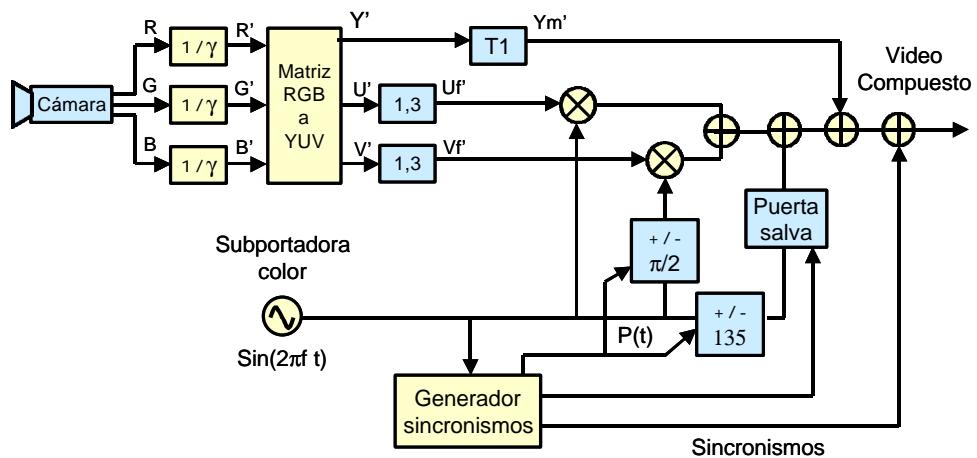


Fig. 3.54 Diagrama de bloques general de un codificador de vídeo en el sistema PAL

Por lo que respecta al decodificador, puede utilizarse el mismo diagrama de bloques que hemos comentado al describir el sistema NTSC y que se ha representado en la figura 3.31. Las variantes más significativas del sistema PAL respecto al NTSC son la estructura del propio decodificador de croma que requiere, además de las dos componentes de la subportadora de color en cuadratura, la señal $P(t)$ para determinar en qué líneas se ha realizado la inversión en la componente $V'(t)$.

Los circuitos de regeneración de sincronismos son también distintos en los dos sistemas de transmisión. La separación de las componentes de luminancia y croma puede realizarse mediante filtros del tipo muesca o filtros en peine. Esta última estrategia puede usarse en el sistema PAL para separar las componentes $U'(t)$ y $V'(t)$ antes de entrar al demodulador de cromo. En efecto, hemos visto que en el caso del PAL las componentes de croma moduladas quedan situadas en posiciones espectrales distintas, por lo que pueden ser separadas mediante el uso de filtros en peine que seleccionen cada una de ellas de forma individual. Esta opción garantiza una mejor separación entre las dos componentes de color y la desaparición prácticamente total de los efectos de cruce de color (*cross-color*).

3.5 Normas utilizadas en la transmisión de señales de TV en color analógicas

A pesar de que sólo existen tres sistemas básicos para la obtención de señales en vídeo en color compatibles, el ITU-R (anteriormente CCIR) establece distintas normas en las que se especifican el número de líneas, los anchos de banda de los canales de radiofrecuencia, el ancho de banda de la señal, la separación entre la portadora de audio y vídeo, el tipo de modulación que se aplica a la señal de audio, etc. Estas normas se denominan con las letras A a N y sus características básicas se detallan en la tabla 3.2.

Tabla 3.2 Normas de señales de televisión analógicas para la radiodifusión terrestre

Norma	Número de líneas	Ancho de banda canal (MHz)	Ancho de banda de vídeo (MHz)	Portadora de sonido respecto a imagen (MHz)	Banda Lateral Vestigial (MHz)	Modulación de vídeo	Modulación de audio
A	405	5	3	-3,5	0,75	Pos.	AM
B	625	7	5	+5,5 (+5,742)	0,75	Neg.	FM
C	625	7	5	+5,5	0,75	Pos.	AM
D	625	8	6	+6,5	0,75	Neg.	FM
E	819	14	10	+11,15	2	Pos.	AM
F	819	7	5	+5,5	0,75	Pos.	AM
G	625	8	5	+5,5	0,75	Neg.	FM
H	625	8	5	+5,5 (+5,742)	1,25	Neg.	FM
I	625	8	5,5	+6	1,25	Neg.	FM
J	625	8	6	+6,5	0,75	Neg.	FM
K	625	8	6	+6,5	1,25	Neg.	FM
L	625	8	6	+6,5	1,25	Pos.	AM
M	525	6	4,2	+4,5	0,75	Neg.	FM
N	625	6	4,2	+4,5	0,75	Neg.	FM

En principio las normas anteriores son independientes del sistema de codificación del color y sólo establecen los criterios de anchos de banda asignados al vídeo, separación entre canales y formas de modulación de la información de audio y vídeo. Así, en Marruecos utilizan la norma H para la transmisión de señales en sistema SECAM en la banda de UHF, mientras que en Malta se utiliza la misma norma H, pero con el sistema de codificación de color PAL. No obstante, algunas normas sólo pueden utilizarse con un sistema de codificación del color. Uno de estos ejemplos es la norma M que se utiliza en Estados Unidos y Japón.

En España se utiliza la norma G para la transmisión de señales codificadas en PAL en la banda de UHF. Hasta hace poco tiempo también se transmitía en la banda de VHF con la norma B, pero actualmente ya ha desaparecido. Comentemos con algún detalle los parámetros especificados en la norma G, cuya interpretación puede generalizarse fácilmente para el resto de normas.

Se trata de un sistema de 625 líneas (que puede ser usado para transmitir PAL o SECAM, así en Grecia también se usa la norma G pero la señal está codificada en SECAM). El ancho de banda asignado a un canal de televisión, que incluye el vídeo, el audio y las guardas, es de 8 MHz. Esto significa que el resto de canales en la misma banda de frecuencias estarán separados por 8 MHz.

Por otra parte, el ancho de banda de la señal de vídeo en el sistema G es de 5 MHz. Esto significa que antes de realizar la modulación de la señal de vídeo a radiofrecuencia se utiliza un filtro paso bajo de 5 MHz. Este filtro es el que elimina la parte de la información de luminancia que excede los 5 MHz de ancho de banda y el que elimina parte de las bandas laterales superiores de la información de croma.

La portadora de audio se sitúa a 5,5 MHz por encima de la portadora de vídeo y está modulada en FM. La inserción de la señal de audio se realiza, habitualmente, cuando la señal de vídeo en banda base ya ha sido previamente filtrada por el filtro de 5 MHz. A la portadora de audio se le asigna una potencia que está situada unos 10 dB's por debajo de la potencia de la portadora de vídeo.

La señal de vídeo se modula en AM con banda lateral inferior vestigial. Esto significa que, una vez modulada la señal a la frecuencia de transmisión, se aplica un filtro paso banda que elimina parte de la banda lateral inferior. En el sistema G, se permite que el ancho de banda asignado a esta banda inferior sea de 0,75 MHz. Esto supone que el ancho total del vídeo es de 5,75 MHz (5MHz de la banda lateral superior y 0,75 MHz de la inferior). La modulación de la señal de vídeo es negativa, lo que significa que la señal de vídeo compuesto se invierte (impulsos de sincronismo positivos y luminancia negativa) antes de enviarla al modulador de AM. En la figura 4.55 se representa un diagrama esquemático de cómo se sitúan las portadoras de vídeo, de croma y de audio analógico en la banda de 8 MHz asignada a cada canal. En esta figura también se indica la portadora de audio digital NICAM, que está situada a 5,85 MHz de la portadora de vídeo y que se transmite con una potencia de 20 dB por debajo. La señal NICAM es un sistema de audio digital estereofónico que se introdujo a mediados de la década de los 80 para proporcionar sonido de alta calidad. El NICAM admite la radiodifusión de programas estereofónicos o la transmisión de programas en dual (canal principal monofónico más un canal auxiliar con versiones originales o comentarios). También existe una alternativa analógica al NICAM, denominada Zweiton, que se utiliza habitualmente por algunas cadenas en España, en la que se transmite un único canal auxiliar que puede utilizarse para decodificar una señal estéreo o la versión original del programa.

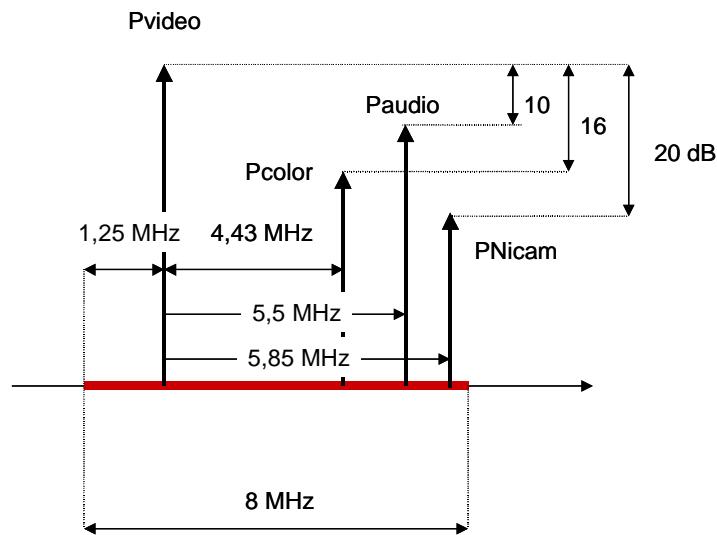


Fig. 3.55 Distribución de portadoras y niveles en un canal de 8 MHz

En la tabla 3.3 se indican las normas y sistemas de color utilizados en diferentes países. La lista incluye las normas utilizadas tanto en las bandas de VHF (actualmente en fase de abandono en un gran número de países; en España, esta banda ya no está en uso desde Noviembre de 1999) como UHF. La lista no es completa y sólo incluye algunos países representativos de cada uno de los sistemas y normas.

Tabla 3.3 Lista de países y normas de TV

País	VHF	UHF	Sistema	País	VHF	UHF	Sistema
Argentina	N	N	PAL	Korea	M		NTSC
Australia	B	H	PAL	Luxemburg	C	L	PAL/SEC
Austria	B	G	PAL	México	M	M	NTSC
Bélgica	B	H	PAL	Mónaco	E	L	SECAM
Bulgaria	D	K	SECAM	Marruecos	B	H	SECAM
China	D	K	PAL	Noruega	B	G	PAL
Chequia	D	K	SECAM	Polonia	D	K	SECAM
Dinamarca	B	G	PAL	Portugal	B	G	PAL
Egipto	B	G,H	SECAM	Suecia	B	G	PAL
Francia	E/L	L	SECAM	España	B	G	PAL
Alemania	B	G	PAL	Suiza	B	G	PAL
Gibraltar	B	H	PAL	Turquía	B	G	PAL
Gran Bret.	I	I	PAL	USA	M	M	NTSC
Italia	B	G	PAL	Malta	B	H	PAL
Japon	M	M	NTSC	Túnez	B	G	SECAM

Actualmente, las bandas definidas para la radiodifusión terrestre de señales de televisión o audio están distribuidas en unos márgenes de frecuencias denominados bandas y que se definen como:

- *Onda Larga (OL)*: 150-285 KHz
- *Onda Media (OM)*: 535-1605 KHz
- *Onda Corta (OC)*: 3,95-26,1 MHz
- *TV Banda I (VHF)*: 41-68 MHz, ancho de canal de 7 MHz
- *Banda II(VHF-FM)*: 87,5-108 MHz, ancho de canal de 100 KHz (emisoras audio FM comercial)
- *TV Banda III(VHF)*: 174-223 MHz, ancho de banda de canal de 7 MHz.
- *TV Banda IV(UHF)*: 470-606 MHz, ancho de banda de canal de 8 MHz.
- *TV Banda V(UHF)*: 606-854 MHz, ancho de banda de canal de 8 MHz.

Las bandas que actualmente se utilizan en España son la IV y la V, donde están definidos hasta un total de 49 canales de televisión, que se numeran del 21 al 69. En la tabla siguiente se proporcionan los márgenes de frecuencias, la frecuencia portadora de vídeo, de croma y de sonido de algunos de estos canales. Evidentemente, pueden obtenerse fórmulas cerradas que determinen estas frecuencias en función de los canales.

Tabla 3.4 Tablas de canales de frecuencia

Banda	Canal	Frecuencia Canal (MHz)	Portadora Vídeo	Portadora Sonido	Subportadora color
IV	34	574-582	575,25	580,75	579,68
IV	35	582-590	583,25	588,75	587,68
IV	36	590-598	591,25	596,75	595,68
IV	37	598-606	599,25	604,75	603,68
V	38	606-614	607,25	612,75	611,68
V	39	614-622	615,25	620,75	619,68
V	40	622-630	623,25	628,75	627,68
V	41	630-638	631,25	636,75	635,68
V	42	638-646	639,25	644,75	643,68
V	43	646-654	647,25	652,75	651,68

4 Televisión digital

4.1 Introducción

La transmisión de señales de televisión en formato digital ha supuesto un cambio significativo tanto en el ámbito tecnológico como en lo que respecta a la producción de programas y servicios que se ofrecen al espectador. Los modernos canales digitales ofrecen multitud de programas en un mismo paquete de televisión y han introducido nuevos conceptos como el pago por canal (*pay per channel*), el pago por programa (*pay per view*), la reemisión periódica de los programas en diferentes franjas horarias, canales temáticos, canales guía, etc. Desde el punto de vista tecnológico, la principal ventaja de la televisión digital es que la codificación de la información de audio y vídeo puede transmitirse en un ancho de banda menor que el empleado por los sistemas analógicos. Junto con la información convencional pueden transmitirse datos de tipo texto sobre el programa (subtítulos o resumen), codificar la señal de audio en estéreo o multicanal, codificar el programa en varios idiomas. Además, la calidad de imagen y sonido es superior, debido a que la naturaleza digital de las señales les proporciona cierto nivel de protección frente al ruido.

El objetivo de este capítulo es proporcionar los principios matemáticos y tecnológicos sobre los que se basa la televisión digital actual. En el proceso de transmisión y recepción de la televisión digital intervienen un gran número de subsistemas entre los que destacan:

- 1) *Sistemas de digitalización de la señal de audio y vídeo.* Entenderemos que se trata de sistemas que digitalizan ambas señales a partir de señales analógicas y que obtienen un formato PCM convencional sin comprimir. En el caso de la señal de audio, el formato PCM es parecido al del sistema *Compact Disc*, con una frecuencia de muestreo de 44,1KHz por canal y 16 bits por muestra. Para la señal de vídeo el formato digital de partida suele ser alguna variante del formato ITU 601 (normalmente 4:2:2 o 4:2:0).
- 2) *Sistemas de compresión / descompresión de la información.* La señal de vídeo en formato ITU-601 tiene un gran volumen de datos que hacen inviable su transmisión directa. La compresión de esta información es la etapa más importante en la transmisión de vídeo digital, ya que determina la calidad final de las imágenes que se reproducen y establece la eficiencia espectral del sistema de transmisión. Actualmente, el formato de codificación utilizado por la mayoría de proveedores es el MPEG-2, tanto en el caso de audio como en el de vídeo, y se prevé que permanezca en el mercado durante un periodo prolongado. Este estándar de codificación combina distintas estrategias, basadas tanto en

la redundancia estadística de los datos como en las características de percepción del sistema visual humano. Está específicamente diseñado para proporcionar una calidad de la señal de vídeo suficiente para su transmisión como señal de televisión. El análisis de la señal para su codificación en tiempo real es extremadamente complejo y se requiere de un hardware dedicado a estas funciones cuyo precio es relativamente elevado. La decodificación se realiza sin que el receptor deba tomar decisiones relevantes, por lo que se trata de sistemas más simples y relativamente económicos. La tasa de bits final que se consigue depende del tipo de señal de vídeo, pudiendo estar entre 2 Mbits/s o 9Mbits/s, lo que significa una compresión que oscila entre un factor de 15 a 40 respecto al formato 4:2:0.

- 3) *Multiplexación / Demultiplexación de señales de vídeo y audio y varios programas.* La trama de bits asociada a un canal puede estar formada por varios programas, cada uno de los cuales está constituido por una fuente de vídeo y uno o varios canales de audio y datos. La información de audio y datos debe intercalarse entre la información de vídeo para poder mantener una sincronía perfecta durante la reproducción. Las especificaciones y normativas para realizar esta multiplexación de las señales están especificadas en el estándar MPEG2 Sistemas.
- 4) *Transmisión / Recepción de las señales.* La transmisión de las señales de televisión digitales está normalizada (en Europa y otras áreas geográficas) por la organización DVB (*Digital Video Broadcasting*). En países como Estados Unidos o Japón, donde el volumen de negocio vinculado a la televisión es muy importante, han aparecido sistemas propietarios ofrecidos por las plataformas productoras en los que se modifica tanto la codificación de la señal como, sobre todo, su formato de transmisión. El DVB cubre toda la normativa de transmisión por satélite, cable y terrena y establece los procedimientos utilizados para el acceso a los programas, códigos de protección, sistemas de modulación, etc.

De todos estos aspectos nos concentraremos fundamentalmente en el proceso de compresión y descompresión de la señal de vídeo, aunque también revisaremos brevemente los conceptos de multiplexación de vídeo y audio y veremos los principios de los sistemas de DVB. No se considerará la codificación de la señal de audio ni los subsistemas de conversión entre los formatos analógicos y digitales.

En la figura 4.1 se representan las distintas etapas conceptuales que intervienen en el proceso de codificación de la señal de vídeo digital. La *codificación de fuente* consiste en extraer toda la redundancia posible en la señal, reduciendo el número de bits con el que se representa la información sin que ello suponga una pérdida aparente de calidad para el sistema visual humano. Los métodos y estrategias empleados para la codificación de fuente son muy variados, y resulta difícil establecer un procedimiento de compresión óptimo puesto que los resultados dependen, en gran medida, de las características de las señales. Generalmente los métodos de compresión avanzados combinan distintas estrategias simultáneamente para codificar la señal. La *multiplexación* de distintas señales básicas es necesaria para formar una única trama binaria que transporte la información y contenidos de todo el canal definido por la plataforma operadora. Por otra parte, la *codificación de canal* se utiliza para adecuar la forma de onda de las señales a las características del canal. En esta etapa se introducen

códigos de protección frente a errores, códigos de encriptación de la información y se adecua la forma de onda de las señales para que puedan ser directamente utilizadas por la etapa de modulación o transmisión. Esta última etapa se encarga de adaptar los datos a las características del medio de transmisión, para garantizar una correcta recepción de las señales. El medio de transmisión condiciona el tipo de modulación que se utiliza y la codificación de canal. Así, en un medio como el cable, los sistemas de detección y corrección de errores pueden ser más simples que en el caso de la comunicación vía satélite, debido a que la relación señal a ruido es mucho mayor y, por tanto, se reduce la probabilidad de error.

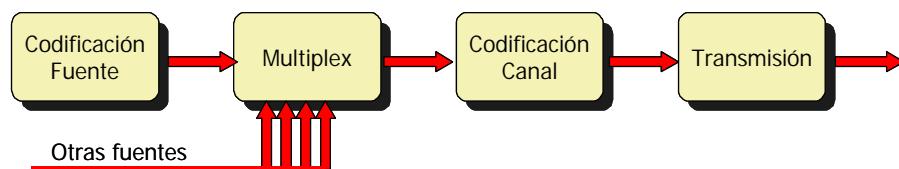


Fig. 4.1 Etapas en la codificación y transmisión de una señal de vídeo digital

Los procedimientos utilizados para la codificación de la señal de vídeo en MPEG2 son bastante complejos y algunos detalles concretos escapan de los propósitos de este texto. Lo mismo ocurre respecto a algunos sistemas de modulación y acceso condicional utilizados en la radiodifusión de la señal. Nuestro objetivo es presentar los conceptos de forma progresiva y que el lector comprenda las ideas generales que subyacen en el sistema de televisión digital para que domine sus principios y conozca sus limitaciones. Por ello, se omiten algunos detalles y formalismos matemáticos que no resultan cruciales en la comprensión y filosofía del sistema.

En la primera parte de este capítulo se introduce la necesidad de comprimir la señal de vídeo digital para posibilitar su transmisión. Se presentan los conceptos básicos de codificación por diferencias y redundancia estadística de los datos. Ambos procedimientos se utilizan en algunos módulos de la codificación de vídeo mediante MPEG2. Posteriormente, se describe la codificación de imágenes estacionarias en formato JPEG. La filosofía del sistema JPEG resulta fundamental para comprender la codificación de imágenes en movimiento y es utilizada en el formato MPEG, con ligeras variantes, para codificar algunos fotogramas de la señal de vídeo. En la siguiente sección se desarrollan los principios del formato MPEG-1, precursor del MPEG-2 en aplicaciones multimedia, y con el que comparten muchos aspectos comunes. Una vez definido el formato MPEG1, resulta bastante simple concentrarse en los aspectos adicionales que proporciona el MPEG2 y que resultan necesarios para obtener una calidad suficiente para su aceptación como señal de televisión. También se comentan brevemente las diferencias entre estos formatos de codificación de vídeo y los formatos H-261 y H-263 utilizados en la transmisión de señales de videoconferencia mediante líneas telefónicas o de datos.

En las últimas secciones se desarrollan algunos aspectos relativos a la multiplexación de señales de audio, vídeo y datos y se establecen los principios de transmisión de las señales de televisión digital por cable, satélite o redes terrenas de acuerdo con la normativa DVB, aportando algunos detalles sobre la codificación de canal de las señales y los métodos de acceso condicional a los programas.

4.2 Compresión de imágenes: necesidad y conceptos básicos

La digitalización de señales analógicas aporta considerables ventajas entre las que pueden citarse la protección frente a ruidos, la facilidad para encriptar las señales, la posibilidad de procesar digitalmente los datos, etc. No obstante, por lo general, la digitalización directa supone un aumento considerable del ancho de banda. En el caso de señales de vídeo analógicas, hemos visto que un ancho de banda de 5 MHz resulta suficiente para mantener una resolución espacial y tasa de refresco de imágenes aceptable. En cambio, la digitalización de la secuencia de vídeo en un formato no comprimido, como el 4:2:0, requiere transmitir una tasa de bits de $124,4 \cdot 10^6$ bits/s (véanse los cálculos en el capítulo 1). Evidentemente, aunque se usen códigos y modulaciones digitales de gran eficiencia espectral, el ancho de banda necesario para transmitir esta tasa de bits es muy superior a los 5 MHz que requiere el sistema analógico. Por esta razón, los formatos digitales derivados del ITU 601 se han mantenido como estándares digitales para el intercambio de información entre equipos y terminales en estudios de grabación y producción de vídeo y no se han extendido a sistemas de radiodifusión de la señal de vídeo.

Tomando estos formatos digitales como punto de partida, es necesario aplicar un proceso de reducción de la tasa de bits que permita la radiodifusión de las señales sin una pérdida aparente de calidad. Un sistema de compresión de información suele estar formado por dos etapas que se ilustran en la figura 4.2. La transformación de los datos es un procedimiento genérico que se utiliza para representar la información en una forma alternativa y en la que, en principio, resulta más evidente la redundancia existente en los datos originales. Es necesario que esta transformación sea invertible, es decir, que a partir de los datos transformados podamos recuperar exactamente la información original.

Consideremos como ejemplo la aplicación de una transformada de Fourier a las señales que representan la información de un fragmento de audio. Es posible que resulte más eficiente codificar los datos resultantes de esta transformación que los datos originales. En efecto, en el dominio transformado de Fourier es más simple identificar las regiones espectrales que tienen una energía por debajo del umbral de audición, o analizar las bandas que resultarán enmascaradas por otras señales y que por lo tanto no serían audibles. Si este análisis de la señal es factible, no tendrá sentido codificar toda aquella información que el sistema auditivo humano es incapaz de escuchar y resulta más rentable dedicar los bits a las zonas audibles. El resultado final es que en este dominio transformado podemos codificar la información de forma más eficaz. No obstante, debemos recordar que al final de la cadena de codificación / decodificación, deben presentarse los datos en el mismo formato de entrada para que puedan ser correctamente interpretados por el sistema auditivo. Esto significa que deberemos aplicar la transformación inversa sobre los datos codificados para devolverlos a su dominio de representación original.

La transformación de los datos puede ser de naturaleza muy distinta, o incluso, no resultar necesario si existen procedimientos eficientes para la compresión directa. En algunos casos, es posible que existan varias transformaciones en paralelo o en serie y que la codificación se realice tomando datos parciales de cada una de ellas. Veremos que este es el caso del sistema MPEG-2, en el cual se combinan distintas estrategias de codificación en función del tipo de información y de las características de la secuencia de vídeo. En cualquier caso, es fundamental mantener el esquema simplificado de la figura 4.2 en mente, teniendo siempre en cuenta que es necesario, independientemente de lo complejas que

sean las transformaciones que utilizamos para optimizar la codificación, disponer de métodos que permitan restablecer la señal codificada al dominio original.

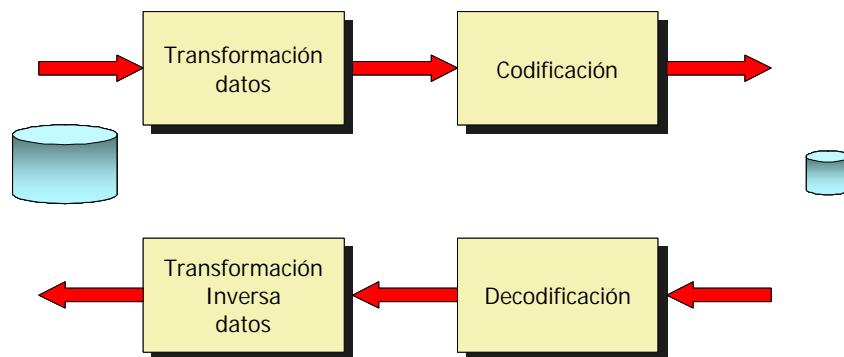


Fig. 4.2 Proceso general de codificación y decodificación de la información

La compresión puede ser con pérdidas (*lossy*) o sin pérdidas (*lossless*) en función de si la información que se recupera coincide exactamente con la original o es sólo una aproximación. Los métodos de codificación sin pérdidas se utilizan sobre todo en aplicaciones de codificación de datos binarios de aplicaciones informáticas en los que es absolutamente necesario recuperar la información original. Los formatos más populares son el zip y el arj. En tratamiento de imágenes los métodos de compresión sin pérdidas encuentran su aplicación en la codificación de imágenes médicas o científicas en las que puede resultar crítico la pérdida de parte de la información. El formato JPEG dispone de una versión de codificación sin pérdidas que suele emplearse en estas aplicaciones.

La compresión con pérdidas es la más habitual en la codificación de señales de vídeo y audio. Evidentemente, las pérdidas son tolerables siempre que la calidad de las señales decodificadas sean aceptables. El principio general sobre el que se sustenta la codificación con pérdidas es que no resulta necesario codificar aquellas componentes de la información que no son observables por los sistemas de percepción humana. Por lo tanto, estos métodos se fundamentan en las características psico-fisiológicas de los sistemas auditivo y visual, que son, en última instancia, los que deben evaluar la calidad del algoritmo de compresión. Por ello, es fundamental comprender las limitaciones y características de estos sistemas de percepción para diseñar codificadores en los que las pérdidas de información resulten poco evidentes o incluso inapreciables. La principal ventaja de estas estrategias de codificación es que consiguen unos factores de compresión muy superiores a los que se obtienen con los métodos sin pérdidas. Además, suelen ser métodos escalables con la aplicación, es decir, el grado de pérdida de calidad que se tolera depende del ámbito al que se destine el codificador. Así, la calidad que se exige en una aplicación de vídeo en directo para internet es muy inferior a la que se exige para la radiodifusión de televisión. Análogamente, un codificador de vídeo para aplicaciones multimedia (MPEG-1) deberá realizar una compresión mucho mayor que un codificador para aplicaciones de televisión (MPEG-2), ya que las tasas de transferencia de bits que admite un CD-ROM son mucho menores que las que admiten los canales de televisión.

4.2 Un ejemplo: la codificación diferencial

En este apartado consideraremos uno de los sistemas más simples e intuitivos que se emplean en la codificación de señales de audio y vídeo. Este método de compresión se utiliza en los formatos JPEG y MPEG para codificar alguno de los coeficientes transformados. Utilizaremos estas ideas para ilustrar los conceptos de codificación con pérdidas y sin pérdidas, así como la transformación de las señales. Además, nos permitirá introducir de forma natural los métodos de compresión basados en el análisis estadístico de los mensajes.

El esquema básico de un codificador diferencial se representa en la figura 4.3. En nuestro caso, la señal de entrada puede ser la luminancia de una línea de la imagen. La transformación utilizada consiste simplemente en calcular la diferencia entre dos muestras consecutivas. En principio, desde un punto de vista intuitivo, esta transformación parece natural debido a la redundancia que existe entre dos elementos consecutivos. Es de esperar que, una vez realizada la transformación, la mayoría de muestras estarán situadas en un entorno próximo al cero, por lo que, seguramente, podemos encontrar compresores que aprovechando esta característica de la señal reduzcan la tasa de bits necesaria para su codificación.

Nótese que la transformación que hemos utilizado es invertible y que podemos recuperar de forma exacta la información original a partir de los resultados transformados. Analíticamente, la transformación diferencial viene dada por:

$$y[n] = x[n] - x[n-1] \quad (4.1)$$

mientras que la transformación inversa se obtiene como:

$$z[n] = y[n] + z[n-1] \quad (4.2)$$

Puede verificarse que una transformación es inversa de la otra sin más que aplicar las ecuaciones anteriores a las tablas de valores que se muestran en la figura 4.3.

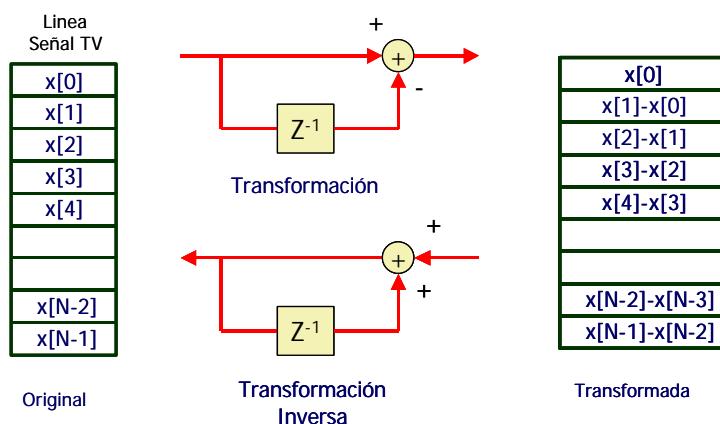


Fig. 4.3 Transformación diferencial

La transformación diferencial puede considerarse un caso particular de la predicción de muestras. Los predictores lineales son uno de los métodos más utilizados para la codificación de señales de audio y vídeo y se basan en estimar el valor de la muestra actual a partir de una combinación lineal de las muestras anteriores. La filosofía general de un predictor se ilustra en la figura 4.4 donde se indica la transformación directa y la inversa. Si la predicción de la muestra es correcta, la señal que se codifica es el error de predicción que, en principio, tendrá un valor próximo a cero, por lo que podrá codificarse de forma bastante eficiente. La ecuación que rige el filtro predictor es:

$$\tilde{x}[n] = \sum_{k=1}^P a_k \cdot x[n-k] \quad (4.3)$$

donde P es el orden del predictor (número de muestras anteriores que se utilizan para realizar la predicción de la muestra actual) y a_k son los coeficientes de predicción. Los coeficientes de predicción pueden calcularse en base a las características estadísticas de la señal.

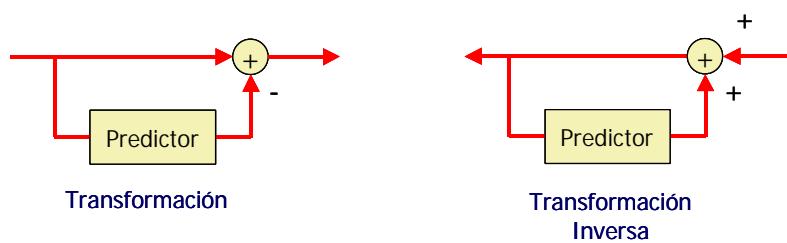


Fig. 4.4 Esquema general de una codificación por predicción de muestras

La codificación diferencial que estamos analizando puede considerarse como un predictor de orden 1, con un coeficiente de predicción igual a la unidad. En apartados posteriores veremos algunos casos particulares de predictores aplicados a la codificación de imágenes donde se utilizan no sólo la muestras anteriores de la línea actual, sino también las muestras de otras líneas.

Los resultados de aplicar el codificador diferencial a una línea de la imagen se muestran en la figura 4.5, donde se representa el nivel de gris de una línea original y los valores resultantes después de la transformación. Obsérvese como los valores transformados se concentran alrededor del cero y que únicamente en los contornos de la imagen se producen niveles significativos en la señal diferencia.

En la figura 4.6 se representan los resultados del codificador diferencial aplicados a una imagen en blanco y negro. La imagen original está codificada con 256 niveles de gris (8 bits por pixel) y se representa asignando el negro al nivel 0 y el blanco al nivel 255. La imagen diferencia se ha representado tomando el nivel 0 como un gris neutro. Los valores de señal negativos son más oscuros que el gris neutro, mientras que los valores positivos son más claros. Esta imagen se ha obtenido aplicando el codificador diferencial a cada una de las líneas de la imagen. A partir de estos resultados visuales, es evidente que hemos conseguido el propósito que perseguíamos. La mayoría de píxeles de

la imagen diferencia están próximos al nivel de gris medio y sólo en los contornos de la imagen aparecen píxeles cuyo valor se aparta apreciablemente del cero.

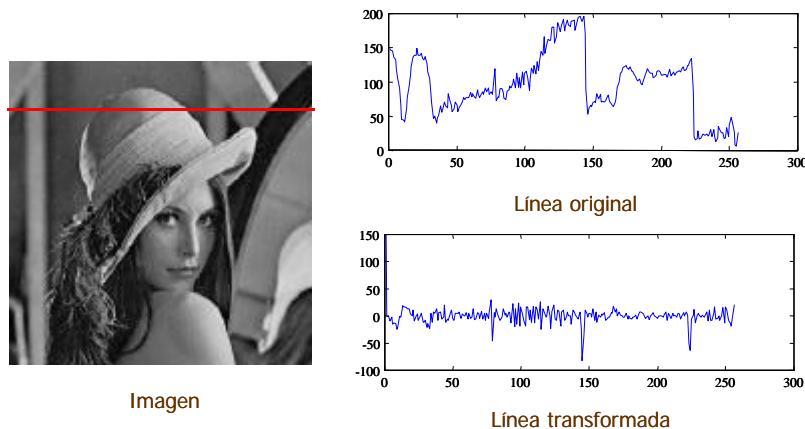


Fig. 4.5 Representación de los niveles de una línea original y de la diferencial



Fig. 4.6 Imagen original e imagen diferencias

Este efecto puede apreciarse más claramente en las gráficas de la figura 4.7, donde se representan los histogramas de la imagen original y de la imagen diferencia. El histograma es una representación bidimensional de la frecuencia relativa con la que aparece cada nivel de gris en la imagen. Así, para cada posible nivel de gris en el eje de abcisas, encontramos el número de píxeles de la imagen que toman ese nivel de gris. Por tanto, el histograma es una representación aproximada de la probabilidad de que un píxel tome un determinado nivel de gris en la imagen. Nótese que en estas gráficas se han normalizado los niveles de gris, de forma que el negro corresponde al cero mientras el blanco corresponde a la unidad. En el histograma de la izquierda, que corresponde a la imagen original, los

niveles de gris se distribuyen dentro de toda la gama sin mostrar preferencia directa por ningún valor. En cambio, en la gráfica de la derecha, que corresponde a la imagen diferencia, existe una gran concentración de elementos de imagen con un nivel de gris próximo a 0.5 (gris neutro con el que hemos representado el valor cero).

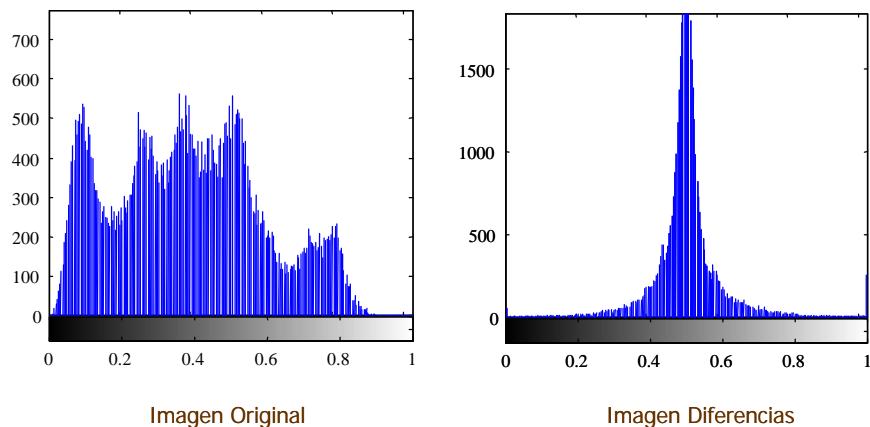


Fig. 4.7 Histogramas de la imagen original y la imagen transformada

Cuando una señal tiene una estadística bien definida, en la que predominan unos pocos valores que se presentan con mucha probabilidad, frente a otros muchos posibles niveles que se presentan con poca probabilidad, suele resultar rentable utilizar códigos de longitud variable en los que los valores más probables se codifican empleando pocos bits, mientras que los valores con poca probabilidad utilizan muchos bits. Esta es la idea básica de los códigos de Huffman y sus variantes, que examinaremos con cierto detalle más adelante.

Es importante subrayar que la codificación que hemos realizado hasta este momento es una codificación sin pérdidas, siempre que mantengamos los cálculos de la diferencia con la precisión adecuada. En este ejemplo, la imagen original estaba codificada con 256 niveles de gris (8 bits por muestra), por lo que las señales diferencia pueden tomar valores comprendidos entre -255 y 255 pues es posible que un píxel tome el valor nulo cuando el anterior era blanco o viceversa. Esto significa que si queremos reconstruir la imagen original con los niveles de gris exactos (sin pérdidas), deberemos mantener los 511 posibles valores que puede tomar la señal diferencia (9 bits por muestra). Si estos valores se aproximan mediante una cuantificación más tosca, ya no será posible recuperar la imagen original de forma exacta.

Cuando se realiza una codificación con pérdidas, es importante seleccionar correctamente la estructura del codificador y el decodificador para que las aproximaciones no se vayan deteriorando a medida que avanza la descodificación. Para ilustrar este problema consideremos el ejemplo de la figura 4.8, en el que se aplica un cuantificador uniforme de 21 niveles directamente en la salida del codificador diferencial. Si la señal está normalizada entre los valores -1 y 1 (que corresponderían con los niveles de gris de -255 y 255), un cuantificador uniforme de 21 niveles significa que los valores originales se

aproximan a la décima (-1, -0.9, -0.8, ..., 0, 0.1,...0.9,1). Si sobre esta aproximación aplicamos el decodificador diferencial, obtenemos la imagen de la figura 4.9, donde puede observarse la presencia de errores significativos en la reconstrucción de la imagen.

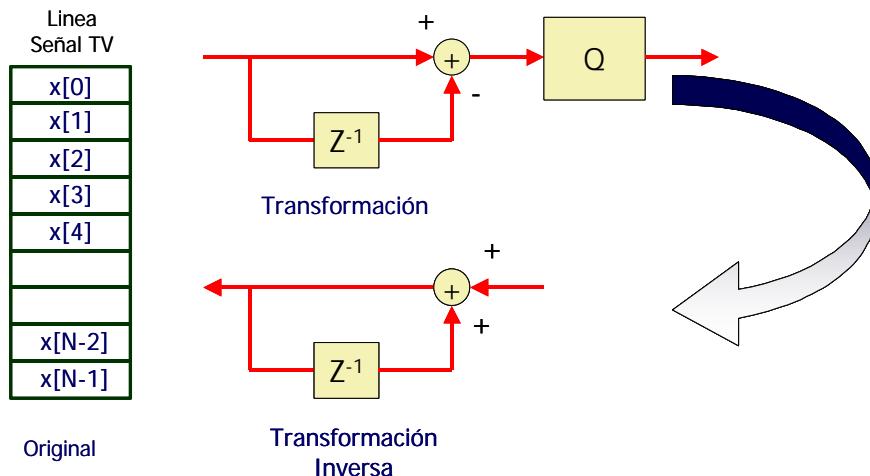


Fig. 4.8 Cuantificación directa de la señal transformada. Codificación con pérdidas

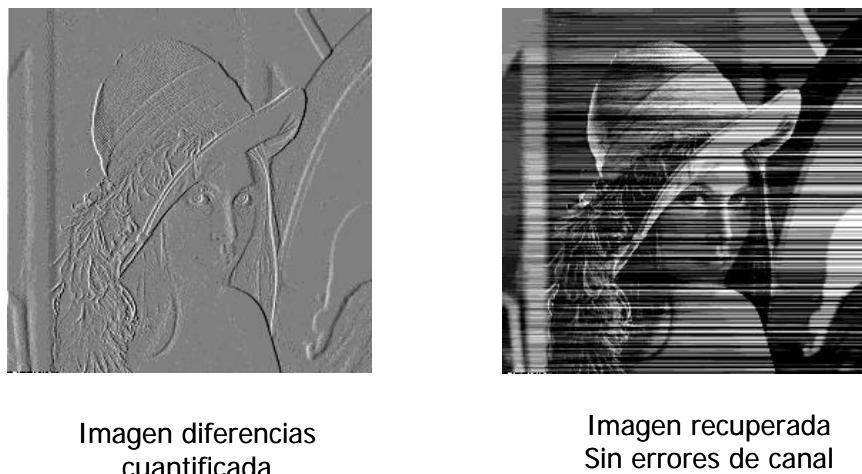


Fig. 4.9 Imagen diferencial cuantificada con 21 niveles e imagen reconstruida

Estos resultados indican que, al tratarse de una codificación diferencial, los errores existentes en el proceso de cuantificación pueden acumularse, provocando una pérdida de calidad considerable en algunas líneas de la imagen. La razón de este problema es que en el esquema de la figura 4.8 el codificador utiliza siempre la imagen original para determinar la señal diferencia, mientras que el

decodificador no dispone de la señal original, por lo que debe utilizar la muestra reconstruida anteriormente para determinar la muestra actual. Si la muestra anterior tenía un error acumulado, este error se mantendrá en las muestras sucesivas.

Una solución a este problema consiste en modificar la estructura del codificador diferencial de manera que las diferencias se calculen entre la señal original y la muestra que puede reconstruirse en el decodificador. Esta nueva estructura de codificación se representa en la figura 4.10. Ahora el cuantificador se aplica a la diferencia entre la muestra original y la muestra que se reconstruye en el lazo de realimentación de la parte inferior del codificador. Nótese que este lazo es idéntico al que utiliza el decodificador para recuperar la señal original. La señal se obtiene en base al error actual y a las señales que se habían reconstruido anteriormente. Con esta nueva estructura, los errores que se cuantifican son un indicativo de la diferencia entre la muestra original y la que hemos reconstruido hasta este momento, de modo que los errores no se acumulan. En la figura 4.11 se representa la señal diferencia cuantificada con 21 posibles niveles y la señal que se reconstruye en el decodificador. Obsérvese que existen pérdidas apreciables en la reconstrucción de la imagen, pero que ahora se deben exclusivamente a que se ha elegido un paso de cuantificación con muy pocos niveles de gris.

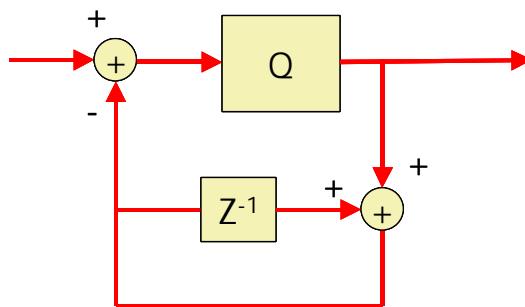


Fig. 4.10 Estructura del codificador diferencial basada en la diferencia con las muestras reconstruidas

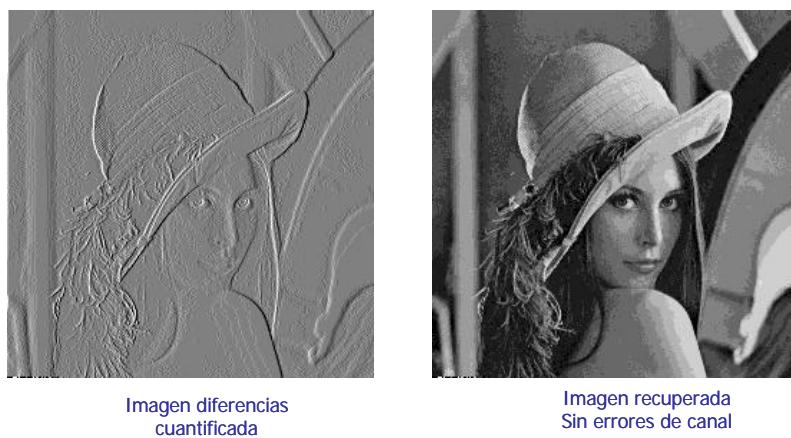


Fig. 4.11 Decodificación de la imagen con la estructura de la figura 4.10

La cuantificación que hemos utilizado en el ejemplo anterior reduce el número total de símbolos asociados a la imagen diferencia a 21. Entre todos estos símbolos, los más probables son los que están situados en un entorno próximo al cero, por lo que sería conveniente que estos se codificaran con un número de bits menor que los que tienen menos probabilidad. En la siguiente sección se presentan los conceptos básicos para establecer procedimientos sistemáticos que permitan asignar códigos eficientes a los distintos mensajes que puede proporcionar un codificador. Estos métodos están basados en la propia estadística de las señales y pueden utilizarse tanto para la compresión con pérdidas como sin pérdidas.

4.3 Medida de la información de una fuente

Una fuente de mensajes se caracteriza por el número de mensajes que puede producir y por la probabilidad con la que se produce cada mensaje. En el ejemplo del codificador diferencial del apartado anterior, podríamos decir que la fuente asociada a la imagen original podía producir 256 posibles mensajes, cuya probabilidad podría estimarse como el histograma dividido por el número total de puntos de la imagen (frecuencia relativa de cada nivel de gris). En primera aproximación también hubiéramos podido postular que todos los niveles de gris eran equiprobables. Por otra parte, la fuente de mensajes asociada a la salida del codificador diferencial contiene, para el caso de una codificación sin error, un total de 511 mensajes con una probabilidad mucho mayor para los mensajes próximos al valor nulo. Análogamente, cuando aplicamos un cuantificador de 21 niveles obtenemos un total de 21 posibles mensajes con una probabilidad determinada. En general, podemos afirmar que la estadística de los mensajes de una fuente se modifica cuando se aplican transformaciones a las señales, lo que nos sugiere concentrar los esfuerzos en buscar aquellas transformaciones que permitan obtener una estadística de mensajes que pueda codificarse con un número reducido de bits.

En la figura 4.12 se representa la idea básica de caracterización de una fuente de mensajes. A partir de los mensajes producidos podemos estimar su función densidad de probabilidad, de modo que la fuente queda completamente determinada si conocemos todos los mensajes posibles y la probabilidad con la que se producen.

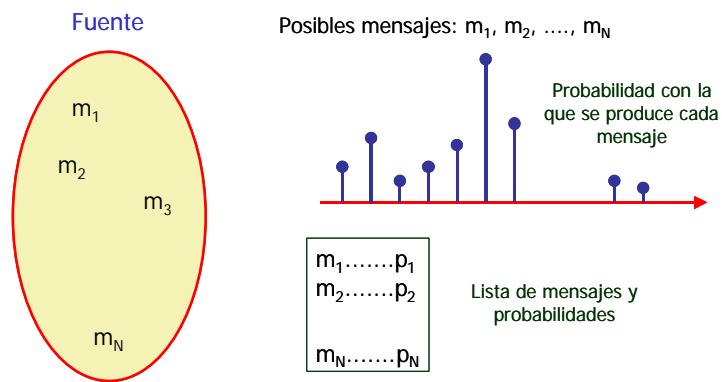


Fig. 4.12 Caracterización de una fuente de mensajes

Una vez especificada la estadística de la fuente, deberemos plantearnos establecer una medida cuantitativa de la cantidad de información asociada a cada mensaje. La medida debe verificar ciertos requisitos que parecen naturales y que generalmente están asociados al concepto cualitativo de información. Estos postulados son:

- 1) *La información asociada a un mensaje es positiva.* Es lógico suponer que siempre que se produce uno de los posibles mensajes aporta información. No tiene sentido que se produzcan pérdidas de información si recibimos nuevos datos.
- 2) *Los mensajes con poca probabilidad aportan mayor información que los mensajes con poca probabilidad.* Los mensajes que se producen frecuentemente nos aportan muy poca información porque son esperados de antemano. Así, cuando se nos informa que hay circulación lenta en los accesos de entrada a Barcelona entre las 8 y las 9 de la mañana, es un mensaje que nos aporta muy poca información.
- 3) *La información que aportan dos mensajes independientes es la misma que la que aportan cuando se producen por separado.* Esto significa que la información que nos aportan dos eventos independientes que se producen simultáneamente coincide con la suma de las informaciones que nos aportarían si se produjeran por separado.

Podemos formular estos postulados mediante las siguientes ecuaciones:

$$I(m_k) > 0 \quad (4.4)$$

$$I(m_i) > I(m_j) \quad \text{si} \quad p_i < p_j \quad (4.5)$$

$$I(m_i, m_j) = I(m_i) + I(m_j) \quad (4.6)$$

En virtud del teorema fundamental del cálculo, únicamente existe una función que verifique las tres propiedades simultáneamente. De este modo, la medida cuantitativa de la información se encuentra directamente a partir de las propiedades que queremos que verifique.

$$I(m_k) = -\log_a p_k \quad (4.7)$$

Cuando la base del logaritmo es 2, la unidad de información se denomina bit. Consideraremos como ejemplo la información que proporciona una moneda al caer. Existen dos posibles mensajes (cara y cruz), los dos con la misma probabilidad (1/2). La información asociada al mensaje cara será:

$$I(\text{cara}) = -\log_2(1/2) = 1 \text{ bit} \quad (4.8)$$

Análogamente, si consideramos una fuente que puede transmitir 8 mensajes equiprobables, la información asociada a cada uno de los mensajes vendrá dada por:

$$I(m_i) = -\log_2(1/8) = \log_2 2^3 = 3 \text{ bit} \quad (4.9)$$

De acuerdo con estos resultados, para estimar la cantidad de información ligada a uno de los mensajes de la fuente, es necesario establecer en primer lugar las características estadísticas de la fuente de mensajes. Estas características pueden estimarse de forma experimental dejando que la fuente genere mensajes y aproximando su probabilidad por la frecuencia relativa de los mismos. En general, si la estimación se realiza sobre una muestra suficientemente alta este procedimiento suele proporcionar resultados satisfactorios.

4.3.1 Entropía de una fuente de mensajes

Conociendo la información asociada a cada mensaje podemos determinar la información media que proporciona la fuente. Para ello debemos ponderar la información de cada mensaje por la probabilidad de que se produzca:

$$H = \sum_{k=1}^N p_k \cdot I(m_k) = -\sum_{k=1}^N p_k \cdot \log_2 p_k \quad (4.10)$$

Esta función se conoce con el nombre de entropía y establece un límite al número de bits medio con el que podrán codificarse los mensajes de la fuente. Ningún codificador podrá obtener códigos cuyo número de bits medio sea inferior a la entropía de la fuente.

Por otra parte, es posible demostrar que para un número de mensajes prefijado N, la entropía es máxima cuando todos los mensajes son equiprobables. Esto significa que si la fuente no tiene ninguna prioridad para producir unos mensajes con mayor frecuencia que otros, resulta más compleja de codificar de forma eficiente. En cambio, si algunos mensajes se producen con mucha frecuencia mientras que otros son poco probables, la entropía disminuye y puede ser posible encontrar códigos que aprovechen esta circunstancia. Las transformaciones de señal se utilizan para sacar provecho de esta idea. Si mediante una transformada de la señal podemos forzar a que los mensajes muestren diferencias de probabilidad considerables, habremos dado un paso importante para reducir la cantidad de información con la que podemos codificar la fuente.

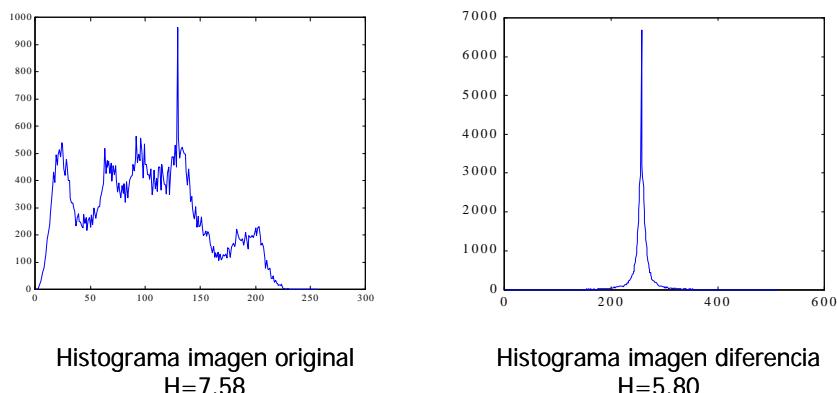


Fig. 4.13 Histogramas y entropías de la imagen original y la imagen diferencias

Desde el punto de vista del histograma, los resultados anteriores indican que conviene obtener gráficos con picos fuertemente marcados que indiquen la existencia de mensajes muy probables. En la figura 4.13 se comparan los histogramas de una imagen original (la que hemos venido considerando en los ejemplos de las secciones anteriores) y de su imagen diferencia. La entropía de la imagen original es de 7.58 bits, próxima a los 8 bits con los que ha sido codificada la imagen. En cambio, la entropía de la imagen diferencia se sitúa sobre los 5.8 bits, lo que es un indicativo de que será posible codificarla de forma más eficiente. El problema ahora se limita a buscar formas eficientes y sistemáticas de codificar esta información.

4.4 Códigos de longitud variable

La idea básica de los códigos de longitud variable es asignar palabras código de longitudes distintas en función de la probabilidad de los mensajes. Los mensajes más probables se codificarán con palabras con un menor número de bits que los mensajes menos probables. Podemos considerar como ejemplo los mensajes de la fuente que se resume en la tabla siguiente y en la que los códigos asignados a cada mensaje han sido elegidos sin seguir ningún procedimiento sistemático.

Tabla 4.1 Ejemplo de códigos de longitud variable

mensajes	probabilidad	código
m_1	0.4	0
m_2	0.3	10
m_3	0.1	1100
m_4	0.1	1101
m_5	0.1	1110

Podemos calcular el número medio de bits que se utilizan para codificar esta fuente con este código arbitrario calculando el número de bits utilizado para cada mensaje ponderado por la probabilidad de que se produzca el mensaje.

$$N_{medio} = \sum_{k=1}^N p_k \cdot N_k = 1 \cdot (0.4) + 2 \cdot (0.3) + 4 \cdot (0.3) = 2.2 \text{ bits} \quad (4.11)$$

Evidentemente, el número de bits medio está por encima de la entropía:

$$H = -\sum_{k=1}^N p_k \cdot \log_2 p_k = 2.04 \text{ bits} \quad (4.12)$$

Para que los códigos de longitud variable resulten eficaces es necesario disponer de procedimientos sistemáticos para generar estos códigos directamente a partir de las características de las fuentes. Además, es fundamental que los códigos obtenidos resulten eficientes, es decir, que permitan obtener

una tasa media de bits próxima a la entropía de la fuente. En las secciones siguientes examinamos algunas alternativas basadas en los códigos de Huffman y sus variantes más utilizadas.

4.4.1 Códigos de Huffman

El código de Huffman (1951) es la técnica más utilizada para la eliminación de la redundancia en la codificación. El procedimiento para obtener el código de Huffman asociado a una fuente es conceptualmente simple, aunque, como veremos, desde el punto de vista computacional resulta complejo. Es posible demostrar que el código de Huffman es óptimo en el sentido que consigue el número más pequeño de símbolos de código por símbolo de fuente, con la restricción de que los mensajes deben codificarse uno a uno. También puede demostrarse que el número de bits medio que se consigue obtener con un código de Huffman está siempre comprendido entre la entropía de la fuente y la entropía más uno:

$$H < \bar{N}_{Huffman} < H + 1 \quad (4.13)$$

El procedimiento para obtener el código de Huffman puede resumirse en las siguientes etapas:

- 1) Ordenar los mensajes por probabilidades decrecientes.
- 2) Crear un árbol combinando siempre los dos nodos con menor probabilidad. El resultado de la combinación de los dos nodos es un nuevo nodo cuya probabilidad es la suma de las probabilidades de los nodos que lo han construido. Continuar con este procedimiento hasta obtener un único nodo.
- 3) Para asignar los códigos a los mensajes basta con recorrer el árbol desde la raíz a cada uno de los mensajes, asignando un 0 o un 1 en función de si tomamos una rama hacia arriba o hacia abajo (la asignación de ceros y unos a las ramas puede ser totalmente arbitraria)

El procedimiento anterior resulta bastante más claro si se ilustra con un ejemplo sencillo, como el que se muestra en la figura 4.14. Obsérvese como los nodos van combinándose dos a dos, eligiendo siempre los de probabilidad más pequeña. Si existen varios nodos con la misma probabilidad, pueden combinarse dos cualesquiera. La probabilidad asignada al nodo raíz deberá ser siempre la unidad.

En la figura 5.15 se muestra el mismo árbol, pero ahora se utiliza para la asignación de códigos a los mensajes. Así, para determinar el código que asignamos al mensaje m_4 deberemos partir del nodo raíz, realizar dos pasos hacia abajo y después dos pasos más hacia las ramas superiores. El código resultante para este mensaje es 1100. Nótese que si seguimos esta construcción del árbol de Huffman el mensaje con menor probabilidad siempre queda codificado con todos los símbolos igual a la unidad.

Una vez determinado el código de Huffman para esta fuente, podemos calcular el número medio de bits que se utilizan obteniendo un total de 2,4 bits. Si calculamos la entropía de la fuente original obtendremos 2,344 bits, lo que da una idea de la calidad del código.

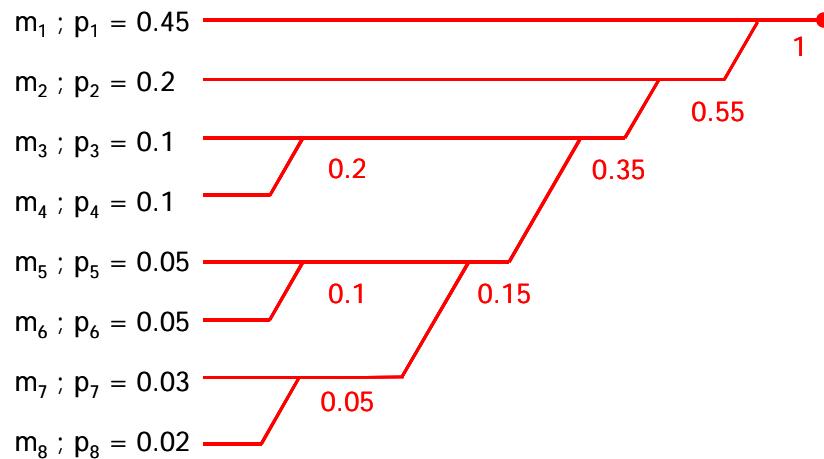


Fig. 4.14 Proceso de creación del código de Huffman

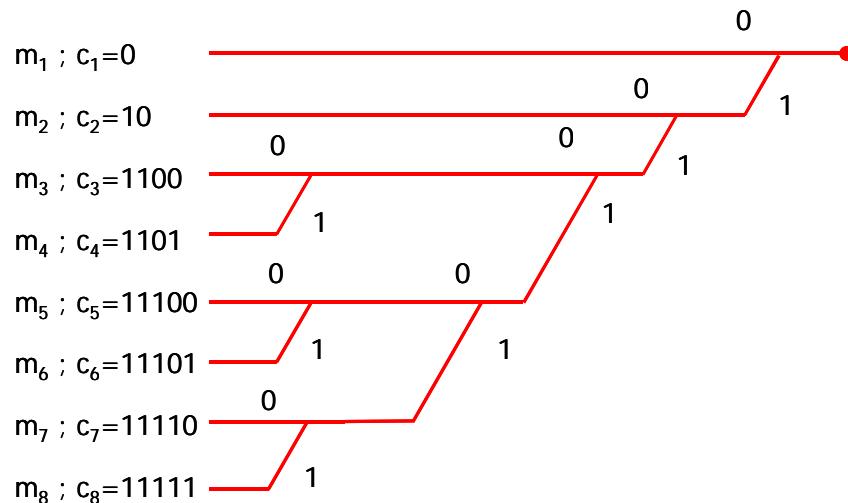


Fig. 4.15 Proceso de generación de código a partir del árbol de Huffman

Una de las características de los códigos de Huffman es que permiten decodificar la secuencia de datos directamente a partir de la tabla de códigos y sin necesidad de señalizar el principio y el final de un mensaje. Esto puede no parecer trivial, ya que los códigos son de longitud variable, pero se cumple debido a que, por construcción del árbol, ninguno de los códigos largos tiene como subcadena una palabra código más corta. Para ilustrar el procedimiento de decodificación consideremos la cadena de bits 111011101001011111..., cuyo proceso de descodificación se representa esquemáticamente en la figura 4.16. La decodificación empezaría comprobando si el primer bit (1) es una palabra código, buscándolo en la tabla de palabras código con un único bit. Como no es así, comprobaríamos si los

dos primeros bits (11) pueden ser una palabra código. Continuaríamos con este procedimiento hasta que encontramos la palabra 11101, que sí se encuentra en la tabla, y que corresponde al mensaje m_6 . Procediendo de este modo llegaríamos a decodificar la secuencia de mensajes completa que se corresponde con los códigos c_6 , c_4 , c_1 , c_1 , c_2 y c_8 .

TABLA		Secuencia de bits
$c_1=0$		111011101001011111...
$c_2=10$		1 1 1 0 1 → $c_6=11101$
$c_3=1100$		111011101001011111...
$c_4=1101$		1 1 0 1 → $c_4=1101$
$c_5=11100$		111011101001011111...
$c_6=11101$		111011101001011111...
$c_7=11110$	0	→ $c_1=0$
$c_8=11111$		

Fig. 4.16 Ejemplo de decodificación de una secuencia de mensajes

Por otra parte, cuando se produce algún error de un bit en el canal, puede producirse un cambio de palabra código que supone la pérdida del sincronismo y la posibilidad de que aparezcan varios errores consecutivos. Eventualmente, el receptor puede detectar de forma automática la aparición de los errores, debido a que no encuentra las palabras código en las tablas o que la estadística de aparición no se corresponde con la esperada. Los algoritmos para la detección de que se han producido errores y la eventual sincronización con la cadena de datos son sumamente complejos.

En aplicaciones de almacenamiento de los datos en ficheros, es recomendable que la codificación y la decodificación se realicen empleando *buffers* temporales de memoria con una longitud que sea un múltiplo de 8 bits. Con ello, cuando el *buffer* está lleno, se pueden almacenar la secuencia de bits, aprovechando una palabra de un byte o múltiplo del byte.

La obtención del código de Huffman, una vez se dispone de la estadística de la señal, es conceptualmente muy sencilla. No obstante, los algoritmos asociados son complejos y requieren realizar bastantes simplificaciones y comparaciones en la tabla de mensajes. Otro de los problemas importantes es la estimación de las estadísticas con las que se producen los mensajes. Las estimaciones de los mensajes menos probables son poco fiables y se requiere trabajar con muestras muy grandes para obtenerlos. A menudo, se obtiene que la probabilidad de algunos mensajes es nula, debido a que no se han producido ninguna vez en la secuencia que hemos analizado. Aunque el procedimiento de Huffman puede aplicarse de la misma forma en estos casos, es posible que los

valores de la tasa de bits media con que se codifica la fuente no sean tan óptimos como cabía esperar. Existen varias alternativas para simplificar el diseño de la tabla de codificación de Huffman, algunas de las cuales se exponen con cierto detalle en el siguiente apartado. Todas ellas son soluciones subóptimas que simplifican el análisis de la fuente y la construcción de la tabla de codificación.

4.4.2 Modificaciones del código de Huffman

Consideraremos las dos variantes más comunes del método de Huffman: los códigos truncados y los códigos de desplazamiento. Los códigos truncados se basan en descartar los mensajes menos probables del proceso de análisis del código de Huffman. Estos mensajes se codifican posteriormente utilizando palabras de longitud fija precedidas de un prefijo proporcionado por el mismo análisis de la fuente simplificada. Las etapas de análisis del código se resumen en los siguientes puntos:

1. Se ordenan todos los N mensajes de la fuente por orden de probabilidad decreciente.
 2. Se eligen los M mensajes más probables y se genera un nuevo mensaje L que representa los (N-M) mensajes descartados. La probabilidad de L es la suma de los mensajes con baja probabilidad que representa.
 3. Se construye el código de Huffman asociado a la nueva fuente de M+1 mensajes.
 4. Se construye el código de los N-M mensajes de menor probabilidad, empleando como prefijo el código de Huffman asignado al nuevo mensaje L, seguido de un código binario convencional de longitud constante.

El procedimiento se resume esquemáticamente en la figura 4.17, donde se supone que se descartan los 5 mensajes con menor probabilidad. La suma de las probabilidades de estos mensajes se inserta en la tabla de mensajes previamente ordenados (en nuestro ejemplo es el 5 mensaje más probable) y se construye el código de Huffman con esta nueva estadística. El código resultante se utiliza como prefijo para asignar las palabras que representan los mensajes originales. En este ejemplo debemos añadir tres bits adicionales para poder representar los 5 mensajes.

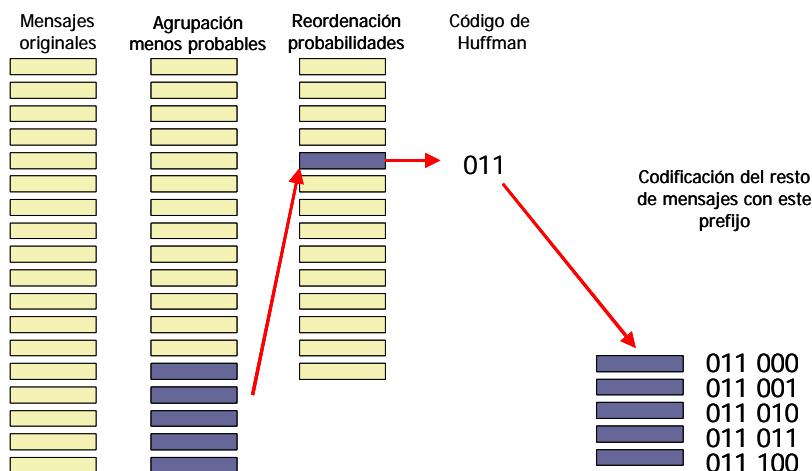


Fig. 4.17 Construcción de un código de Huffman truncado

Los códigos de Huffman de desplazamiento tienen una filosofía de construcción bastante más compleja, aunque simplifican considerablemente el proceso de construcción de la tabla sin degradar excesivamente las propiedades del código. El procedimiento de construcción del código de desplazamiento se representa en la figura 4.18. Los mensajes o símbolos de la fuente original se ordenan por probabilidades decrecientes y se dividen en K bloques, que contienen, todos ellos, el mismo número R de mensajes (en nuestro ejemplo K=3 y R=7). Las probabilidades de todos los mensajes de los K-1 bloques con menor probabilidad se suman y se construye un nuevo mensaje L, al que se le asigna esta probabilidad. El procedimiento de Huffman convencional se aplica a esta nueva fuente modificada, que contiene un total de R+1 mensajes.

La construcción del código definitivo es ahora bastante simple. El código asignado al nuevo mensaje L es utilizado como código de desplazamiento. Esto significa que, si no aparece, entenderemos que hacemos referencia a un mensaje del primer bloque, si aparece una vez, referencia a un mensaje del segundo bloque, si se repite dos veces, hace referencia a un mensaje del tercer bloque, y así sucesivamente. Los mensajes dentro de cada uno de los bloques quedan codificados con los códigos que han obtenido los mensajes del primer bloque, en el mismo orden e independientemente de sus probabilidades. Evidentemente, los códigos de estos mensajes no han sido optimizados, pero, como su probabilidad suele ser pequeña, no tienen excesiva incidencia sobre la eficiencia del código. Nótese que aunque el número de bloques sea pequeño, el análisis de Huffman se reduce notablemente, ya que sólo debe realizarse para R+1 mensajes de los KxR mensajes de la fuente.

En la figura 4.18 se representa la construcción de un código para una fuente sencilla. En este caso, el nuevo mensaje L tiene una probabilidad alta, por lo que el código de desplazamiento asignado es bastante corto. Este resultado suele ser general, sobre todo si el número de bloques en que se divide la fuente es elevado. Probablemente, resulta más simple comprender la construcción del código a partir de esta gráfica que a partir del texto.

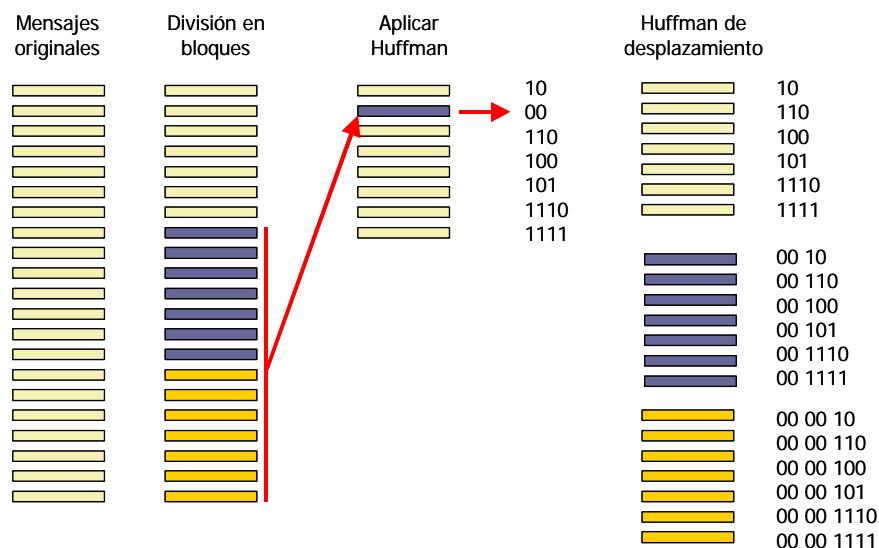


Fig. 4.18 Construcción de los códigos de Huffman de desplazamiento

4.5 Otros tipos de códigos

Los códigos de longitud variable no son los únicos que pueden conseguir codificar de forma eficiente una fuente de información y extraer su redundancia estadística. En función de las características de la fuente, existen códigos más o menos complejos que pueden conseguir eficiencias considerables. En este apartado describiremos brevemente los códigos RLE (*Run Length Encoding*), los códigos LZW y los principios básicos de la codificación aritmética. El RLE y la codificación aritmética se emplean, conjuntamente con los códigos de Huffman, en distintas partes de la codificación de imágenes mediante JPEG o de secuencias de vídeo mediante MPEG.

4.5.1 Codificación por longitud de series (*Run Length Encoding – RLE*)

La codificación por longitud de series es conceptualmente muy simple y empezó a utilizarse a mediados de la década de los 50 para la codificación de documentos digitalizados. Actualmente, junto con sus extensiones bidimensionales, es el método estándar para la codificación de documentos del tipo facsímil. Para documentos de texto o gráficos binarios, que únicamente contienen los niveles blanco y negro, podemos codificar cada línea indicando, de forma alternada, cuantos píxeles toman el nivel blanco y cuantos el nivel negro en palabras sucesivas. Para ello, es necesario establecer que siempre empezamos por uno de los niveles (p.ej. blanco). En el supuesto de que en realidad la línea empezara por negro, podríamos codificar que hay una longitud inicial de cero blancos. Esta forma de codificación es bastante eficiente y comprime de forma notable este tipo de documentos.

A este procedimiento básico, podemos añadir un código de Huffman que tenga en cuenta la estadística con la que se producen los mensajes que indican la longitud de las series de blancos y negros. Aquellas longitudes que se producen con mayor frecuencia serán codificadas con palabras de menor longitud obteniendo una mejora global del código.

La idea básica del RLE puede extenderse a fuentes con varios mensajes o símbolos, pero en las que predomine uno de los mensajes con gran probabilidad. Esta variante del RLE se utiliza en la codificación de los coeficientes de la transformada coseno de los bloques de imagen en los estándares JPEG y MPEG. En este caso, muchos de los coeficientes de la transformada coseno, una vez cuantificados, toman el valor nulo, por lo que resulta más eficiente especificar la cantidad de coeficientes consecutivos que toman este valor.

En imágenes multínivel, en las que todos los niveles tienen probabilidades parecidas, puede emplearse la codificación RLE aplicándola a los distintos planos de bits de la imagen. En la figura 4.19 se representa una imagen de 256 niveles descompuesta en los planos de bits de cada píxel. La primera imagen representa únicamente el bit más significativo, la segunda imagen el segundo bit, etc. Como ahora las imágenes son binarias, podemos aplicar el RLE como si se tratara de documentos de texto. Evidentemente, la ganancia que obtenemos para la imagen asociada al bit más significativo es importante, debido a que existen series bastante largas de píxeles blancos o negros. A medida que disminuimos el peso del bit, las series se reducen y la eficiencia de la codificación baja. Una posible alternativa es descomponer la imagen en planos de bits, pero representados en forma de código Gray en vez de binario natural. El código Gray tiene la propiedad de que entre dos niveles consecutivos sólo se produce un cambio en uno de los bits. Así, cuando la imagen pasa del nivel 127 al 128 en codificación binaria natural, cambian los valores de todos los planos de bits. En cambio, si utilizamos

la codificación del nivel de gris mediante el código de Gray, sólo cambiaría uno de los dígitos. El resultado es que todos los planos de bits presentan zonas más o menos uniformes en las que las series de blancos y negros pueden codificarse con ganancias significativas.

El código Gray puede obtenerse con cierta facilidad a partir del código binario. Si la representación de un determinado nivel de gris en código binario natural es $b_7b_6b_5b_4b_3b_2b_1b_0$, donde b_7 representa el bit más significativo y b_0 el bit menos significativo, la representación en código Gray $g_7g_6g_5g_4g_3g_2g_1g_0$ puede obtenerse aplicando la siguiente relación:

$$\begin{aligned} g_7 &= b_7 \\ g_k &= a_k \oplus a_{k+1} \quad \text{si } 0 \leq k \leq 6 \end{aligned} \quad (4.14)$$

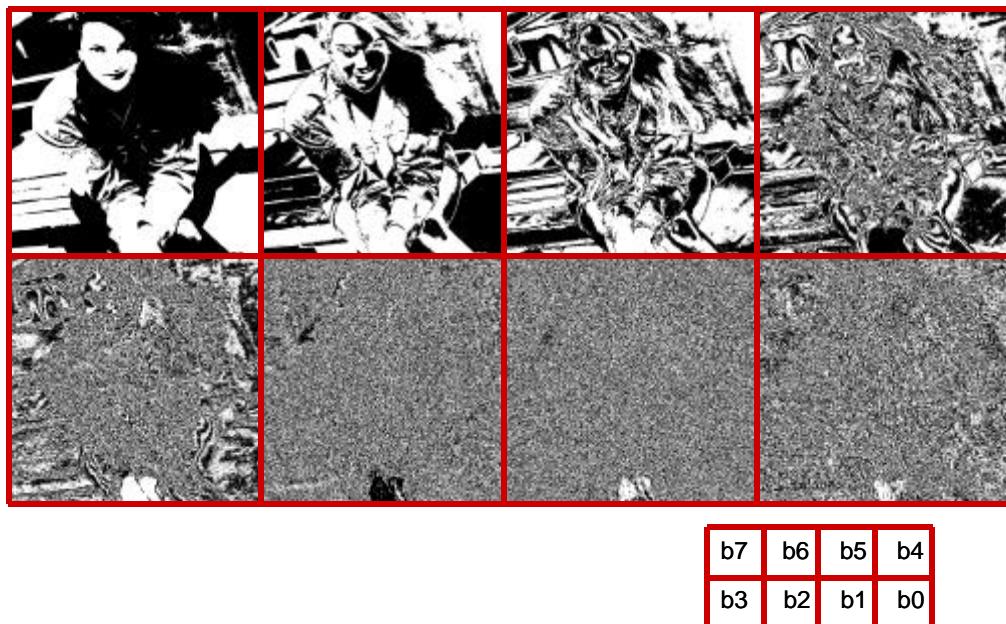


Fig. 4.19 Representación de una imagen multinivel en planos de bits en binario natural

4.5.2 Codificación LZW

Esta codificación fue introducida por Lemple y Ziv y posteriormente mejorada por Welch. Se utiliza en muchos compresores de datos con aplicaciones informáticas. La idea básica es sustituir los símbolos de 8 bits (bytes) que se encuentran en los ficheros convencionales por símbolos de 12 o más bits. Al representar las palabras con símbolos de 12 bits, disponemos de 4096 posibilidades, de las cuales 256 se asignan a las palabras de 8 bits que disponíamos en la fuente original. Las 3840 palabras restantes se asignan dinámicamente a las series de bits que se producen más a menudo. La primera vez que se produce una cadena se registra tal cual, junto con una palabra código que la representará a partir de aquel momento. La tabla de códigos se genera dinámicamente y se incluye en el fichero. Este

código es muy eficiente para comprimir ficheros de tipo texto o binarios en los que las palabras más habituales quedan codificadas con 12 bits. Existen multitud de variantes sobre el algoritmo básico.

4.5.3 Codificación aritmética

La codificación aritmética se basa en asociar números reales comprendidos entre 0 y 1 a secuencias o cadenas de mensajes de la fuente. Aunque en la práctica es difícil de implementar, suele conseguir resultados de compresión excelentes, superiores, en la mayoría de casos, a la codificación de Huffman. Recordemos que Huffman era un código óptimo sólo cuando se restringía que cada mensaje de la fuente debía ser codificado con un símbolo. En el caso de la codificación aritmética se consiguen mejores resultados debido a que los símbolos corresponden a cadenas de mensajes.

Existen diversas variantes de los códigos aritméticos, sobre todo en lo que respecta a su implementación y los formatos de representación de los números reales con aritmética finita. Esencialmente, el proceso de codificación se ilustra en la figura 4.20 para la siguiente cadena de mensajes: $a_1a_3a_1a_2a_4$. Para simplificar el procedimiento, suponemos que los mensajes proceden de una fuente con 4 mensajes posibles a_1, a_2, a_3 y a_4 con probabilidades 0,4; 0,3; 0,2 y 0,1. El mensaje a_4 se utiliza exclusivamente para indicar el fin de cadena. El número real asignado al mensaje se va determinando progresivamente a medida que se codifican los distintos elementos de la cadena. Para codificar el primer elemento, el intervalo real $[0; 1]$ se divide en 4 porciones proporcionales a la probabilidad de cada uno de los mensajes. El mensaje a_1 quedará codificado como el intervalo real $[0; 0,4)$, el mensaje a_2 como el intervalo $[0,4; 0,7)$ y así sucesivamente. En la figura se indica el intervalo en el que ha sido codificado el primer elemento de la cadena. Para codificar el segundo elemento, tomamos el intervalo definido por el primero y se vuelve a dividir en 4 intervalos con una longitud proporcional a sus probabilidades, aplicando progresivamente este procedimiento a medida que se incorporan más elementos de la cadena.

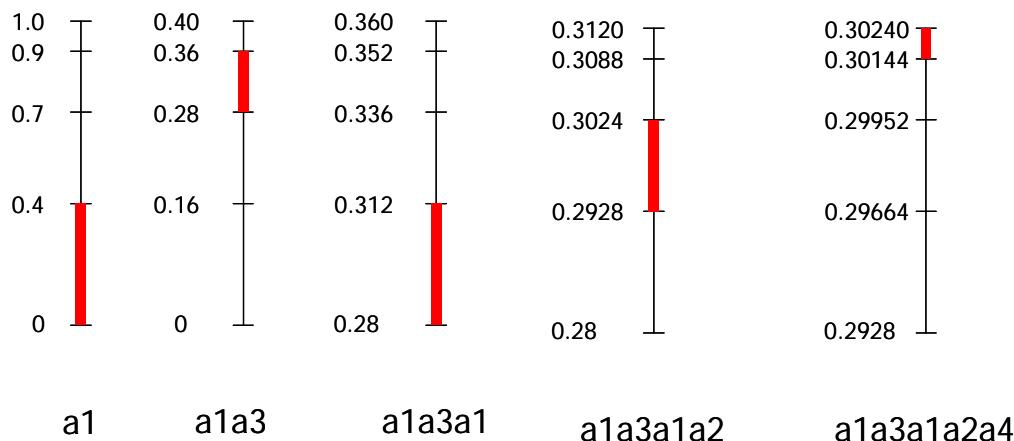


Fig. 4.20 Ejemplo de codificación aritmética de una cadena de mensajes

En la gráfica se observa que la secuencia que pretendíamos codificar quedaría asignada al intervalo [0,30144; 0,30240]. Cualquier número real dentro de este intervalo, por ejemplo el 0,302, codificaría de forma única el mensaje. En este ejemplo, es suficiente con tres cifras decimales para codificar la cadena de 5 mensajes, lo que representa una codificación bastante eficiente de la fuente. Nótese que los códigos resultantes son de longitud variable, puesto que en función de la cadena de mensajes podemos necesitar más o menos dígitos para codificar el mensaje. La implementación del codificador y del decodificador en tiempo real o mediante *software* no son evidentes y en la práctica existen distintas estrategias para optimizar los tiempos de cálculo cuya descripción detallada escapa de los objetivos de este texto.

4.6 Codificación y predictores

Hemos visto que la codificación diferencial se basa en utilizar el píxel anterior para realizar una estimación del píxel actual, con lo que obtenemos una señal diferencia cuya estadística permite realizar una codificación eficiente de la fuente. La extensión natural de la codificación diferencial son, como ya hemos comentado, los predictores lineales. En general, los predictores utilizan una combinación lineal de las muestras anteriores para obtener una aproximación a la muestra actual. Suelen obtenerse excelentes resultados cuando la señal puede modelarse como una superposición de señales sinusoidales y, en principio, la estimación es tanto mejor cuanto mayor es el número de muestras que se utilizan en la predicción. En codificación de audio, donde la señal depende de una única variable temporal, suelen utilizarse modelos de predicción que tienen en cuenta entre 6 y 12 muestras anteriores de la señal. Normalmente no se utiliza un mayor número de muestras debido a que la correlación entre la muestra actual y las pasadas disminuye apreciablemente con el transcurso del tiempo y ya no resulta eficiente utilizarlas para estimar el valor actual de la señal. Los coeficientes del predictor se pueden calcular de antemano, conociendo las características estadísticas de la señal de audio, o calcularlos de forma adaptativa, de forma que vayan modificándose en el tiempo a medida que varían las características de la señal. Existen algoritmos eficientes para ambas alternativas.

En codificación de imagen y vídeo, la naturaleza espacial y temporal de la señal proporciona mucha mayor libertad al diseñador para elegir las muestras de señal que se utilizarán en la predicción. Pueden tomarse las muestras anteriores correspondientes a la misma línea, las muestras de la línea anterior, las muestras de la misma posición en la imagen o el cuadro anterior, etc. Decimos que realizamos *predicción intra-trama* cuando únicamente se utilizan muestras dentro de la misma imagen del píxel que pretendemos predecir. Cuando se usan muestras de imágenes anteriores se dice que la *predicción es inter-trama*. En la figura 4.21 se muestran algunas posibles variantes del predictor diferencial en su modalidad intra-trama. Los coeficientes situados encima de la plantilla indican cómo se pondera la muestra en esta posición para realizar la predicción. Al igual que en el caso de señales de audio podríamos plantearnos obtener los coeficientes óptimos para una determinada estadística de los píxeles de la imagen. Para un predictor de orden unidad (por ejemplo la codificación diferencial que hemos considerado anteriormente) puede demostrarse que el valor con el que debe ponderarse la muestra anterior viene dado por el coeficiente de correlación entre la muestra que se pretende estimar y la muestra que utilizamos para la estimación:

$$a_1 = \frac{r(1)}{r(0)} \quad (4.15)$$

El valor de este coeficiente suele situarse muy próximo a la unidad, siendo típicos valores entre 0,99 y 0,98. Este resultado indica que el peso unidad que hemos estado utilizando hasta este momento es totalmente adecuado.

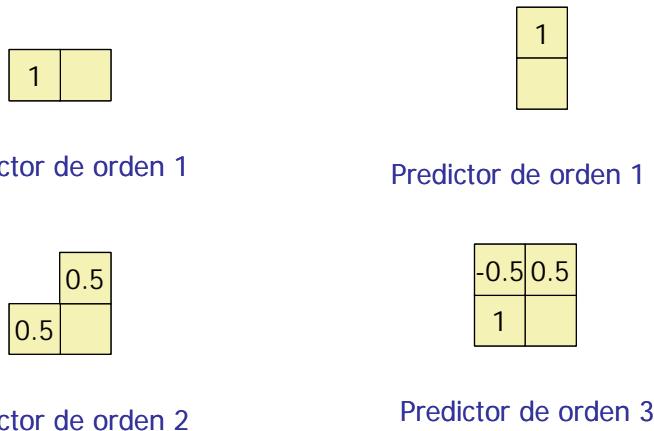


Fig. 4.21 Ejemplos de predictores intra-trama

Además, debemos tener en cuenta que en tratamiento de imagen y vídeo, el número de muestras que deben tratarse por segundo es mucho más elevado que en tratamiento de audio. Este hecho justifica que en muy raras aplicaciones (normalmente sólo cuando la codificación puede realizarse en tiempo diferido) se utilicen los coeficientes óptimos y que normalmente se sustituyan por números sencillos como la unidad, 1/2, 3/4, etc, cuyos cálculos pueden implementarse de forma eficiente. Los coeficientes óptimos son, en general, valores reales que dificultan excesivamente el cálculo y que, en la mayoría de los casos, no justifican las mejoras que introducen.

Para mejorar la estadística de la señal de error, en ocasiones se utilizan soluciones adaptativas simplificadas. En tratamiento de señal de audio, una solución adaptativa significa que los coeficientes del predictor se van cambiando muestra a muestra o cada cierto periodo de tiempo. Existen algoritmos que permiten ir actualizando los coeficientes del predictor en función de la propia señal. En tratamiento de imagen, las soluciones adaptativas normalmente consisten en elegir entre varias estructuras de predicción predefinidas aquella que mejor se adapta a las características de la señal. Para ello, la imagen se divide en bloques o subimágenes de tamaño prefijado (4x4, 8x8 o 16x16 son los tamaños más utilizados) y en cada bloque se evalúan los errores que se obtienen utilizando distintas estructuras de predicción, eligiendo aquella estructura que obtiene una suma de errores absolutos mínima. Para ilustrar este procedimiento consideremos el diagrama de la figura 4.22, donde se representan los píxeles más próximos al elemento de imagen que pretendemos estimar. El error total que obtenemos en el bloque de análisis viene dado por:

$$E^{(k)} = \sum_{(i,j) \in I} |e^{(k)}(i, j)| \quad (4.16)$$

donde $e^{(k)}(i,j)$ representa el error obtenido en el píxel (i,j) utilizando la estructura de predicción k e I representa el intervalo de validez de la subimagen que estamos analizando. $E^{(k)}$ nos proporciona el error total que obtenemos utilizando la estructura de predicción k . Este error total es la suma de los errores absolutos obtenidos en cada píxel de la subimagen. Las diferentes estructuras de predicción que podríamos utilizar vienen definidas, entre otras posibilidades, por:

$$\begin{aligned} e^{(1)}(i,j) &= (X - A)_{(i,j)} \\ e^{(2)}(i,j) &= (X - B)_{(i,j)} \\ e^{(3)}(i,j) &= (X - C)_{(i,j)} \\ e^{(4)}(i,j) &= (X - (0.5A + 0.5C))_{(i,j)} \end{aligned} \quad (4.17)$$

En este ejemplo, se evalúan las diferencias entre el píxel a estimar con el de la izquierda, el de arriba, el de arriba a la izquierda o con el promedio entre el de arriba y el de la derecha. El análisis podría extenderse a otras combinaciones dentro de la misma imagen o incluso a estimaciones inter-trama. Una vez evaluados los diferentes errores $E^{(k)}$ obtenidos para una subimagen, se elige el valor mínimo y se proporciona la información al decodificador de cuál ha sido la estructura utilizada para realizar la predicción de este bloque de imagen. Esto significa que todos los píxeles dentro del bloque se codifican con la misma estructura, aunque previamente se determina que estructura es la que produce un error mínimo. Obsérvese que en el supuesto de que la subimagen corresponda a un contorno vertical, los errores obtenidos cuando utilizamos el píxel de la izquierda pueden ser considerables debido a los cambios de nivel de gris asociados al contorno. En cambio, si utilizamos el píxel situado en la línea de arriba, el error de predicción seguramente se reducirá apreciablemente.

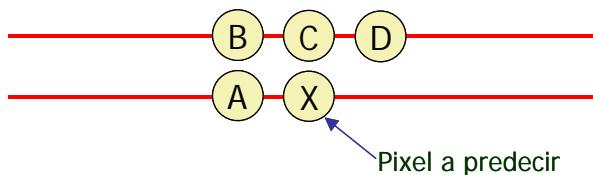


Fig. 4.22 Identificación de elementos en estructuras de predicción adaptativas

4.7 La transformada coseno discreta

La transformada coseno se utiliza en muchos estándares de compresión de imágenes estacionarias y de vídeo como el JPEG y el MPEG. Las razones por la que se ha extendido su uso son muy variadas e intentaremos explicar las más significativas en este apartado. La codificación de imagen utilizando algún tipo de transformadas es bastante popular y se extiende a otras transformaciones como Fourier, Walsh-Hadamard, Hart, Karhunen-Loeve, Wavelet, etc. Estos procedimientos de codificación se conocen con el nombre genérico de métodos transformados y se distinguen de los métodos predictivos que hemos descrito en los apartados anteriores por realizar un cambio de dominio en la representación de la imagen. La codificación mediante predicción lineal también realiza una transformación de los datos, pero en este caso, la transformación se realiza en el mismo dominio espacial o temporal de la

imagen mediante las diferencias entre los elementos reales y los estimados. No puede considerarse, por tanto, un cambio de dominio de representación de la imagen.

La codificación mediante métodos transformados y en concreto, mediante la transformada coseno, se basa en el esquema de la figura 4.23. La imagen a comprimir se divide en bloques o subimágenes de tamaño reducido sobre las que se aplica la transformación. El resultado de la transformada de cada bloque se cuantifica y posteriormente se aplican códigos eficientes para transmitir o almacenar esta información. La compresión de imagen se produce principalmente durante el proceso de cuantificación, donde parte de la información se descarta o se cuantifica con un número de niveles muy reducido. Veremos que la información de cada subimagen puede quedar muy bien empaquetada en unos pocos coeficientes transformados, de modo que resulta suficiente con determinar que coeficientes son los más representativos y codificar la imagen utilizando únicamente estos elementos. Evidentemente, también existe alguna ganancia de compresión debido a la codificación de los coeficientes mediante algoritmos eficientes, aunque en ningún modo es tan significativa como la debida a la parte de cuantificación.

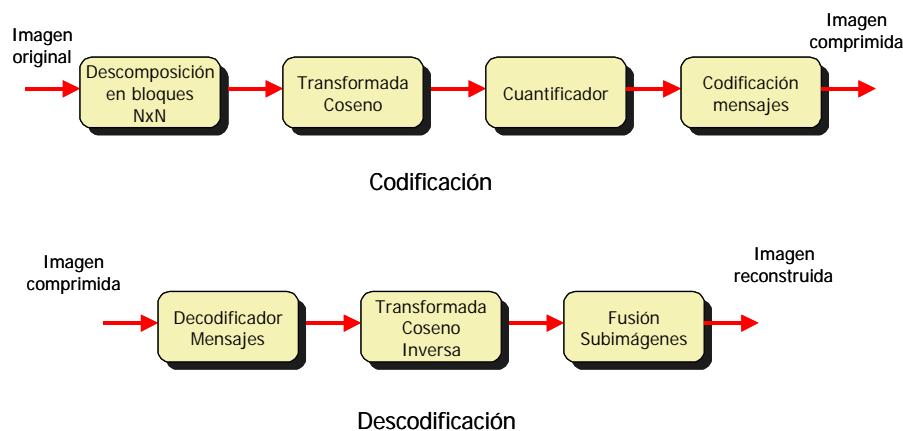


Fig. 4.23 Proceso general de codificación y descodificación de imágenes mediante la transformada coseno

La decodificación de la imagen se realiza aplicando el proceso inverso: los coeficientes se descodifican y se aplica la transformada inversa, de manera que se recuperan los elementos de imagen en el dominio espacial original. Posteriormente, es necesario recomponer la imagen a partir de los bloques en los que se ha dividido originalmente. Esta recomposición de la imagen suele ser uno de los principales problemas de los métodos transformados. En efecto, debido a la cuantificación de los coeficientes transformados, la subimagen se recupera con cierto error respecto a la original. Este error es muy crítico si se produce en los límites de las subimágenes, debido a que al fusionar todos los bloques aparecerán cambios de nivel de gris bruscos que pueden reproducir la descomposición en bloques cuando se observa la imagen descomprimida. El problema es particularmente notorio con algunas transformadas como la de Fourier o Walsh-Hadamard. La transformada coseno presenta excelentes propiedades en la codificación de los contornos de las subimágenes que, de hecho, ha sido uno de los motivos principales por los que se ha elegido esta transformada en casi todos los estándares

de codificación. La transformada Wavelet, que se incorporará en estándares más recientes como el JPEG-2000 o el MPEG-4, también dispone de estas propiedades.

Otro de los aspectos cruciales en la codificación por métodos transformados es la elección del tamaño de las subimágenes. En principio es aconsejable elegir los tamaños de los bloques para que exista cierto grado de correlación entre bloques adyacentes. Otra restricción importante es que los tamaños de las subimágenes deberían ser una potencia de 2 para facilitar el uso de algoritmos rápidos en la implementación de la transformación. El factor de compresión que puede obtenerse aumenta a medida que se utilizan bloques más grandes, aunque también se incrementa la carga computacional. En la mayoría de aplicaciones pensadas para una resolución espacial similar a la de los sistemas de televisión, los bloques son de tamaño 8x8 o 16x16.

Antes de introducirnos en los detalles de la transformada coseno analizaremos sus características principales, que justifican su amplio uso en codificación de imagen y que, posteriormente, nos ayudarán a comprender algunas de sus particularidades. Estos aspectos se tratan sin ningún orden preestablecido y únicamente pretendemos proporcionar una idea general de sus ventajas e inconvenientes respecto a otros métodos transformados.

- 1) *Capacidad de compactación de la energía en el dominio transformado.* La transformada coseno discreto (*DCT- Discrete Cosinus Transform*) consigue concentrar la mayor parte de la información en unos pocos coeficientes transformados. Esto permite obtener importantes ventajas para una codificación eficiente de la imagen, puesto que basta con codificar de forma precisa estos coeficientes principales para obtener una buena representación de todo el bloque de la imagen. Debe tenerse en cuenta que la capacidad de compactación de la energía en unos pocos coeficientes es un parámetro puramente estadístico, lo que significa que siempre es posible encontrar un bloque de imagen en la que la energía en el dominio transformado esté dispersada entre todos los coeficientes. No obstante, esto tiene una probabilidad de ocurrencia muy baja y no suele producirse nunca si trabajamos con imágenes naturales. La capacidad de compactación de la energía de la DCT es muy superior a la que se obtiene con las transformadas de Walsh-Hadamard, Hart y también Fourier. La transformada Karhunen-Loeve consigue una compactación óptima, aunque al precio de un elevado coste computacional. La transformada Wavelet también consigue concentrar la energía en unos pocos coeficientes, aunque en este caso, la posición de los coeficientes con mayor energía depende de la imagen.
- 2) *Es una transformada independiente de los datos.* Esto significa que el algoritmo es independiente del contenido de la imagen. Prácticamente todas las transformadas que hemos venido comentando son independientes de los datos, con la excepción de la Karhunen-Loeve, donde las matrices asociadas a la transformación deben calcularse en función de las características de la propia imagen. Por este motivo, la transformada de Karhunen Loeve es óptima en el sentido de compactación de energía.
- 3) *Existen algoritmos eficientes para el cálculo rápido.* Existen algoritmos análogos al de la FFT para realizar la transformación. Los algoritmos se encuentran disponibles en circuitos integrados especializados en realizar esta transformación. La restricción para

poder utilizar estos algoritmos es que los bloques tengan un tamaño que sea múltiplo de una potencia de dos. Si no es así, deberán añadirse ceros a las subimágenes, aunque esto supone que aumenta notablemente el riesgo de que aparezca el efecto de bloques al decodificar la imagen.

- 4) *Errores reducidos en los contornos de los bloques.* Probablemente éste es uno de los motivos principales por los que se ha elegido la DCT frente a otras transformadas. La presencia de errores de codificación muy pequeños en los límites de la subimagen es necesaria para evitar la aparición del efecto de bloques en la imagen reconstruida. Compararemos las prestaciones de la DCT con la transformada de Fourier para ilustrar esta característica.
- 5) *Identificación e interpretación frecuencial de los componentes transformados.* Para aprovechar al máximo la capacidad de compresión de un método es conveniente que los coeficientes puedan interpretarse desde el punto de vista frecuencial con facilidad. Esto permite introducir conceptos psico-visuales, permitiendo dedicar un menor número de bits a aquellas componentes que de antemano sabemos que no resultan críticas para el sistema visual.

4.7.1 Definición de la transformada coseno unidimensional

Nuestro objetivo en este apartado es presentar los conceptos elementales de la transformada coseno unidimensional y razonar su uso en la compresión de información. No pretendemos realizar un análisis exhaustivo de las propiedades de esta transformada ni justificar matemáticamente los procedimientos utilizados en la codificación de señales. Se supone no obstante, que el lector dispone de conocimientos básicos sobre la transformada discreta de Fourier, ya que la utilizaremos para realizar algunas comparaciones entre ambas.

La transformada coseno discreta de una secuencia unidimensional $x[n]$ de N muestras de longitud se define como:

$$C[k] = \alpha[k] \cdot \sum_{n=0}^{N-1} x[n] \cdot \cos\left(\frac{(2n+1)k\pi}{2N}\right) \quad 0 \leq n \leq N-1 \quad 0 \leq k \leq N-1 \quad (4.18)$$

donde el coeficiente $\alpha[k]$ viene determinado por:

$$\alpha[k] = \begin{cases} \sqrt{1/N} & \text{si } k = 0 \\ \sqrt{2/N} & \text{si } k = 1, \dots, N-1 \end{cases} \quad (4.19)$$

La transformada inversa permite determinar la secuencia original a partir de los coeficientes $C[k]$ mediante el uso de la ecuación:

$$x[n] = \sum_{k=0}^{N-1} \alpha[k] \cdot C[k] \cdot \cos\left(\frac{(2n+1)k\pi}{2N}\right) \quad 0 \leq n \leq N-1 \quad 0 \leq k \leq N-1 \quad (4.20)$$

Existen otras variantes de esta transformada en las que las ecuaciones y factores numéricos en la función coseno son ligeramente distintos, pero que no serán consideradas en este texto.

Las ecuaciones anteriores tienen cierto parecido con las de la transformada discreta de Fourier, con la salvedad de las expresiones en los índices en la función y, sobre todo, en que las exponenciales complejas se sustituyen por funciones coseno. La diferencia más importante entre ambas funciones es que la transformada discreta de Fourier (*DFT, Discrete Fourier Transform*) es compleja, mientras que la DCT es real. Esto supone, ya de entrada, una mejor compactación de la información, puesto que de N muestras reales mantenemos un total de N coeficientes reales que las representan. En la figura 4.24 se muestran los resultados de aplicar la DCT y la DFT a un fragmento de 16 muestras de una línea de una imagen. Obsérvese que para representar la información mediante la DFT se requiere disponer tanto del módulo como de la fase mientras que con la DCT toda la información queda compactada en los 16 coeficientes.

Ambas transformadas realizan una compactación de la energía en unos pocos coeficientes que se concentran en las zonas de baja frecuencia. En el caso de la DCT, los coeficientes con mayor energía son los que corresponden con el origen (valor medio de la señal), el primer y el segundo coeficiente. En la DFT tienen una magnitud considerable el coeficiente de continua, el primer coeficiente y el último. Recuérdese que en la DFT los coeficientes centrales corresponden a alta frecuencia y que los situados al final de la secuencia se corresponderían con las frecuencias negativas. Así, como la señal es real, el primer y último coeficiente tienen el mismo módulo y fases cambiadas de signo.

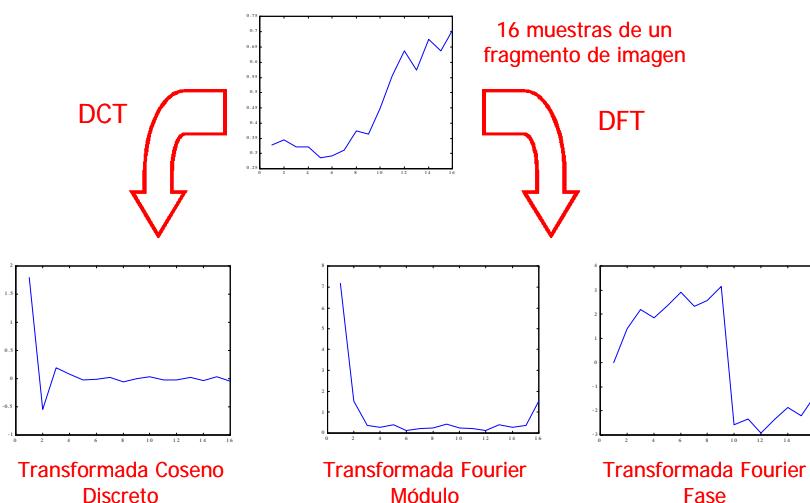


Fig. 4.24 Transformadas de Fourier y Coseno de una secuencia de 16 muestras

De acuerdo con estos resultados, podemos cuestionarnos cómo se reconstruiría la señal original si los coeficientes con valor más pequeño se suponen nulos y no se codifican. En principio, como estos coeficientes tienen un valor pequeño, representan únicamente una pequeña fracción de la energía de la señal, por lo que los errores de reconstrucción deberán ser también pequeños. Podría argumentarse que si únicamente se codifican los coeficientes con mayor energía, nunca podremos codificar correctamente los cambios rápidos, es decir, las componentes de alta frecuencia en la imagen. Esto no tiene por qué ser cierto si nuestra política es la de codificar todos aquellos componentes que tengan una magnitud representativa, sean de baja o de alta frecuencia, y eliminar los que tengan una energía reducida. Esta es la política que se utiliza en la mayoría de estándares de compresión que consideraremos.

En la figura 4.25 representamos las señales reconstruidas utilizando únicamente 11 y 9 coeficientes en las ecuaciones de reconstrucción de la señal mediante los coeficientes transformados (ecuaciones de la transformada inversa). Como puede observarse, en las dos transformadas, las aproximaciones obtenidas son bastante buenas. Los coeficientes que se han descartado son los asociados a la parte de alta frecuencia tanto para la DFT como para la DCT.

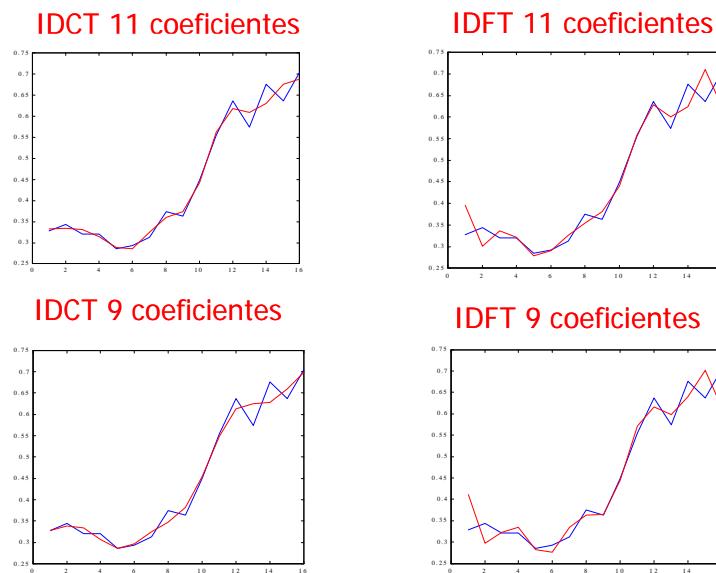


Fig. 4.25 Reconstrucción de la información original utilizando sólo parte de los coeficientes transformados

Para justificar estos resultados es necesario indicar que tanto la DFT como la DCT son transformadas ortonormales. Esto significa que la energía de la señal puede determinarse, bien como la suma de las muestras al cuadrado, o bien como la suma de los coeficientes transformados al cuadrado. La energía del error de reconstrucción será, por tanto, proporcional a la energía de los coeficientes que no hayan sido utilizados en la reconstrucción. Si estos coeficientes son pequeños, el error de reconstrucción

también lo será. Por otra parte, conviene insistir en una diferencia importante entre los resultados que se obtienen con cada una de las transformadas. Aunque los errores totales de la aproximación sean prácticamente iguales para las dos transformadas, la DCT consigue un error prácticamente nulo en los dos extremos de la señal, mientras que en la DFT, ambos extremos son recuperados con errores considerables. Este resultado es extremadamente importante y supone que la DFT apenas tenga aplicación práctica en la codificación de imágenes por bloques. En efecto, si codificamos la señal por bloques y luego la reconstruimos superponiendo los errores resultantes, los errores en la codificación de los puntos extremos producirán cambios de nivel abruptos que serán fácilmente visibles, produciendo el mencionado efecto de bloques en la imagen reconstruida.

La razón por la que la DCT produce resultados excelentes en los puntos extremos no es trivial ni sencilla de explicar. Para ello, deben interpretarse la DCT y la DFT como si se tratara de desarrollos en serie de Fourier de señales periódicas. La idea de esta interpretación es sencilla, se trata de determinar una señal periódica, cuyo desarrollo en serie de Fourier, produce los mismos coeficientes que los que obtenemos aplicando la DCT o la DFT. Los resultados se ilustran en la figura 4.26, donde se representa una señal de duración finita $x[n]$ junto con sus extensiones periódicas equivalentes a la DCT y a la DFT. En el segundo caso, la extensión periódica consiste simplemente en repetir la forma básica de la señal en el tiempo. El desarrollo en serie de Fourier de esta señal periódica producirá los mismos coeficientes transformados que la DFT. Sin embargo, para obtener los coeficientes de la DCT se debe realizar la extensión periódica de la señal base de una forma algo más compleja, reflejando la señal base respecto a la última muestra y recorriéndola en sentido inverso. El periodo fundamental de la señal es ahora el doble que su duración original. Nótese que, como esta señal base tiene simetría respecto a los coeficientes centrales, las exponenciales complejas de la serie de Fourier cancelarán sus componentes imaginarios dando lugar a coeficientes reales.

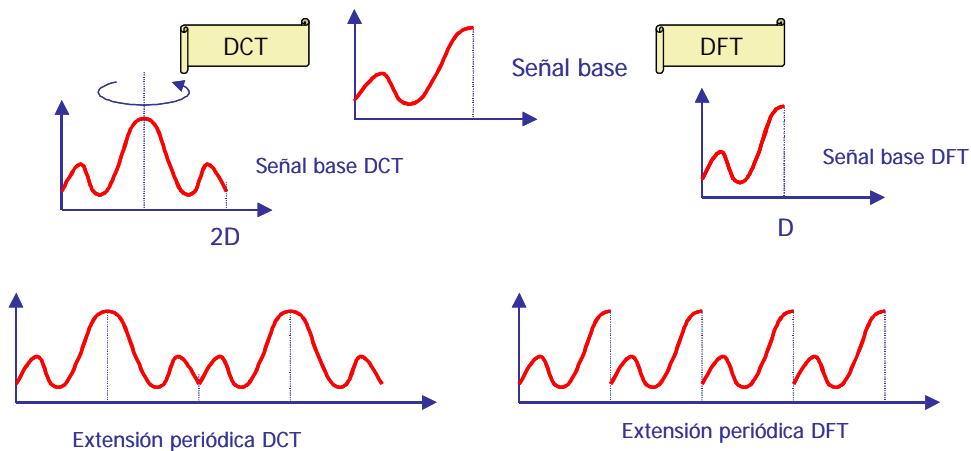


Fig. 4.26 Desarrollos periódicos implícitos en la DCT y la DFT

Con estos resultados es evidente que las mejoras en la aproximación de los puntos extremos en la DCT se deben a que la forma en que se realiza su extensión periódica implica que no aparezcan

cambios bruscos en la secuencia. En la extensión periódica asociada a la DFT, los cambios entre el origen y el final de la secuencia base sólo pueden aproximarse correctamente si se tienen en cuenta todos los términos de alta frecuencia, ya que la aparición de las discontinuidades sólo pueden aproximarse mediante el uso de estos componentes. Si estos no intervienen en la reconstrucción de la señal, se producen errores significativos en estas regiones.

En este punto conviene recordar que hemos comentado que es conveniente imponer que los tamaños de los bloques sean una potencia de 2, para poder utilizar algoritmos de cálculo rápido. En el contexto de la DFT, es habitual introducir ceros adicionales a la señal para que la longitud final sea una potencia de 2; sin embargo, con la DCT, la introducción de ceros tiene como coste adicional que se rompe con la naturaleza continua de su extensión periódica, por lo que perderemos la propiedad de aproximar con precisión las muestras extremas del bloque de señal. En este caso, pues, es extremadamente importante que los tamaños de los bloques sean una potencia de 2.

4.7.2 Extensión de la transformada coseno a dos dimensiones

La transformada coseno puede extenderse de forma directa a dos dimensiones. La ecuación que se utiliza para la transformada directa es:

$$C[k, l] = \alpha[k] \cdot \alpha[l] \cdot \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} x[n, m] \cdot \cos\left(\frac{(2n+1)k\pi}{2N}\right) \cdot \cos\left(\frac{(2m+1)l\pi}{2M}\right) \quad (4.21)$$

donde suponemos que el bloque de imagen $x[n, m]$ tiene unas dimensiones de N filas por M columnas, lo que define el ámbito de validez de los índices n y k entre 0 y $N-1$ y de los índices m y l entre 0 y $M-1$. En general sólo trataremos con bloques cuadrados cuyos tamaños son, como hemos comentado en el apartado anterior, una potencia de 2.

La transformada inversa puede obtenerse de forma análoga:

$$x[n, m] = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} \alpha[k] \cdot \alpha[l] \cdot C[k, l] \cdot \cos\left(\frac{(2n+1)k\pi}{2N}\right) \cdot \cos\left(\frac{(2m+1)l\pi}{2M}\right) \quad (4.22)$$

Las transformadas bidimensionales pueden obtenerse a partir de las transformadas unidimensionales. Para ello, es necesario aplicar la transformada unidimensional a las filas del bloque de imagen y posteriormente, sobre los resultados obtenidos, aplicar nuevamente la transformada unidimensional sobre las columnas. Este esquema de tratamiento de la imagen mediante transformadas unidimensionales se ilustra en la figura 4.27. Evidentemente, puede intercambiarse el orden del análisis entre filas y columnas. En general, los algoritmos que procesan la imagen de acuerdo con la expresión bidimensional de la ecuación 4.22 suelen ser más eficientes que los algoritmos unidimensionales.

En la figura 4.28 se muestra el resultado de aplicar la transformada coseno a los niveles de gris de un bloque de la imagen. Obsérvese cómo los coeficientes de mayor peso quedan concentrados alrededor de la esquina superior izquierda. Esta figura pretende mostrar el concepto de los métodos transformados desde un punto de vista puramente computacional. La matriz de la izquierda, en el dominio espacial y la matriz de la derecha, en el dominio de la transformada coseno, representan

exactamente la misma información y puede pasarse de una matriz a la otra aplicando la transformada directa o la transformada inversa. Es evidente que, si los resultados obtenidos con este bloque de imagen son generalizables a cualquier otro bloque, la representación en el dominio transformado realiza una compactación considerable de la energía de la señal en unos pocos coeficientes.

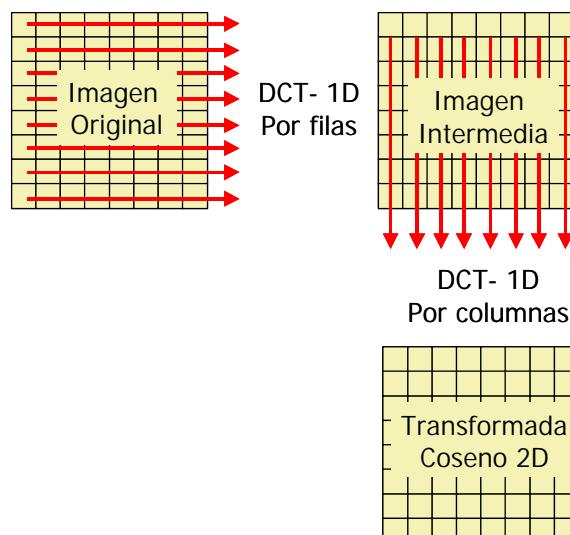


Fig. 4.27 Cálculo de la transformada coseno de una imagen a partir de transformadas 1D

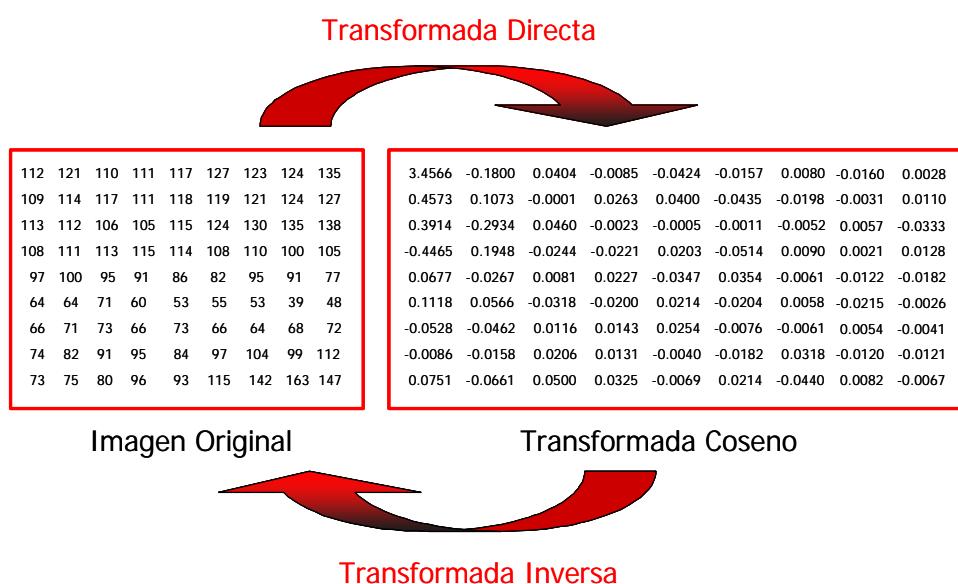


Fig. 4.28 Ejemplo numérico de una transformada coseno de un bloque de imagen

En la figura 5.29 se representan, utilizando niveles de gris, una subimagen de 16x16 elementos y su transformada coseno. En este ejemplo, los niveles altos de la transformada coseno se representan como blancos, mientras que a los niveles bajos se les ha asignado el negro. También se representa un diagrama en tres dimensiones de los coeficientes transformados. Nuevamente, los niveles de los coeficientes decrecen a medida que nos alejamos del extremo superior izquierdo, siendo la mayoría próximos a cero.

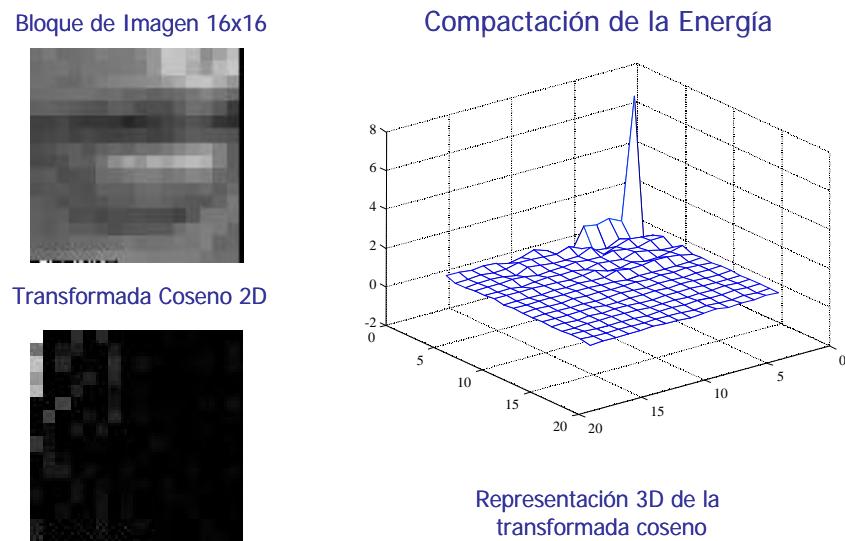


Fig. 4.29 Representación de un bloque de imagen y su transformada coseno

La ecuación 4.22 sugiere una interpretación interesante de la transformada coseno. Podemos agrupar los factores constantes y las funciones coseno en funciones genéricas y expresar la imagen como:

$$x[n, m] = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} C[k, l] \cdot H_{k,l}[n, m] \quad (4.23)$$

donde $H_{k,l}[n,m]$ representa:

$$H_{k,l}[n, m] = \alpha[k] \cdot \alpha[l] \cdot \cos\left(\frac{(2n+1)k\pi}{2N}\right) \cdot \cos\left(\frac{(2m+1)l\pi}{2M}\right) \quad (4.24)$$

y pueden interpretarse como un base de subimágenes de tamaño $N \times M$ cuya forma depende de los coeficientes k y l . De acuerdo con este punto de vista, los coeficientes transformados $C[k, l]$ nos indican en qué medida participa cada una de estas subimágenes en la obtención de la subimagen original. Por tanto, los bloques originales pueden considerarse como una superposición de estas imágenes base donde los coeficientes transformados indican el peso de la imagen base en la

reconstrucción de la imagen final. Los coeficientes próximos a cero pueden, por tanto, eliminarse sin que afecten a la calidad de la imagen reconstruida, ya que su matriz asociada apenas tiene contribución en la imagen original.

En la figura 4.30 se representa la imagen base obtenida para $k=0$ y $l=2$, tanto en forma de niveles de gris como en gráfico tridimensional. La subimagen es constante en el sentido horizontal y tiene una oscilación lenta en el vertical, concretamente de un periodo completo. Por tanto, el coeficiente asociado a esta matriz base tendrá en cuenta las variaciones lentas en el sentido vertical de la imagen original. Nótese que en el ejemplo de la figura 5.29, en el que la imagen era básicamente un contorno horizontal (variaciones de baja frecuencia en el sentido vertical), los coeficientes transformados de mayor peso eran los que se obtenían para los primeros valores de la variable l ($l=0, l=1, l=2, l=3$ y $l=4$) manteniendo la variable $k=0$. Este resultado indica que predominaban los cambios lentos en el sentido vertical.

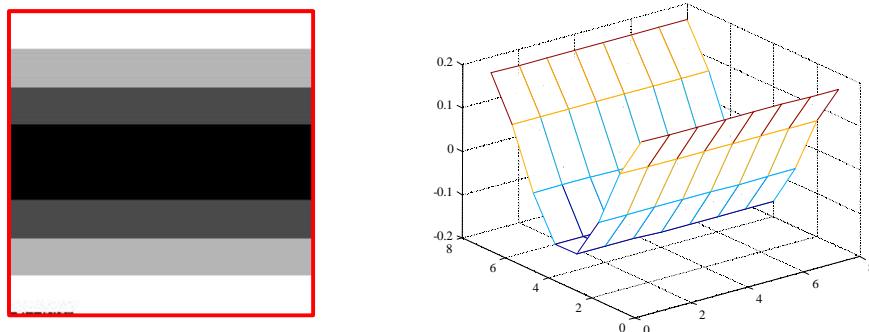


Fig. 4.30 Representación de la imagen base $H_{02}[n,m]$

Tal y como se desprende de la ecuación 4.24, las imágenes base están formadas por productos de dos funciones cosenoideales cuyas frecuencias están directamente relacionadas con las variables k y l . La variable k actúa directamente sobre el eje vertical (componente n de la imagen) mientras que la l lo hace sobre el eje horizontal. A medida que k o l aumentan, aumenta la frecuencia de la función coseno, por lo que se producen un mayor número de oscilaciones. En consecuencia, los valores de k y l pequeños representan zonas de baja frecuencia mientras que los valores altos representan alta frecuencia.

Las imágenes base que obtenemos para todos los pares de valores k y l con una transformada coseno de 8x8 muestras se representan en la figura 4.31. Si observamos la primera fila de imágenes base ($k=0$) vemos que la frecuencia horizontal de la señal va aumentando a medida que aumenta el valor de l . En cambio, si mantenemos $l=0$ y nos desplazamos en el sentido vertical, aumenta la frecuencia vertical manteniendo la imagen constante en el eje horizontal. Las zonas donde tanto l como k son elevados corresponden a zonas de alta frecuencia en los dos sentidos. Es importante que el lector retenga dónde se sitúan las imágenes base y sus contenidos frecuenciales para interpretar correctamente los resultados de una transformada coseno.

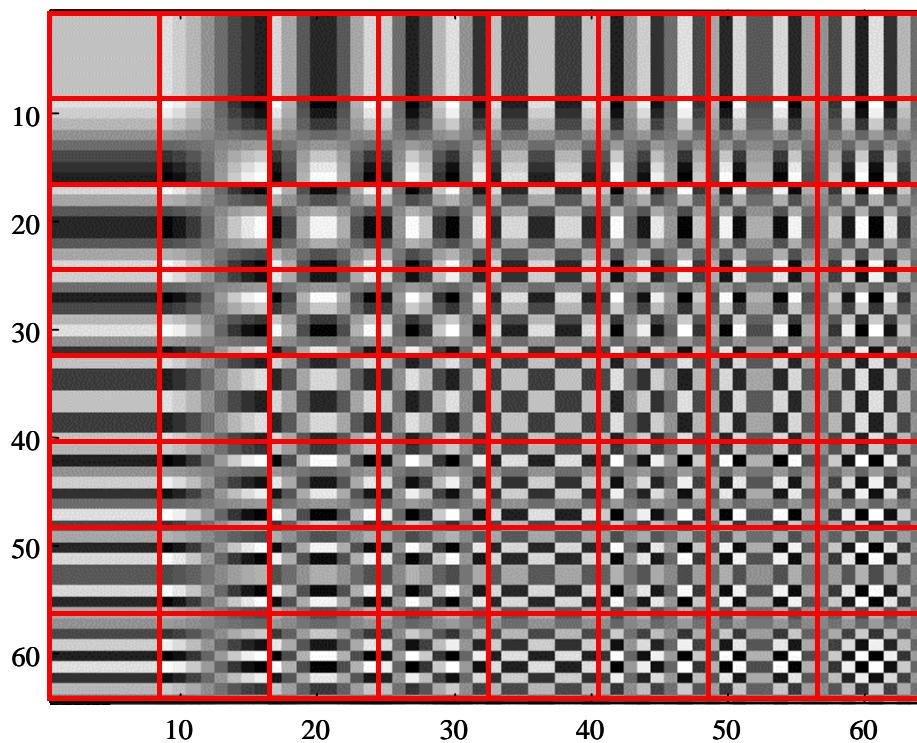


Fig. 4.31 Representación de las imágenes base de la transformada coseno 2D de 8x8

El resultado de una transformada puede interpretarse directamente a partir del conjunto de las imágenes base. En la figura 4.32 se muestran los valores numéricos que se obtienen en la transformada 4x4 de un bloque de imagen. Estos valores indican el grado en que interviene cada imagen base en la reconstrucción del bloque de imagen original. La posición de las imágenes base es la misma que la de los coeficientes en la transformada, por lo que los resultados de la transformada pueden proporcionar una idea inmediata del contenido frecuencial de la señal. En la figura 4.33 se representa una imagen y los resultados de aplicar la transformada coseno a los bloques de 16x16 píxeles que la constituyen. Las transformadas coseno se representan invertidas en nivel de gris, es decir, los valores próximos a cero se muestran en blanco y los valores altos en negro. Los resultados de las transformadas se representan en la misma posición que el bloque de imagen al cual corresponden.

Adviértase que en las zonas que en la imagen original son uniformes se obtiene una transformada coseno en la que prácticamente sólo intervienen los primeros coeficientes de la zona superior izquierda, mientras que en las zonas con detalles de alta frecuencia (plumas, sombrero, etc) también son importantes los coeficientes de alta frecuencia, aumentando el contenido de las matrices transformadas en estas regiones. Obsérvese también cómo en las zonas con presencia de contornos verticales únicamente son significativos los coeficientes de alta frecuencia situados en la parte superior izquierda de las matrices transformadas.

Un buen método de codificación deberá ser capaz de advertir todas estas diferencias y codificar más o menos coeficientes en función de las características de la transformada. La propiedad de que la posición de los coeficientes transformados está directamente relacionada con el contenido frecuencial de la imagen base puede ser aprovechada para mejorar la compresión, basándonos en las características del sistema visual.

7	5	0.4	0.2
4	1	0.25	0.1
0.3	0.2	0.08	0.01
0.1	0.05	0.02	0.003

La imagen $x[n,m]$ puede representarse como:

$$7 H_{00} + 5 H_{01} + 4 H_{10} + 0.3 H_{02} + 1 H_{11} + 0.4 H_{02} + \dots$$

Transformada Coseno

Fig 4.32 Interpretación de la transformada coseno como combinación lineal de imágenes base



Fig. 4.33 Representación de las transformadas coseno de los bloques de una imagen

El sistema visual es bastante crítico con codificaciones imprecisas de las componentes de baja frecuencia, por lo que es conveniente que estas regiones sean codificadas con bastante exactitud. Respecto a la alta frecuencia, el sistema visual tolera bastante bien que su codificación sea imprecisa, siempre que esté presente en la imagen decodificada. Si no está presente, se observa una pérdida en

los detalles de la imagen y un desenfoque aparente de la misma que resulta molesto. La codificación imprecisa de la alta frecuencia significa que pueden utilizarse cuantificadores con pocos niveles. El sistema visual acepta bastante bien que los cambios de nivel de gris en un contorno no coincidan exactamente con los de la imagen original, siempre y cuando existan estos cambios bruscos que permiten definir los objetos y aportan nitidez a la imagen resultante. Básicamente, existen dos estrategias para codificar los coeficientes de la transformada coseno: la codificación por zonas y la codificación por umbral, cuya filosofía examinaremos con cierto detalle en los siguientes apartados.

4.7.3 Codificación por zonas

La codificación por zonas se basa en establecer un procedimiento para determinar qué coeficientes transformados son codificados y cuáles no. El procedimiento para determinar las zonas que se codifican puede ser de aplicación general, es decir, invariable para todas las imágenes que se codifican, o depender del contenido específico de la imagen. En el primer caso, suele suponerse que las imágenes pueden modelarse como algún proceso estocástico predefinido (normalmente Markov) con el que puede realizarse una estimación *a priori* de las varianzas y márgenes dinámicos de los coeficientes transformados. A partir de este modelo y en función de la compresión deseada, se decide qué coeficientes serán codificados. Esto da lugar a una máscara como la de la figura 4.34, en la que los coeficientes que se codifican se representan como un uno, mientras que los que no se codifican se representan como un cero. En general, sólo se codifican los coeficientes para los que se estima una varianza significativa en función del contenido de la imagen, que suelen ser de baja frecuencia y se sitúan cerca del coeficiente de continua.

En la mayoría de aplicaciones es habitual utilizar distintas resoluciones del cuantificador en función de la posición de los coeficientes. Así, los coeficientes más importantes se codifican con un mayor número de bits, mientras que los menos significativos o los que el sistema visual no percibe con tanta sensibilidad se codifican con menor precisión. Una forma de indicar el número de bits con que se codifica cada coeficiente es mediante una máscara como la de la figura 4.34, pero en la que se indica específicamente el número de bits utilizado para cada coeficiente. En la figura 4.35 se muestra una posible distribución de los bits en función de la posición de los coeficientes.

1	1	1	1	1	1	0	0	0
1	1	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Fig. 4.34 Ejemplo de máscara de coeficientes para la codificación por zonas

12	8	6	4	2	2	0	0
8	6	2	2	0	0	0	0
6	2	2	0	0	0	0	0
4	2	0	0	0	0	0	0
2	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Fig. 4.35 Codificación por zonas con número de bits variable en función de la importancia del coeficiente

Una variante de la codificación por zonas fija consiste en evaluar, para cada imagen, la dinámica de las transformadas coseno de todos los bloques que la constituyen y decidir, posteriormente, cómo debe realizarse la asignación de zonas a codificar en función de la imagen. El procedimiento consiste en determinar las transformadas coseno de cada bloque de la imagen y posteriormente estimar la potencia o la varianza de cada uno de los coeficientes. Para codificar la imagen se utilizarán los coeficientes con los que se haya obtenido una varianza mayor.

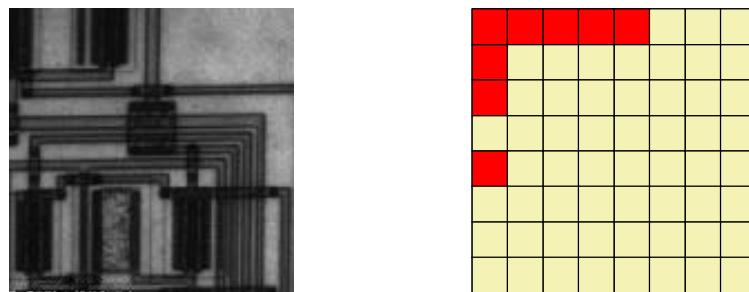


Fig. 4.36 Una imagen y los 8 coeficientes transformados más representativos

Esta estrategia tiene la ventaja de que, si la imagen tiene mucho contenido en una determinada frecuencia espacial o dirección, estos coeficientes tendrán mayor importancia y serán codificados correctamente. Generalmente, se dispone de un número total de bits con el que debe realizarse la codificación de la imagen y éstos se asignan en función de la importancia relativa de cada coeficiente. Este procedimiento tiene la ventaja de que la imagen se codifica de forma más eficiente, aunque es necesario realizar un estudio previo de la varianza de los coeficientes transformados, lo que supone

una mayor carga computacional. También deberemos almacenar, junto con los coeficientes, la máscara zonal que ha sido utilizada y los bits asignados a cada uno de los coeficientes. En la figura 4.36 se representa una posible máscara obtenida utilizando este procedimiento y utilizando únicamente los 8 coeficientes más representativos. En este caso, la imagen tiene muchos detalles en el sentido horizontal y vertical, por lo que los coeficientes con mayor varianza se concentran en la primera fila y primera columna.

La selección entre un método de codificación zonal fijo o uno adaptado a la imagen depende de las características de las imágenes con que operemos, el tiempo de proceso que podamos dedicar y el factor de compresión que deseemos obtener.

4.7.4 Codificación por umbral

Esta estrategia suele proporcionar mejores resultados que la codificación por zonas y consiste en utilizar únicamente aquellos coeficientes que superen un determinado umbral. Los coeficientes por debajo de este límite se suponen nulos. El análisis de umbral se realiza de forma independiente para cada bloque. En general, la mayoría de los coeficientes de un bloque transformado no superarán el umbral y se codificarán como ceros. El inconveniente es que, al no conocer a priori qué coeficientes superarán o no el umbral, deberemos codificarlos todos. Sin embargo, como la mayoría son nulos, pueden codificarse de forma muy eficiente utilizando alguna variante de los métodos de codificación mediante series de ceros.

Para optimizar el número de ceros consecutivos que aparecen al codificar la matriz transformada, suelen ordenarse los coeficientes en la secuencia que se representa en la figura 4.37. El recorrido en zigzag diagonal permite que los coeficientes de baja frecuencia sean los primeros en codificarse, independientemente de que correspondan a frecuencias horizontales o verticales. Los últimos coeficientes siempre corresponden a frecuencias elevadas, por lo que, lo más probable, es que la mayoría de ellos sean nulos y se produzcan series largas de ceros consecutivos.

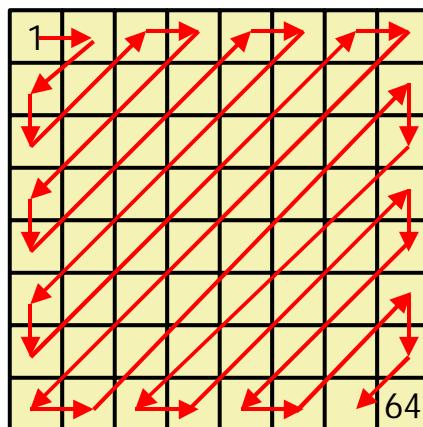


Fig. 4.37 Secuencia de ordenación de los coeficientes transformados en la codificación por umbral

La ventaja de este procedimiento es que, si en un bloque aislado de la imagen existe una componente de alta frecuencia significativa, su coeficiente transformado superará el umbral y será codificado correctamente, proporcionando una excelente calidad en la imagen reconstruida. En general, la codificación por umbral suele obtener unos factores de compresión algo inferiores a la codificación zonal. No obstante, la calidad final de las imágenes es bastante superior. Los estándares JPEG y MPEG utilizan variantes de la codificación por umbral de la transformada coseno.

En la figura 4.38 se representa una posible codificación de un bloque transformado mediante la técnica de umbral. Los resultados se representan de forma esquemática, indicando el nivel para los coeficientes que superan el umbral y el número de elementos para las secuencias de ceros consecutivos (se muestran precedidas de una C). Para optimizar la compresión es habitual insertar un código especial, que indica que todos los elementos restantes de la matriz no superan el umbral y que por lo tanto pueden aproximarse por cero. Este código se indica con el símbolo END en este ejemplo. Como en el caso de la codificación zonal pueden utilizarse cuantificadores con distinta precisión en función de la posición del coeficiente, con lo que conseguimos adecuar los bits dedicados a cada término a las características del sistema visual humano. En las secciones siguientes describiremos cómo se realizan estas asignaciones de bits a los coeficientes en los estándares de codificación más populares.

72	12	11	7	1	3	0	0
18	14	5	2	1	0	0	0
10	6	0	0	0	7	0	0
8	1	0	0	0	0	0	0
2	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Código: 72, 12, 18, 10, 14, 11, 7, 5, 6, 8, 2, 1, C1, 2, 1, 3, 1, C3, 1, C9, 7, END

Fig. 4.38 Ejemplo de codificación por umbral

4.8 Codificación de imágenes estacionarias: el estándar JPEG

El estándar de codificación de imágenes estacionarias *JPEG* ha tenido un gran éxito y actualmente es soportado por prácticamente todas las aplicaciones informáticas que requieren el manejo de imágenes. También se utiliza como formato de almacenamiento básico en cámaras de fotografía digital y en algunas tarjetas de digitalización de vídeo en las que cada fotograma se codifica en JPEG de forma individual, lo que permite obtener una secuencia de imágenes comprimidas de elevada calidad que,

posteriormente, puede recodificarse mediante sistemas que aprovechen no sólo la redundancia espacial, sino también la temporal. Hasta la aparición de otros codificadores más adecuados, el formato JPEG también se utilizó para transmitir imágenes de vídeo conferencia de baja calidad en tiempo real.

El formato JPEG recibe su nombre del grupo IEC JTC1/SC29/WG10 de la ISO, conocido como el *Joint Photographic Experts Group*, que en 1982 inician las sesiones para establecer un conjunto de formatos que permitan obtener una codificación eficiente de imágenes estacionarias. El objetivo principal del grupo de trabajo es establecer un conjunto de recomendaciones, de uso público, que mitiguen la aparición de formatos propietarios para la codificación de imágenes en soporte informático. En 1986 se incorpora el grupo SGVIII del CCITT para trabajar conjuntamente en la definición básica del formato. Las primeras recomendaciones aprobadas del estándar aparecen en 1991.

Actualmente el JPEG es un formato de alta difusión en multitud de aplicaciones. Periódicamente se incorporan extensiones, revisiones, mejoras y adiciones con la intención de actualizar sus prestaciones a la tecnología actual. Así, recientemente se ha redefinido parte del estándar (JPEG 2000) introduciendo métodos basados en la transformada *wavelet* que permiten mantener la calidad de imagen para factores de compresión muy elevados. También se ha propuesto una nueva estrategia para la compresión sin pérdidas conocida como JPEG LS (JPEG Lossless), que está adquiriendo una amplia difusión.

Los codificadores JPEG se orientan a imágenes multinivel monocromas o de color (múltiples espacios de coordenadas de color) y pueden considerarse como un conjunto de herramientas que admiten diferentes modos y estrategias de compresión en función de las características del sistema en el que debe utilizarse. El JPEG no está diseñado para imágenes binarias de documentos gráficos o texto, para las que existe un conjunto de métodos específicos que se agrupan bajo el estándar JBIG.

Los objetivos que se propusieron la comisión encargada de definir el estándar JPEG pueden sintetizarse en los siguientes puntos:

- a) Conseguir unos codificadores y decodificadores cuyas tasas de compresión y calidades de imagen se situaran cerca de los límites de la tecnología actual.
- b) Los métodos propuestos debían ser útiles para codificar imágenes monocromáticas en niveles de gris, imágenes de color o imágenes multicomponente (p.ej. imágenes obtenidas por satélites de inspección que proporcionan componentes en bandas distintas al espectro óptico). Además, los métodos propuestos debían ser independientes del tamaño y resolución espacial de la imagen.
- c) La implementación de los métodos debía poder realizarse mediante *software* que pueda ejecutarse en múltiples plataformas de forma eficiente y rápida. Los algoritmos deben poder realizarse mediante circuitos integrados dedicados de coste reducido o medio.

Los resultados y la difusión del estándar JPEG confirman que todos estos objetivos básicos han sido ampliamente logrados.

Los modos de compresión definidos por el estándar JPEG son:

- a) *Modo de codificación secuencial.* Define los métodos utilizados para comprimir las imágenes siguiendo un orden de codificación de los bloques de izquierda a derecha y de arriba abajo. El decodificador recupera la imagen en el mismo orden. Existen diversas variantes de este método, la más popular es el denominado método base (*baseline*) que describiremos con cierto detalle posteriormente.
- b) *Modo de codificación progresivo.* La imagen se codifica en múltiples exploraciones utilizando siempre la misma resolución espacial. El decodificador puede obtener, de forma rápida, una primera aproximación de la imagen, que posteriormente se refina al avanzar el proceso de decodificación. Es particularmente útil para el acceso a imágenes localizadas en centros remotos en los que las comunicaciones son lentas.
- c) *Modo de codificación sin pérdidas.* Originalmente el JPEG definió un modo sin pérdidas basado en la codificación diferencial y en estructuras de predicción simples con características muy parecidas a las que hemos considerado en apartados anteriores. Este modo de codificación obtiene factores de compresión relativamente bajos (1:2) cuyo uso en aplicaciones informáticas o de transmisión de datos ha sido moderado. Actualmente se ha propuesto un nuevo método sin pérdidas o con pérdidas muy pequeñas (JPEG LS) que está creando grandes expectativas.
- d) *Modo de codificación jerárquica.* La imagen se codifica con distintos niveles de resolución, lo que facilita que pueda ser mostrada en distintos tipos de *display*. La resolución más baja es la primera que se obtiene en el decodificador. A partir de esta imagen preliminar puede irse decodificando el resto de datos hasta obtener la resolución deseada. El proceso de decodificación puede interrumpirse en cualquier momento.

En esta sección analizaremos con detalle el denominado modo secuencial base del estándar JPEG, que es el algoritmo más utilizado y que forma la base para la codificación de las imágenes intra en el estándar MPEG. Proporcionaremos ejemplos detallados de cómo se realizaría la codificación de imágenes en blanco y negro y su extensión a imágenes de color. También veremos cómo puede extenderse el modo secuencial base a un modo de codificación progresivo. Los modos de codificación sin pérdidas y la codificación jerárquica no serán considerados.

4.8.1 JPEG: modo secuencial base (*baseline*)

Este modo de codificación es, como hemos comentado, el más ampliamente difundido y constituye la base para la codificación de las imágenes de referencia en los estándares de transmisión para videoconferencia (H.261, H.263), multimedia (MPEG-1) y televisión digital (MPEG-2). Por el momento consideraremos únicamente la codificación de imágenes monocromáticas con 256 niveles de gris (8 bits). Otros modos de codificación del JPEG admiten imágenes de 12 bits. Desde el punto de vista conceptual, los procedimientos utilizados son análogos.

En el esquema de la figura 4.39 se representa el procedimiento básico para codificar la imagen, que se descompone en bloques de 8x8 píxeles ordenados de forma secuencial, primero por columnas (de

izquierda a derecha) y posteriormente por filas (de arriba abajo). Se extrae la componente de nivel de gris medio (se resta el nivel -128) a cada uno de los bloques, con lo que las imágenes resultantes quedan codificadas entre los niveles -128 y 127. A estas subimágenes se les aplica la transformada DCT bidimensional, obteniendo los resultados con una precisión de 3 bits adicionales respecto a la precisión de la entrada (11 bits). Los coeficientes resultantes quedan, por tanto, representados dentro del margen dinámico [-1024, 1023]. A estos coeficientes se les aplican tablas de recuantificación, que pueden ser definidas por el propio usuario (deberán incluirse las tablas en las cabeceras del fichero) o utilizar las tablas que por defecto recomienda el estándar (tablas de Lohscheller). Finalmente, se realiza una codificación de longitud variable de los coeficientes resultantes.

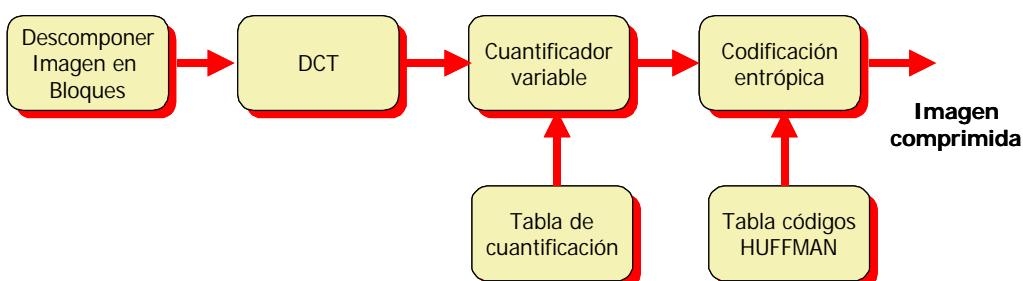


Fig. 4.39 Procedimiento de codificación del modo secuencial base (baseline)

La cuantificación de los coeficientes obtenidos mediante la transformada coseno presenta ciertas peculiaridades que deben ser consideradas con detalle. El tratamiento del coeficiente de continua (DC), que corresponde al coeficiente (0,0) de la matriz transformada, se debe codificar con mayor precisión que los coeficientes de alterna (AC). Esto es debido a que los errores en este coeficiente supondrían una variación del nivel de gris medio de todo el bloque, con lo que la estructura de codificación por bloques podría resultar visible en la imagen decodificada. Por este motivo, los coeficientes DC se codifican en modo diferencial, respecto al coeficiente de continua del bloque anterior. Este tipo de codificación supone una extensión del posible margen dinámico de la diferencia que se extenderá entre los valores de [-2047, 2047], ya que, aunque con poca probabilidad, es posible que se produzcan las diferencias entre dos valores extremos de la componente de continua, es decir: 1023-(-1024) ó -1024-(1023).

Las tablas de cuantificación indican la importancia relativa de cada coeficiente transformado y suelen representarse como el paso de cuantificación con el que se recodificará cada elemento. En la figura 4.40 se representa la tabla de cuantificación de Lohscheller, que se recomienda para la cuantificación de la componente de luminancia en imágenes estáticas con una resolución igual a la del formato ITU 601. Esta tabla se obtiene de forma experimental, a partir de pruebas de calidad subjetivas que tienen en cuenta la respuesta del sistema visual humano. Obsérvese como el paso de cuantificación aumenta, es decir, se utilizan menos niveles, a medida que aumentan las frecuencias. Existen diferencias de simetría entre las componentes horizontales y verticales que sólo se justifican a partir de la naturaleza experimental con la que se han obtenido los resultados. También debe observarse que el paso de cuantificación utilizado para el coeficiente de continua es mayor que los pasos de cuantificación utilizados para los componentes AC de baja frecuencia. Este hecho podría parecer contradictorio con

la mayor precisión que requiere la codificación de la componente de continua. Sin embargo, debe tenerse en cuenta que esta componente se codifica de forma diferencial por lo que, en realidad, el número de niveles con el que se codifica es mayor que el número de niveles utilizado para las componentes AC. Así, teniendo en cuenta que el margen dinámico de la componente AC es el doble que el del resto de componentes, el número de niveles de cuantificación que se utilizan para codificar esta componente también es mayor que el que se utiliza para el resto de componentes (es decir; $2V_{\max}/16 > V_{\max}/11$).

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Matriz de cuantificación de luminancia Q_{uv}

Fig. 4.40 Matriz de Lohscheller para la cuantificación de los coeficientes transformados de luminancia

En general, las matrices de cuantificación dependen del tipo de componente de color que se está considerando. Las matrices de Lohscheller para las señales de diferencia de color para imágenes con una resolución equivalente a la del estándar ITU 601 se proporcionan en la figura 4.41.

Desde un punto de vista operativo, las matrices de cuantificación se utilizan para determinar el nivel real con el que se codificará cada coeficiente transformado. El cálculo del nivel real se realiza dividiendo los coeficientes de la DCT por el número de pasos de cuantificación especificados por la matriz de Lohscheller y aproximando el resultado obtenido por el entero más próximo.

$$V_{uv} = \text{round}\left(\frac{DCT_{uv}}{Q_{uv}}\right) \quad (4.25)$$

donde V_{uv} representa los coeficientes definitivos que deberemos codificar, DCT_{uv} los coeficientes originales obtenidos mediante la transformada coseno y Q_{uv} la matriz de cuantificación. En la figura 4.42 se muestra un ejemplo del proceso de cuantificación completo que se aplica a un bloque de luminancia de 8x8 píxeles de la imagen original. La matriz superior izquierda muestra los valores originales de los píxeles de la imagen, mientras que en la de la derecha se ha realizado la sustracción

del nivel de gris medio. En la matriz izquierda de la línea inferior se muestra el resultado de aplicar la transformada coseno al bloque de datos. La componente continua obtenida para este bloque de datos tiene el valor 1002. En este ejemplo, suponemos que en el bloque anterior habíamos obtenido una componente continua de valor 945, por lo que el resultado que finalmente debemos codificar es la diferencia entre las dos componentes $1002 - 945 = 57$. La matriz de la parte inferior derecha es el resultado de la cuantificación mediante la matriz de Lohscheller.

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Matrices de cuantificación de crominancia Q_{uv}

Fig 4.41 Matriz de Lohscheller para la cuantificación de los coeficientes transformados de crominancia

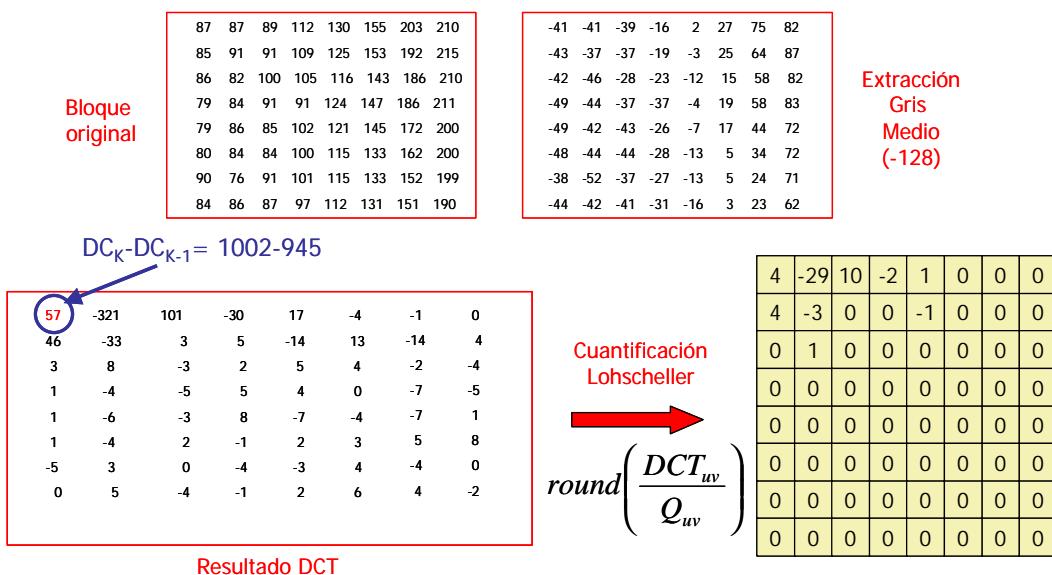


Fig. 4.42 Ejemplo de cuantificación de una subimagen mediante las matrices de Lohscheller

Obsérvese que, tal y como esperábamos, la mayoría de coeficientes cuantificados toman el valor cero, de modo que la codificación combinada entre códigos de longitud variable y métodos de longitud de series puede resultar muy eficiente. Debemos notar también que en ningún momento se presupone que el resultado de los coeficientes de alta frecuencia será nulo. Esto significa que si en una subimagen estos coeficientes fueran realmente significativos, sus valores serían cuantificados y, por lo tanto, las componentes de alta frecuencia podrían recuperarse de la información comprimida. En consecuencia, la compresión JPEG no necesariamente presupone una pérdida de los detalles y componentes de alta frecuencia de la imagen. Las matrices de cuantificación tienen en cuenta la importancia relativa de cada coeficiente y su incidencia en la visualización de la imagen.

Por otra parte, las matrices de cuantificación permiten un control directo sobre la calidad de la imagen. Si utilizamos matrices con unos pasos mayores, obtendremos una concentración de toda la energía de los coeficientes alrededor de la componente DC, por lo que la imagen se codificará utilizando un menor número de bits. Sin embargo, un mayor paso de cuantificación supone que los coeficientes se aproximan de forma más tosca, por lo que la calidad de la imagen reconstruida empeora. En general, cuando se disminuye el factor de calidad del compresor JPEG, se observa una pérdida de las componentes de alta frecuencia. Esto es fácilmente comprensible si tenemos en cuenta que a medida que aumentamos el paso de cuantificación es posible que todas las componentes de alta frecuencia queden aproximadas como ceros, por lo que en realidad no resultan codificadas.

Una vez los coeficientes han sido cuantificados, debe procederse a su codificación mediante códigos de longitud variable (Huffman) y códigos de longitud de series (RLE). El estándar JPEG proporciona diversas recomendaciones sobre las tablas de Huffman que pueden ser utilizadas. También es posible, utilizar extensiones del método base, que permiten que el usuario proporcione sus propias tablas de Huffman, especialmente adaptadas al tipo de imágenes que se pretende comprimir. La ganancia media que se obtiene por el uso de códigos de Huffman propios se sitúa entre el 2,4 % y el 7,9 % de mejora en el factor de compresión, por lo que, teniendo en cuenta la mayor carga computacional asociada a esta filosofía, suele emplearse en muy pocas implementaciones. También es posible sustituir los códigos de longitud variable de Huffman por códigos aritméticos. En este caso puede obtenerse un factor de mejora adicional del 2,5 % respecto al uso de códigos de Huffman propios.

Consideraremos la codificación de los coeficientes utilizando las tablas recomendadas por el JPEG. Los procedimientos y tablas utilizados para codificar los coeficientes AC y DC son conceptualmente distintos, por lo que serán considerados de forma separada.

Los coeficientes AC se ordenan mediante una exploración en zigzag como la que hemos ilustrado en la figura 4.37. La estrategia concreta de codificación puede considerarse una mezcla entre códigos de Huffman modificados y códigos de longitud variable. Cada coeficiente se codifica utilizando dos tipos de palabras que denominaremos A y B. La palabra B se utiliza para indicar el nivel de los coeficientes no nulos en binario natural en complemento a 1. El número de bits que se utilizan para la codificación depende del nivel del coeficiente y se especifica en la palabra del tipo A. Así, para codificar un coeficiente con nivel 3, utilizaremos únicamente 2 bits, con el código 11. Un coeficiente de nivel 23 se codificaría con 5 bits y su código binario natural 10111. Si el coeficiente fuera negativo (-23), se seguirían utilizando 5 bits, pero ahora el código se expresaría en complemento a 1, es decir, 01000. En resumen, las palabras del tipo B expresan directamente el valor del coeficiente no nulo en binario con complemento a uno, utilizando un número de bits variable que previamente está definido en la palabra

de tipo A que la precede. El contenido de las palabras de tipo A es un tanto más complejo, ya que se utilizan para codificar la longitud de la serie de ceros y el número de bits del siguiente coeficiente no nulo. La información que contiene una palabra de tipo A se descompone, por tanto, en dos componentes que suelen denominarse R/S (Run/Size – que se asocia a la longitud de la serie de ceros y al tamaño en bits del siguiente dato no nulo).

Para analizar con mayor profundidad como se combinan los dos tipos de palabras en la codificación de una serie de coeficientes AC, consideremos el siguiente ejemplo, que correspondería a una lectura de los coeficientes transformados de una subimagen ordenados en zigzag:

12, 156, -13, 0, 0, 0, 5, 0, 0, 0, 0, 3, 0, 0, 0,resto de coeficientes nulos

El primer coeficiente AC a codificar toma el valor 12 y no tiene ninguna serie de ceros que le preceda. Por tanto, el código de la palabra A tendrá un valor R=0 y S=4. El primer parámetro indica que no existe ningún cero que precede al elemento y el segundo representa el número de bits necesario para codificar el nivel del coeficiente (12). Este número de bits puede calcularse de forma muy simple como el número de bits necesario para representar el valor absoluto del coeficiente en binario natural, en el que el bit más significativo toma el valor 1. Así, para representar el nivel -37 deberemos considerar el código binario del número 37 (10011) y contar el número de bits con el que puede realizarse la codificación (5). Como el margen dinámico de los coeficientes AC está situado entre [-1023, 1023], el valor máximo de S queda limitado a 10 bits (véase la tabla 4.2).

Siguiendo con este esquema, la secuencia de palabras A y B con las que codificaríamos la secuencia anterior sería:

A(0/4), B₁₂, A(0/7), B₁₅₆, A(0/4), B₋₁₃, A(3/3), B₅, A(4/2), B₃, A(EOB)

Nótese cómo las palabras del tipo A van indicando simultáneamente la longitud de ceros consecutivos en la secuencia de coeficientes que preceden a la codificación de un coeficiente. Los códigos de las palabras B pueden determinarse con facilidad expresando el valor del coeficiente en complemento a 1 con el número de bits definido en la palabra A. Así, B₁₂=1100, B₁₅₆=10011100, B₋₁₃=0010 (en complemento a 1), B₅=101, B₃=11.

Es interesante subrayar que la representación en complemento a uno permite especificar de forma unívoca el número de bits en función del nivel de los coeficientes, sean positivos o negativos. En la tabla 4.2 se muestra una relación entre el número de bits y el margen de valores de coeficientes que pueden representarse.

Tabla 4.2 Relación entre número de bits (S) y niveles de los coeficientes

Número de bits (S)	Niveles positivos	Niveles Negativos
1	{1}	{-1}
2	{3,2}	{-2,-3}
3	{7,6,5,4}	{-4,-5,-6,-7}
4	{15,14,13,12,...8}	{-8,-9,...,-15}
5	{31,30,...,16}	{-16,-17,...,-31}

Las palabras del tipo A también pueden utilizarse para indicar que el resto de coeficientes de la secuencia son nulos. Para ello se utiliza una palabra especial que se indica con el código EOB. Con objeto de optimizar los códigos de longitud variable con que se codificarán las palabras del tipo A el número máximo de una secuencia de ceros que se permite codificar es 15. En el supuesto de que un bloque de coeficientes tuviera más de 15 ceros consecutivos, se utiliza la palabra A(15/0), que indica la presencia de 16 ceros consecutivos. Pueden concatenarse varias palabras de este tipo. Así, si un coeficiente de valor no nulo estuviera precedido por 34 coeficientes nulos, se utilizaría la secuencia A(15/0), A(15/0), A(2/S), B para codificarlo.

Las palabras del tipo A se codifican mediante un código de Huffman que tiene en cuenta la estadística con la que se producen estos mensajes. Las tablas pueden ser confeccionadas por el propio usuario o utilizar las que se recomiendan en el estándar. Los códigos y tablas para codificar las componentes de luminancia o de color son distintos. En la figura 4.43 se proporciona un fragmento de los códigos de Huffman recomendados para codificar la componente de luminancia.

A(R/S)	Código Huffman	A(R/S)	Código Huffman
0/0 EOB	1010	1/8	111111110000110
0/1	00	1/9	111111110000111
0/2	01	1/10	111111110001000
0/3	100	2/1	11100
0/4	1011	2/2	11111001
0/5	11010	2/3	1111110111
		2/4	11111110100
0/9	111111110000010	2/5	111111110001001
0/10	111111110000011	2/6	1111111110001010
1/1	1100	2/7	1111111110001011
1/2	11011	2/8	1111111110001100
1/3	1111001	2/9	1111111110001101
1/4	111110110	2/10	1111111110001110

Fig. 4.43 Tablas de Huffman de los códigos (R/S) para las componentes de luminancia (fragmento)

La codificación de los componentes DC es algo más simple. También se utilizan dos tipos de mensajes A y B. Los mensajes del tipo B codifican el valor del coeficiente de continua en binario natural en complemento a 1 de forma totalmente análoga a como se codificaban los coeficientes AC. No obstante, los mensajes del tipo A son más simples, puesto que sólo se requiere codificar el número de bits (S) que se utilizan para representar los niveles de los coeficientes. Los mensajes del tipo A se codifican mediante tablas de Huffman que utilizan la estadística de los componentes DC. Las tablas de Huffman recomendadas por el estándar para la señal de luminancia se proporcionan en la gráfica de la figura 4.44.

En la figura 4.45 se proporciona un ejemplo completo de codificación de los coeficientes DC y AC de los coeficientes transformados de una subimagen de 8x8 elementos. En la parte de la derecha se indican los códigos que se obtienen para cada uno de los coeficientes ordenados según la exploración en zigzag convencional. Los códigos utilizados se han obtenido de las tablas de Huffman recomendadas. Nótese que toda la matriz puede codificarse utilizando únicamente 59 bits.

A(S)	Código
0	00
1	010
2	011
3	100
4	101
5	110
6	1110
7	11110
8	111110
9	1111110
10	11111110
11	111111110

Fig. 4.44 Tablas de Huffman para la codificación del número de bits de los coeficientes DC (luminancia)

<table border="1"> <tr><td>4</td><td>-29</td><td>10</td><td>-2</td><td>1</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>4</td><td>-3</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table>	4	-29	10	-2	1	0	0	0	4	-3	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Coeficiente de continua: S=3		100;	valor 4 100
4	-29	10	-2	1	0	0	0																																																																													
4	-3	0	0	-1	0	0	0																																																																													
0	1	0	0	0	0	0	0																																																																													
0	0	0	0	0	0	0	0																																																																													
0	0	0	0	0	0	0	0																																																																													
0	0	0	0	0	0	0	0																																																																													
0	0	0	0	0	0	0	0																																																																													
0	0	0	0	0	0	0	0																																																																													
0	0	0	0	0	0	0	0																																																																													
0	0	0	0	0	0	0	0																																																																													
Coeficientes AC:		100100																																																																																		
-29		A(0/5) 11010	valor -29 00010																																																																																	
4		1101000010																																																																																		
0, -3		A(0/3) 100;	valor 4 100																																																																																	
10		100100																																																																																		
0, 1		A(1/2) 11011	valor -3 01																																																																																	
-2		1101101																																																																																		
0, 0, 0, 0, 0, 1		A(0/4) 1011	valor 10 1010																																																																																	
0, -1		10111010																																																																																		
		Total: 100100 1101000010 100100 1101101 10111010 0101 11001 11110101 11000																																																																																		

Fig. 4.45 Ejemplo de codificación de un bloque de coeficientes transformados previamente cuantificados

El procedimiento de decodificación del JPEG es bastante trivial, al menos desde un punto de vista conceptual, una vez comprendida la codificación. En la figura 4.46 se muestra el esquema general de un decodificador, que, en esencia, es una réplica invertida del esquema del codificador que se ha presentado en la figura 4.39. En la figura se indica específicamente que las tablas de Huffman y los cuantificadores utilizados pueden recuperarse directamente de la imagen comprimida.

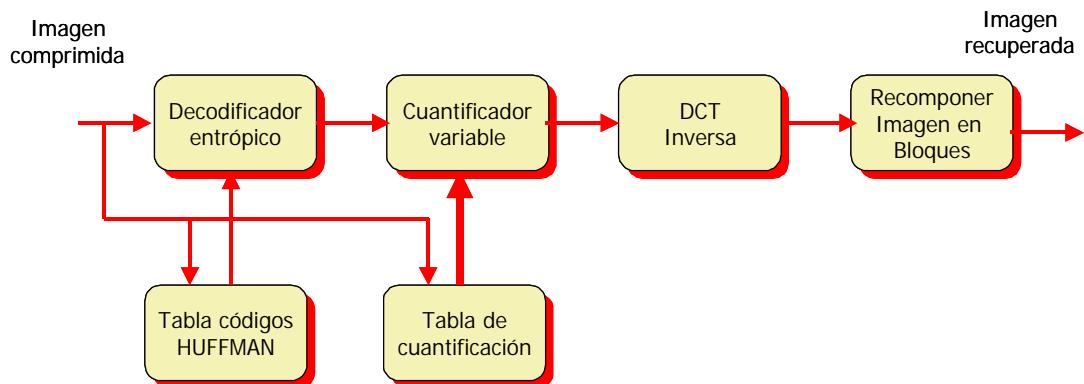


Fig. 4.46 Diagrama general del proceso de decodificación JPEG

La descripción de los procedimientos de codificación utilizados en el JPEG que se han presentado en este apartado pretende, únicamente, proporcionar una perspectiva global de la sistemática y metodología utilizada para la codificación de imágenes estáticas. El estándar completo proporciona los detalles sobre cómo se estructura la información de las tablas en las cabeceras de los ficheros y otras extensiones y variantes. Existen programas de dominio público que pueden encontrarse en diversos web's y ftp's oficiales del ISO o paralelos.

4.8.2 Codificación de imágenes con múltiples componentes.

Aunque el objetivo básico en la creación del estándar JPEG era la interoperabilidad entre distintas aplicaciones o sistemas de almacenamiento y transmisión de imágenes, muchos aspectos formales no han sido especificados en las recomendaciones, proporcionando cierto grado de libertad a los desarrolladores y fomentando la libre competencia entre distintas opciones de implementación. Probablemente, esta ha sido una de las causas principales de su amplia aceptación en el mercado, ya que, aun siendo un estándar, los desarrolladores disponen de un entorno genérico en el que pueden competir para optimizar la calidad de sus productos.

Existen muchos detalles que no están especificados en el estándar, como las dimensiones de las imágenes, los procedimientos para convertir imágenes que no están codificadas con 8 o 12 bits de resolución o los espacios de color idóneos para la representación de las imágenes en distintos tipos de dispositivos. No obstante, el estándar JPEG proporciona un amplio conjunto de especificaciones para trabajar con imágenes con múltiples componentes e incluso con distintas resoluciones en cada una de

las componentes, proporcionando métodos flexibles para la codificación. El estándar permite trabajar con un total de hasta 255 componentes, con 8 o 12 bits de resolución para los métodos con pérdidas y con 2 a 16 bits para los métodos sin pérdidas. Un ejemplo de imágenes con varias componentes son las imágenes en color. Otro ejemplo son las imágenes de inspección de la tierra por satélite en las que pueden capturarse imágenes en distintas bandas espectrales fuera del margen visible.

Las distintas componentes de la imagen pueden tener resoluciones espaciales distintas. En imágenes en color, es habitual que la componente de luminancia presente una resolución doble a las componentes diferencia de color, tanto en el sentido horizontal como en el vertical. El estándar admite distintas posibilidades para codificar varias componentes. Una primera alternativa sería codificar cada una de las componentes de forma secuencial. Así, en una imagen en color, primero podríamos codificar la componente de luminancia, después la componente de color C_r y finalmente la componente C_b . Otra alternativa consiste en intercalar (*interleaving*) los bloques correspondientes a las distintas componentes. En el caso de imágenes en color, en las que la relación de muestras entre la componente de luminancia y las señales diferencia de color es de 2:1, es habitual codificar un bloque de 8x8 elementos de la componente de luminancia, seguido de un bloque de 4x4 elementos de la componente C_r y de un bloque de 4x4 de la componente C_b . Todas estas alternativas y posibilidades son un claro ejemplo de la flexibilidad y capacidad de adaptación del estándar JPEG a las necesidades del usuario.

4.8.3 Modo progresivo

El modo progresivo es una alternativa del JPEG pensada para facilitar el acceso a bases de datos de imágenes a través de líneas de baja velocidad o en aplicaciones de internet. Cuando un usuario remoto accede a una imagen JPEG codificada en el modo básico, recuperará la imagen en el mismo orden en que ha sido codificada, es decir, de izquierda a derecha, por filas, y de arriba abajo. A medida que va decodificándose la imagen, el usuario la observa con la misma calidad con la que ha sido codificada. Esto significa que, en muchos casos, sólo nos daremos cuenta de que la imagen no es la que nos interesa cuando el proceso de transmisión de los datos está en una fase avanzada.

El modo progresivo resuelve este problema transmitiendo primero sólo una parte de la información de cada bloque y, gradualmente, transmitiendo el resto de los datos hasta proporcionar la imagen completa. Con ello, el usuario remoto puede obtener una primera aproximación global de la imagen desde los inicios de la transmisión. Esta estrategia de transmisión requiere utilizar un buffer de memoria considerable para almacenar todos los coeficientes de la imagen. A menudo el modo progresivo se confunde con el modo jerárquico. La diferencia entre los dos es que en el primero la resolución (número de píxeles) de la imagen transmitida es siempre igual a la de la imagen original, simplemente se varía el orden en el que se transmite la información. En el modo jerárquico se transmite primero una versión de la imagen con menor resolución y posteriormente se va refinando, aumentando la resolución en el dispositivo de representación. Con este modo podemos adaptarnos a las características del *display* del usuario y simultáneamente facilitar el proceso de selección de imágenes. Sin embargo, es algo más complejo que el modo progresivo.

Existen dos alternativas para realizar la transmisión de los bloques de los coeficientes que se conocen como la *selecciónpectral* y las *aproximaciones sucesivas*. La idea básica de la selecciónpectral se

ilustra en la figura 4.47 y consiste en transmitir primero los coeficientes más significativos de cada uno de los bloques que forman la imagen (por ejemplo el coeficiente de DC y los dos primeros coeficientes AC). Una vez finalizada la transmisión de estos coeficientes para todos los bloques de la imagen, se transmiten los siguientes más significativos (por ejemplo el 3º, 4º y 5º de la exploración en zigzag de los AC) y así sucesivamente hasta haber transmitido todos los coeficientes. La realización de esta codificación es bastante sencilla y, teniendo en cuenta la nueva ordenación de los coeficientes, pueden utilizarse variantes de los métodos de Huffman y RLE considerados en el modo básico. Su problema básico es que las primeras imágenes que se obtienen en el receptor no contienen componentes de alta frecuencia, por lo que se verán bastante borrosas.

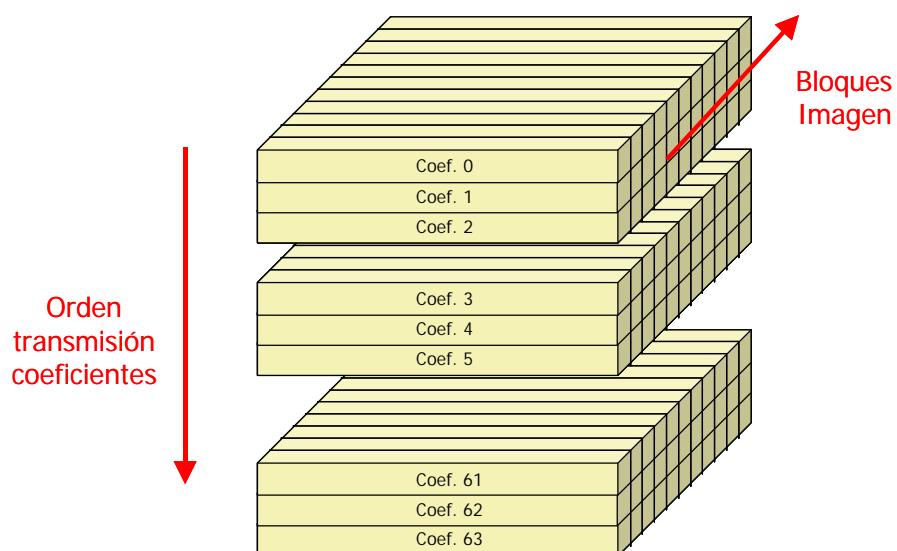


Fig. 4.47 Idea básica de la selección espectral en el modo progresivo

En la ordenación mediante aproximaciones sucesivas se mejora algo la calidad de la imagen recibida, ya que se transmiten todas las componentes frecuenciales de todos los bloques de la imagen. La tasa de transmisión se mantiene baja debido a que, en una primera fase (*scan*), sólo se transmiten los bits más significativos de los coeficientes. Posteriormente se va refinando la imagen del receptor transmitiendo el resto de bits. En la figura 4.48 se muestra una estrategia de transmisión basada en las aproximaciones sucesivas. En este caso, en la primera pasada se envían los 4 bits más significativos. En las fases sucesivas se van enviando los bits adicionales, uno a uno.

Mediante el uso de estrategias que combinan la selección espectral con las aproximaciones sucesivas, pueden obtenerse calidades muy aceptables en las primeras fases de la transmisión. Una estrategia muy utilizada consiste en transmitir el coeficiente de continua con toda la precisión posible y 3 o 4 coeficientes AC con sólo los bits más significativos. Posteriormente se van transmitiendo el resto de bits de los coeficientes AC.

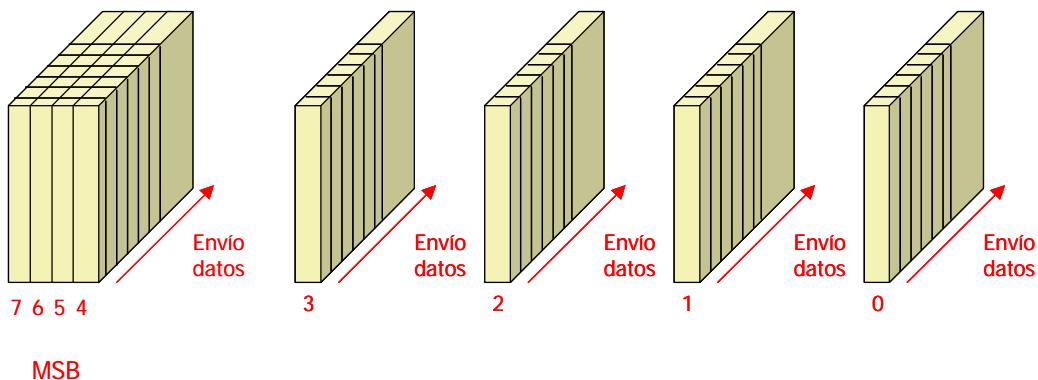


Fig. 4.48 Aproximaciones sucesivas en el modo progresivo

4.8.4 Calidades subjetivas obtenidas con el JPEG

El modo básico del JPEG obtiene unos excelentes compromisos entre factor de compresión y calidad de imagen. Las pruebas de calidad subjetiva suelen realizarse con usuarios especialistas que deben valorar entre 0 y 5 la calidad de la imagen reconstruida. Las valoraciones numéricas se corresponden con las siguientes descripciones:

Pobre:	0-1
Moderada:	1-2
Buena:	2-3
Muy buena:	3-4
Excelente:	4
Indistinguible:	5

Entre las calidades *muy buena* y *buena* suelen aparecer efectos de pérdidas en detalles de alta frecuencia. En la calidad *moderada* empieza a apreciarse ligeramente el efecto de bloque y las pérdidas en altas frecuencias son evidentes. En la calidad *pobre*, los efectos de bloque son muy notables.

Generalmente, la compresión de una imagen se expresa en bits por píxel (bpp). Una imagen en color que utilice las tres componentes R, G y B con una resolución de 8 bits por componente tiene un total de 24 bpp. El formato más habitual de las imágenes sin comprimir es el formato 4:2:2, en el que la luminancia tiene el doble de muestras que cada una de las señales diferencia de color. Estas imágenes están expresadas con 16 bpp. Las imágenes en el formato 4:2:0 tienen 12 bpp. Las calidades subjetivas obtenidas con el JPEG en función de los bpp son:

0,25-0,5 bpp	Moderada-Buena
0,5-0,75 bpp	Buena-Muy Buena
0,75-1,5 bpp	Excelente
1,5-2 bpp	Indistinguibles

Con el estándar JPEG 2000, basado en la transformada *wavelet*, se consiguen aún mejores valoraciones, especialmente para tasas de compresión por debajo de 0,1 bpp, que suelen ser valoradas con calidades entre moderadas y buenas. La transformada *wavelet* es una alternativa a la transformada coseno que mejora los factores de compresión y en la que, en baja calidad, los defectos típicos de bloque u oscilaciones alrededor de las zonas de alta frecuencia son menos visibles. Esta transformada se usa en aplicaciones de vídeo telefonía (MPEG-4) y en algunos *codecs* de vídeo para aplicaciones informáticas (ficheros de vídeo para Windows).

En la figura 4.49 se muestran algunos ejemplos de imágenes codificadas con JPEG a distintas tasas de compresión. Como puede observarse los defectos son evidentes en las dos imágenes con mayor factor de compresión.



Fig. 4.49 Ejemplos de imágenes codificadas con JPEG a distintos factores de compresión

En nuestra descripción del formato JPEG no se han considerado algunas extensiones del método secuencial base que suelen usarse con cierta frecuencia. Entre estas extensiones destaca la existencia de variantes que permiten el uso de hasta 4 tablas de códigos de longitud variable para los coeficientes DC y 4 tablas para los coeficientes AC. Recuérdese que en el método base sólo existen dos tablas para cada tipo de coeficiente (unas para la luminancia y otras para las señales diferencia de color). Otro tipo de extensión es el uso de códigos aritméticos en sustitución de los códigos de longitud variable, que permiten una compresión adicional de entre el 2.5 % y el 5 %. También, como ya hemos comentado, existen extensiones para tratar imágenes cuantificadas con 12 bits.

4.9 Codificación de secuencias de vídeo: MPEG

El éxito del estándar ISO JPEG motivó la creación de un grupo de expertos dedicado a considerar y evaluar distintas estrategias de codificación de secuencias de vídeo digital y a promover la estandarización de metodologías adaptadas a distintos tipos de aplicaciones. El Moving Pictures Experts Group es un comité de la ISO (ISO IEC/JTC1/SC29/WG11) creado en 1988 y que ha sido el responsable de la aparición de varios estándares para la codificación de vídeo. El primer estándar propuesto se conoce con el nombre de MPEG-1 y está orientado al almacenamiento y reproducción de señales de vídeo y audio en soporte CD-ROM para su uso en aplicaciones multimedia. Posteriormente se propuso un estándar de codificación, conocido como MPEG-2, con una calidad suficiente para la transmisión de señales de televisión digital (calidad *broadcasting*) que ha sido ampliamente aceptado como sistema de compresión de vídeo y audio en distintas aplicaciones y formatos de registro y transmisión de vídeo. Así, el MPEG-2 fue aprobado como formato de compresión para la transmisión de señales de televisión digitales por el organismo DVB (*Digital Video Broadcasting*) y fue aceptado como formato de transmisión de vídeo en redes ATM por el ATM *Video Coding Experts Group*. También se ha incluido en el DVD-Video (*Digital Versatile Disc*) para la codificación de vídeo y audio y se utiliza en distintos productos para el almacenamiento de vídeo en formato comprimido bajo soportes de disco duro o cinta magnética.

El comité MPEG continua dedicado a la elaboración de nuevos formatos como el MPEG-4 (de reciente aparición) o el MPEG-7 y sigue aportando modificaciones, alternativas, mejoras y adiciones a los formatos ya existentes para mantenerlos constantemente actualizados a las capacidades tecnológicas de los sistemas actuales. Las tareas de estandarización del MPEG han tenido un gran éxito comercial debido a que la filosofía usada para establecer las normativas sólo especifican la sintaxis y la semántica del flujo de bits que debe recibir el decodificador, proporcionando libertad sobre los procedimientos de codificación y decodificación utilizados para producir o interpretar este flujo. Esto ha facilitado la libre competencia entre distintos productos, que pueden conseguir mejores o peores relaciones entre calidad de imagen y sonido y factor de compresión en función de los algoritmos y arquitecturas *hardware* utilizadas para la compresión. Uno de los propósitos iniciales de la creación del MPEG era promover estándares de codificación de vídeo de libre uso con el objetivo de facilitar la libre competencia y minimizar el impacto en el mercado de productos propietarios que dominaran el mercado o que dificultaran la compatibilidad entre distintos productos de vídeo digital. Actualmente, después de la amplia implantación del MPEG-2, puede decirse que se ha logrado este objetivo inicial.

La inclusión de una señal de vídeo y audio en un soporte como el CD, que originalmente había sido pensado para incluir únicamente la información asociada a un canal de audio estereofónico, representó un importante problema tecnológico durante el desarrollo del estándar MPEG-1, ya que los factores de compresión que debían utilizarse, si se partía de un formato de vídeo digital como el ITU 601 4:2:2, se situaban entre 100 y 200. El flujo de datos que admite un CD-ROM de simple velocidad (150 Kbytes/s) es de unos 1,2 Mbps mientras que el flujo de bits implícito en una señal de vídeo 4:2:2 es de aproximadamente 160 Mbps.

La solución al problema se obtuvo reduciendo la calidad de la imagen en un doble proceso de reducción espacial y temporal. Este proceso significa una reducción inicial de los datos basada en el remuestreo temporal y espacial, por lo que no puede considerarse como una verdadera compresión de

vídeo. Por una parte, la resolución de las imágenes se redujo a la mitad, tanto en la dirección horizontal como en la vertical (formato SIF), produciendo imágenes de 360x240 para NTSC (30 imágenes por segundo) y de 360x288 para PAL (25 imágenes por segundo). Nótese que con esta selección las dos versiones de vídeo tienen el mismo número de píxeles por segundo ($2,592 \times 10^6$). También son habituales los formatos de píxel cuadrado 320x240 para NTSC y 384x288 para PAL. Los formatos SIF estándar suelen reducirse ligeramente en el sentido horizontal a 352x240 para NTSC y 352x288 para PAL, con el objeto de que estas dimensiones sean un múltiplo de 16 y faciliten la compensación de movimiento. Estos formatos reducidos se conocen con el nombre de SPA (*Significant Pel Area*). En cualquiera de los casos la resolución espacial de las imágenes es aproximadamente la cuarta parte de la del estándar ITU 601. Esta reducción supone una pérdida de calidad considerable en las imágenes si se comparan con el formato estándar de radiodifusión. La resolución que se obtiene para estas dimensiones de la imagen es equivalente a la del formato de registro en cinta magnética VHS. El lector puede argumentar que, en general, un vídeo en cinta magnética VHS se ve mejor que la mayoría de vídeos en formato MPEG-1. Esto puede deberse a varios motivos. Por una parte, muchos vídeos MPEG-1 son producidos a partir de registros en cinta magnética que, en algunos casos, son también VHS, por lo que existen ruidos previos a la digitalización que dificultan considerablemente las tareas de compresión. Además, aún en el caso que el vídeo MPEG-1 se produzca a partir de copias de calidad, el sistema visual es mucho más crítico con el formato digital, tal y como se presenta en la pantalla del ordenador, debido a la mayor claridad con que aparece cualquier defecto en el proceso de digitalización o codificación.

Otra simplificación importante en el MPEG-1 es que no admite vídeo entrelazado, por lo que sólo se requiere codificar uno de los dos campos de la señal original. Esta reducción tiene bastante sentido si tenemos en cuenta que la presentación de estas secuencias suele realizarse sobre soporte informático, de modo que la propia tarjeta de vídeo es la que se encarga de aumentar la frecuencia de refresco de la pantalla por encima del número de imágenes y evitar la aparición del parpadeo (*flicker*). Veremos que una de las dificultades adicionales del MPEG-2 es que debe ser capaz de codificar vídeos entrelazados, lo que impone ciertas restricciones a los procesos de compresión de las imágenes. Es importante mencionar que la sintaxis del MPEG-1 admite imágenes de hasta 4095x4095 con 60 imágenes por segundo. No obstante, estos tipos de secuencias están pensadas para usos específicos y no permiten obtener flujos de información próximos a la velocidad del CD-ROM, de modo que, en la práctica, las restricciones más habituales sobre el formato de la señal de vídeo son las que hemos descrito.

A pesar de la reducción espacial y temporal de la señal de vídeo, el flujo resultante sigue siendo elevado (unos 40 Mbps), por lo que es necesario aplicar procedimientos de compresión de gran eficiencia. Veremos que la compresión basada en la extracción de la redundancia espacial de cada fotograma no es suficiente y deben utilizarse técnicas de reducción basadas en la redundancia temporal entre las imágenes. Las tasas más habituales del MPEG-1 se sitúan por debajo de los 1,8 Mbps.

El estándar MPEG-1 fue publicado en noviembre de 1992, mientras ya se estaba desarrollando un nuevo formato de compresión de vídeo para aplicaciones que exigían una mayor calidad y que en principio estaba orientado a cubrir las necesidades de la radiodifusión de señales de televisión digital con definición convencional (equivalente al PAL o al NTSC). El estándar MPEG-2, mantiene cierto grado de compatibilidad con las secuencias de bits producidas por el MPEG-1 y pretende cubrir otras

aplicaciones de mayor ancho de banda. Produce unos flujos de bits que se sitúan entre los 2 Mbps y los 15 Mbps. Su diseño es sumamente flexible y cubre un amplio espectro de sistemas de vídeo que van desde los formatos digitales ITU 601 4:2:2, ITU 601 4:2:0, HDTV (televisión de alta definición) 4:2:0, EDTV (televisión con definición mejorada) 4:2:0, formatos CIF y SIF, formatos HHR 4:2:0, algunos de ellos en versiones de exploración progresiva, otros con exploración entrelazada y algunos con ambos tipos de exploración. Para que el MPEG-2 pueda cubrir todo este amplio espectro de aplicaciones ha sido necesario definir subconjuntos de la sintaxis y semántica de la secuencia de bits en la forma de distintos *perfíles* y *niveles*. Muchos de los aspectos de la codificación de vídeo utilizada por el MPEG-1 y el MPEG-2 son comunes, diferenciándose fundamentalmente en las dimensiones de las imágenes que deben tratar, por lo que durante nuestra exposición serán tratados ambos estándares conjuntamente. La diferencia más importante entre el MPEG-1 y el MPEG-2 es que este último debe ser capaz de tratar señales de vídeo entrelazadas. Los algoritmos de cálculo de la DCT y la compensación de movimiento deben contemplar los sistemas entrelazados de forma distinta, por lo que comentaremos estos detalles específicos de forma separada.

Recientemente ha aparecido el estándar MPEG-4 que se orienta a la transmisión de señales de vídeo y audio con velocidades muy bajas (64 kbps), fundamentalmente para aplicaciones de vídeo telefonía, aplicaciones multimedia, intranets e internet. Las secuencias de bits están especialmente diseñadas para hacer frente a las altas variabilidades del canal este tipo de comunicaciones, en donde es fácil que se produzcan reducciones de la velocidad, pérdidas parciales de datos, etc. Por ello, la trama o *stream* de vídeo y audio suele estar escalada de modo que, en función de la tasa de bits recibida, el receptor puede visualizar las imágenes con mayor o menor calidad. El escalado permite realizar comunicaciones fiables desde los pocos kbps de algunas conexiones de internet hasta los 4 Mbps que pueden obtenerse para intranets, pero deben soportarse complejos protocolos de multiplexación y demultiplexación de datos para hacer frente a la alta variabilidad de las redes. El MPEG-4 proporciona mayor flexibilidad que el MPEG-2 para el uso de técnicas de compresión basadas en la forma de onda permitiendo el uso de la transformada *wavelet*, la codificación vectorial y la realización de transformaciones afines (que admiten rotación de los bloques) para la compensación de movimiento. Además permite realizar una codificación orientada a objetos, pudiendo utilizar codificadores fractales o morfológicos, modelado de objetos y descriptores de escenas en las que incluso se prevé la posibilidad de manipular parte de los contenidos por parte de los usuarios. Actualmente también se está desarrollando el MPEG-7, que pretende dar soporte para el manejo de descriptores asociados a información de vídeo y audio para multimedia. Este estándar está orientado a aplicaciones de búsqueda en bases de datos de contenidos audiovisuales y especifica los procedimientos para insertar los descriptores de imagen, vídeo, gráficos 3D, etc. Ni el MPEG-4 ni el MPEG-7 están pensados para aplicaciones de difusión de vídeo para televisión digital convencional.

En los apartados siguientes describiremos los procedimientos básicos utilizados para la compresión de la señal de vídeo en el MPEG-1 y el MPEG-2. Nuestro objetivo es que el lector disponga de una idea global sobre los métodos utilizados para la compresión y de cómo se forma la secuencia de bits que aporta la información de las señales de vídeo, audio y control. Las descripciones serán bastante más superficiales que las utilizadas en la descripción del JPEG debido, por una parte, a la mayor complejidad del sistema de codificación y, por otra, a la gran cantidad de posibilidades, formatos de imagen y tipos de bloques que existen en el MPEG, cuya exposición detallada requeriría, como mínimo, un libro completo dedicado a los pormenores del estándar. Creemos que esta presentación será suficiente para que el lector adquiera una idea suficientemente clara de los distintos

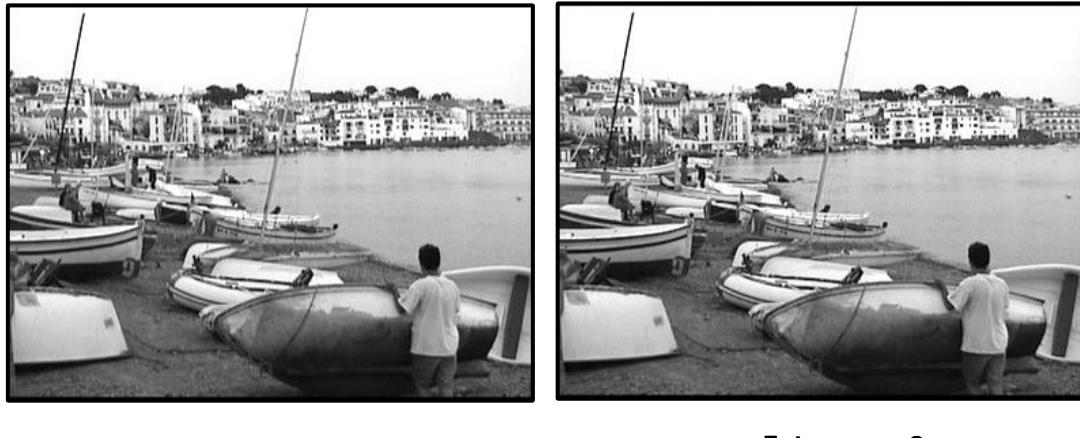
procedimientos que intervienen en la compresión de la señal de vídeo y de la interrelación entre ellos. Pueden encontrarse detalles adicionales en numerosos textos dedicados a la codificación MPEG que se proporcionan en el apartado de bibliografía. También existe gran cantidad de información en distintos servidores de *internet* en la forma de artículos divulgativos o programas ejemplo de la codificación MPEG. El estándar oficial MPEG-1 puede encontrarse en los documentos ISO/IEC 11172-1 (estructura de multiplexación de datos), ISO/IEC 11172-2 (codificación de vídeo) y ISO/IEC 11172-3 (codificación de audio). El MPEG-2 se encuentra en los documentos ISO/IEC 13818-1 (Múltiplex), ISO/IEC 13818-2 (Vídeo) y ISO/IEC 13818-3 (Audio).

4.9.1 Uso de la redundancia temporal para la compresión de vídeo.

El JPEG aprovecha la redundancia existente entre los píxeles adyacentes de una imagen para realizar una considerable reducción en el volumen de datos sin que se produzcan pérdidas aparentes de calidad. En una secuencia de vídeo, podríamos aplicar el JPEG a cada uno de los fotogramas para obtener una señal comprimida. De hecho, este procedimiento se utiliza en muchas tarjetas de digitalización de vídeo para PC y permite obtener una primera versión de la secuencia que deseamos almacenar en el disco duro. Téngase en cuenta, que el flujo que requiere la digitalización de una señal de televisión en el formato ITU-601 4:2:0 es de unos 30 Mbytes/s, bastante por encima de los flujos que permiten los discos duros actuales, incluso discos con bus SCSI II. Por ello, es necesario aplicar una primera compresión mediante *hardware* o *software* que permita el almacenamiento de la señal. Los métodos basados en comprimir la señal de vídeo fotograma a fotograma mediante métodos parecidos al JPEG se denominan Motion-JPEG (MJPEG) y suelen ser métodos propietarios de cada fabricante o desarrollador. El nombre genérico que reciben estos algoritmos es un tanto confuso, por cuanto no aprovechan la redundancia temporal existente en las imágenes ni realizan ninguna estimación del movimiento de los objetos contenidos en las mismas. El principal problema de este procedimiento de compresión es que, si deseamos mantener una calidad aceptable, se obtiene un flujo de datos del orden de 8-10 Mbytes/s (60-80 Mbps, para imágenes de 720x576), que aun siendo suficiente para su almacenamiento en una unidad de disco duro, es excesivo para su transmisión como señal de televisión digital o su almacenamiento en un soporte CD-ROM. La codificación mediante M-JPEG tiene ciertas ventajas durante el proceso de edición y manipulación del vídeo, pues todos los fotogramas se obtienen con la misma calidad y su decodificación es independiente del resto de fotogramas de la secuencia. Por ello, estos formatos se seguirán manteniendo, a pesar de la aparición del MPEG y su optimización en la compresión de la señal, en este tipo de aplicaciones.

Para obtener factores de compresión adecuados a la transmisión como señal de televisión es necesario aprovechar la redundancia temporal existente entre las sucesivas imágenes. Téngase en cuenta que, dada la naturaleza de la mayor parte de escenas, gran parte de los objetos y elementos permanecen en la misma posición y con las mismas características de un fotograma a otro. Esto indica que es posible obtener una excelente aproximación de la información contenida en un fotograma aprovechando la información que disponemos del fotograma anterior. En la figura 4.50 se muestran dos fotogramas consecutivos de una secuencia de vídeo. En este ejemplo, las diferencias sólo son apreciables cuando se realiza la resta entre las dos imágenes. En la figura 4.51 se muestra el resultado de realizar esta diferencia. Sobre la imagen diferencia hemos calculado el valor absoluto y para hacer más evidentes la diferencia entre niveles se ha multiplicado por un factor 3, finalmente se ha representado en vídeo inverso. La imagen resultante indica de forma clara el resultado que esperábamos, ya que la mayor

parte de los elementos están próximos al cero, por lo que se ha reducido la entropía de forma notable, pudiendo codificar la imagen de forma eficientemente mediante el uso de códigos RLE, de longitud variable o aritméticos. Existen algunas zonas de la imagen en las que se observa que las diferencias aumentan de nivel y que se corresponden, aproximadamente, con los contornos de los objetos. Esto se debe al movimiento de la cámara, que está realizando un ligero movimiento horizontal y al de los propios objetos (el barquero está dando la vuelta a la barca).



Fotograma 1
(25 imágenes/seg)

Fotograma 2
(25 imágenes/seg)

Fig. 4.50 Ejemplo de dos fotogramas consecutivos de una secuencia de vídeo con poco movimiento



Fig. 4.51 Resultado de la imagen diferencia entre los dos fotogramas

La idea de utilizar la imagen anterior para predecir la actual se utiliza en diversos *codecs* para la producción de vídeo en soporte informático (AVI, Internet). La ventaja de esta estrategia es que puede desarrollarse sin necesidad de un soporte *hardware* adicional en la mayoría de plataformas y en un tiempo reducido (en algunos casos tiempo real). El procedimiento general suele conocerse con el nombre de codificación de fotogramas en diferencias (*frame differencing*) y, como no existe ningún estándar al respecto, existen multitud de variantes que son propietarias del desarrollador. Normalmente, la secuencia de vídeo se descompone en fotogramas clave (*key frame*) y fotogramas diferencia (*delta frames*) tal y como se representa en la figura 4.52. Los fotogramas clave deben codificarse de forma independiente del resto de fotogramas, teniendo en cuenta únicamente la redundancia espacial de la imagen. Generalmente se utilizan variantes del JPEG o métodos de codificación vectorial (como en el Cinepack o el Indeo 3.2) para codificar estos fotogramas. Las imágenes delta se codifican utilizando como base de predicción las imágenes anteriores. Es conveniente que periódicamente se inserte un fotograma clave en la secuencia codificada para evitar la propagación de errores de codificación en las imágenes delta.

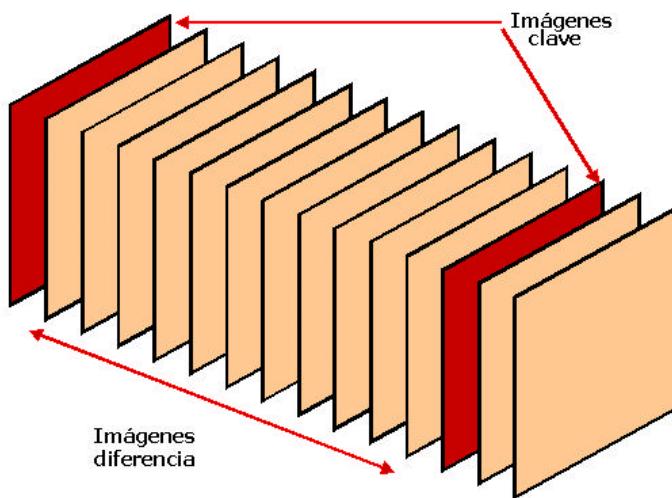


Fig. 4.52 Descomposición de una secuencia de vídeo en fotogramas clave y diferencia (*delta*)

Una posible estrategia de codificación de las diferencias consiste en descomponer las imágenes delta en bloques de 8x8 o 16x16 elementos y comparar cada uno con el bloque equivalente de la imagen clave precedente. Si la diferencia entre ambos bloques no supera un determinado umbral, el valor de este bloque en la imagen delta se mantiene igual al que tenía en la imagen clave. En cambio, cuando la diferencia es significativa debe volver a codificarse. La ventaja del tratamiento por bloques es que, si se utilizan métodos transformados para codificar las imágenes clave, no es necesario descomprimirlas y pueden compararse directamente los coeficientes de ambas subimágenes. El uso de un umbral para determinar si es necesario recodificar los bloques de las imágenes delta es bastante crítico. Si el

umbral es excesivamente bajo, la probabilidad de superar el umbral es alta por lo que la codificación pierde eficiencia. Si el umbral es alto, es posible que no se detecten movimientos pequeños, por lo que aparecen errores significativos en la imagen. Además, estos errores resultan a veces muy molestos, ya que aparecen en las partes que tienen movimiento, produciendo deformaciones de los objetos. Consideremos por ejemplo el desplazamiento horizontal de una cara sobre un fondo constante. En función del valor del umbral, es posible que en algunos bloques correspondientes a la zona de la cara se detecte movimiento, mientras que en otros no. Esto puede llegar a producir que el sujeto aparezca con la boca deformada o con tres ojos (situaciones que en realidad se producen en algunos codecs para videoconferencia en internet). De todos modos, la resolución espacial del sistema visual a los objetos en movimiento es más baja que para los objetos estáticos y los errores suelen desaparecer al cabo de uno o dos fotogramas o, a lo sumo, con el siguiente fotograma clave. La ganancia de compresión de este tipo de codificadores está directamente relacionada con el movimiento que presentan las imágenes. En imágenes de televisión donde existen abundantes cambios de plano y movimientos de cámara, estos codificadores proporcionan muy poca compresión. En cambio, suelen presentar excelentes resultados en aplicaciones de videoconferencia en las que la cámara permanece estática y el sujeto presenta un movimiento muy reducido o en aplicaciones de vídeo vigilancia de cámaras de seguridad en las que sólo se producen movimientos significativos durante las entradas y salidas de individuos. La carga computacional asociada al codificador es muy reducida. Por ello, también son utilizados con frecuencia en aplicaciones de transmisión de vídeo a baja velocidad de imágenes con poco movimiento.

Una mejora respecto a la estrategia anterior consiste en enviar la señal de error entre la predicción y la imagen real. Con ello, se mejora notablemente la calidad de las imágenes al precio de una menor eficiencia de compresión y mayor complejidad de cálculos, tanto en el compresor como en el descompresor.

Uno de los parámetros que también suele resultar crítico es la frecuencia con la que se intercalan los fotogramas clave, que, evidentemente, depende del tipo de imágenes que se transmitan. En aplicaciones de vídeo conferencia suele recomendarse utilizar un fotograma clave por cada 10 o 15 fotogramas reales.

4.9.2 Compensación de movimiento

La compensación de movimiento es una técnica de predicción temporal que intenta optimizar los resultados obtenidos por la codificación diferencial analizando el movimiento que realizan los objetos dentro de la imagen y compensándolos. Evidentemente, los cálculos involucrados en la realización de esta estrategia son sumamente más complejos, pero la mejora en factor de compresión es muy notable. Si consideramos las imágenes del ejemplo de las figuras 4.50 y 4.51 observamos que la mayor parte de diferencias entre las dos imágenes se deben al propio movimiento de la cámara, por lo que podrían reducirse con un ligero desplazamiento relativo entre ellas. En general, la codificación diferencial presenta problemas cuando se producen desplazamientos de los objetos dentro de la imagen o aparecen nuevos motivos que pueden ser debidos al movimiento de la cámara, el cambio de ángulo de visión (*zoom*) el desplazamiento real de los objetos o los cambios de plano; situaciones, todas ellas, que son bastante habituales en la realización de programas de televisión, documentales o películas. En la figura 4.53 se muestra un ejemplo de un desplazamiento de cámara en una imagen simplificada. La

figura muestra cuál sería la zona de la imagen utilizada para predecir el bloque de interés mediante una codificación diferencial. Obviamente, el error de predicción es en este caso igual a la imagen que deseamos predecir, por lo que la ganancia del codificador es nula. La figura 4.54 muestra el bloque que utilizaríamos para la predicción si utilizamos técnicas de compensación de movimiento. Es evidente, que, en este ejemplo, el error de predicción puede llegar a ser exactamente nulo, por lo que no resultaría necesario volver a codificar el bloque, únicamente tendríamos que transmitir el vector de desplazamiento que tenemos que utilizar en la imagen de referencia para construir este fragmento de imagen.

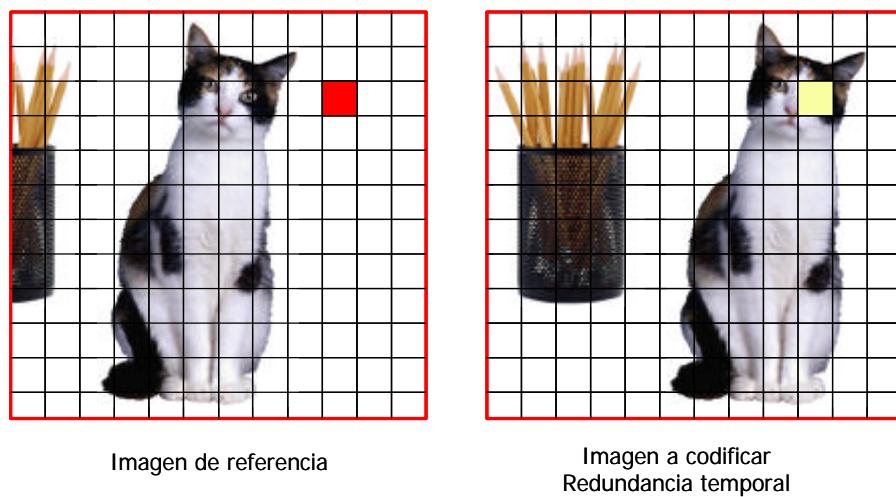


Fig. 4.53 Ejemplo de bloques utilizados para realizar la predicción en una codificación diferencial

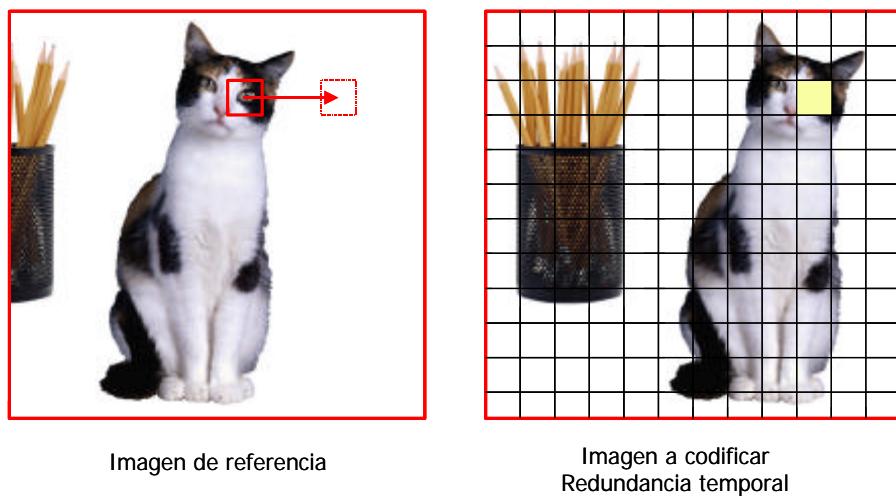


Fig. 4.54 Ejemplo del bloque utilizado para la predicción en una codificación por compensación de movimiento

En la práctica las cosas no suelen ser tan simples y no resulta trivial encontrar en la imagen precedente un bloque que se corresponda exactamente con el fragmento que deseamos codificar. Los objetos pueden sufrir deformaciones, cambios de perspectiva o de tamaño (ángulo de visión) o incluso trasladarse en el espacio mediante rotaciones. Todos estos cambios no quedan cubiertos con la simple introducción de un vector de translación, por lo que las ganancias de predicción y los factores de compresión no serán tan ideales como sugerimos en este ejemplo. Incluso en nuestro problema simplificado se observa que una parte de los objetos que están visibles en la segunda imagen no eran accesibles en la primera. El procesador deberá ser capaz de identificar las regiones de la imagen con nuevos componentes y codificarlos basándose únicamente en la redundancia espacial, como si se tratara con una imagen estacionaria. Estas consideraciones pretenden insistir sobre la dificultad de predecir con exactitud todos los fragmentos de la imagen. No obstante, resulta evidente que la nueva estrategia puede permitir obtener resultados muy eficientes en un gran número de secuencias de vídeo reales.

En síntesis, el proceso para codificar un bloque de una imagen mediante la compensación de movimiento consiste en determinar qué zona o bloque de la imagen de referencia puede utilizarse de forma más eficiente para realizar la predicción del bloque actual. Una vez determinada esta región, deberemos transmitir el vector de desplazamiento asociado (únicamente las dos componentes de desplazamiento horizontal y vertical) y el error que se produce al utilizar esta estimación de la imagen en sustitución de la imagen original. El decodificador, que se supone que de algún modo ya ha decodificado la imagen que utilizamos como referencia, puede recuperar la información del bloque actual utilizando el bloque desplazado de la imagen de referencia y añadiendo el error que se ha transmitido. La señal de error puede transmitirse de forma directa o utilizando métodos transformados. El MPEG determina la transformada coseno de la señal de error, la cuantifica y la transmite mediante códigos de longitud variable, con lo que se consigue una compresión adicional sobre la forma de onda de esta señal. El decodificador, por tanto, debe realizar la transformación inversa sobre la señal de error antes de reconstruir el fragmento de imagen. El procesador del codificador debería ser capaz de decidir en qué casos resulta más conveniente utilizar la compensación de movimiento o codificar directamente el bloque original sin ningún tipo de predicción. Para ello debe evaluar el número de bits final con el que se transmitiría cada opción. Estos cálculos son sumamente complejos, pues deberían evaluarse todas las alternativas antes de realizar la decisión, por lo que a menudo se simplifican teniendo en cuenta resultados parciales de la estimación de movimiento y la varianza de la imagen. En el caso en que se decida transmitir la información original sin hacer uso de ninguna predicción o compensación de movimiento, deberá advertirse al decodificador de nuestra decisión. Obsérvese que la codificación diferencial puede considerarse un caso particular de la codificación con compensación de movimiento, en la que todos los vectores de desplazamiento son nulos.

En la práctica, las direcciones y módulos de los vectores de desplazamiento asignados a bloques próximos suelen ser parecidos. Téngase en cuenta que los bloques suelen ser más pequeños que los objetos, por lo que si un objeto se desplaza en una determinada dirección todos los bloques que lo constituyen también se desplazan en la misma dirección y distancia. Por ello, el MPEG no codifica directamente los vectores de movimiento, sino las diferencias entre vectores de movimiento de bloques adyacentes. Esto permite que muchos vectores sean próximos a cero, por lo que pueden usarse códigos de longitud variable para optimizar la tasa de bits final. Nótese que, durante el proceso de compresión, cualquier detalle, por pequeño que sea, debe ser optimizado con el objeto de reducir la información total que debe transmitirse.

El estándar MPEG no establece ninguna recomendación sobre el método mediante el que puede realizarse la estimación del vector de desplazamiento o sobre los procedimientos para decidir cuándo un bloque debe codificarse mediante la transmisión del error y su compensación de movimiento o transmitir directamente el bloque utilizando métodos de compresión intra-trama. Los procedimientos y algoritmos utilizados para realizar estos cálculos corresponden al desarrollador de productos y por ello existen multitud de codificadores MPEG con distintas prestaciones, velocidades y calidades de codificación. Es importante subrayar que toda la carga computacional del proceso de estimación de los vectores de movimiento y las decisiones de cómo debe codificarse cada bloque recae sobre el codificador. El receptor únicamente deberá interpretar correctamente las informaciones transmitidas y utilizarlas realizar la compensación de movimiento y restablecer una aproximación a la imagen original. En la figura 4.55 se muestra gráficamente cómo se utiliza un fragmento de la imagen de referencia para aproximar un bloque de la imagen actual. La figura no indica explícitamente que sobre esta primera aproximación del bloque deberá transmitirse la señal de error transformada, cuantificada y codificada mediante códigos de Huffman para que el receptor pueda recuperar, aunque sea con pérdidas, los datos originales.

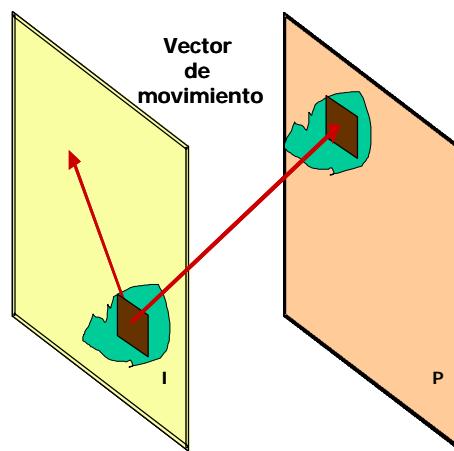


Fig. 4.55 Transmisión de los vectores de movimiento para realizar la predicción de un bloque de la imagen

Una extensión de la compensación de movimiento que se utiliza en el MPEG es la compensación de movimiento bidireccional. La idea básica consiste en codificar una imagen utilizando no sólo las imágenes de referencia del pasado, sino también imágenes de referencia futuras. En principio podría argumentarse que esta extensión no es causal y por tanto irrealizable, pero, evidentemente, siempre podemos introducir un retardo prefijado en el codificador y el decodificador que nos permita utilizar imágenes de referencia futuras. La solución en el codificador consiste, simplemente, en utilizar un buffer de memoria suficientemente grande como para disponer de imágenes clave o de referencia futuras durante el proceso de codificación de la imagen actual. El problema es que el decodificador deberá disponer también de las imágenes clave futuras para poder decodificar una imagen que ha sido construida a partir de vectores de desplazamiento basados en imágenes futuras. Para ello, es suficiente con intercambiar el orden de transmisión de las imágenes, que no necesariamente debe coincidir con

el orden con el que son presentadas en el *display*. Si el decodificador recibe primero las imágenes clave, puede almacenarlas en memoria interna y utilizarlas para decodificar las imágenes en las que se ha utilizado este tipo de compensación de movimiento. Una vez decodificadas todas las imágenes pueden presentarse en el orden que corresponde.

En la figura 4.56 se muestra un esquema que ilustra la codificación mediante compensación bidireccional. Para codificar un bloque de la imagen B, puede utilizarse como referencia la imagen I y la imagen P (de momento, los símbolos utilizados para describir las imágenes son simplemente notación, veremos que esta notación describe distintos tipos y características de imagen en el MPEG). El procesador debe ahora calcular dos vectores de movimiento que se denominan *forward* (hacia delante) y *backward* (hacia atrás). El cálculo del vector de movimiento *hacia adelante* se realiza buscando la región de la imagen de referencia previa que más se aproxima al bloque que deseamos codificar. Análogamente, el vector *hacia atrás* se determina utilizando la imagen clave siguiente. A partir de estos dos vectores, el procesador puede considerar distintas posibilidades para codificar el bloque de interés:

- Codificar el error de predicción forward:* Esta señal de error se obtiene como la diferencia entre los píxeles del bloque original y los píxeles del bloque que se ha determinado como idóneo en la imagen de referencia precedente.
- Codificar el error de predicción backward:* De forma análoga al caso anterior, se obtiene como la diferencia entre el bloque original y el bloque desplazado de la imagen de referencia posterior.
- Codificar el error de predicción bidireccional:* En este caso, el bloque original se compara con el promedio entre los bloques estimados por el vector de movimiento *forward* y el *backward*.
- Codificar el bloque sin utilizar ninguna predicción:* Si los errores obtenidos en los casos anteriores requieren un número de bits considerable, puede resultar rentable codificar directamente el fragmento de imagen utilizando procedimientos intra-trama.

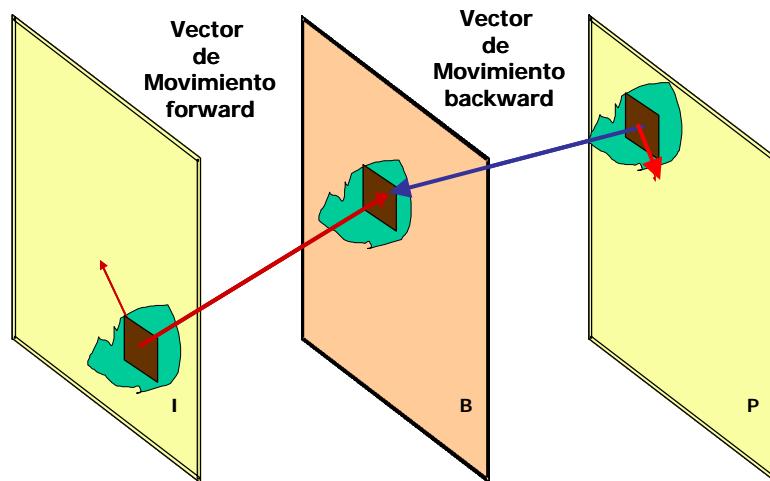


Fig. 4.56 Uso de la compensación de movimiento bidireccional

Evidentemente, la compensación bidireccional representa un incremento considerable de la complejidad computacional del algoritmo. El procesador debe considerar distintas posibilidades para codificar cada bloque de la imagen resultante y decidir cuál de ellas es la óptima. La decisión debería basarse en el número de bits total que se requiere para codificar el bloque, considerando todas las posibles alternativas, lo que significa que no sólo deberían calcularse los errores de predicción sino que además deberían transformarse mediante la DCT, cuantificarse y determinar el número de bits necesario para codificarlos utilizando las tablas de Huffman. No obstante, para simplificar los cálculos, los codificadores de MPEG suelen interrumpir los cálculos y tomar decisiones en fases intermedias, basándose en las características de los distintos errores de predicción obtenidos y de los píxeles del bloque de imágenes.

El uso de técnicas de compensación de movimiento bidireccionales fue introducido en el estándar MPEG-1 y aporta bastante ventajas respecto a la compensación *forward*, que ya había sido utilizada en el estándar H-261 para la transmisión de videoconferencia por redes de baja velocidad. Entre las ventajas de la compensación de movimiento en los dos sentidos temporales deben citarse:

- a) La predicción *backward* permite predecir objetos que no aparecen en la imagen de referencia previa, pero que sí están presentes en la imagen que estamos codificando y en la imagen de referencia futura. Sería el caso de un movimiento horizontal de cámara en el que aparecen nuevos objetos en la escena. Estos nuevos objetos están disponibles en la referencia futura, por lo que pueden utilizarse para nuestra predicción. Otro ejemplo más drástico es cuando se produce un cambio de plano entre las dos imágenes de referencia. En este caso, si la imagen que estamos considerando ya forma parte del nuevo plano deberemos utilizar únicamente predicciones *backward*.
- b) En el caso en que el bloque de imagen se encuentre tanto en la imagen de referencia anterior como en la posterior, puede resultar rentable utilizar el promedio entre los dos bloques para reducir el ruido propio de la imagen y mejorar la codificación.

Los modelos de movimiento utilizados en el MPEG-1 y el MPEG-2 cubren únicamente las traslaciones de los objetos. Estos movimientos son los más habituales en escenas de cine y televisión. El MPEG-4 permite también utilizar modelos de movimiento basados en transformaciones afines que incorporan las posibles rotaciones de los objetos.

4.9.3 Algoritmos para la estimación de los vectores de movimiento

La estimación de los vectores de movimiento es una de las tareas que requieren mayor dedicación por parte del codificador. Los estándares MPEG no establecen ningún método para la estimación de estos vectores, dejando libertad a los desarrolladores de *hardware* y de *software* para que utilicen los algoritmos que consideren más adecuados. Una correcta estimación de los vectores de movimiento puede significar una drástica reducción en la tasa de bits y un aumento de la calidad final con la que se visualizan las imágenes.

Existen distintas procedimientos alternativos para realizar la estimación de los vectores de movimiento. La primera decisión importante suele ser determinar el área de la imagen de referencia en la que se realizará la búsqueda del vector de movimiento. Tal y como se muestra en la figura 4.57, el área de búsqueda estará situada en las proximidades de la posición del bloque que deseamos codificar, pero en las coordenadas de la imagen de referencia. El área de la zona de búsqueda depende de los modelos de movimiento máximo que se admite. También es una función de la diferencia de tiempos entre la imagen a predecir y la de referencia. Cuanto mayor es esta diferencia, mayor es el desplazamiento que pueden experimentar los objetos entre las dos imágenes.

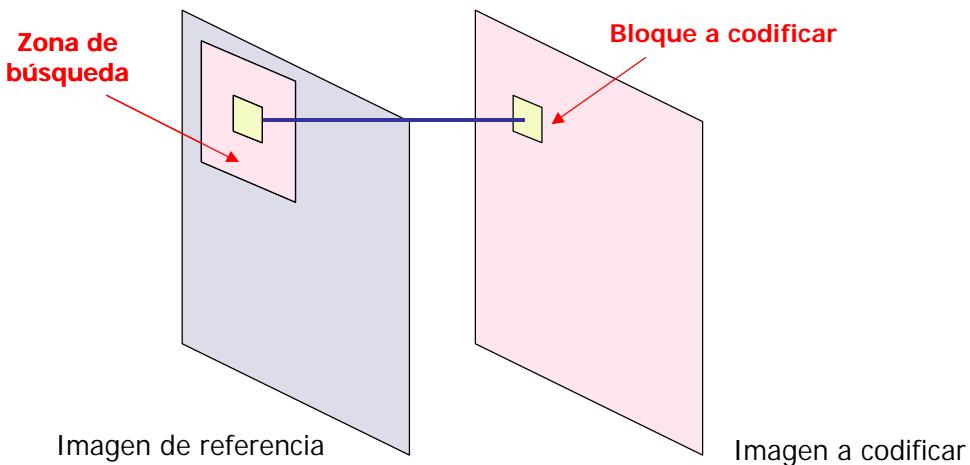


Fig. 4.57 Región de búsqueda del vector de movimiento en la imagen de referencia

La mayoría de las variantes utilizadas en los productos MPEG actuales corresponden a la categoría de los denominados algoritmos de *block matching* (ensamblaje de bloques). Los algoritmos de *block matching* comparan los píxeles del bloque que se pretende estimar con los de la imagen de referencia y establecen una medida de distancia entre los dos bloques. El vector de movimiento óptimo se obtiene en la posición de la región de análisis en la que se encuentra un valor mínimo de la distancia entre los dos bloques. La figura 4.58 muestra la zona por la que debe desplazarse el bloque que deseamos codificar dentro de la región de búsqueda de la imagen de referencia. La búsqueda exhaustiva supone que debemos calcular la distancia entre las dos subimágenes para todos los posibles desplazamientos. La carga computacional aumenta por tanto con el área de la región de búsqueda y con el tamaño de los bloques. En el MPEG-1 se utilizan bloques de imagen de 16x16 píxeles (denominados *macrobloques*), por lo que en cada cálculo de distancia deben evaluarse un total de 256 píxeles. En el MPEG-2 también se utilizan macrobloques de 16x16 píxeles o de 16x8. Comentaremos y justificaremos estos detalles más adelante.

Las funciones de distancia más utilizadas son el error cuadrático, el error absoluto y la correlación cruzada entre las dos imágenes. Las siguientes ecuaciones indican los cálculos involucrados en cada una de estas distancias:

Error cuadrático:

$$VM(i, j) = \sum_{(n,m) \in I} (b[n, m] - u_r[i + n, j + m])^2 \quad (4.26)$$

Error absoluto:

$$VM(i, j) = \sum_{(n,m) \in I} |b[n, m] - u_r[i + n, j + m]| \quad (4.27)$$

Correlación cruzada:

$$VM(i, j) = \frac{\sum_{(n,m) \in I} b[n, m] \cdot u_r[i + n, j + m]}{\left(\sum_{(n,m) \in I} b^2[n, m] \right)^{1/2} \cdot \left(\sum_{(n,m) \in I} u_r^2[n + i, m + j] \right)^{1/2}} \quad (4.28)$$

donde $b[n, m]$ representa el bloque de imagen que estamos codificando y $u_r[n, m]$ la imagen de referencia. Las coordenadas (i, j) deben variarse para toda la región de búsqueda. El vector de desplazamiento se obtiene como los valores (i, j) que producen un mínimo en la función $VM(i, j)$. De todas estas funciones, la que se utiliza con mayor frecuencia es la del error absoluto, ya que produce unos resultados similares a las demás y tiene una implementación más simple. Nótese que el uso de funciones de distancia entre los píxeles no es más que una simplificación para determinar el bloque óptimo. En rigor, el procedimiento ideal para determinar el vector de movimiento consistiría en evaluar el número total de bits con que puede codificarse el bloque de interés en función de la zona utilizada como referencia. No obstante, es evidente que esta estrategia requeriría unos cálculos sumamente complejos y además no aportaría excesivas mejoras.

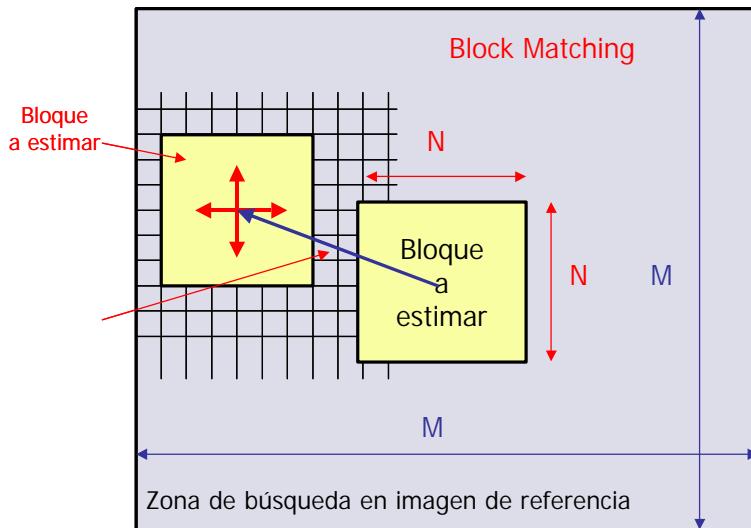


Fig. 4.58 Procedimiento de búsqueda exhaustiva del vector de movimiento con block-matching

Las medidas de distancia que nos proporcionan las ecuaciones anteriores pueden utilizarse para evaluar si la compensación de movimiento nos permitirá obtener una buena predicción del bloque. Obsérvese que si la distancia mínima que hemos calculado es excesivamente grande, significa que el error entre el bloque original y la referencia desplazada también será grande (de hecho, estamos evaluando el error entre ambas imágenes). Cuando la varianza de la señal de error es excesiva, puede resultar más rentable codificar el bloque con métodos convencionales de intra-trama. Así pues, la decisión de codificar el bloque mediante compensación de movimiento o como intra-trama puede tomarse una vez determinada la distancia mínima que obtenemos en el proceso de cálculo del vector de movimiento.

Una vez determinado el valor del vector de movimiento, podemos refinar la estimación utilizando desplazamientos fraccionales. Los desplazamientos fraccionales suponen que la estimación del movimiento puede realizarse con una precisión superior al píxel. Para ello, podemos interpolar el bloque a codificar en un factor 2 en cada dirección y compararlo con la zona de la imagen de referencia que se ha detectado como válida, también interpolada en un factor 2. El procedimiento de *block matching* puede aplicarse ahora a los 8 píxeles vecinos al punto inicial y detectar si la medida de distancia obtenida es menor que la que disponíamos inicialmente. En la figura 4.59 se muestran los píxeles de la imagen de referencia iniciales, los interpolados y los elementos sobre los que se calculará la distancia con el bloque a codificar. El gráfico presupone que el vector de movimiento inicial se sitúa en el centro de la matriz interpolada y que esta matriz se extiende, como mínimo, a las dimensiones del bloque interpolado (32x32 píxeles en MPEG-1). La interpolación de la imagen que se suele utilizar es lineal, lo que significa que se realiza primero una interpolación por filas, tomando el valor medio entre dos muestras como valor interpolado y posteriormente una interpolación por columnas con la misma operación. El método propuesto podría extenderse a factores de interpolación superiores, pero los resultados obtenidos no suponen una mejora significativa.

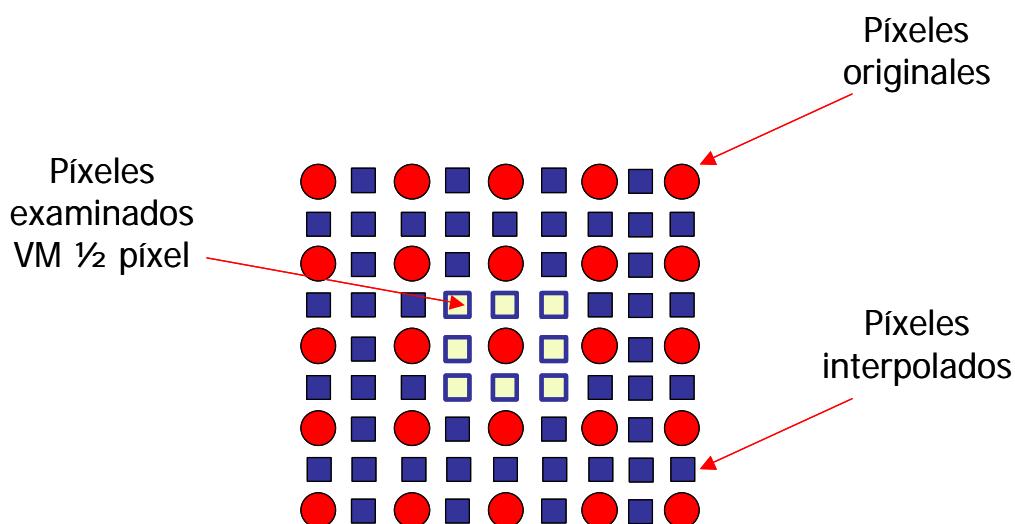


Fig. 4.59 Cálculo del vector de movimiento con desplazamientos fraccionales

La búsqueda exhaustiva supone un coste computacional excesivo, por lo que la mayoría de algoritmos implementan alguna simplificación del problema, orientando la búsqueda del vector óptimo sin necesidad de realizar todos los cálculos de distancia posibles. Los métodos más usados son la búsqueda logarítmica, la búsqueda en la dirección del gradiente conjugado, la búsqueda en tres etapas y la búsqueda jerárquica, que es la que suele proporcionar un mejor compromiso entre el resultado obtenido y la complejidad computacional. Las primeras alternativas que se han citado se basan en realizar la búsqueda del vector de desplazamiento óptimo examinando únicamente un conjunto reducido de puntos en las primeras etapas del cálculo. A partir de los resultados obtenidos, se va refinando la búsqueda hasta determinar la posición en la que obtenemos una distancia mínima. La ventaja de estos métodos es que no es necesario examinar todos los posibles vectores de desplazamiento para obtener el mínimo. No obstante, es probable que la convergencia del método se produzca en un mínimo local, cuyo valor no coincide exactamente con el mínimo absoluto que obtendríamos con la búsqueda exhaustiva. Sin embargo, la simplificación en los cálculos y la reducida probabilidad de que se produzcan estos errores justifican el uso de estas estrategias.

El método de búsqueda en tres pasos se representa esquemáticamente en la figura 4.60. Inicialmente se calcula la distancia del bloque a codificar situado con un desplazamiento nulo respecto a la imagen de referencia. También se calculan las 8 distancias indicadas como el paso 1 y que están situadas en los diferentes cuadrantes de la región de búsqueda. El paso de búsqueda, es decir, la distancia que separa los distintos píxeles que son examinados, en esta primera iteración es de 4 (tanto en la dirección horizontal, vertical y diagonal). Con estos cálculos, se determina el punto para el que hemos obtenido distancia mínima y se examinan sus 8 vecinos situados a un paso de búsqueda mitad (es decir 2). Nuevamente se toma el valor mínimo y se calculan las distancias entre el bloque a codificar y las imágenes de referencia con los 8 vecinos situados a un paso de búsqueda unidad. La posición en la que obtenemos el valor mínimo se toma como el vector de desplazamiento.

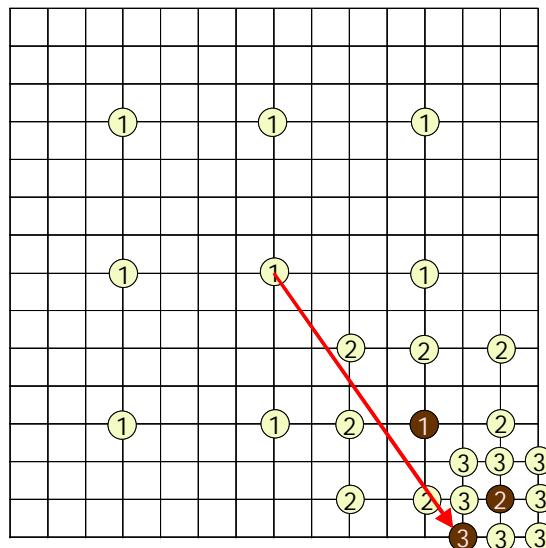


Fig. 4.60 Procedimiento de búsqueda en tres pasos

La búsqueda en tres etapas requiere realizar el análisis de un número fijo de puntos. En cada etapa necesita determinar 8 distancias entre el bloque a codificar y la imagen de referencia. En nuestro ejemplo, el número total de puntos examinados es 25 (8 por cada iteración más el punto central). Con ello, la distancia máxima a la que puede obtenerse el vector de desplazamiento respecto al centro es de (7,7). Por tanto, la región de búsqueda se extiende en un margen de (-7,7) respecto a la posición central, tanto en la dirección horizontal como la vertical. Si deseamos ampliar la región de búsqueda, puede utilizarse un paso de búsqueda inicial de 8 o 16. La filosofía del procedimiento sería la misma, aunque ahora el número de etapas necesarias para determinar el punto final aumentaría (5 etapas para un paso inicial de 16). Nótese que esta estrategia supone un ahorro considerable respecto a la búsqueda exhaustiva. En nuestro ejemplo la búsqueda exhaustiva hubiera requerido examinar un total de 225 puntos (15x15), mientras que éste método permite obtener una solución examinando tan sólo 25 puntos. Otra de las ventajas de la búsqueda en tres etapas es que, por su propia construcción, siempre converge.

Otra estrategia de búsqueda simplificada es usar el algoritmo de direcciones conjugadas, cuya idea general se muestra en la figura 4.61. En este caso, la búsqueda se inicia en el píxel central y se examinan los dos píxeles adyacentes en la dirección horizontal. Si la distancia mínima se encontrara para el píxel central, se daría por terminada la búsqueda en esta dirección. En caso contrario, se toma como nuevo punto de partida el que ha obtenido el valor mínimo y se continúan examinando sus vecinos horizontales (de hecho, sólo es necesario examinar uno de ellos, ya que el otro ya habrá sido calculado previamente). La búsqueda en el eje horizontal termina cuando encontramos un punto cuyos dos vecinos presentan una medida de distancia mayor. Por tanto, el procedimiento consiste simplemente en buscar un mínimo local en la dirección horizontal examinando píxeles vecinos. Una vez localizado este mínimo se realiza la búsqueda en el sentido vertical hasta encontrar otro mínimo local. Finalmente, se aplica el mismo procedimiento en el sentido diagonal hasta obtener el vector de desplazamiento final.

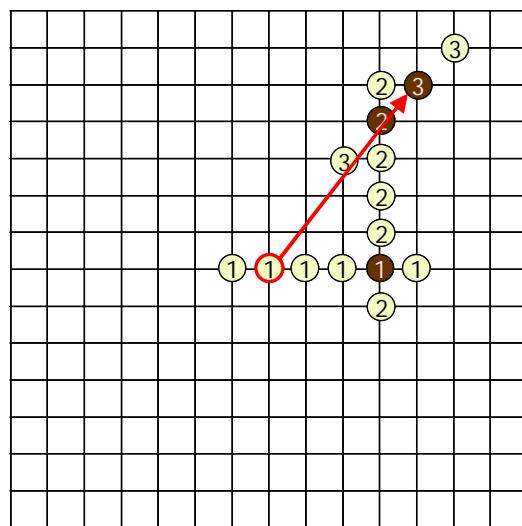


Fig. 4.61 Búsqueda del vector de desplazamiento mediante direcciones conjugadas

La búsqueda por direcciones conjugadas no tiene un número predeterminado de puntos a evaluar. El cálculo total dependerá de la posición del vector de movimiento, siendo reducido cuando el mínimo se sitúe cerca del punto inicial. Para mejorar la velocidad del algoritmo, puede cambiarse el punto inicial por una estimación previa del vector de movimiento, que normalmente es el vector de movimiento que hemos obtenido en el bloque analizado anteriormente (recuérdese que bloques próximos tendrán vectores de movimiento parecidos). El algoritmo no tiene ninguna restricción en cuanto al tamaño de la región de inspección, que puede ser tan grande como deseemos. No obstante, es posible que no encuentre un mínimo local dentro del área especificada, en cuyo caso deberemos tomar como mínimo los valores obtenidos en los límites de la zona de búsqueda.

La búsqueda jerárquica suele obtener unos resultados excelentes con una complejidad computacional moderada. El procedimiento consiste en reducir la resolución espacial de la zona de búsqueda y del bloque que estamos comparando, con el objeto de simplificar los cálculos y obtener una primera aproximación del vector de desplazamiento. Suponiendo que la zona de búsqueda se reduce a una matriz de 64x64 y que utilizamos un bloque de 16x16 píxeles, podemos realizar, tal y como se indica en la figura 4.62, una primera reducción a 32x32 y 8x8 y posteriormente una segunda simplificación a 16x16 y 4x4 respectivamente. Con este tamaño de imágenes reducido se puede aplicar un procedimiento de búsqueda exhaustiva o uno de búsqueda simplificada para determinar una primera estimación del vector de movimiento. Posteriormente, se reconstruyen las imágenes con resolución doble y se inspeccionan los 8 píxeles vecinos del vector candidato, quedándonos con el mínimo como nuevo candidato. El procedimiento puede aplicarse hasta obtener las imágenes con la resolución original o incluso continuarlo una etapa más si queremos obtener un vector de movimiento final con una resolución fraccionaria. El número de niveles de descomposición de las imágenes depende de los tamaños del área de búsqueda y del bloque. Dos o tres etapas de descomposición suelen resultar suficientes. Los resultados proporcionados por la búsqueda jerárquica suelen ser bastante mejores que los de los métodos precedentes, sobre todo en los casos en los que el área de búsqueda es de dimensiones considerables.

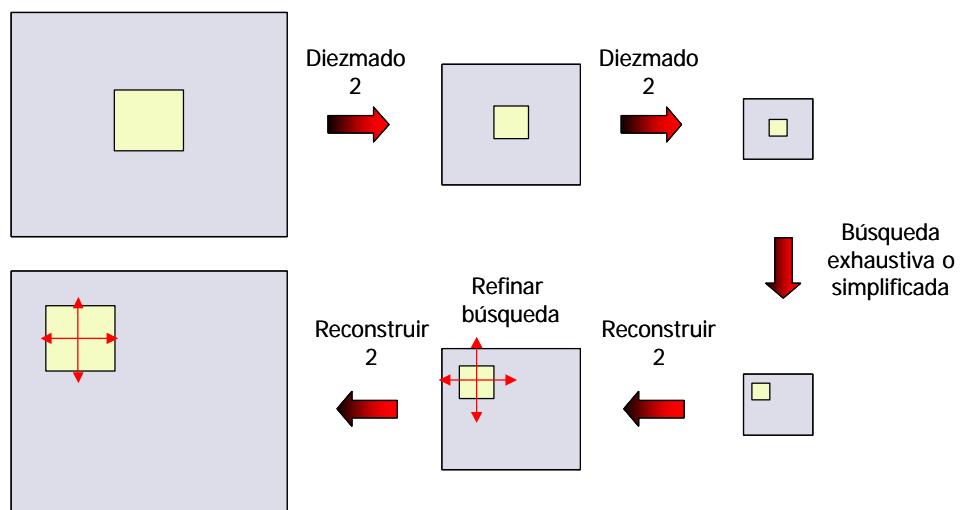


Fig. 4.62 Idea general del procedimiento de búsqueda jerárquica del vector de movimiento

4.9.4 Tipos de imágenes en el MPEG

El MPEG-1 y el MPEG-2 definen distintos tipos de imágenes dentro de la secuencia de vídeo en los que se utilizan técnicas de codificación distintas. El objetivo final es conseguir un sistema robusto, que pueda recuperarse automáticamente de los posibles errores en la recepción de los datos y que a su vez presente un factor de compresión considerable. Por ello, se definen subsecuencias o grupos de imágenes en los que se combina la compresión intra-trama, con compresiones basadas en la compensación de movimiento *forward* y compresiones basadas en la compensación bidireccional. Los tipos de imágenes que pueden aparecer en la secuencia codificada son:

- a) *Imágenes I (Intra-trama)*. El procedimiento de compresión utilizado es muy parecido al del JPEG. Las imágenes se descomponen en bloques de 8x8 píxeles que son transformados mediante la transformada coseno. Los coeficientes transformados son cuantificados mediante tablas pre-establecidas y finalmente se codifican mediante códigos de longitud variable. Por tanto, estas imágenes se codifican de forma independiente sin aprovechar ningún tipo de redundancia temporal. Se utilizan para el acceso aleatorio a distintas partes de la secuencia de vídeo en CD-ROM (MPEG-1) o en el DVD-Vídeo (MPEG-2). Las imágenes I pueden utilizarse para realizar la predicción mediante compensación de movimiento de las imágenes P y B que se describen a continuación.
- b) *Imágenes P (Predicted)*. La codificación de estas imágenes se realiza utilizando compensación de movimiento *forward* a partir de una imagen I u otra imagen P previa. Estas imágenes pueden propagar errores y se utilizan para realizar la predicción de otras imágenes P o B.
- c) *Imágenes B (Bidireccionales)*. Se obtienen usando compensación de movimiento bidireccional a partir de imágenes I o P. No se utilizan para predecir ninguna otra imagen por lo que no propagan los errores.
- d) *Imágenes D*. Estas imágenes sólo se utilizan en el MPEG-1. Son imágenes intra-trama en las que únicamente se codifica el coeficiente de continua de la transformada coseno. Se utilizan únicamente en los modos de avance rápido del vídeo para obtener una secuencia de alta velocidad pero baja calidad.

La relación y proporción de cada tipo de imágenes en la secuencia MPEG puede ser definida por el codificador, admitiéndose diversas posibilidades. Una de las secuencias utilizadas con mayor frecuencia se ilustra en la figura 4.63 y está formada por una secuencia básica de 12 imágenes en las que interviene una imagen del tipo I, dos del tipo P y 6 del tipo B. Las posibles secuencias de imágenes se denominan grupo de imágenes (*Group Of Pictures – GOP*) y suelen identificarse mediante dos parámetros N y M que identifican el número total de imágenes de la secuencia y el número de imágenes que hay entre dos imágenes que pueden utilizarse como imágenes de referencia (I o P). En nuestro ejemplo, la secuencia GOP tiene los parámetros N=12 y M=3.

Atendiendo a la estructura de la figura 4.63, la imagen P1 se codificaría utilizando compensación de movimiento *forward* sobre la imagen I1. Análogamente, la imagen P2 utilizaría la compensación de movimiento tomando como referencia a P1 y la P3 tomaría como referencia P2. Obsérvese que este proceso es muy sensible a los posibles errores en la decodificación de la secuencia que se propagaran al resto de imágenes. En efecto, si debido al ruido se produce un error en la decodificación de una de

las imágenes I o P, esta imagen quedará almacenada en la memoria del decodificador y se utilizará para reconstruir el resto de imágenes a partir de los vectores de movimiento y los errores de predicción recibidos. Aunque los nuevos errores de predicción se reciban correctamente, las nuevas imágenes no estarán exentas de errores, por cuanto se están utilizando como referencia una imagen incorrecta. Por ello, es importante dedicar un número considerable de bits a las imágenes que se utilizan como referencia para garantizar que no existe una propagación de errores considerable. Además, las transmisiones digitales de los datos deberán contener códigos protectores de error para garantizar la correcta decodificación de la información y mantener una tasa de bits erróneos baja. Los códigos de protección de error deberán adaptarse a las características del canal y están definidos por la normativa de transmisión regulada por el DVB. También se incluyen potentes códigos de protección de errores en los formatos de almacenamiento digital como el CD-ROM o el DVD. El MPEG no se ocupa de las protecciones que deben asignarse a los datos pues siempre dependen de las características de la aplicación y del ruido existente en el canal de transmisión (radiodifusión terrena, satélite, cable) o en el soporte físico de los datos (DVD, cinta magnética, etc).

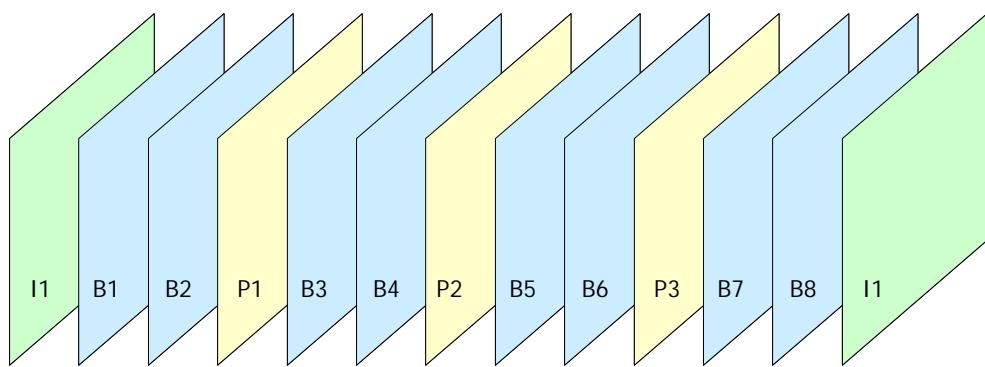


Fig. 4.63 Una posible secuencia de imágenes en MPEG ($N=12$, $M=3$)

Las imágenes B se codifican utilizando la compensación de movimiento bidireccional sobre las imágenes de referencia (I o P) más próximas. Así, la imagen B6 se codifica utilizando como referencia las imágenes P2 o P3. Esta forma de codificar la información representa ciertos problemas en los sistemas de acceso aleatorio basados en el MPEG, como el DVD-Vídeo. Para decodificar una imagen del tipo B, como por ejemplo la B6, deberemos decodificar primero las imágenes del tipo I, después P1, que permite decodificar P2 y esta a su vez permite decodificar P3. Una vez disponemos de P2 y P3 podemos proceder a decodificar B6. Por ello, las únicas imágenes que permiten el acceso aleatorio real son la imágenes del tipo I, que son las que siempre quedan referenciadas cuando queremos acceder a una escena concreta. Dependiendo del modelo de decodificador es posible que un DVD Vídeo, en el modo de avance rápido, sólo muestre las imágenes de tipo I o de tipo I y P, omitiendo las del tipo B. Además, la parada de imagen suele realizarse también en una imagen I o P. Las imágenes de tipo B suelen tener asignado un número de bits muy reducido, por lo que la calidad es muy baja. En modo de reproducción normal el sistema visual interpola estos problemas no advirtiendo la falta de calidad de las imágenes. No obstante, cuando la imagen está parada estos problemas pueden resultar demasiado evidentes. La reproducción de la secuencia en sentido inverso a velocidad normal (es decir decodificando todas las imágenes) suele ser una opción que sólo

incorporan los reproductores de DVD de gama más alta. En la figura 4.64 se muestra un grupo de imágenes (GOP) con parámetros $N=9$ y $M=3$ en el que se indican qué imágenes se utilizan como referencia para la codificación de cada una.

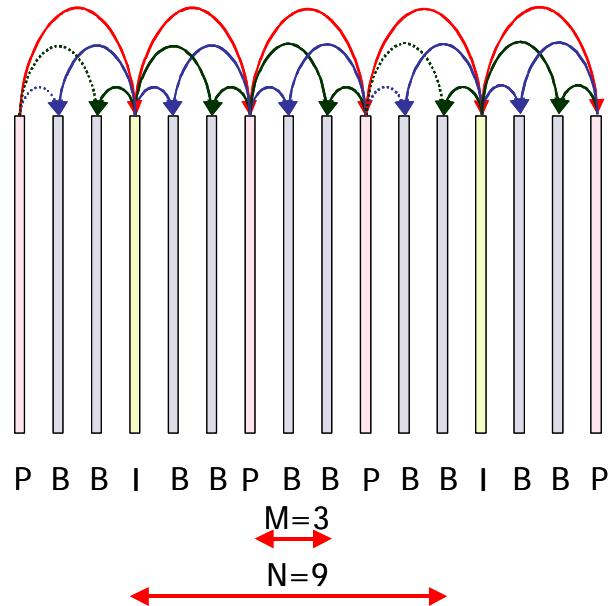


Fig. 4.64 Grupo de imágenes con $M=3$, $N=9$ en el que se indican las relaciones de predicción

Como las imágenes B no se utilizan como referencia de ninguna otra imagen, no propagan los errores. Esto significa que, en la práctica, los errores cometidos en estas imágenes no tienen tanta incidencia sobre las prestaciones del sistema, por lo que generalmente son las imágenes en las que se aplica la mayor compresión. Aunque el MPEG permite una gran libertad en la asignación de los bits, podríamos decir que las relaciones de bits que se dedican a cada uno de los tipos de imágenes son aproximadamente de 5, 3 y 1 para las imágenes I, P y B respectivamente. No obstante, estas relaciones pueden cambiar en función del contenido de la secuencia de vídeo y de las características del codificador.

4.9.5 Orden de transmisión de las imágenes

El uso de predicción bidimensional, junto con el objetivo de reducir al máximo posible la memoria del decodificador, impone que el orden en el que se transmiten las imágenes sea distinto del orden en el que deben visualizarse. Si no existieran estas restricciones sobre la memoria del decodificador, éste podría almacenar en un *buffer* las imágenes recibidas y decodificarlas cuando hubiera recibido toda la información necesaria. El MPEG establece que no se necesite almacenar más de dos imágenes de referencia en el decodificador, más una memoria adicional para la imagen que se esté codificando en cada momento. Para ello, es necesario cambiar el orden de transmisión de manera que las imágenes

del tipo B se reciban siempre después de haber recibido las referencias que se han utilizado en su codificación. En la figura 4.65 se muestra el orden de transmisión de las imágenes junto con la información que dispone el decodificador en cada una de las dos memorias de imágenes referencia y en la memoria de decodificación de imágenes B. También se indica cómo se reconstruye la secuencia de visualización y el retardo que ésta tiene respecto a la señal transmitida.

Orden Natural:	P ₋₃	B ₋₂	B ₋₁	I ₀	B ₁	B ₂	P ₃	B ₄	B ₅	P ₆	B ₇	B ₈	I ₉
Orden Transmisión:		I ₀	B ₋₂	B ₋₁	P ₃	B ₁	B ₂	P ₆	B ₄	B ₅	I ₉		
Memoria 1 Decodificador:	P ₋₃	P ₋₃	P ₋₃	P ₃	P ₃	P ₃	P ₃	P ₃	P ₃	P ₃	I ₉		
Memoria 2 Decodificador:		I ₀	I ₀	I ₀	I ₀	I ₀	I ₀	P ₆	P ₆	P ₆	P ₆		
Memoria Imag. Decodif:			B ₋₂	B ₋₁			B ₁	B ₂		B ₄	B ₅		
Imagen mostrada display:	P ₋₃	B ₋₂	B ₋₁	I ₀	B ₁	B ₂	P ₃	B ₄	B ₅	B ₆			

Fig. 4.65 Orden de transmisión y de visualización de la secuencia de imágenes MPEG ($N=9, M=3$)

El MPEG define el concepto de grupo cerrado de imágenes (*closed GOP*) cuando la codificación de la secuencia permite presentar imágenes en el *display* justo una imagen después de recibir una imagen intra-trama (I). El grupo cerrado de imágenes tiene interés en algunos sistemas de acceso aleatorio como el DVD, en los que las imágenes I se utilizan como referencia de inicio de pasajes. Una de las características del DVD es que debe permitir el salto de determinados pasajes (p.ej. visionado del vídeo sin mostrar escenas violentas o de contenido erótico) sin que el espectador advierta ningún tipo de discontinuidad. Esto supone que cuando se produce un salto en la secuencia normal puedan seguir viéndose las imágenes sin ningún tipo de interrupción.

Para obtener un grupo cerrado es necesario realizar algunas restricciones sobre la forma en la que se codifican las imágenes. Obsérvese en la figura 4.65, que después de recibir la imagen I₀, deberíamos mostrar por pantalla las imágenes B₋₂ y B₋₁. Sin embargo, en una codificación convencional, estas imágenes no podrían decodificarse, debido a que nos hemos saltado la imagen P₋₃ que se utiliza para la compensación de movimiento *forward*, por lo que deberíamos esperar dos cuadros adicionales hasta poder empezar con la decodificación de la secuencia.

Para evitar este problema, la definición de grupos cerrados se restringe a que todas las imágenes B que preceden a una imagen I sean codificadas utilizando únicamente compensación de movimiento *backward*. Con ello, las imágenes B₋₂ y B₋₁ podrán decodificarse utilizando únicamente I₀. En la figura 4.64 que muestra las relaciones de predicción entre las distintas imágenes, hemos indicado mediante líneas discontinuas las relaciones que no se utilizan en la definición de los grupos cerrados. En la transmisión de señales de televisión, donde el flujo de imágenes es constante y no puede ser configurado por el usuario no suelen utilizarse los grupos cerrados. La definición de grupo cerrado

significa que todas las imágenes existentes en el GOP puede decodificarse sin necesidad de recurrir a otras imágenes de un GOP anterior.

4.9.6 Algunos ejemplos con secuencias y codificadores ideales

Antes de entrar en los detalles de la codificación usada para cada tipo de imagen y de la estructura de la trama MPEG consideraremos algunos aspectos de carácter general sobre las propiedades y características que deberían presentar los codificadores ideales trabajando con secuencias simplificadas.

La secuencia de vídeo más simple es un plano fijo estacionario. En este caso, toda la información de vídeo estaría contenida en las imágenes I, cuya codificación se repetiría periódicamente durante la transmisión de estas imágenes en la secuencia MPEG. Evidentemente, la trama MPEG contendría informaciones de cabecera sobre sincronismos, sobre los tipos de codificación utilizados en los distintos tipos de imagen y otros detalles relativos a la estructura de la trama de datos, pero esencialmente, la totalidad de la información de vídeo estaría dedicada a la codificación de imágenes intra-trama. Probablemente, el codificador detectaría esta situación ideal y podría decidir dedicar un gran número de bits a la codificación de las imágenes intra, con el objeto de aumentar la calidad de la secuencia de vídeo. Esta situación ideal no significa que el procesador del codificador esté exento de realizar los complejos cálculos de estimación de movimiento y predicción bidireccional. Simplemente significa que los resultados de los cálculos serían perfectos, es decir, todos los vectores de movimiento serían nulos, las diferencias entre vectores de movimiento también serían nulas, la predicción de las imágenes sería exacta y, por lo tanto, los errores de predicción serían cero, con lo que su codificación supondría sólo un número de bits mínimo. Obsérvese, no obstante, que si el plano fijo estacionario tiene cierto ruido aditivo, la situación dejará de ser ideal, por cuanto los errores de predicción de las imágenes P y B coincidirán con las diferencias entre los ruidos en las imágenes I, P y B. Esto significará que parte del flujo de información deberá dedicarse a codificar un ruido que no aporta ninguna información visual. Por esta razón, suele decirse que es muy importante eliminar el ruido existente en las imágenes antes de proceder a su codificación, ya que se trata de una señal de naturaleza totalmente aleatoria, que, por lo tanto, no puede predecirse, y que requiere un número de bits que, en cualquier caso, empeoran la calidad con la que se visualiza la imagen.

Otra secuencia simple es un *travelling* horizontal sobre una escena estacionaria. En este caso, la cámara se desplaza a una velocidad constante, por lo que todos los objetos experimentan un movimiento de la misma magnitud y sentido. Las imágenes intra-trama se codificarán de forma independiente del resto de imágenes y por lo tanto contendrán una cantidad de información parecida a la que contienen en otro tipo de escenas. Una parte de los bloques de las imágenes P pueden predecirse de forma exacta, puesto que corresponden a los mismos objetos que ya estaban presentes en la imagen I anterior, desplazados por el movimiento que ha experimentado la cámara. Además, todos estos vectores de movimiento serían idénticos, por lo que su codificación como diferencias puede resultar muy eficiente. No obstante, existirán algunos bloques de las imágenes P que no puedan ser estimados basándose en las imágenes anteriores, ya que se corresponden con los nuevos objetos que han quedado descubiertos por el movimiento de la cámara. Un codificador ideal debería identificar estos bloques, reconocer que no resulta apropiado utilizar la compensación de movimiento y codificarlos como bloques intra-trama. La relación entre el número de bloques que se codifican

mediante compensación de movimiento y los que se codifican como intra-trama dependerá de la velocidad con la que se realice el *travelling*. Los bloques del tipo B podrán codificarse con un número de bits prácticamente nulo, pues el uso de la compensación bidireccional permitirá obtener predicciones ajustadas de todos los objetos utilizando las referencias pasadas o futuras. Los objetos de la imagen B que aún no hayan aparecido en la referencia futura se codificarán utilizando predicción *forward*, mientras que los que ya hayan desaparecido en la imagen anterior se codificarán utilizando compensación de movimiento *backward*. Los objetos que estén presentes en las dos referencias se codificarán como el promedio entre las dos predicciones, minimizando de este modo la incidencia del ruido en las pérdidas de ganancia de compresión.

Finalmente, consideramos la codificación de un cambio de plano entre dos escenas estacionarias. Supondremos que estamos trabajando con una secuencia GOP con $N=9$ y $M=3$ y que el cambio de plano se produce en la imagen B4. En este caso, la imagen P1 se codificará utilizando la imagen I0 como referencia y utilizando predicción de movimiento *forward*. La imagen P2 también intentará codificarse del mismo modo tomando como referencia P1. No obstante, al haberse producido el cambio de plano, el error de predicción aumentará considerablemente, por lo que la mayoría de bloques se codificarán como bloques intra-trama (en teoría todos los bloques deberían codificarse como intra-trama, pero es posible que la compensación de movimiento encuentre zonas parecidas entre las dos imágenes, que aunque en realidad corresponden a objetos distintos pueden utilizarse para la predicción). Las imágenes B4 y B5 se codificarán utilizando compensación de movimiento que en su mayor parte será *backward*, ya que debido al cambio de plano la referencia ideal a esta imágenes se encuentra en el futuro (P2).

Debemos insistir que las ideas presentadas en esta sección pretenden proporcionar un punto de vista general sobre la filosofía de codificación del MPEG y que no necesariamente todos los codificadores deben comportarse del modo descrito. Durante el análisis de las imágenes es posible que aparezcan errores en la estimación de movimiento, tomando como referencia zonas de imagen que se parecen a las que queremos codificar, pero que no corresponden a los mismos objetos. No obstante, si existe parecido entre las regiones, podremos obtener alguna ganancia de codificación. Por otra parte, el MPEG admitiría que un determinado codificador no realizará la compensación de movimiento *backward* y todas las imágenes B fueran codificadas con compensación *forward* o como imágenes intra-trama. El resultado sería una pérdida en el factor de compresión y probablemente una pérdida en la calidad de las imágenes. No obstante, el decodificador de MPEG sería compatible con este formato ya que la trama de datos que proporcionaría el codificador coincide con la que se hubiera transmitido en el caso de considerar que la codificación *backward* no resultaba eficiente.

4.9.7 Diagramas generales del proceso de codificación y decodificación de imágenes

En las figuras 4.66 y 4.67 se muestran los esquemas genéricos del proceso de codificación y decodificación de imágenes en el MPEG. Los esquemas son muy simplificados y deben interpretarse únicamente como la idea general implícita en todo el proceso de codificación donde intervienen múltiples decisiones sobre el procedimiento con el que se codifica cada macrobloque, las escalas de cuantificación, los códigos de longitud variable y la codificación de las componentes de croma y de luminancia por separado. Atendiendo al esquema del codificador, el bloque de preprocesado incluye los posibles filtros de interpolación, diezmado o eliminación de ruido que se aplica a la señal de vídeo

a codificar. El bloque de estimación de movimiento proporciona los vectores de desplazamiento que deben aplicarse en la predicción de las imágenes. Suponemos que este bloque es suficientemente sofisticado como para no actuar en el caso de que se trate de imágenes del tipo I o indicar al resto de bloques cuándo resulta conveniente aplicar la compensación de movimiento y cuándo es preferible codificar los macrobloques como intra-trama.

La salida del estimador de movimiento se introduce en los códigos de longitud variable para optimizar la codificación de estos vectores. El bloque encargado de realizar la DCT se aplica sobre la señal error de predicción. Esta señal será la diferencia entre la imagen actual y la referencia desplazada. En el caso en que se realice una codificación intra-trama del bloque, la señal en la salida del predictor por compensación es cero de modo que se calcula directamente la DCT del bloque a codificar. Los coeficientes transformados se cuantifican, generalmente utilizando las tablas de cuantificación por defecto aunque también es posible utilizar tablas propias; que en el caso del MPEG-1 deben ser las mismas para toda la secuencia de vídeo, pero que para el MPEG-2 pueden variarse de imagen a imagen.

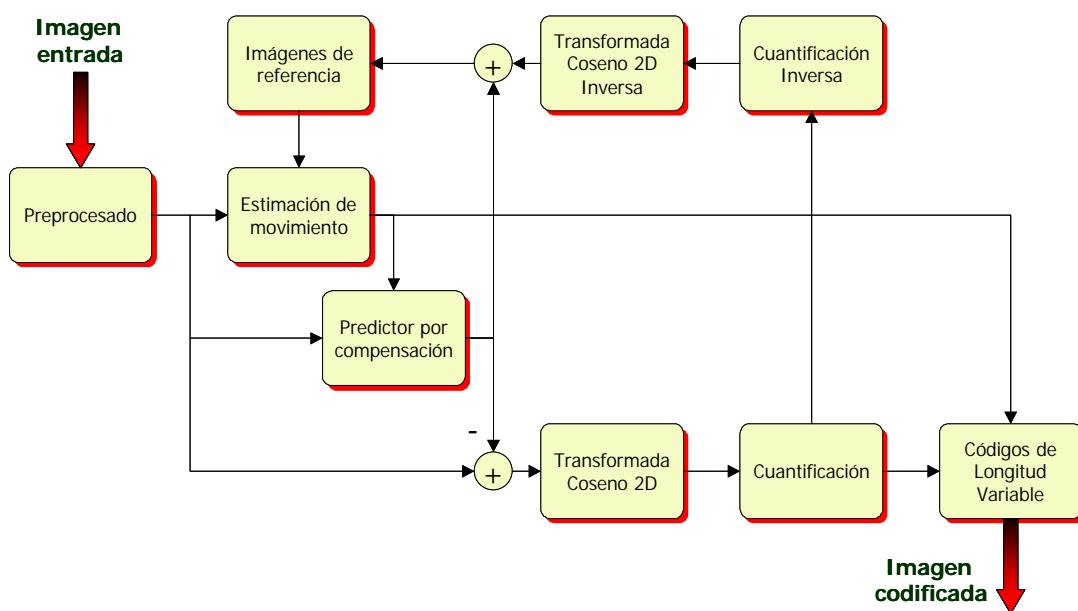


Fig. 4.66 Esquema básico del proceso de codificación empleado en el MPEG

Por otra parte, el sistema puede actuar directamente sobre la tasa de bits con que se codifica la secuencia de vídeo, modificando en tiempo real los factores de escala del cuantificador. Si un determinado fragmento de la secuencia de vídeo contiene excesiva información no predictible, puede reducirse la tasa de bits utilizando cuantificadores con menos pasos. Evidentemente, esto incide sobre la calidad de la secuencia.

El codificador almacena en memoria las imágenes de referencia que recuperará el decodificador, basándose en los valores de los coeficientes cuantificados. El esquema es parecido al utilizado en los sistemas de predicción lineal estudiados en los primeros apartados de este capítulo y es necesario para evitar que los errores de cuantificación se vayan acumulando e impidan que el receptor recupere una imagen de calidad aceptable.

El esquema del decodificador, que se representa en la figura 4.67, realiza la conversión de los códigos de longitud variable a coeficientes cuantificados y vectores de movimiento y restablece las imágenes al formato de vídeo convencional. Pueden aplicarse filtros de postproceso para mejorar la calidad de la imagen (por ejemplo filtros de interpolación de líneas para mejorar la visualización en el *display*)

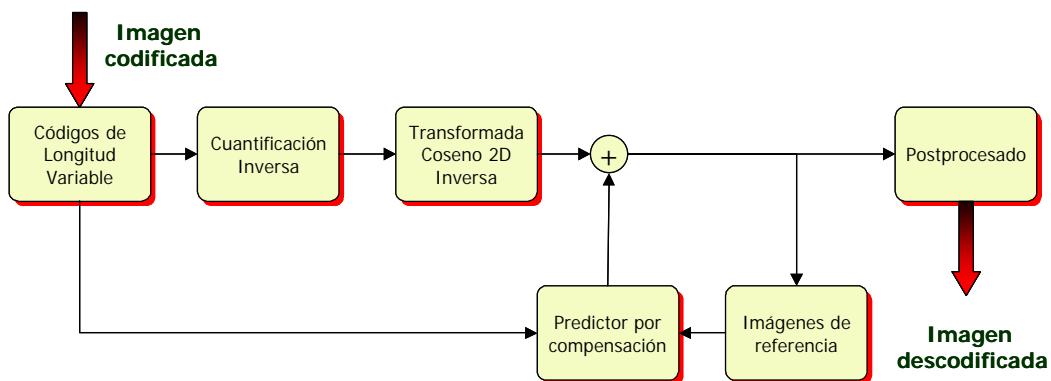


Fig. 4.67 Esquema del proceso de decodificación empleado en el MPEG

4.9.8 Estructura jerárquica de la trama MPEG

La trama MPEG está estructurada de forma jerárquica con el objetivo de aumentar la flexibilidad del sistema y la intercomunicación entre el codificador y el decodificador. Mediante esta estructura de trama pueden resultar compatibles codificadores que realicen procesos lógicos distintos. La estructura de la trama empieza con una cabecera de la capa superior y continúa con las cabeceras e informaciones de las capas siguientes. Las capas utilizadas, en orden jerárquico descendiente, son las de secuencia, grupo de imágenes, imagen, *slice*, macrobloque y bloque. Las relaciones de estas capas en una secuencia de vídeo y en cada una de las imágenes se representan de forma esquemática en la figura 4.68. En los siguientes apartados describimos algunos de los detalles que se incorporan en las cabeceras de cada capa.

- Capa de secuencia.* La secuencia está formada por uno o más Grupos de Imágenes (GOP). En esta cabecera se definen los parámetros básicos de la secuencia de vídeo como las dimensiones de la imagen, la frecuencia de imagen, los tamaños de buffer esperados del decodificador, la relación de aspecto del píxel, etc. En el MPEG-1 la cabecera de secuencia puede contener matrices de cuantificación distintas a las matrices

- por defecto. En el MPEG-2, las posibles variantes de las matrices de cuantificación se definen en la capa de imagen y por lo tanto pueden variarse de imagen a imagen.
- b) *Capa de Grupo de Imágenes (GOP)*. El grupo de imágenes puede definirse como una secuencia de imágenes que se transmiten en un orden distinto al que serán presentados en el *display* y que facilita el acceso aleatorio a la secuencia. Los diferentes paquetes de GOP que forman una secuencia se transmiten en el mismo orden en que serán representados en pantalla, el cambio de orden entre la transmisión y la representación sólo se produce en las imágenes que forman cada GOP. El grupo de imágenes siempre empieza con una imagen del tipo I. La cabecera aporta información sobre si se trata de un GOP abierto o cerrado. También indica el código de tiempos SMPTE de la primera imagen I. En el MPEG-1 debe transmitirse el tipo de estructura del grupo (es decir, los parámetros M y N). En el MPEG-2 esta información es opcional y puede tomar valores por defecto.
 - c) *Capa de Imagen (Picture)*. Es una unidad de codificación que incluye la luminancia y las dos componentes de croma de la imagen. La cabecera define el tipo de imagen (I, P, B, D) así como el rango de valores máximos del vector de movimiento en esta imagen. También indica el orden de la imagen dentro del GOP.
 - d) *Capa de Slice*. Es la unidad básica de resincronización del sistema y por lo tanto desempeña un papel fundamental en el soporte de errores. El decodificador puede saltarse un *slice* corrupto e ir directamente a decodificar el siguiente. El número de *slices* en una imagen es variable y depende de las características del ruido en el que se aplica el MPEG. Puede variar entre 1 *slice* por imagen o 1 *slice* por macrobloque. En la cabecera se incluye el código de inicio de *slice* y el factor de escala del cuantificador.
 - e) *Capa de Macrobloque*. Es la unidad básica sobre la que se realiza la compensación de movimiento. Está formado por una región de 16x16 píxeles en la luminancia y regiones de 8x8 en las componentes de croma. Proporciona información sobre el tipo de codificación utilizado en el macrobloque, la escala del cuantificador y el vector de movimiento.
 - f) *Capa de bloque*. Es la unidad de codificación a la que se aplica la transformada coseno. Está formada por bloques de 8x8 píxeles en la componente de luminancia y de 4x4 píxeles en la de croma. Contiene la información relativa a la componente de continua (DC), los coeficientes AC y la información de finalización de la codificación de bloque (EOB).

4.9.9 Algunos detalles sobre la codificación de las imágenes del tipo I

La codificación de las imágenes del tipo I es muy parecida a la que se utiliza en el JPEG. La diferencia más importante es la que se refiere a las tablas de cuantificación y a los procedimientos que se utilizan para realizar cambios sobre las escalas de cuantificación. En principio, las tablas de cuantificación pueden proporcionarse en el inicio de la secuencia en el MPEG-1 o en la cabecera de imagen en el MPEG-2, aunque normalmente se utilizan las tablas de cuantificación por defecto. Sólo en casos en los que las dimensiones en número de píxeles de las imágenes sean muy distintas de las habituales pueden obtenerse ventajas considerables por el cambio de tablas de cuantificación. El tamaño de los *slices* no viene especificado en el estándar y por lo tanto deberá fijarlos el codificador. La estructura de macrobloques implícita en la trama MPEG impone que el orden en que se codifican

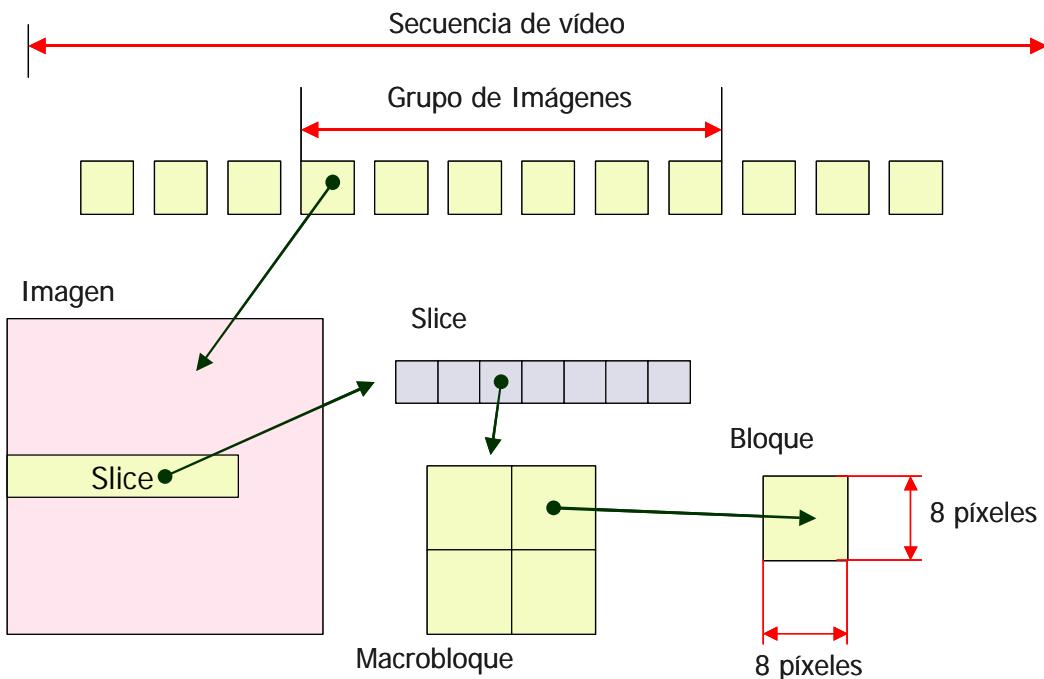


Fig 4.68 Estructura jerárquica de la trama MPEG

los bloques intra-trama no coincide con el que se utiliza en el JPEG (por filas de 8x8), sino que se codifiquen por filas dentro del macrobloque (es decir, dos filas de 8x8).

La transformada coseno se aplica directamente a los valores de los píxeles sin extraer el nivel de gris neutro. La DCT se calcula con una precisión de 11 bits, por lo que el rango de la componente de continua va de 0 a 2047 (siempre es positiva, ya que no hemos eliminado la componente de continua) y el de los AC va de -1024 a 1023. El coeficiente de continua se codifica de forma diferencial respecto al coeficiente de continua del bloque anterior. La codificación diferencial se reinicia a cero cada vez que empieza un nuevo *slice*. Siempre se codifica con un paso de cuantificación fijo de 8, utilizando códigos de longitud variable parecidos a los del JPEG en los que se emplea un símbolo para indicar el margen de niveles del coeficiente y otro símbolo para indicar su valor.

Los coeficientes AC se exploran en zigzag y se pasan a través de cuantificadores uniformes. La tabla de cuantificación por defecto para la componente de luminancia se representa en la figura 4.69. El MPEG permite, como ya hemos comentado, variar la escala del cuantificador para ajustar la tasa de bits que se produce en la salida del codificador. Si observamos que la tasa de bits es excesivamente alta, podemos aumentar el paso de cuantificación, por lo que se reducirá la cantidad de bits con que se codifica la información y se perderá calidad. La escala del cuantificador se indica con una palabra de 5 bits incorporada en la cabecera del macrobloque y se aplica a los 4 bloques que lo forman. Esta

palabra de 5 bits define un número entre 1 y 31 que, a partir de la tabla de cuantificación nos permite determinar el nivel del coeficiente, mediante la ecuación:

$$S_{quv} = \text{round}\left(\frac{8 \cdot S_{uv}}{qs \cdot Q_{uv}}\right) \quad (4.29)$$

donde S_{uv} es el valor del coeficiente transformado, Q_{uv} el valor de la tabla de cuantificación, qs es el factor de escala y S_{quv} es el coeficiente transformado una vez cuantificado. De acuerdo con esta ecuación, cuando el factor de escala toma el valor 8, se utiliza la tabla de cuantificación por defecto. Si disminuimos el factor de escala, utilizaremos un paso de cuantificación menor (más calidad de imagen y mayor tasa de bits) y viceversa. En el MPEG-2 es posible definir, además de la escala de cuantificación, si se utilizan curvas de cuantificación lineales o no lineales.

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Fig. 4.69 Tabla de cuantificación por defecto en la componente de luminancia para la codificación intra-trama

Para los coeficientes AC se utilizan códigos de longitud variable en los que se indica el número de ceros consecutivos y el margen dinámico del siguiente coeficiente no nulo en una primera parte del mensaje y el valor del coeficiente en una segunda parte. La idea de fondo es análoga a la que se utiliza en el JPEG.

4.9.10 Algunos detalles sobre la codificación de las imágenes de tipo P

La codificación de las imágenes P es algo más compleja que las imágenes intra-trama, pues debe realizarse la estimación de movimiento y decidir cuál es la forma más eficiente de codificar un macrobloque en función de los resultados obtenidos. En este apartado definiremos los distintos tipos de macrobloques que puede producirse en la codificación de las imágenes P.

Una de las primeras peculiaridades de las imágenes P es que puede saltarse la codificación de algunos macrobloques (*skipped macroblock*). Esto ocurre cuando el procesador considera que no es necesario codificar un macrobloque, puesto que la información que contiene puede aproximarse con suficiente precisión basándonos en la información que tenemos en la misma posición en la imagen de referencia

anterior. En este caso, los 4 bloques que forman el macrobloque se sustituyen por los mismos píxeles que teníamos en la imagen anterior, lo que significa que, en el fondo, estamos realizando una compensación de movimiento con un vector de movimiento nulo y que el error de predicción obtenido es suficientemente próximo a cero como para que no sea preciso codificarlo.

Además de los macrobloques *skipped* es posible tener 7 tipos adicionales que se denotan como *pred-mc*, *pred-c*, *pred-m*, *intra-d*, *pred-mcq*, *pred-cq* y *intra-q*. Las abreviaturas que se utilizan para identificar los posibles tipos de macrobloque corresponden a:

- pred*: Macrobloque codificado utilizando predicción. La codificación no es intra-trama.
- m*: Macrobloque codificado utilizando compensación de movimiento *forward*. Se transmite el vector de movimiento asociado.
- c*: Como mínimo se transmite la información de error asociada a uno de los bloques que forman el macrobloque.
- d*: Se utiliza el cuantificador por defecto.
- q*: Se realiza un cambio en la escala del cuantificador.
- skipped*: Macrobloques que no son transmitidos.

De acuerdo con estas abreviaturas, un macrobloque del tipo *pred-mcq* estará codificado utilizando predicción por compensación de movimiento *forward*, lo que significa que se transmite el vector de movimiento asociado. Además, el sufijo *c* indica que, como mínimo, se transmite la información de error asociada a uno de los bloques y que se realiza un cambio de escala en el cuantificador. Un macrobloque del tipo *pred-c* indicaría que se utiliza predicción, pero sin transmitir el vector de movimiento, y que como mínimo se envía la información asociada a uno de los bloques.

16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16

Fig. 4.70 Matriz de cuantificación inter-trama para la componente de luminancia

La misión del codificador es elegir entre las distintas posibilidades de codificar un macrobloque aquella que proporcione la mejor opción. Generalmente, las decisiones no son evidentes y deben probarse varias posibilidades antes de llegar a una decisión. Las matrices de cuantificación utilizadas en el caso de realizar codificación mediante predicción tienen un paso de cuantificación uniforme para todos los coeficientes. Esto es debido a que el contenido frecuencial de las señales de error no muestra

ninguna preferencia para concentrar su energía en unas frecuencias particulares y en principio todos los coeficientes tienen la misma importancia en la reconstrucción de las imágenes. En la figura 4.70 se muestra la matriz de cuantificación por defecto que se utiliza en la codificación inter-trama.

Insistimos de nuevo en que la complejidad del MPEG es asimétrica. Las decisiones que debe tomar el codificador sobre cómo se realiza la compresión de un determinado macrobloque son muy complejas e involucran un gran número de parámetros, pero todas estas decisiones se proporcionan al decodificador a través del formato de trama de la señal. El decodificador únicamente requiere aplicar de forma sistemática los procedimientos inversos a los que ha decidido utilizar el decodificador. Además, como el MPEG no establece los procedimientos mediante los que deben tomarse estas decisiones, es posible que existan codificadores que puedan obtener una mayor calidad de imagen utilizando una tasa de bits inferior a otros codificadores.

4.9.11 Detalles sobre la codificación de las imágenes del tipo B

Las imágenes del tipo B son las más complejas de codificar, debido a que requieren un mayor número de decisiones para determinar qué tipo de compensación de movimiento se aplica: *forward*, *backward* o interpolación bidireccional. En cualquier caso, deben estimarse dos posibles vectores de movimiento que se codificarán o no en función de los resultados y el análisis de las imágenes realizado por el codificador. Los vectores de movimiento se codifican de forma diferencial reinicializando el vector anterior a cero cada vez que se inicia un *slice* o que un macrobloque se codifica como intra-trama.

Existen hasta 12 posibles tipos de macrobloques que, al igual que en las imágenes P, se denotan mediante sufijos y prefijos. La notación empleada es la misma que la de los macrobloques para imágenes P, con la salvedad de que aparecen tres nuevos sufijos que corresponden con los macrobloques que utilizan interpolación bidireccional (sufijo -i), los macrobloques con compensación de movimiento *forward* (sufijo -f) y los que usan compensación de movimiento *backward* (sufijo -b). Los posibles tipos de macrobloques son:

- pred-i*: Se usa compensación de movimiento con interpolación bidireccional. El error resultante en todos los bloques es suficientemente pequeño y no se codifica.
- pred-ic*: Igual que el caso anterior pero ahora, al menos uno de los bloques tiene un error significativo, que deberá ser codificado.
- pred-b*: Se usa compensación de movimiento *backward*. El error resultante no se codifica.
- pred-bc*: Igual que el caso anterior, pero ahora se codifica el error de al menos uno de los bloques.
- pred-f*: Se usa compensación de movimiento *forward* para predecir el macrobloque y no resulta necesario codificar los errores.
- pred-fc*: En este caso se emplea la compensación de movimiento *forward* y es necesario codificar al menos uno de los errores.
- intra-d*: Codificación intra-trama del macrobloque utilizando el cuantificador por defecto.
- pred-icq*: Igual que el *pre-ic* pero indicando un cambio de escala en el cuantificador.
- pred-fcq*: Idem que el caso anterior, pero para compensación de movimiento *forward*.
- pred-bcq*: Igual que el caso anterior, pero con compensación de movimiento *backward*.

intra-q: Codificación intra-trama proporcionando la información de un cambio de escala en el cuantificador.

skipped: No es necesario codificar este macrobloque y puede aproximarse por el de la imagen de referencia con un vector de movimiento nulo.

En la figura 4.71 se representa un posible diagrama lógico que puede utilizar el codificador para la toma de decisiones respecto a la clasificación del bloque en uno de los tipos anteriores. El algoritmo de decisión es sólo una posible alternativa, puesto que, como ya sabemos, el MPEG no establece los criterios de decisión ni los procedimientos de estimación. La primera decisión que se toma es el tipo de compensación de movimiento que resulta más rentable utilizar. Para ello deben estimarse los vectores de movimiento *forward* y *backward* y evaluar el error que se comete cuando se predice el bloque utilizando compensación de movimiento *forward*, *backward* o bidireccional. Normalmente, la decisión sobre la compensación óptima se toma basándose en el valor medio de los errores de predicción elevados al cuadrado. Aquella compensación que produzca un error cuadrático medio mínimo es la que tomaremos como óptima. Otra posibilidad consiste en utilizar la suma de los valores absolutos del error de predicción en todos los píxeles del macrobloque y optar por la compensación que obtenga un mínimo en esta medida.

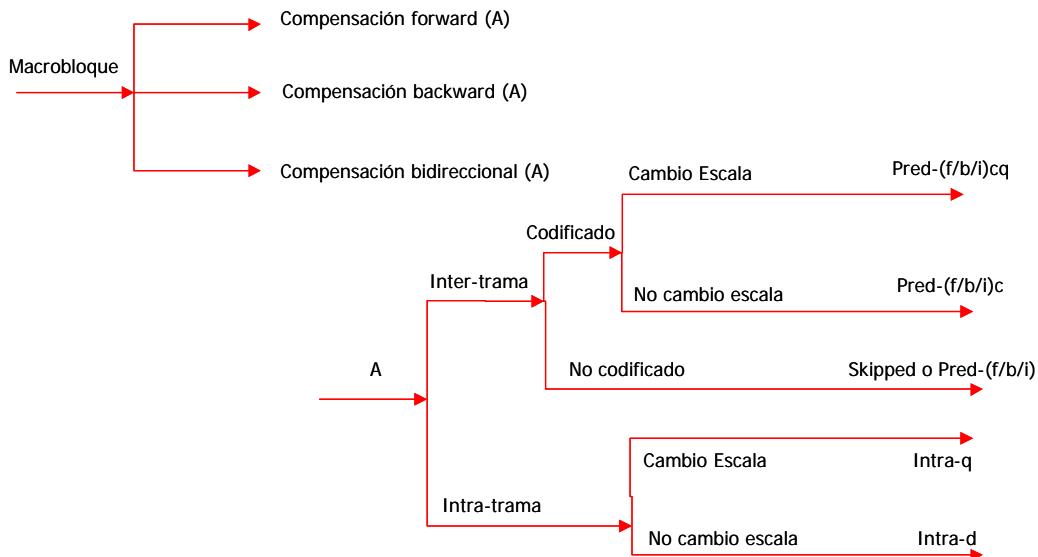


Fig. 4.71 Algoritmo de decisiones para la clasificación del macrobloque

La siguiente decisión consiste en determinar si se aplicará una codificación intra-trama o una codificación inter-trama. Los resultados obtenidos en la etapa anterior pueden proporcionar una valiosa ayuda en esta decisión. Si las magnitudes de los errores cuadráticos o absolutos del error de predicción han resultado muy grandes, probablemente no resultará rentable codificar el macrobloque utilizando predicción, por lo que puede resultar aconsejable codificarlo con procedimientos intra-

trama convencionales. En cambio, si los errores son muy pequeños, significa que la predicción es correcta y es recomendable utilizar la técnica de predicción elegida en el primer apartado.

En el supuesto de que la decisión sea utilizar una codificación inter-trama, el codificador debe decidir, seguramente también en función del error obtenido, si es necesario codificar el error obtenido en el macrobloque o es suficiente con que en el receptor se reconstruya el macrobloque utilizando la misma información de que ya se dispone en una imagen de referencia. Finalmente, cuando es necesario enviar alguna información relativa a los errores de predicción o a la codificación intra-trama del macrobloque deberá decidirse si se sigue utilizando la misma escala del cuantificador que en los bloques anteriores o resulta aconsejable realizar un cambio de escala. La decisión del cambio de escala suele venir determinada por la tasa de bits que estamos obteniendo en cada momento.

4.9.12 Control de la tasa de bits

Una de las características implícitas al MPEG es que la tasa de bits con el que se codifica una secuencia de bits depende no sólo de los parámetros propios del codificador, sino del contenido de las imágenes. Hemos visto que cuando las imágenes contienen abundante movimiento y objetos pequeños el número de bits necesario para codificarlas con una calidad predeterminada es muy superior al que se requiere cuando se trabaja con imágenes casi-estacionarias y fácilmente predecibles. Esto significa que si deseamos utilizar este sistema de codificación en un canal de datos con velocidad de transmisión constante, deberemos ir adaptando los parámetros del codificador en función del contenido de las imágenes. Análogamente, si el codificador debe trabajar en un canal con tasas de transmisión variables, será necesario ajustar dinámicamente sus parámetros hasta obtener los flujos de transmisión deseados.

La forma básica para controlar el flujo de datos consiste en modificar la escala del cuantificador. Para ello, el codificador suele disponer de un modelo del *buffer* del decodificador que va actualizando a medida que se van codificando nuevos fotogramas (entrando los bits correspondientes a estos fotogramas en el *buffer*) y se van decodificando las imágenes (extrayendo los bits de las imágenes decodificadas). El codificador informa del tamaño del *buffer* de memoria que requiere el decodificador mediante una información situada en la cabecera de la secuencia MPEG y deberá realizar la codificación de la secuencia garantizando que no se produce rebosamiento del *buffer* (el decodificador no podría almacenar los nuevos datos recibidos) ni que queda temporalmente vacío (no podrían decodificarse nuevas imágenes).

Cuando se detecta que por las características de la secuencia de vídeo el *buffer* está excesivamente lleno, es habitual reducir los pasos del cuantificador realizando aproximaciones más toscas del error de predicción o de los macrobloques que se codifican en modo intra-trama. Con ello, se reduce la tasa de bits de la secuencia al coste de bajar la calidad de la señal de vídeo. La modificación del paso de cuantificación no es la única estrategia posible para controlar la velocidad del tren de bits. También es posible realizar el control introduciendo cambios en los algoritmos de decisión y clasificación de macrobloques. Así, en el supuesto de que el estado del *buffer* del decodificador esté excesivamente lleno, podríamos decidir no contemplar la posibilidad de codificar macrobloques del tipo intra-trama en las imágenes B o incluso no utilizar bits adicionales para codificar los macrobloques de estas imágenes y aproximarlos siempre por los valores predichos. En general, el control de la tasa de bits a

este nivel del algoritmo de decisión es mucho menos directa que a través del paso de cuantificación y sólo se emplea en casos en los que el desbordamiento del buffer es inminente. En la radiodifusión de señales de televisión es habitual asignar tasas de codificación distintas a los diferentes programas en función del contenido de los mismos. Así, los noticiarios o programas de entrevistas, en los que existe poco movimiento de las imágenes, suelen tener asignada una menor tasa de bits que los programas de deportes. La plataforma que gestiona el servicio se encarga de asignar las tasas de codificación a cada programa en función de los contenidos, repartiendo el ancho de banda del canal entre los distintos programas.

4.9.13 Tratamiento de imágenes entrelazadas en el MPEG-2

Las imágenes entrelazadas ofrecen ciertos problemas adicionales en la compresión de vídeo. El problema básico es que si el formato entrelazado procede de una cámara de vídeo, cada campo corresponde a instantes de tiempo distintos, por lo que los objetos que se desplazan en el sentido horizontal pueden aparecer en distintas posiciones en cada uno de los campos. Esto provoca que aparezca un efecto de serrado entre las líneas pares e impares de un mismo objeto que reduce la correlación existente entre las líneas adyacentes. En televisión analógica convencional este efecto es tolerado por el sistema visual humano, ya que sólo se produce de forma significativa en los objetos que tienen un movimiento considerable dentro de la imagen. La resolución espacial del ojo disminuye cuando aumenta el movimiento de los objetos, por lo que este problema puede pasar inadvertido cuando se presenta la secuencia de vídeo en tiempo real. No obstante, si pudiéramos detener la imagen analógica, mostrando los dos campos simultáneamente, observaríamos todos estos problemas existentes en la imagen.

Debe mencionarse que cuando el material procede de la conversión de una película cinematográfica a vídeo entrelazado, no se produce este problema. En este caso, los conversores de película a vídeo (telecinemas) parten de un único fotograma para obtener los dos campos, de forma que ambos corresponden al mismo instante de tiempo.

En la figura 4.72 se muestra visualmente el problema del entrelazado en una imagen. La imagen ha sido obtenida con un movimiento horizontal de la cámara y se representan los dos campos superpuestos. La figura muestra también las imágenes obtenidas para cada uno de los campos por separado donde podemos observar que la correlación entre líneas adyacentes es ahora muy alta.

El problema del entrelazado en las técnicas de compresión de imagen es evidente puesto que la correlación entre las líneas sucesivas puede disminuir apreciablemente, resultando en imágenes más complejas de codificar tanto en modo intra-trama como inter-trama. Recordemos que en la codificación intra-trama, la compresión se basaba en extraer la redundancia espacial existente en la imagen y que esta redundancia se reduce cuando existen diferencias considerables entre píxeles que están próximos.

El MPEG-1 no considera un tratamiento específico para las imágenes entrelazadas, pues se supone que el vídeo o material audiovisual del que se parte ha sido tratado previamente para poder realizar una codificación progresiva. Téngase en cuenta que los tamaños de imagen más habituales en MPEG-1 son de 352x288 píxeles, lo que significa que basta con muestrear uno solo de los campos (288

líneas) para obtener una imagen o *frame* completo. En cambio, el MPEG-2 debe ser capaz de tratar con resoluciones de imagen de calidad *broadcasting*, por lo que deberemos ser capaces de tratar con material previamente entrelazado. Por ello, en el MPEG-2 se definen algunos modos adicionales que permiten tratar este tipo de material con el objeto de lograr mejores factores de compresión que los que se obtendrían considerando los dos campos como una única imagen. Es importante insistir que los modos entrelazados del MPEG-2 se han definido específicamente para poder tratar con este tipo de formatos y que sólo tienen sentido cuando se procesan imágenes adquiridas con este tipo de exploración. Si se dispone de un formato de partida no entrelazado (p.ej. el cine), deberán aplicarse las técnicas convencionales, por cuanto, en general, pueden conseguir un factor de compresión más elevado.

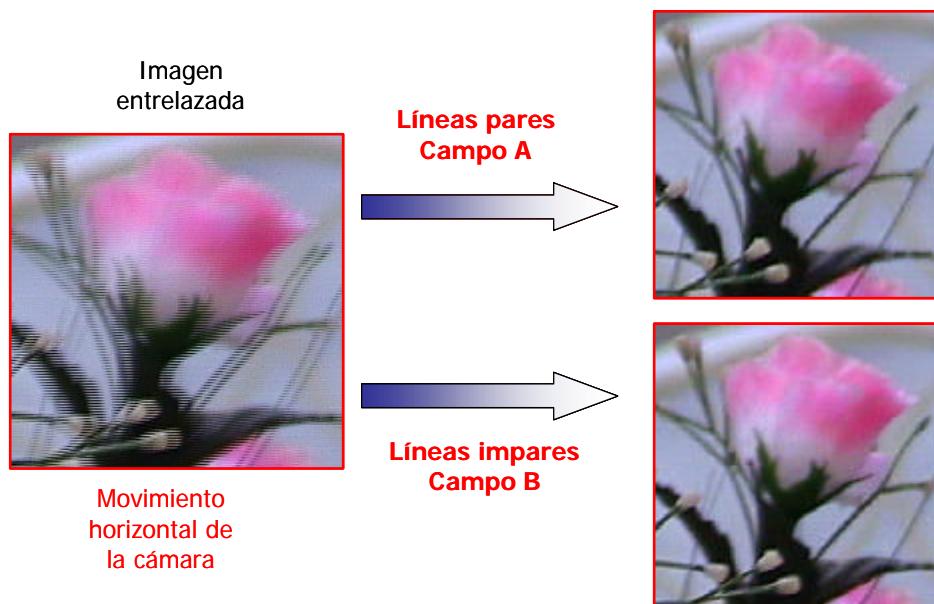


Fig. 4.72 Ejemplo de imágenes entrelazadas en un movimiento de cámara horizontal

El tratamiento de imágenes entrelazadas en el MPEG-2 afecta tanto a la definición de los macrobloques y la definición de los bloques que intervienen en la transformada coseno, como a la exploración en zig-zag de los coeficientes transformados y a los procedimientos de estimación de movimiento. Examinaremos todos estos aspectos detenidamente.

En la codificación de una imagen, la selección de los modos entrelazados o de los modos progresivos convencionales puede realizarse en cada macrobloque de forma individual. Con ello podemos obtener los beneficios de un mayor factor de compresión mediante el uso de técnicas progresivas en aquellas zonas de la imagen que permanecen estacionarias, mientras que en regiones con elevado movimiento horizontal podemos utilizar los modos entrelazados.

Un macrobloque convencional de 16x16 píxeles se descompone en 4 bloques de 8x8 píxeles tomando los píxeles adyacentes tal y como se muestra en la figura 4.73. Un macrobloque entrelazado también

se descomponen en 4 bloques de 8x8 píxeles, pero ahora los bloques se toman utilizando líneas alternadas. Es decir, en un macrobloque entrelazado, los bloques están formados por líneas que corresponden siempre al mismo campo mientras que en un macrobloque convencional las líneas se van alternando entre los dos campos. La figura 4.74 muestra claramente las diferencias partiendo de la imagen entrelazada que habíamos utilizado en la figura 4.72. En este ejemplo se ha utilizado un macrobloque que representa toda la imagen con objeto de visualizar con más claridad como deformadas las imágenes cuando los bloques se dividen por campos, con una menor resolución espacial en el sentido vertical.

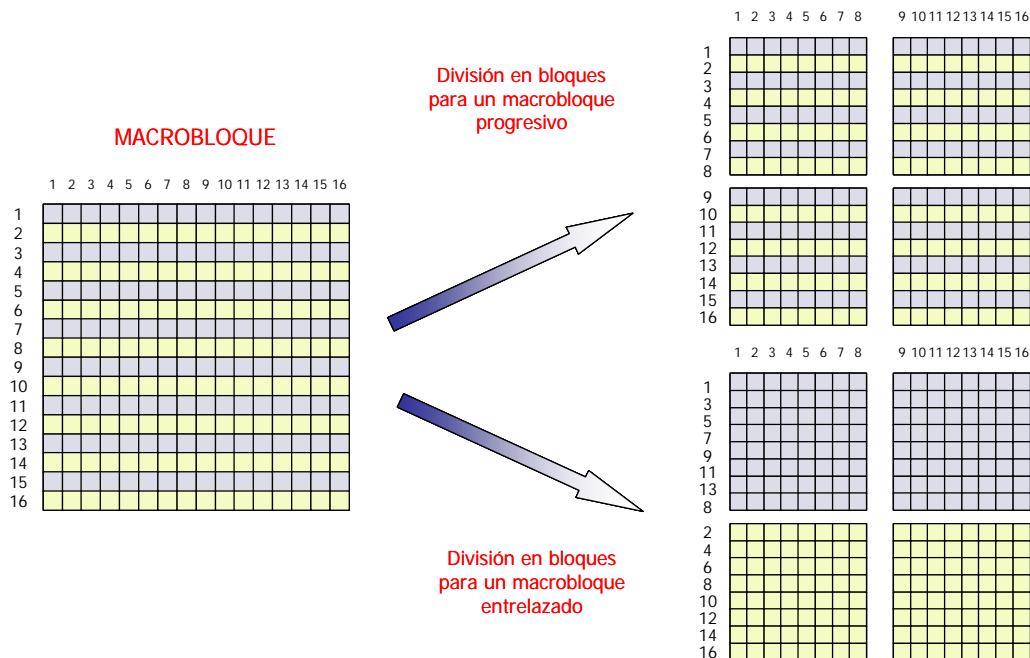
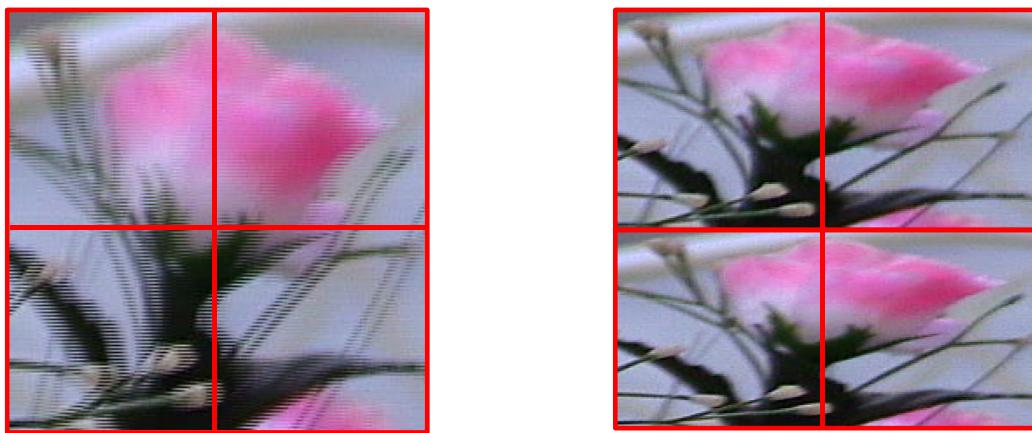


Fig. 4.73 División en bloques de un macrobloque para imágenes progresivas y entrelazadas

Cuando a un bloque del tipo entrelazado le aplicamos la transformada coseno debemos tener en cuenta las diferentes resoluciones espaciales en cada uno de los sentidos. La subimagen ha sido submuestreada en el eje vertical por lo que deberemos esperar un mayor contenido frecuencial en esta dirección. Esta circunstancia resulta evidente si tenemos en cuenta, que por el hecho de aumentar el espacio entre las muestras, los niveles de los píxeles cambiarán de forma más rápida y no estarán tan correlados en la dirección vertical. Estas consideraciones deben ser tenidas en cuenta al ordenar los coeficientes transformados. La ordenación en zigzag convencional presupone que las componentes frecuenciales son iguales en los dos ejes. Con este nuevo muestreo la situación ha cambiado y las frecuencias verticales son más importantes que las horizontales. Por ello es conveniente redefinir la ordenación de los coeficientes de forma que los primeros coeficientes sigan siendo los más significativos. La nueva ordenación se representa en la figura 4.75 y, como puede comprobarse, los coeficientes verticales y los horizontales van alternándose, aunque primando los primeros en las

primeras posiciones de la exploración. Esta curiosa forma de explorar los coeficientes suele conocerse en el entorno MPEG con el nombre de *Yeltsin Walk*.



Descomposición de un macrobloque progresivo en 4 bloques

Descomposición de un macrobloque entrelazado en 4 bloques

Fig. 4.74 Descomposición en bloques en el modo progresivo y el entrelazado

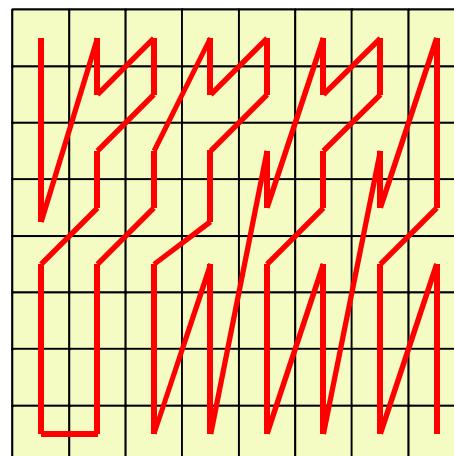


Fig. 4.75 Orden de exploración de los coeficientes transformados en bloques entrelazados

Por lo que respecta a la compensación de movimiento, también puede realizarse mediante campos separados o basándose en la imagen global. La forma más evidente de tratar con campos separados consiste en descomponer las imágenes de referencia en dos campos A y B y utilizar el campo A para predecir la parte del macrobloque del campo A y viceversa. En este caso, las predicciones son

independientes para cada uno de los campos que se tratan siempre por separado. No obstante, el MPEG-2 también admite otros tipos de predicción o compensación de movimiento, entre los que destacan la predicción de movimiento basada en frames (convencional) y la denominada compensación de movimiento 16x8, donde para cada macrobloque se determinan dos vectores de movimiento independientes que se determinan a partir de los dos campos de las imágenes de referencia. Existe una predicción alternativa denominada *dual frame*, en la que los vectores de movimiento se estiman tomando como referencia el promedio de los dos campos. Una vez determinado el vector de movimiento se realiza del vector de movimiento del campo A y del campo B suponiendo un modelo de movimiento lineal de la imagen. Esta estrategia ofrece excelentes resultados, pero sólo puede usarse cuando la secuencia MPEG sólo se compone de imágenes P e I.

Todas las observaciones relativas al tratamiento de imágenes entrelazadas que hemos descrito se refieren tan sólo a la componente de luminancia. Las componentes de color siguen un tratamiento parecido pero su organización concreta depende del tipo de formato de la señal de partida. Así, en un formato como el 4:2:2 en que las componentes de croma sólo están submuestreadas en la dirección horizontal, podemos definir las componentes de croma del macrobloque en función del campo par o impar de la misma forma en que se define para las componentes. En el formato 4:2:0, las señales diferencia de color están también submuestreadas en el sentido vertical, por lo que, en cada campo, cada dos líneas de luminancia deberán compartir una misma información de croma

4.9.14 Restricción de parámetros, niveles y perfiles

Los algoritmos MPEG proporcionan una gran flexibilidad en lo que se refiere a tamaños de imagen que pueden tratarse, frecuencias de muestreo de imagen, complejidad de los algoritmos, tasas de bits y anchos de banda del flujo de señal, etc. Todo ello indica que no tiene sentido que una aplicación sencilla debeat soportar los costes de procesado y tratamiento de señal que sólo se requieren para tratar las imágenes más complejas. Por ello, se definen un conjunto de grupos de compatibilidad distintos que deben cumplir los distintos codificadores y decodificadores en función de la aplicación a la que estén orientados.

En el MPEG-1 únicamente se define un posible nivel de compatibilidad, que se conoce con el nombre genérico de *restricción de parámetros* del MPEG-1. Todos los codificadores y decodificadores para aplicaciones multimedia dentro del flujo de datos de 1 a 3 Mbps deben tener este nivel de compatibilidad (el orden de magnitud para el que fue optimizado el MPEG-1). El estándar genérico admite sistemas mucho más complejos, pero prácticamente todos los productos comerciales se restringen a las características proporcionadas por este nivel. Las restricciones de parámetros del MPEG-1 son:

- a) El número de píxeles horizontales de la imagen debe poder llegar hasta 720.
- b) El número de píxeles verticales debe poder llegar hasta que 576.
- c) El número de macrobloques por segundo debe ser menor o igual que 9900 (396x25, 330x30).
- d) El número de macrobloques por imagen debe poder llegar hasta 396.
- e) La frecuencia de imagen debe poder llegar hasta 30 imágenes por segundo.
- f) La tasa de bits debe poder llegar hasta 1,86 Mbps.

- g) El tamaño del *buffer* del decodificador debe poder almacenar un mínimo de 376.832 bits.

Las dos primeras restricciones son un tanto confusas, pues parecen indicar que, en este nivel de compatibilidad, el MPEG-1 puede admitir resoluciones de imagen de calidad de radiodifusión. El objetivo de estas restricciones es que puedan admitirse codificadores con relaciones de aspecto muy rectangulares (panorámicas o alargadas). No es necesario que se cumplan las dos necesariamente. Las restricciones más importantes son la c) y la d), que establecen el número de macrobloques por segundo y el número de macrobloques por imagen que necesariamente deben proporcionarse. Según estos requisitos, cualquier codificador o decodificar que pretenda pasar este nivel de compatibilidad deberá ser capaz de codificar o decodificar imágenes de resolución SIF.

El mayor ámbito de aplicaciones a las que se dirige el MPEG-2 hace necesario que se definan varios grados de compatibilidad distintos. Los distintos grados de compatibilidad se estructuran en dos parámetros que se denominan *niveles* y *perfils*. El nivel hace referencia, fundamentalmente, al tamaño de las imágenes, mientras que el perfil establece las restricciones sobre los algoritmos de compresión. Están definidos un total de 4 niveles y 5 perfils. Los posibles grados de compatibilidad se obtienen combinando un nivel y un perfil, aunque no todas las combinaciones son posibles. En el gráfico de la figura 4.76 se muestran las distintas opciones definidas en este momento.

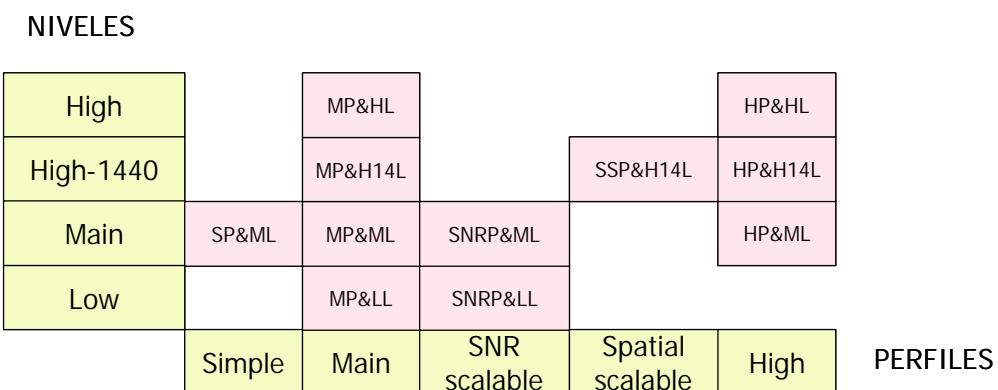


Fig. 4.76 Niveles y perfils definidos en MPEG-2

Los distintos grados de compatibilidad se obtienen combinando un nivel y un perfil. Así, con el perfil principal y el nivel principal se obtiene el MP&ML (Main Profile, Main Level) que en este caso concreto corresponde al estándar para la radiodifusión de televisión digital elegido por el DVB.

Los diferentes niveles tienen las siguientes características:

- a) *Low*. Se utiliza para imágenes de aproximadamente un cuarto de cuadrante del ITU-601, es decir, imágenes del tipo SIF como las utilizadas en el MPEG-1. La tasa máxima de bits es de 4 Mbps.

- b) *Main*. Se utiliza para imágenes del tipo ITU 601, es decir, 720x576 para 25 imágenes por segundo y 720x480 para 30 Hz. La tasa de codificación debe mantenerse por debajo de los 15 Mbps.
- c) *High-1440*. Este nivel está pensado para soportar imágenes de TV de alta definición (HDTV). Los tamaños de imagen son de 1440x1080 en 30 imágenes por segundo y de 1440x1152 en formatos de 25 imágenes por segundo. La tasa de bits máxima es de 60 Mbps que pueden aumentar hasta 80 Mbps en el perfil *High*.
- d) *High*. Está orientado a formatos de televisión de alta definición panorámicos. Los tamaños de las imágenes son de 1920x1080x30 fps y de 1920x1152x25fps. El número máximo de bits está limitado a 80 Mbps que pueden aumentar hasta los 100 Mbps si se usa el perfil *High*.

Los perfiles están definidos de acuerdo con las siguientes características:

- a) *Simple*. No requieren el uso de imágenes del tipo B. Esto permite simplificar tanto el codificador como el decodificador al coste de una mayor tasa de bits. El decodificador puede funcionar con una memoria de sólo 1 Mbit.
- b) *Main*. Es el mejor compromiso entre tasa de compresión y coste. Utiliza los tres tipos de imágenes (I, P y B). Este es el nivel más utilizado actualmente y al que hemos restringido el material proporcionado en este capítulo.
- c) *SNR scalable*. Se trata de una transmisión jerárquica que en principio está pensada para usos futuros. La secuencia se transmite en dos flujos de datos. Uno de ellos (*base layer*) proporciona imágenes de calidad aceptable al receptor mientras que la segunda (*enhanced layer*) proporciona una resolución adicional sobre los coeficientes de la DCT, de forma que cuando ambas secuencias de bits son decodificadas simultáneamente se obtienen unos resultados de reconstrucción excelentes. Este tipo de escalabilidad está pensado para que decodificadores con distintas complejidades y costes puedan acceder a la misma secuencia.
- d) *Spatial scalable*. La idea de fondo es similar a la de la escalabilidad en SNR, aunque ahora una de las secuencias aporta información sobre la secuencia de vídeo con una resolución espacial convencional, mientras que la otra aporta un incremento de la resolución.
- e) *High*. Está orientado a la radiodifusión de emisiones de televisión digital de alta definición tanto en el formato 4:2:0 como en el 4:2:2. Admite tanto la escalabilidad espacial como la de SNR junto con otros tipos de escalabilidad en la velocidad de transmisión de las secuencias. En principio, este perfil está pensado para hacer compatibles las difusiones de televisión de alta definición en distintos tipos de decodificadores, tanto de definición convencional como de alta definición. La escalabilidad en velocidad de transmisión puede utilizarse para proporcionar un nivel básico de recepción en las zonas donde las señales se reciben con dificultad y con una calidad mejorada en áreas con mejores condiciones de recepción. Esta última característica puede ser de gran interés en la transmisión de televisión digital terrestre.

Los distintos perfiles tienen un nivel de compatibilidad ascendente, es decir, un decodificador de un determinado perfil debe ser capaz de decodificar el perfil que le corresponde y todos los perfiles de menor complejidad. La combinación que se utiliza en la mayoría de aplicaciones actuales es el

MP@ML (*Main Profile, Main Level*), que permite la codificación de imágenes entrelazadas en el formato 4:2:0, utilizando tasas de bits que en función de los contenidos y la calidad de las imágenes pueden estar situadas entre los 4 Mbps y los 9 Mbps. El MP@ML ha sido elegido por el DVB como sistema básico para la radiodifusión de señales de televisión digital. También es, con algunas diferencias de parámetros mínimas, el grado de compatibilidad que se emplea en el almacenamiento de vídeo digital utilizado por el DVD-Vídeo.

4.10 Multiplexación de secuencias MPEG

En los apartados anteriores hemos analizado los principios generales del proceso de compresión de la secuencia de vídeo digital en una trama de datos digital. Hemos visto que la trama de vídeo básica tiene una estructura jerárquica en la que se van definiendo progresivamente los parámetros de la secuencia de vídeo, del grupo de imágenes, de cada una de las imágenes, los macrobloques y los tipos usados en la codificación y finalmente se proporcionan los coeficientes transformados. Esta trama de datos se conoce como la trama elemental (*Elementary Stream - ES*) y aporta toda la información relativa a un programa de vídeo. La estructura básica de una ES en el MPEG se representa de forma muy simplificada en la figura 4.77.

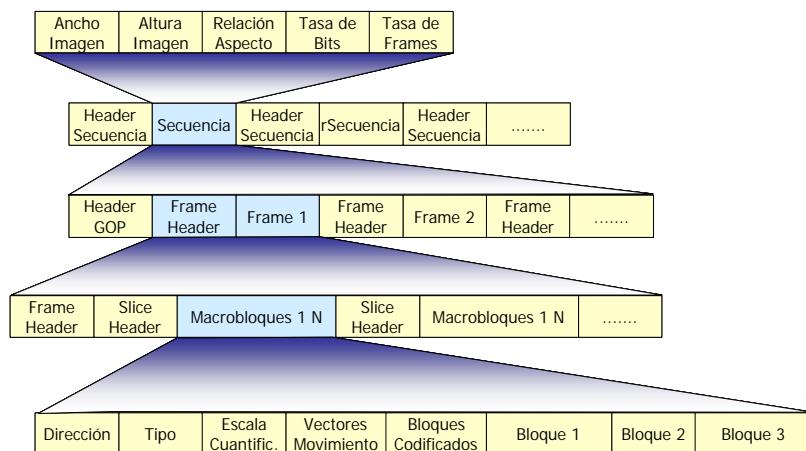


Fig. 4.77 Diagrama simplificado de la Elementary Stream

La ES sólo es el primer nivel de señalización entre el codificador y el decodificador. Contiene toda la información relativa a cómo se ha realizado la codificación del vídeo, pero carece de información detallada de cómo sincronizar el vídeo con el audio y con los datos adicionales. Además, la ES es una trama de bits continua, que sólo aporta información sobre las imágenes de un determinado programa y que, por lo tanto, debe intercalarse con datos de audio (ES Audio, *Elementary Stream Audio*), datos de usuario (ES Data, *Elementary Stream Data*) y, en el caso de radiodifusión digital, también deberá multiplexarse con otros programas. Toda la problemática de intercalar y sincronizar audio, vídeo y datos correspondientes a distintos programas en una única trama de datos se considera en el estándar *MPEG-Systems* y se conoce como el problema de la multiplexación de tramas en MPEG. Las tramas

de datos que finalmente se obtienen están organizadas en diferentes niveles que permiten que el procesador del receptor pueda identificar los datos de audio y de vídeo asociados a los distintos programas y decodificarlos. En este apartado veremos los conceptos generales de la multiplexación de secuencias MPEG. Como siempre, la presentación sólo pretende proporcionar una idea general de la filosofía utilizada, dejando los detalles para estudios más avanzados que pueden encontrarse en los textos proporcionados en la bibliografía o en el propio estándar MPEG.

Existen dos posibles alternativas para construir la secuencia MPEG multiplexada, conocidas como la trama de programa (*Program Stream*) y la trama de transporte (*Transport Stream*). Las dos posibilidades permiten multiplexar video, audio y datos sincronizando los decodificadores de las distintas fuentes. Las diferencias entre ambas alternativas se deben a que la trama de programa está orientada a aplicaciones de almacenamiento de vídeo en las que se producen pocos errores de lectura, mientras que la trama de transporte se orienta a sistemas de comunicaciones con lo que las condiciones de recepción de la trama pueden llegar a variar considerablemente, produciendo ráfagas de errores significativas. La trama de programa está compuesta por paquetes de gran longitud de datos mientras que en la de transporte los paquetes son de longitud pequeña para poder realizar un control de errores más eficaz. En el MPEG-1 sólo está definida la trama de programa, ya que todas las aplicaciones son para el almacenamiento de vídeo en CD-ROM o en disco duro. El MPEG-2 dispone de las dos opciones, la trama de programa se utiliza en aplicaciones como el DVD-Vídeo, mientras que la trama de transporte se usa en la radiodifusión de señales de televisión. Debemos señalar que en ninguno de los dos casos se incluyen códigos de control o corrección de errores. La inclusión de estos códigos se realiza en otras capas del sistema de comunicaciones y no están reguladas por el MPEG. En televisión digital, la inclusión de redundancia para la protección de errores está especificada por el propio sistema DVB del que hablaremos más adelante.

4.10.1 Estructura del paquete básico (*Packetized Elementary Stream*)

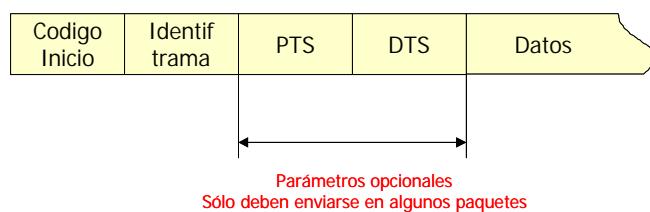


Fig. 4.78 Estructura simplificada de la Packetized Elementary Stream

Antes de llegar a la trama de programa o de transporte la ES, se agrupa en paquetes elementales que convierten el flujo continuo de información de la ES en bloques discretos, que permiten un mejor control de transmisión o almacenamiento. La agrupación en paquetes de la ES se denomina PES (*Packetized Elementary Stream*) y se define para cada uno de los posibles tipos de señal, es decir, vídeo, audio y datos. La estructura simplificada de una PES se representa en la figura 4.78 y se compone de una cabecera que contiene un código de inicio de paquete y un código que identifica si la información del paquete corresponde a señal de vídeo, audio o datos. Los identificadores PTS y DTS son opcionales y por lo general no se encuentran en todos los paquetes. No obstante, son

extremadamente importantes puesto que se utilizan como unidades de sincronización entre la información de vídeo y de audio. Sin ellos, sería imposible obtener un buen sincronismo entre el sonido y movimiento de los labios de la fuente de vídeo.

El PTS es una referencia temporal del orden de presentación de los fotogramas (*Presentation Time Stamp*) mientras que el DTS es una referencia del orden de decodificación. Cuando no se transmiten imágenes del tipo B o cuando se trata de PES de audio o datos, ambos índices coinciden. La información que contienen es un número entero de 33 bits que indica el estado de un contador de pulsos de reloj del codificador. El reloj del codificador es de 27 MHz y se divide por un factor de 300 para atacar un contador binario que reinicia la cuenta a cero cuando se desborda. Los índices enviados en el PTS y el DTS son los valores de este contador.

Los valores del PTS y del DTS no requieren enviarse en cada paquete por cuanto existe bastante redundancia en la recepción de las imágenes debido que el receptor sabe que la diferencia de tiempos entre todas ellas es constante (el periodo de imagen). Cuando los PES deben incorporarse en una trama de programa, los PTS y DTS deben de enviarse con un periodo no superior a los 0,7 segundos. En cambio, cuando se trata de una trama de transporte, la tasa de transmisión no puede ser inferior a los 0,1 s.

El esquema de la figura 4.78 es sólo una simplificación de la estructura de un paquete PES. En la figura 4.79 se muestra el contenido algo más detallado con los diferentes campos opcionales. La figura corresponde al contenido de los paquetes PES para el MPEG-2, que varía ligeramente respecto al MPEG-1. El código de inicio es una palabra fija de 24 bits que toma el valor 00 00 01 en hexadecimal. La siguiente palabra de 8 bits es un identificador de la trama. Los 4 primeros bits se utilizan para indicar si se trata de una trama de vídeo, audio o datos. Los 4 bits menos significativos asignan un número entero para identificar cada una de las secuencias.

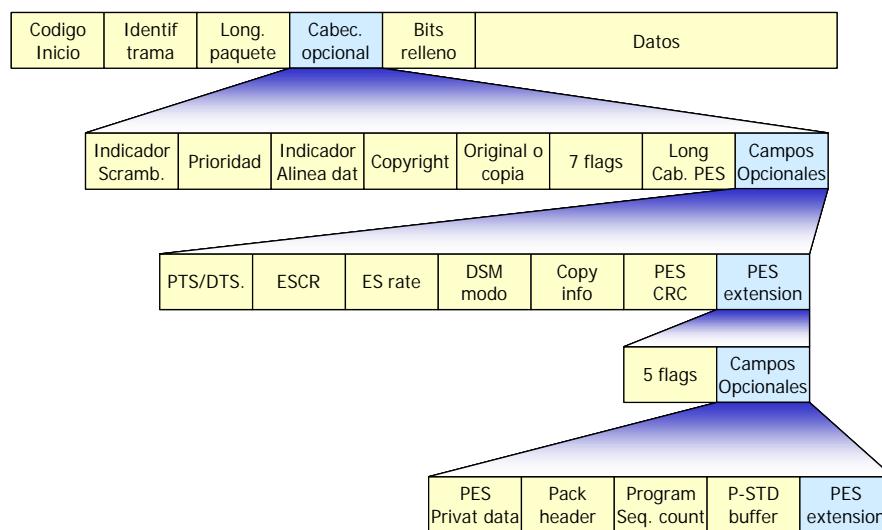


Fig. 4.79 Detalle del contenido de una trama PES para MPEG-2

El tamaño total del paquete es variable y depende de la aplicación a la que se destina. La longitud del paquete se indica en la cabecera y puede llegar hasta un máximo de 64 kBytes (16 bits que indican la longitud en bytes del paquete). A continuación se envía una cabecera opcional que contiene los códigos PTS y DTS, a los que ya hemos hecho referencia. Esta cabecera también contiene otro tipo de datos, como el indicativo del acceso condicional, la prioridad de la trama, el *copyright*, contadores de secuencia del paquete, etc. El estándar también prevé la posibilidad de añadir hasta un total de 16 bytes de relleno (todos los bytes toman el valor hexadecimal FF). Estos bytes pueden usarse para que la longitud total del paquete sea un múltiplo exacto de los sectores físicos del soporte de almacenamiento.

4.10.2 La trama de programa (*Program Stream*)

Esencialmente, la trama de programa se compone de una multiplexación de varias PES que se estructuran en unos nuevos paquetes (*packs*) de nivel superior. La figura 4.80 muestra el contenido básico de un *pack* de la trama de programa. La cabecera del *pack* contiene una palabra identificativa del inicio de cabecera, la tasa de multiplexación y una referencia del reloj del sistema. Discutiremos el procedimiento con el que se recupera la referencia del reloj del sistema cuando consideremos la trama de transporte, ya que en ambos casos la idea es la misma. La cabecera del sistema (*system header*) proporciona información general sobre el número y tipo de secuencias PES que componen el paquete. La finalización del paquete se indica con un código específico.

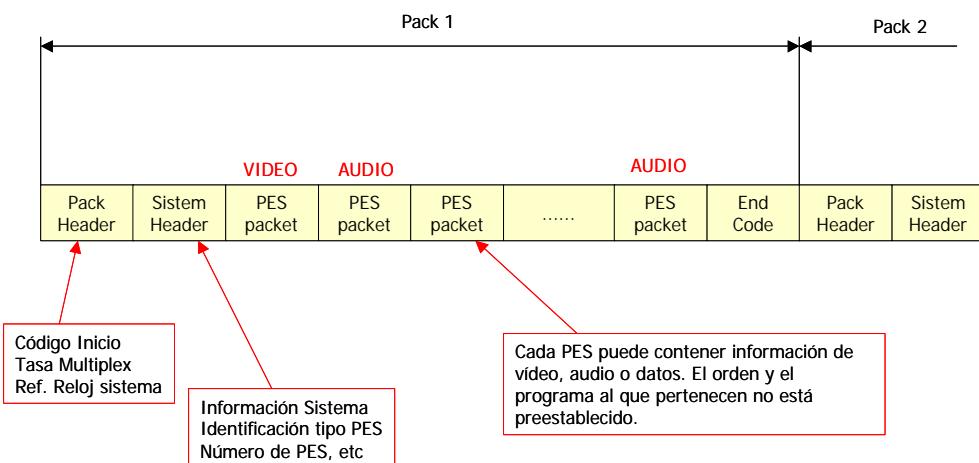


Fig. 4.80 Diagrama básico de la estructura de la trama de programa

4.10.3 La trama de transporte

La multiplexación de varias señales y programas para la radiodifusión de programas de televisión digital exige que los paquetes en los que se estructura la información sean de corta duración. Con ello se facilita la tarea de insertar códigos de protección de errores estrictamente necesarios debido a que las tasas de error típicas con las que se trabaja son del orden de 10^{-4} o superiores. Los paquetes

utilizados en la trama de transporte son de una longitud fija de 188 bytes y contienen una cabecera y un fragmento de datos de una trama PES. Es importante subrayar que aunque los paquetes de transporte son más cortos que las tramas PES (por ejemplo de 2048 bytes), se utilizan para transportar toda la información de la PES de audio, vídeo y datos de varios programas.

Para aumentar la eficiencia de la transferencia de información pueden utilizarse dos tipos de cabecera. La que se usa normalmente es de una longitud muy corta (32 bits), aunque puede extenderse ocasionalmente en determinados paquetes para proporcionar información adicional. En el caso de utilizar la cabecera extendida, se reduce la carga de datos PES del paquete, manteniendo siempre su longitud total constante. La estructura básica de los paquetes de la trama de transporte se ilustra en las figuras 4.81 (estructura de cabecera más datos) y 4.82 (información de cabecera).

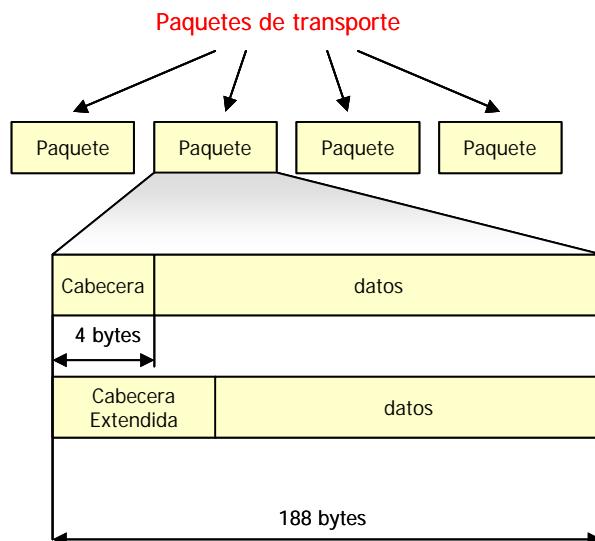


Fig. 4.81 Estructura de los paquetes de transporte

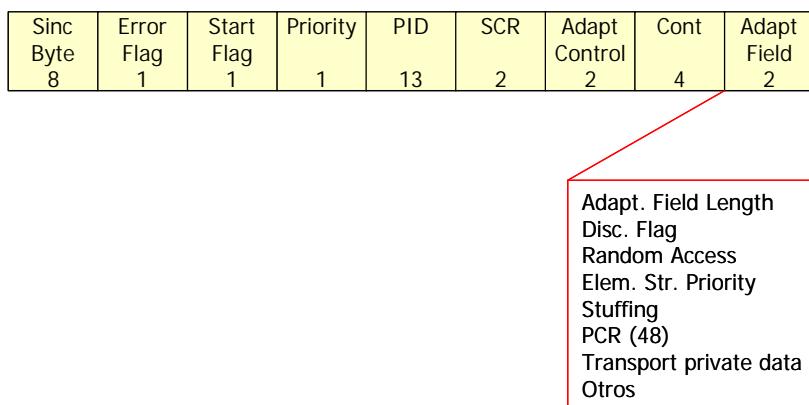


Fig. 4.82 Detalle de la información de cabecera de los paquetes de transporte

El byte de inicio de paquete contiene el código hexadecimal 47. El *flag* de error se activa cuando se ha producido un error en las etapas previas. El *flag* de inicio se utiliza para indicar que este paquete contiene el inicio de una trama PES. La prioridad de la trama de transporte únicamente admite los valores alta y baja (1 bit). En la trama de transporte están multiplexados los PES de varios programas y para cada uno de ellos, los PES asociados al vídeo, audio y datos. El PID (*Packet IDentification code*) es un código de 13 bits que se utiliza para identificar cada una de las secuencias PES que componen la trama de transporte. Un demultiplexor deberá examinar este código en las cabeceras de entrada y proporcionar al decodificador todos los paquetes que correspondan a la trama PES que se pretende decodificar. Para ayudar al decodificador se incluye un código de contador secuencial (cont) de 4 bits que se incrementa para cada nuevo paquete con un mismo PID.

Esta filosofía de marcar el contenido de los paquetes permite que los multiplexadores trabajen de forma estadística, asignando un mayor o menor flujo de datos a cada canal en función de los contenidos. Hemos visto que por la propia naturaleza de compresión del MPEG es posible que para mantener la calidad estable sea necesario aumentar la tasa de bits en algunas secuencias con movimiento elevado. Esto significa que es difícil mantener la tasa de bits en las tramas ES o las PES. Un multiplexor estadístico puede examinar las necesidades de los distintos programas y asignar una mayor frecuencia de paquetes a las PES que así lo requieran. Para mantener la trama de transporte con un flujo de datos constante se suelen utilizar paquetes de relleno, que no contienen información útil. Los paquetes de relleno se identifican con el código PID 8191 y son directamente descartados por el demultiplexor cuando los identifica.

Cuando se utiliza la cabecera extendida, se proporcionan distintos tipos de información adicional sobre la trama de transporte y las secuencias PES que la forman. Entre las distintas opciones destacaremos el código PCR (*Program Clock Reference*) y el *Stuffing* (Relleno). El relleno se utiliza en los paquetes que se corresponden con el final de una trama PES. Hemos comentado que la trama PES se transmite en varios paquetes del nivel de transporte por lo que, normalmente, el último fragmento de la PES no coincidirá exactamente con el tamaño de datos del paquete de transporte. Para solventar este problema se utilizan los bits de relleno que sean necesarios hasta el paquete coincida con el final de la trama PES. El siguiente paquete correspondiente a este mismo código PID indicará el inicio de un nuevo PES en uno de los bits de cabecera.

Por otra parte, al compartir varios programas en una misma trama, los codificadores no estarán sincronizados, por lo que es importante proporcionar algún mecanismo de control que permita que el receptor se sincronice con el procesador que está codificando los datos correspondientes al programa que tiene sintonizado. La sincronía entre ambos se establece mediante un código binario de 48 bits denominado PCR y que se corresponde al valor de un contador que está directamente conectado al reloj de 27 MHz del codificador. El valor del contador se incrementa en cada ciclo de reloj y se reinicializa a cero cada vez que llega al final de la cuenta. Este valor se va transmitiendo periódicamente en los paquetes de transporte y es utilizado por el decodificador para sincronizarse con los datos recibidos. El mecanismo de sincronización que se utiliza es parecido al de un PLL, pero implementado de forma digital. El receptor dispone de un oscilador interno de 27 MHz que se controla a través de un VCO. El reloj interno activa un contador cuyo valor se compara con el valor recibido en el paquete. La diferencia entre el contador interno y el valor recibido se utiliza para controlar el VCO y poner en fase ambos subsistemas. Evidentemente, la precisión en la sincronía entre ambos sistemas está limitada por el *jitter* y la estabilidad con la que se reciban los códigos PCR.

4.10.4 Información específica de programas (*Program Specific Information -PSI*)

Como varios programas comparten la misma trama de transporte, debe de proporcionarse algún mecanismo que permita que el receptor pueda saber qué PES de vídeo, audio y datos están asociadas a cada uno de los programas. El código de identificación PID permite discriminar las distintas tramas pero no identifica la asociación entre ellas. Para solventar este problema se han incorporado las tramas de información PSI (*Program Specific Information*) que se transmiten en los paquetes de transporte con un código PID igual a cero.

Cuando el receptor se conecta no conoce la distribución de los programas en la trama de transporte y debe esperar a recibir tramas con el código de identificación de PSI (PID = 0) con las que será posible construir una tabla de asignación de programas (*Program Asignation Table*, PAT). La PAT se transmite de forma periódica y contiene los códigos PID en los que se encuentra el mapa de informaciones para cada programa (*Program Map Table*, PMT). La estructura general es un tanto compleja y se comprende mejor con el ejemplo de la figura 4.83.

La trama con el PID igual a cero informa que los canales asociados al programa 1 pueden encontrarse en los paquetes que tengan un PID igual a 25, los asociados al programa 2 en los paquetes que tengan un PID igual a 32 y así sucesivamente para todos los programas de la trama de transporte. Si el usuario quiere decodificar el programa 2 deberá esperar a recibir paquetes de transporte con el código de identificación 32. En estos paquetes se proporciona la PMT que indica en qué tramas pueden encontrarse la PES de vídeo, audio y datos. En nuestro ejemplo, el programa número 2 tiene una PES de vídeo, que se identifica con el PID 49; 4 PES de audio con los PID's 51, 55, 56 y 57 y una de datos (PID=72). Las PES de audio pueden corresponder a transmisiones estereofónicas en dos idiomas distintos.

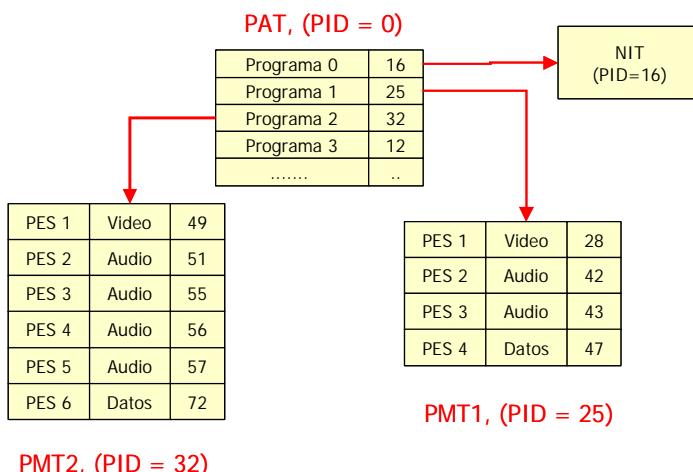


Fig. 4.83 Tablas de información sobre referencias de programa

Además de las tablas PAT y PMT descritas se envían, usando procedimientos parecidos, otros tipos de informaciones auxiliares como la CAT (*conditional access table*) que define los parámetros y

condiciones para el acceso condicional. El identificador que se utiliza para transmitir esta información es el PID 0x0001 y debe estar presente siempre que al menos uno de los programas de la trama de transporte disponga de acceso condicional.

El estándar MPEG permite que se incorporen informaciones de programa y gestión cuyo contenido puede establecerse en función de la aplicación a la que se destina la trama de datos. Así, el estándar de transmisión de televisión digital DVB introduce algunas informaciones adicionales para gestionar la configuración automática del receptor y para proporcionar información adicional a los usuarios sobre los contenidos de la programación, facilitando la navegación a través de los distintos programas que componen la trama MPEG. Los contenidos de estas informaciones complementarias no son regulados por el estándar MPEG sino por la aplicación que los define. En el caso de la transmisión de señales de televisión en formato digital se conocen con el nombre de tablas DVB-SI (*service information*) y son definidas por el propio estándar DVB. Las tablas complementarias se transmiten usando procedimientos análogos a los descritos para las PSI. Los servicios utilizados por el DVB se componen de cuatro tablas básicas más tres tablas opcionales. Los contenidos de las tablas básicas se describen a continuación:

- a) *Tabla de información de red (NIT - Network Information Table)*. Se proporciona toda la información relativa a la red. El PID en el que se transmite esta información es, por definición, el número 0. Los datos proporcionados se utilizan en sistemas por vía satélite para informar sobre detalles de la transmisión, como por ejemplo la posición orbital del satélite, el tipo de polarización empleado, la frecuencia, el tipo de modulación, etc. Esta información puede ser utilizada por el IRD o el Set-Top-Box para controlar la posición de los satélites y conmutar entre distintas tramas de transporte del mismo servicio. En el caso de transmisiones terrenas puede informar de los distintos canales de RF que componen una misma red. En este caso, el proveedor del servicio puede utilizar varios canales de transmisión (varias tramas de transporte) para la difusión de todos los canales. Con esta información, el receptor puede disponer de información de antemano de los canales RF en los que se encuentra cada programa.
- b) *Tabla de descripción de servicio (SDT – Service Description Table)*. En esta tabla se listan los nombres de los distintos servicios usados en la trama de transporte.
- c) *Tabla de información de sucesos (EIT – Event Information Table)*. Se proporciona información sobre los programas y acontecimientos que se transmiten en la trama de transporte asociada u otras tramas de la misma red.
- d) *Tabla de hora y fecha (TDT – Time Date Table)*. Se usa para actualizar el reloj del *set-top box*.

4.10.5 Multiplexado de paquetes elementales en la trama de transporte

Uno de los subsistemas más sofisticados que intervienen en el proceso de transmisión de señales de televisión digital es el multiplexor que se encarga de asignar los paquetes elementales (PES) de cada una de las fuentes de audio, vídeo o datos a la trama de transporte. El problema radica en que la trama de transporte debe tener un flujo de datos constante mientras que los factores de compresión utilizados en cada programa pueden variar en función del contenido. En la figura 4.84 se muestra un diagrama de bloques en el que se incluyen los distintos aspectos que pueden incidir en la multiplexación de los

paquetes elementales para formar la trama de transporte. Los paquetes elementales asociados a una determinada fuente de datos se insertan en un *buffer* de memoria de donde deben ser periódicamente extraídas para enviarse a través de la trama de transporte. Obsérvese que si, debido a su contenido, una fuente de vídeo requiere un flujo de datos elevado deberá ser leída con mayor frecuencia que el resto de fuentes. En el proceso de multiplexación también debe tenerse en cuenta que deben enviarse códigos adicionales como el PCR o las informaciones específicas PSI.

El multiplexor debe garantizar que tanto los *buffers* de los codificadores como el *buffer* integrado en el receptor se mantienen alejados del estado de desbordamiento. Para ello, es necesario que los PES se distribuyan en la trama de transporte de la forma más dispersa posible, espaciando al máximo la transmisión entre paquetes sucesivos. Con ello, se evita la aparición de grandes ráfagas de paquetes asignadas a un mismo programa que podrían saturar el *buffer* del receptor. Cuando las distintas fuentes tienen tasas de bits distintas, debe tenerse en cuenta el estado de ocupación de los *buffers* del transmisor, que pueden usarse como un indicador de la probabilidad de transmitir los PES asociados a cada una de las fuentes. Así, cuando el estado de ocupación de un buffer está próximo al desbordamiento, es necesario que se envíen los paquetes al canal con una mayor frecuencia que cuando está prácticamente vacío. Mediante esta técnica puede permitirse que las fuentes modifiquen en tiempo real sus tasas de transmisión en función del contenido. Si un codificador de fuente detecta que los errores de predicción aumentan puede incrementar su tasa de bits. Con ello, el *buffer* asociado se llenará más rápidamente y el multiplexor deberá asignar paquetes elementales con mayor frecuencia a la trama de transporte.

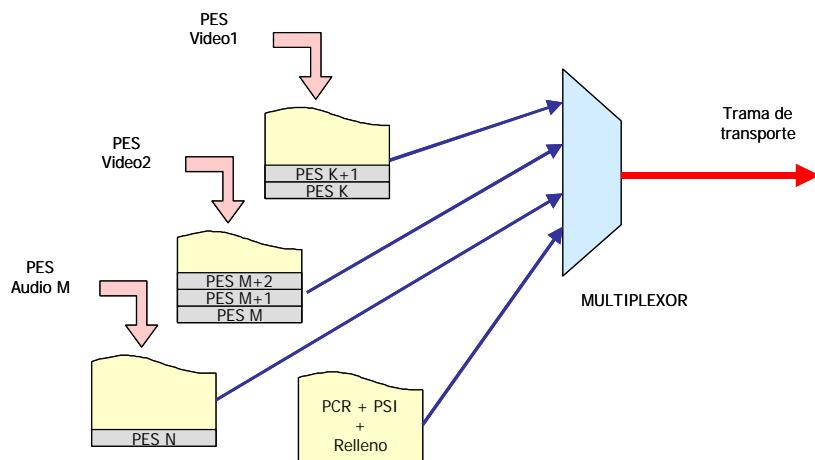


Fig. 4.84 Diagrama genérico del proceso de multiplexación de paquetes en la trama de transporte

Evidentemente, es necesario que la suma de todas las tasas de transmisión de las fuentes más las informaciones propias de sistema sea inferior a la carga admitida por la trama de transporte. En el supuesto que los contenidos de los programas no exijan toda la carga disponible en la trama de transporte, el multiplexor puede aumentar las tasas de transmisión de códigos PCR o de informaciones

PSI. También pueden enviarse paquetes de relleno adicionales en los que no se incorpora ninguna información significativa para los decodificadores.

Otro elemento de especial importancia en la distribución de programas es el *transmultiplexor*. En la práctica, un creador de programas puede proporcionar una o varias tramas de transporte en las que se incluyen todos los programas que distribuye. No obstante, un proveedor de servicios puede desear seleccionar sólo algunos programas de distintos creadores para formar nuevas tramas de transporte con los programas seleccionados. La gestión de recombinar los contenidos de varias tramas de transporte en tramas nuevas es la que se encarga de realizar el transmultiplexor. Su principio de funcionamiento, una vez se han restaurado los PES deseados en los respectivos *buffers* de memoria, es parecido al del multiplexor.

4.11 Principios del DVB

El DVB (*Digital Video Broadcasting*) es un organismo encargado de regular y proponer los procedimientos para la transmisión de señales de televisión digitales compatibles. Está constituido por más de 220 instituciones y empresas de todo el mundo y los estándares propuestos han sido ampliamente aceptados en Europa y casi todos los continentes, con la excepción de Estados Unidos y Japón donde coexisten con otros sistemas propietarios. Todos los procedimientos de codificación de las fuentes de vídeo y audio están basados en los estándares definidos por MPEG. No obstante, hemos visto que los estándares MPEG sólo cubren los aspectos y metodologías utilizados en la compresión de las señales de audio y vídeo y los procedimientos de multiplexación y sincronización de estas señales en tramas de programa o de transporte. Una vez definida la trama de transporte es necesario definir los sistemas de modulación de señal que se utilizarán para los distintos tipos de radiodifusión (satélite, cable y terrena), los tipos de códigos de protección frente a errores y los mecanismos de acceso condicional a los servicios y programas.

El DVB ha elaborado distintos estándares en función de las características del sistema de radiodifusión. Los estándares más ampliamente utilizados en la actualidad son el DVB-S y el DVB-C que contemplan las transmisiones de señales de televisión digital mediante redes de distribución por satélite y cable respectivamente. La transmisión de televisión digital a través de redes de distribución terrestres utilizando los canales VHF convencionales se contempla en el estándar DVB-T, que actualmente está en pruebas en distintos países y cuya implantación parece que será inmediata. Además de estos estándares también están especificados sistemas para la distribución de señales de televisión digital en redes multipunto, sistemas SMATV (*Satellite Master Antenna Televisión*). También existen estándares que definen las características de la señalización en el canal de retorno en sistemas de televisión interactiva, la estructura de transmisión de datos para la encriptación y desencriptación de programas de acceso condicional, la transmisión de subtítulos, y la radiodifusión de datos (nuevos canales de teletexto) mediante sistemas digitales.

En este apartado revisaremos brevemente las características fundamentales de los estándares para la radiodifusión de señales de televisión por vía satélite, cable y terrena. Las diferencias fundamentales entre estos tres sistemas radican en el tipo de modulación digital utilizado y en los códigos de protección frente a errores incorporados a la trama de transporte. Estas modulaciones y códigos

protectores están adecuados a las características del sistema de comunicación por lo que son tanto más robustos cuanto más sensible es el sistema al ruido.

4.11.1 Estándar DVB-S

El estándar para la transmisión de televisión digital por satélite es, sin duda, el más ampliamente utilizado y, probablemente, su éxito, ha sido el detonante de la progresiva implantación del resto de sistemas DVB. Actualmente está siendo utilizado por proveedores de servicios en todos los continentes. En Europa es el único sistema de televisión digital por satélite implantado.

Podemos considerar que el sistema DVB-S parte de la trama de transporte proporcionada por el MPEG-2, introduciendo distintas capas de protección a la señal para adecuarla a las características del canal por el que debe transmitirse. Las etapas sucesivas en las que se introducen nuevas características a la trama de transporte se resumen a continuación:

- a) *Inversión de los bits de sincronismo* en uno de cada ocho paquetes de la trama de transporte. Cada paquete de la trama de transporte es de 188 bytes, lo que significa que se la inversión de signo en los bits de sincronismo se repite cada 1504 bytes.
- b) *Inserción de un código aleatorio* a la trama resultante. La adición de éste código pretende garantizar que las características estadísticas de los datos sean prácticamente aleatorias. La aleatorización se obtiene realizando una suma *OR* exclusiva entre la secuencia de datos y una secuencia obtenida mediante un generador por registros de desplazamiento. La secuencia aleatoria se reinicializa cada 8 paquetes de la trama de transporte.
- c) *Adición de un código de detección y corrección de errores de Reed-Solomon*. Este código se denomina código externo y es común en todos los estándares del DVB. Introduce 8 bytes de redundancia para cada paquete de 188 bytes.
- d) Aplicación de un *entrelazado convolucional (Fournay)* cuyo objetivo es dispersar las ráfagas de errores de canal. De este modo, si se produce una ráfaga de errores, debida a un desvanecimiento del canal, los errores afectarán a paquetes distintos y, probablemente, podrán eliminarse usando las propiedades correctoras de los códigos interno y externo.
- e) *Inserción de un segundo código protector de errores*. Este código recibe el nombre de *código interno* y es de naturaleza convolucional. El grado de redundancia que introduce éste código no está fijado de antemano y puede configurarlo el proveedor del servicio para adaptarse a las características del sistema que deseé utilizar (potencia de transmisión, tamaño de las antenas transmisoras y receptoras, tasa de datos disponible, etc.).
- f) *Modulación de la portadora mediante QPSK (Quadrature Phase Shift Keying)*.

En el receptor deben realizarse las operaciones inversas para tener acceso a la trama de transporte que soporta la información de los programas de televisión en formato MPEG-2. El estándar DVB-S proporciona suficiente flexibilidad como para que el operador del servicio pueda decidir el contenido de los programas de televisión que proporciona. Por tanto, es posible que se transmitan señales con definición mejorada, de alta definición o de definición convencional (ML@MP). En función de las características del enlace puede seleccionarse el código de protección de errores interno. En transmisiones con un ancho de banda de 36 MHz es habitual utilizar códigos internos 3/4 (3 bytes de datos, 1 de redundancia), con lo que se consiguen tasas de datos de unos 39 Mbps.

4.11.2 DVB-C

El estándar de transmisión de televisión digital por cable está basado en el DVB-S y comparte muchos aspectos comunes. Las diferencias fundamentales entre ambos estándares es que en la transmisión por cable se sustituye la modulación QPSK por una modulación 64-QAM y que no se utilizan los códigos de corrección internos. La supresión de estos códigos de protección se debe a que la relación señal a ruido que se obtiene en un sistema por cable es muy superior a la de un sistema por satélite de modo que, es suficiente con los códigos externos para garantizar una correcta recepción, exenta de errores, de los datos.

Para canales de 8MHz de ancho de banda (equivalentes a los canales cable analógicos) puede obtenerse una capacidad de unos 38.5 Mbps (usando modulaciones del tipo 64 QAM). El estándar también contempla el uso de modulaciones digitales con tasas de datos superiores 128 QAM y 256 QAM así como modulaciones con menores tasas (16 QAM, 32 QAM).

4.11.3 DVB-T

El estándar DVB-T comparte el mismo proceso de aleatorización, protección externa e interna de datos y códigos convolucionales de entrelazado que el DVB-S. La diferencia fundamental es que, en este caso, se utiliza una modulación COFDM (*Coded Orthogonal Frequency Division Multiplexing*) con un total de 1705 (modo 2K) o 8817 portadoras (modo 8K). Esta modulación permite, mediante el uso de múltiples portadoras que dispersan los datos de la trama a transmitir, operar en escenarios con un elevado índice de señal multicamino. La protección frente a este tipo de interferencias se consigue insertando intervalos de guarda entre los datos que, eventualmente, reducen la capacidad del canal. Esta reducción de la capacidad es tanto menor cuanto mayor es el número de portadoras utilizadas. No obstante, el aumento del número de portadoras incrementa la complejidad del receptor.

El modo 2K está pensado para transmisiones simples que cubran áreas geográficas reducidas (potencias reducidas) mientras que el modo 8K puede usarse para áreas geográficas extensas usando una única frecuencia portadora por canal, común en toda la red. La fuerte protección del COFDM permite que el sistema pueda operar manteniendo la misma frecuencia portadora en toda una región geográfica extensa (cubierta por varios radioenlaces). El receptor interpreta la señal procedente del radioenlace más débil como una señal multicamino y puede rechazarla. Si los radioenlaces están muy alejados, las diferencias de tiempo de recepción pueden ser considerables y es necesario aumentar los intervalos de guarda entre los datos digitales, perdiendo cierta eficiencia en la transmisión.

5 Cámaras

5.1 Introducción

La cámara constituye uno de los elementos más importantes en cualquier sistema de comunicación visual. Su función básica es convertir la escena tridimensional en una señal temporal, analógica o digital, que pueda ser tratada, almacenada en soporte óptico o magnético, o modulada para transmitirse a un receptor remoto. En este capítulo consideraremos varios aspectos relacionados con la cámara y los elementos que la constituyen. Aunque la presentación está orientada fundamentalmente a cámaras de vídeo, también consideraremos con cierto detalle los principios de funcionamiento de la cámara fotográfica y se proporcionaran algunos ejemplos de cámaras orientadas a aplicaciones industriales de reconocimiento de imágenes.

La primera parte del capítulo se centra en el estudio del sistema de formación de imágenes o grupo óptico. El objetivo es introducir los conceptos de enfoque, variación de la distancia focal, ángulo de visión, profundidad de campo de la imagen, etc., y aprender cómo interaccionan todos ellos en la captación de imágenes. Para ello, se utiliza un modelo de lente muy simplificado, cuya limitación es que no permite realizar cálculos avanzados sobre aberraciones de imagen o posiciones relativas entre varias lentes, pero que resulta eficiente para comprender la incidencia de los distintos componentes del grupo óptico en la imagen final. La presentación de estos conceptos se centra en el contexto de una cámara fotográfica del tipo *reflex* y se particulariza a cámaras de vídeo en aquellos puntos donde aparecen diferencias entre ambos equipos.

En la siguiente sección se definen las unidades fotométricas elementales como la *candela*, el *lumen* y el *lux*. Estas unidades se utilizan con frecuencia en catálogos de cámaras para describir la sensibilidad de los elementos sensores de imagen o las características de los focos o fuentes de luz. Este análisis también permite redefinir el concepto de luminancia de un objeto en función de unidades físicas y representan una alternativa a los parámetros eléctricos que hemos usado en los capítulos anteriores.

La conversión de la proyección plana de la escena a una señal eléctrica requiere el uso de transductores de luz a carga eléctrica, que se conocen como *sensores de imagen*. El sensor de imagen más utilizado en las cámaras de vídeo modernas, tanto en aplicaciones profesionales como domésticas, es la célula CCD (*Charge Coupled Device*). Se trata de un dispositivo de estado sólido de reducido tamaño, elevada resolución de imagen y bajo consumo que ha desplazado completamente el uso de otros tipos de captadores como los tubos de imagen. Dedicaremos cierta atención a comentar los principios de funcionamiento de estos sensores y veremos las distintas alternativas tecnológicas

que suelen usarse en función de la aplicación y sus necesidades. La descripción de sensores de imagen incluye sólo los dispositivos CCD, puesto que el impacto en el mercado de otras tecnologías para la captura de señal de vídeo es cada vez más reducido y tiende a desaparecer. En fotografía analógica y cine se utiliza una película sensible cuyos principios físicos escapan de los propósitos de este texto.

Una vez definidos los principios de los sistemas de captura de imagen se describen las distintas funciones que suelen encontrarse en cámaras de vídeo o fotografía, como el *zoom* digital, el enfoque automático, el ajuste del balance de blanco o los subsistemas de estabilización de imagen. La exposición se centra principalmente en las distintas alternativas tecnológicas que se utilizan para realizar cada una de estas funciones.

En la última sección de este capítulo se incluye un resumen de los distintos tipos de cámaras que pueden encontrarse en el mercado, clasificadas en función de su ámbito de aplicación y sus características. También se proporciona un breve compendio de los distintos formatos de señal de vídeo no estándar que se utilizan en aplicaciones de TV de alta definición, video-vigilancia, visión por ordenador, etc.

5.2 Diagrama de bloques de una cámara

Los elementos que integran una cámara de vídeo o de fotografía pueden ser muy diversos y dependen de la aplicación y del mercado al que se destina la cámara. En esta sección nos concentraremos exclusivamente en los bloques básicos que, con mayor o menor complejidad y calidad, son comunes a prácticamente todas las cámaras.

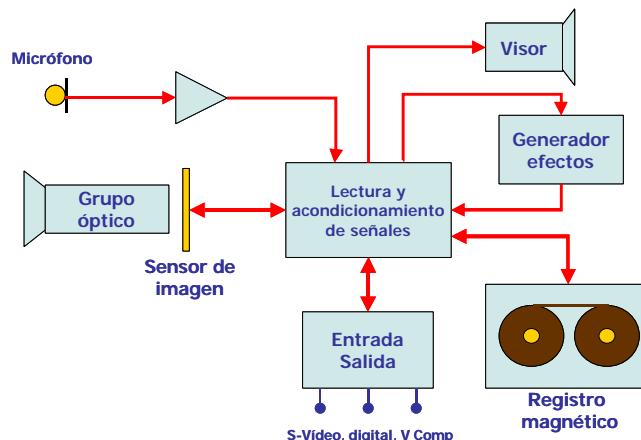


Fig.5.1 Diagrama de bloques de una cámara

En la figura 5.1 se representa un posible diagrama de bloques para una cámara de vídeo doméstica. El grupo óptico, el sensor de imagen y el subsistema de lectura y acondicionamiento de señal están

presentes en todos los tipos de cámaras de vídeo, independientemente de la aplicación a la que se destinan. El resto de bloques pueden estar disponibles o no en función de la aplicación. Así, las cámaras de vídeo-vigilancia no suelen incorporar ningún visor, puesto que la imagen se visualiza directamente desde el centro de control. Examinemos con cierto detalle las funciones de cada uno de los bloques:

- a) *Grupo óptico.* Su función es proporcionar una proyección plana de la escena sobre la superficie del sensor de imagen. Para ello, incluye varias lentes que permiten variar la región del espacio a la que enfoca la cámara. En el caso de cámaras con *zoom*, el grupo óptico puede modificar la posición relativa entre las lentes para controlar el ángulo de la escena que se proyecta al sensor. El grupo óptico también incluye elementos que regulan la cantidad de luz que incide sobre el sensor (iris o diafragma).
- b) *Sensor de imagen.* Es el dispositivo encargado de convertir la luz incidente en una señal eléctrica. En las cámaras de vídeo, la imagen debe ser adquirida a intervalos periódicos (frecuencia de imagen) y debería facilitarse la lectura de la imagen descompuesta en las líneas que constituyen la información de vídeo. Esto no siempre es así, por lo que puede ser necesario un acondicionamiento posterior de las señales al estándar de vídeo con el que trabaja la cámara. En algunos casos, el propio sensor de imagen tiene la capacidad de ajustar la cantidad de luz incidente con la que se estima la señal eléctrica (obturador electrónico). En cámaras de vídeo domésticas suele utilizarse un único sensor de imagen para capturar las tres componentes de color, mientras que en cámaras profesionales es habitual utilizar un sensor distinto para cada componente. En general, los sensores de imagen de estado sólido suelen tener asociado un complejo circuito de control que proporciona las señales necesarias para la lectura de la información.
- c) *Lectura y acondicionamiento de señal.* Las funciones que se realizan dentro de este bloque de tratamiento de señal dependen de la aplicación a la que se destina la cámara. En el caso de cámaras domésticas, se realiza el control de la lectura del sensor en las líneas que componen la imagen, de forma sincronizada con los impulsos de borrado horizontal y se proporcionan las señales de sincronización de cuadro al sensor de imagen. Las señales proporcionadas por el sensor de imagen se filtran y se acondicionan en ganancia, convirtiéndolas, por una parte, al estándar de vídeo en el que se realiza el registro de la secuencia de vídeo sobre la cinta magnética y, por otra, al formato de señal requerido por el visor de imagen. Este módulo de tratamiento también se encarga del control del grupo óptico y del sistema de adquisición, como en el caso de las funciones de enfoque automático, ajuste del iris y la velocidad de obturación en función del modo de exposición y del tipo de imágenes que se estén registrando o, incluso, funciones más avanzadas como la estabilización del movimiento de la cámara. Adicionalmente se suele realizar la conversión de las señales a formatos de vídeo convencionales (vídeo compuesto, S-Vídeo, RF) que se proporcionan como salida de la cámara y que permiten su conexión a receptores de televisión o sistemas de registro sobre cinta magnética convencionales. Evidentemente, el tipo de formatos de señal de vídeo que se utilizan en el subsistema de acondicionamiento de señal depende de las aplicaciones a las que se destine la cámara. En el ámbito de cámaras de estudio profesionales, únicamente se proporcionan las distintas componentes de color, que posteriormente serán tratados por

unidades de control de cámara externas. En cámaras de video-vigilancia o de circuito cerrado de televisión (CCTV) se suele suministrar una única señal de vídeo compuesto en el formato PAL o NTSC.

- d) *Visor.* Proporciona al usuario o al operador de cámara una réplica de la imagen que se está recibiendo en el sensor. Es útil para realizar el control de enfoque y del ángulo de visión de la escena. El visor suele estar formado por un pequeño tubo de rayos catódicos o dispositivo de cristal líquido y una lente que ajusta la imagen para que pueda visualizarse desde una distancia corta. En las cámaras domésticas más modernas suele incorporarse un visor de cristal líquido auxiliar, situado en el exterior del cuerpo de la cámara y desplegable, que se utiliza tanto para ajustar el grupo óptico como para visionar las imágenes previamente registradas en la cinta. Los visores pueden ser en blanco y negro o en color. Estos últimos sólo se encuentran en las cámaras domésticas de alta gama. En cámaras profesionales se siguen utilizando visores en blanco y negro debido a que la resolución del sistema visual es más alta en luminancia que en imágenes de color, con lo que el control de foco de la escena resulta más fácil si se realiza sobre un visor en blanco y negro.
- e) *Registro magnético.* Sólo está presente en las unidades de cámara que incorporan un sistema de registro de vídeo en cinta. Prácticamente todos los modelos orientados al mercado doméstico integran la unidad de cámara con un sistema de registro y reproducción de vídeo portátil. En el ámbito profesional se utilizan este tipo de cámaras para la realización de reportajes en directo. Existen versiones para todos los formatos de vídeo profesionales. En algunos casos, el subsistema de registro de vídeo es un módulo separado que se añade al cabezal de cámara.
- f) *Generador de efectos.* Únicamente se incorpora en videocámaras domésticas y el objetivo es que el usuario final disponga de cierta capacidad para editar y proporcionar efectos de mezcla en sus vídeos sin necesidad de disponer de un sistema de edición independiente. Los efectos más habituales son el registro de señales en blanco y negro, en color sepia, fundidos entre escenas, etc. En algunos modelos, aunque cada vez es menos frecuente, se incluyen pequeñas unidades de titulación.
- g) *Micrófono integrado.* Se usa para el registro de señal de audio en directo. Únicamente lo incorporan las cámaras domésticas y las profesionales para reportajes exteriores. En el registro y producción de vídeo en estudio se utilizan micrófonos y canales de audio independientes y sólo se mezclan con la señal de vídeo en la fase de producción. En las cámaras más modernas se integran dos micrófonos con diagramas de radiación ligeramente inclinados hacia la derecha y la izquierda del eje de la cámara. Con ello se logra obtener un registro estereofónico de la componente de audio. En la mayoría de formatos de vídeo doméstico que se comercializan actualmente (Hi-8, miniDV, Digital 8, etc.) el registro de la señal de audio es digital.
- h) *Entradas y salidas.* Proporcionan la señal de vídeo y/o audio capturada por la cámara. El número y tipo de entradas y salidas depende también de las aplicaciones de la cámara. Así, en cámaras para CCTV únicamente se proporciona la señal de vídeo compuesto,

mientras que en cámaras domésticas suelen proporcionarse distintos formatos de salida: PAL, S-Vídeo, digital, etc. En algunas cámaras, como las *webcam*, el formato de vídeo de salida está especialmente diseñado para su adaptación a un puerto del ordenador (USB, RS-232). Las entradas suelen corresponder a señales de audio externo, capturadas con micrófonos auxiliares. No suelen proporcionarse entradas de señal de vídeo y muy pocos modelos de cámaras domésticas disponen de ellas.

Las cámaras de fotografía digital (*Digital Still Cameras*) tienen un principio de funcionamiento muy parecido a las cámaras de vídeo, con la salvedad de que únicamente registran el contenido de un fotograma y no la secuencia de vídeo completa. De hecho, la mayoría de modelos también proporcionan una señal de vídeo en tiempo real en la salida, que permite reproducir las fotografías almacenadas en la memoria u observar la señal del visor en un receptor de televisión convencional. En consecuencia, tanto el grupo óptico como los sensores de imagen son del mismo tipo y características que los utilizados en las cámaras de vídeo; en todo caso, estos últimos, tienden a utilizar resoluciones de imagen más altas para competir con los sistemas de fotografía analógica. La unidad de acondicionamiento de señal también es parecida, ya que también debe realizarse la lectura del sensor de imagen y producir una señal de vídeo compuesto. En este caso, la conversión a formato de vídeo para registro en cinta se sustituye por un subsistema de digitalización de imágenes que se almacenan en una memoria RAM interna para su posterior volcado a un sistema de registro informático en disquete, integrado dentro de la propia cámara o directamente a un puerto serie conectado con el ordenador. Para mejorar la eficiencia del sistema y poder registrar múltiples imágenes con una memoria interna reducida, se incluyen procesadores específicos que realizan la compresión de imágenes estacionarias (normalmente JPEG o JPEG2000).

En cámaras de fotografía convencional se utiliza película fotográfica (positiva o negativa) como sensor de imagen. Este tipo de sensor tiene una resolución considerablemente superior a la de los sensores electrónicos, proporcionando una gran definición y calidad de imagen. El margen de colores que pueden reproducirse con la técnica fotográfica es, no obstante, menor que el que proporciona una cámara de vídeo. La simplicidad de la cámara fotográfica analógica reduce notablemente los circuitos electrónicos, ya que sólo se utilizan en los modelos automáticos para el cálculo automático de la exposición y del enfoque de las imágenes. En la figura 5.2 se reproduce un esquema simplificado de una cámara del tipo *reflex*. La película fotográfica se sitúa en la parte posterior de la cámara y está protegida de la luz por el obturador, que no es más que una cortinilla que controla el tiempo de exposición de la película. El grupo óptico tiene la misma función que en las cámaras de vídeo, aunque en este caso suelen ser de un mayor tamaño, debido a que la superficie sobre la que se proyecta la imagen es más grande que en vídeo y exige el uso de lentes con una distancia focal superior. Analizaremos con detalle la relación que existe entre el tamaño del sensor y las distancias focales de las lentes.

La cámara *reflex* tiene la particularidad de que la imagen que se observa por el visor coincide exactamente con la imagen que finalmente incidirá sobre la película. En cámaras de fotografía convencionales el visor tiene un grupo óptico simplificado e independiente, que no está situado en la misma posición, de modo que pueden producirse errores de paralelismo que conducen a que la imagen fotografiada no coincida exactamente con la que se veía por el visor. La ventaja de una cámara *reflex* es que estos errores se eliminan completamente, y además, el fotógrafo puede predecir exactamente cuál será la profundidad de campo de la fotografía, ya que la está observando con el

misma sistema de lentes e iluminación que incidirá sobre la película. Aunque la implementación mecánica de una cámara *reflex* es bastante costosa, su principio de funcionamiento es simple. La imagen, tal y como la produce el grupo óptico, se refleja mediante el uso de un espejo para dirigirla hacia el visor. El pentaprisma situado en el recinto del visor realiza una inversión de la imagen para que pueda observarse sin el giro horizontal y vertical que producen las lentes del grupo óptico. Al realizar la fotografía, el espejo que refleja la imagen hacia el visor se levanta, dejando que la imagen se forme sobre el fondo de la cámara, donde está situada la película. Durante este intervalo de tiempo el visor queda en negro. El control del tiempo de exposición se realiza mediante un obturador de precisión, ya que no es posible retirar el espejo a grandes velocidades. A pesar de este sofisticado movimiento mecánico, se trata de cámaras muy robustas, que son usadas por casi todos los fotógrafos profesionales para la elaboración de reportajes de campo. Su principal inconveniente respecto a otro tipo de cámaras profesionales es que utilizan película fotográfica de miniatura (35 mm, baja resolución si se compara con otros tipos de película de mayor tamaño) y que siempre existe un pequeño retardo entre el momento de disparar y el de tomar la fotografía, que puede ser importante en la toma de instantáneas con objetos que se mueven a alta velocidad.

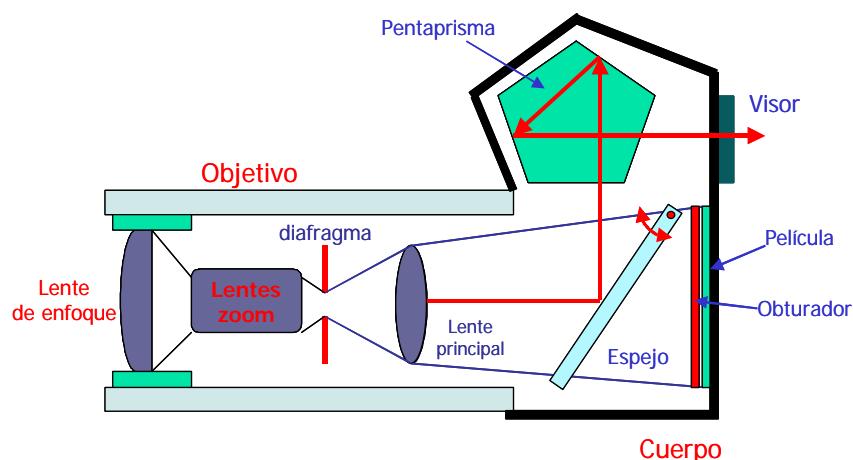


Fig 5.2 Elementos de una cámara fotográfica

5.3 Principios de óptica geométrica y formación de imágenes

El grupo óptico desempeña un papel fundamental en el proceso de adquisición de la imagen, ya que determina la parte de la escena que será capturada y los elementos que permanecerán correctamente enfocados. También se encarga de regular la cantidad de luz que incide sobre el sensor, de modo que es posible operar en un amplio margen de condiciones de iluminación sin necesidad de modificar la sensibilidad del sensor. En este apartado se proporcionan las herramientas básicas para comprender la función de los distintos elementos que forman el grupo óptico: lentes, diafragma y obturador, así como su incidencia sobre la imagen final. El objetivo es comprender las relaciones existentes entre velocidad de obturación, apertura del diafragma, distancia focal y profundidad de campo. Para simplificar el análisis supondremos que todas las lentes del grupo óptico son ideales y que pueden

modelarse por una única lente equivalente. Esta simplificación no nos permite evaluar las aberraciones de las lentes, pero no supone ninguna restricción importante sobre los conceptos que pretendemos manejar.

5.3.1 Lentes, grupo óptico y la lente ideal

Una lente es una pieza de vidrio u otro material transparente que se utiliza para la formación de imágenes de objetos reales mediante el desvío de los rayos de luz que proceden del objeto. Suelen tener forma circular y sus superficies frontal y posterior están pulidas con curvaturas cóncavas o convexas que casi siempre tienen forma esférica. En la figura 5.3 se representa el esquema de una lente.

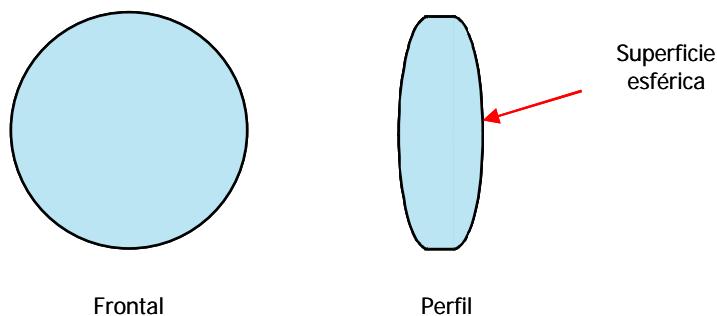


Fig.5.3 Esquema de una lente

Las superficies frontal y posterior introducen un cambio en la dirección de propagación de los rayos de luz, debido al cambio de medio (aire-vidrio-aire), conocido como refracción. Al tratarse de curvaturas esféricas, los rayos que entran perpendiculares al plano de la lente convergen en un único punto denominado *foco*. La distancia entre el foco y el plano de la lente es la *distancia focal* y depende de la curvatura de las superficies frontal y posterior. El plano paralelo al plano de la lente, que pasa por el foco, se denomina *plano focal*. Además, cuando los rayos de luz proceden de una dirección distinta a la orthogonal también convergen en un mismo punto que se sitúa sobre el plano focal. En la figura 5.4 se representan el foco, la distancia focal y el plano focal de una lente.

Debe observarse que el pulido de las superficies de las lentes debe ser muy preciso, puesto que cualquier imperfección modificará la dirección en la que se refractan los rayos y significará que no todos ellos convergen en un único punto. Este problema se conoce como *aberración* y puede introducir distintos tipos de distorsión en la imagen, como la falta de enfoque, aparición de distorsiones geométricas o cromáticas, etc. En nuestro caso supondremos que las lentes son ideales y que no producen este tipo de problemas.

Las lentes pueden ser convergentes o divergentes, dependiendo de la forma cóncava o convexa de sus superficies y de sus radios de curvatura. Las lentes convergentes se comportan tal y como hemos

definido anteriormente, es decir, los rayos que atraviesan la lente convergen en un punto denominado *foco* que está situado por detrás de la lente. En las lentes divergentes, los rayos se separan al atravesar la lente. No obstante, tal y como se ilustra en la figura 5.5, los rayos resultantes parecen proceder de un punto situado por delante de la lente que se denomina *foco virtual*. Las lentes convergentes se utilizan en la captura o proyección de imágenes. Las aplicaciones típicas son las cámaras de vídeo, de fotografía o cine, los proyectores de vídeo o película, lupas de bolsillo, etc. Las imágenes obtenidas con lentes convergentes se denominan *imágenes reales*, en contraposición con las que se obtienen mediante lentes divergentes que se denominan *imágenes virtuales*. En este último caso los rayos parecen proceder de un punto inexistente o virtual. Son ejemplos de imágenes virtuales las que se observan a través de un espejo o de un microscopio.

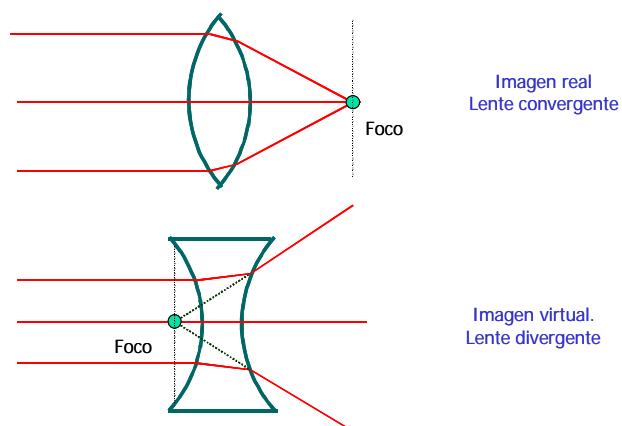


Fig. 5.4 Foco, plano de la lente y plano focal en lentes convergentes



Fig. 5.5 Lentes convergentes y divergentes. Imagen real y virtual.

El carácter convergente o divergente de una lente depende de los radios de curvatura de las superficies frontal y posterior. Cuando ambas superficies son convexas la lente es convergente, mientras que si

son cóncavas, la lente resulta ser divergente. En los casos en que una de las superficies es cóncava y la otra convexa, la lente puede resultar convergente o divergente en función de los radios de curvatura de cada superficie.

La calidad de una lente se mide mediante su *poder de resolución*, que indica la capacidad de la lente para formar correctamente las imágenes de dos puntos o líneas muy próximos. Esta medida puede realizarse por diferentes métodos. Uno de los procedimientos más utilizados es utilizar un patrón de líneas blancas y negras que van alternándose en el sentido horizontal o vertical. El poder de resolución de la lente se considera como el número de pares de línea por milímetro que pueden discernirse en la imagen. Este valor nos proporciona una primera idea aproximada de la calidad de la lente. En general, para obtener grupos ópticos con un poder de resolución aceptable para fotografía o vídeo es necesario utilizar varias lentes.

Existen diversas aplicaciones que utilizan una única lente para la formación de imágenes. Entre estas destacan las lentes de contacto, las gafas, los visores de las cámaras de vídeo, los semáforos, las lupas de bolsillo, los proyectores de transparencias, etc. Sin embargo, las imágenes que se obtienen con una única lente no resultan de suficiente calidad en campos como la fotografía o la microscopía, que requieren imágenes de gran precisión. El problema se debe a que, a pesar de que el proceso de fabricación de la lente se cuide en extremo, resulta inevitable la aparición de cierto grado de aberración.

Las aberraciones en una lente pueden ser de distintos tipos en función de la naturaleza del problema que las origina. La *aberración esférica* aparece cuando los rayos de luz no convergen en el foco de la lente y se debe a las imprecisiones en la fabricación y pulido de las superficies esféricas de la lente. Las *aberraciones cromáticas* aparecen cuando los planos en los que se forman las imágenes dependen de la longitud de onda de la luz. En este caso, las distintas componentes de color no convergen en el mismo punto, de modo que los colores de la imagen aparecen falseados. La *aberración de curvatura de campo* se produce cuando los rayos procedentes de distintas direcciones no convergen sobre el plano focal, sino sobre una superficie con una ligera forma esférica. El resultado es que si enfocamos la parte central de la imagen, los extremos aparecerán desenfocados. La *aberración de coma* aparece en los extremos de la imagen y se debe a que los rayos no convergen en un único punto para los objetos que están alejados del eje de la lente. El *astigmatismo* se produce cuando el plano de convergencia de las imágenes para un patrón de líneas horizontales no coincide con el plano donde converge un patrón de líneas verticales. Por ello, aparece pérdida de definición en los contornos verticales u horizontales de la imagen. La *distorsión geométrica* consiste en que las líneas rectas aparecen curvadas a medida que nos alejamos del centro de la lente. Puede ser de *cojín* o de *barril* en función de la curvatura.

Las aberraciones pueden reducirse si se sustituye la lente por un conjunto de lentes con la forma y la separación apropiadas. La reducción de la aberración se consigue provocando que las lentes individuales tengan aberraciones contrapuestas que se cancelen entre sí. En el grupo óptico de una cámara de fotografía pueden encontrarse entre 15 y 20 lentes individuales, cuyos parámetros geométricos y posiciones relativas han sido calculados con precisión y que resultan equivalentes a una única lente con un nivel de aberración reducido. Pueden combinarse lentes convergentes y divergentes (véase la Fig. 5,6) aunque el equivalente resulta ser siempre una lente convergente. Algunas lentes pueden modificar su posición relativa respecto a otras lentes para permitir modificar la distancia focal

(objetivos con *zoom*) o enfocar correctamente la imagen sobre el sensor. El coste de un grupo óptico está directamente relacionado con la ausencia de aberraciones del conjunto de lentes. Así, los precios de objetivos que tienen una gran apertura (mucha entrada de luz) son muy superiores a los de los objetivos con poca apertura. La razón es que para obtener aperturas grandes se requieren también lentes de gran tamaño, en las que deben eliminarse todas las aberraciones que puedan aparecer en la periferia de la lente.

Grupo óptico

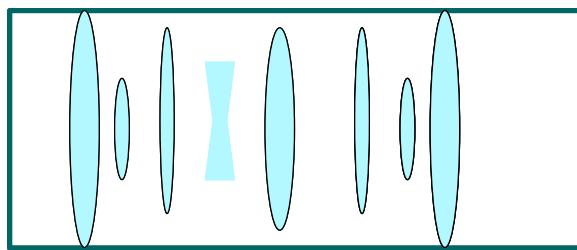


Fig. 5.6 Un grupo óptico está formado por varias lentes para reducir la aberración.

Una buena cámara para fotografía en miniatura de 35 mm puede utilizar grupos ópticos con un poder de resolución de unas 80-100 líneas por milímetro. Esto significa que con una película de suficiente resolución pueden obtenerse hasta un total de 2400 líneas en el sentido vertical (la altura de la película es de 24mm), lo que representa una resolución unas 4 ó 5 veces superior a la de los sistemas de televisión de definición convencional.

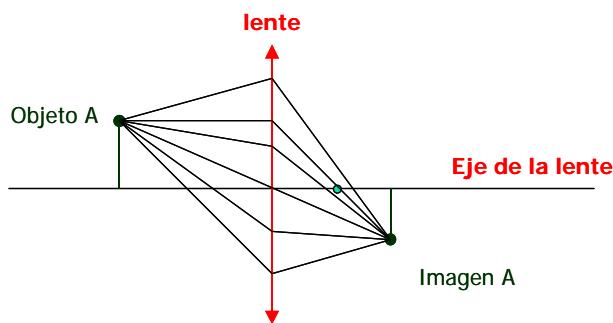


Fig. 5.7 Formación de imágenes en una lente ideal.

La lente ideal puede utilizarse como primera aproximación para estudiar las propiedades y características de un grupo óptico. La lente ideal queda caracterizada por su distancia focal (F) y se representa como una línea vertical con dos flechas en sus extremos. En la figura 5.7 se representa una

lente ideal y se indica cómo se obtiene la imagen del punto A. Todos los rayos que proceden de este punto son desviados por la lente hacia el punto imagen. La imagen siempre se forma sobre un plano paralelo al plano de la lente que se denomina *plano de imagen* y que está situado por detrás del plano focal. Si deseamos enfocar el objeto A, deberemos situar el sensor de imagen sobre el plano de imagen. Para realizar el enfoque en cámaras fotográficas o de vídeo, en las que el sensor está en una posición fija, es necesario modificar la posición de la lente, adelantándola o atrasándola hasta que el objeto quede correctamente enfocado. Es importante tener en cuenta que las imágenes de todos los objetos situados a una misma distancia de la lente (sobre el plano de objeto) se forman en el mismo plano de imagen. Esto significa que los objetos situados a la misma distancia resultarán enfocados simultáneamente. En el siguiente apartado veremos como puede determinarse la distancia a la que se sitúa el plano de imagen.

5.3.2 Formación de imágenes con lentes ideales

Para determinar el punto donde se forma la imagen de un objeto pueden usarse las tres reglas siguientes:

- 1) El rayo procedente del objeto que pasa por el centro de la lente no es desviado.
- 2) El rayo procedente del objeto que entra perpendicular al plano de la lente se desvía hacia el foco.
- 3) La imagen del objeto se obtiene en el punto de intersección de los dos rayos anteriores.

En la figura 5.8 se representan gráficamente las reglas anteriores cuya aplicación es de carácter general, exceptuando aquellos objetos que se encuentran sobre el eje de la lente. En este caso concreto, las dos primeras reglas representan el mismo rayo, por lo que no puede calcularse el punto de intersección. No obstante, resulta trivial calcular la distancia a la que se formará la imagen si consideramos que todos los objetos situados a la misma distancia convergen en el mismo plano de imagen.

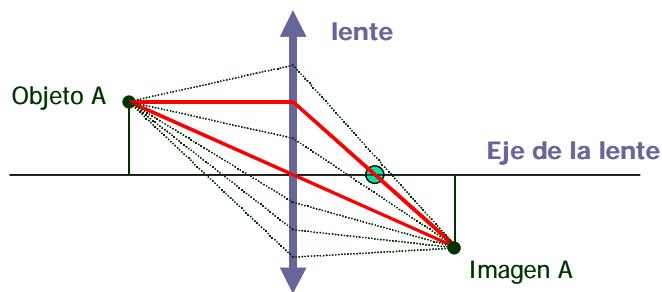


Fig. 5.8 Reglas para el cálculo del punto en el que se forma la imagen

Las reglas anteriores proporcionan un método simple e intuitivo para determinar el punto donde se forma la imagen. Por ello, en muchas representaciones, sólo se utilizarán estos dos rayos para indicar el punto donde aparece la imagen. Sin embargo, no debe olvidarse que del objeto A salen rayos en

todas las direcciones y que todos ellos, cuando pasan a través de la lente, son desviados hacia el punto imagen. La consideración de todos estos rayos será fundamental cuando analicemos la función del diafragma y su incidencia en la profundidad de campo.

A partir de estas consideraciones resulta bastante sencillo determinar analíticamente las relaciones entre las distancias y alturas del objeto y su imagen. En la figura 5.9 se representan los diferentes parámetros que deseamos relacionar donde H y L representan la altura y la distancia respectivamente del objeto real, F la distancia focal y h y l la altura y distancia del punto donde se forma la imagen.

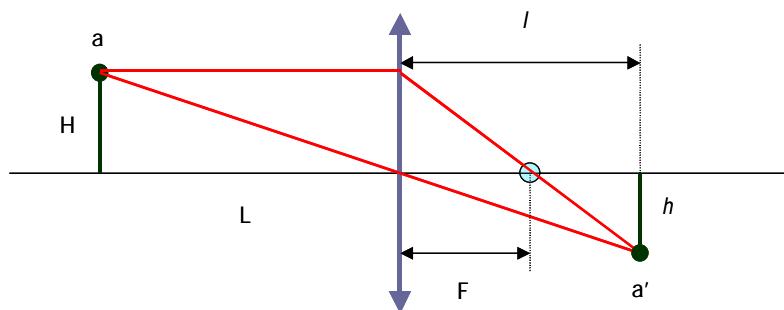


Fig. 5.9 Relaciones trigonométricas en la lente ideal

La ecuación de la recta que pasa por el centro de la lente puede expresarse:

$$y = -\frac{H}{L} \cdot x \quad (5.1)$$

La recta que pasa por el foco viene dada por:

$$y = H - \frac{H}{F} \cdot x \quad (5.2)$$

Igualando ambas expresiones para $x = l$ obtenemos:

$$\frac{1}{L} + \frac{1}{l} = \frac{1}{F} \quad (5.3)$$

que relaciona la distancia a la que se encuentra el objeto de la lente con la distancia en la que se forma la imagen. Análogamente, si sustituimos para $y = h$ obtenemos:

$$\frac{1}{H} + \frac{1}{h} = \frac{1}{F} \quad (5.4)$$

Las ecuaciones 5.3 y 5.4 proporcionan las relaciones básicas entre un objeto y su imagen en función de la distancia focal de la lente. De ellas podemos extraer varias conclusiones que nos ayudan a comprender el fenómeno de formación de imágenes:

- a) *Las imágenes de los objetos muy lejanos se forman sobre el plano focal.* En efecto, a medida que L aumenta, el factor $1/L$ en la ecuación 5.3 se reduce, con lo que el plano de imagen (situado a una distancia l) coincide con el plano focal. Existen algunas cámaras de fotografía muy económicas que no tienen la capacidad de ajustar el enfoque de la imagen y que sitúan la película sobre el plano focal. De este modo, siempre enfocan las imágenes que se encuentran en el infinito. Estas cámaras suelen dar resultados aceptables cuando se realizan fotografías de paisajes o de objetos alejados. Veremos que suelen utilizar diafragmas muy cerrados y grandes angulares, con lo que consiguen una profundidad de campo considerable que permite que los objetos situados a una distancia media permanezcan más o menos enfocados.
- b) *A medida que un objeto se acerca a la lente, el plano de imagen se aleja del plano focal.* Resulta evidente que al aumentar el término $1/L$ debe reducirse $1/l$ para mantener la contribución de ambos constantes e igual a $1/F$. Como una consecuencia directa de esta relación, debe observarse que cuando el objeto se sitúa a una distancia de la lente igual a la distancia focal, no se obtiene ninguna imagen, ya que el rayo que pasa por el centro de la lente y el que se desvía hacia el foco son paralelos y no se intersecan en ningún punto. En la práctica, la distancia mínima para la que puede enfocar una lente depende de la máxima separación que podemos obtener entre la lente ideal equivalente y la posición del sensor de imagen. Esta distancia depende de las características del grupo óptico.
- c) *Si aumenta la distancia focal también aumenta el tamaño de la imagen.* Si aislamos h en la ecuación 5.4 obtenemos:

$$h = \frac{H}{H/F - 1} \quad (5.5)$$

de donde se deduce inmediatamente que un aumento de F supone que el denominador disminuye y por lo tanto que aumenta el tamaño h de la imagen del objeto. Así pues, los teleobjetivos se corresponderán con distancias focales grandes mientras los grandes angulares u ojos de pez con distancias focales pequeñas.

- d) *Si aumenta la distancia focal también aumenta la distancia del plano de imagen.* La distancia a la que se obtiene el plano de imagen puede despejarse de la ecuación 5.3, obteniendo:

$$l = \frac{L}{L/F - 1} \quad (5.6)$$

Por tanto, si F aumenta el denominador disminuye y aumenta la distancia del plano focal. Debe notarse que la relación sólo es válida cuando L/F es mayor que la unidad, ya que en el

caso contrario no se forma ninguna imagen. Esta relación nos indica que las lentes de gran distancia focal requieren una separación considerable entre el plano de la lente y el sensor. Por ello, todos los teleobjetivos suelen ser largos, ya que se trata de mantener la lente equivalente alejada del sensor.

5.3.4 Ángulo de visión

Se entiende como campo o ángulo de visión de una fotografía o secuencia de vídeo el ángulo que abarcan los objetos que aparecen en la imagen cuando son observados desde el punto donde se ha realizado la toma. Un ángulo de visión reducido se corresponde con imágenes tomadas desde un teleobjetivo. Con un gran angular o un ojo de pez pueden obtenerse campos de visión muy amplios. El campo de visión depende de la distancia focal del objetivo y del tamaño del sensor. Para calcularlo de forma simple es conveniente suponer que estamos capturando imágenes que se encuentran en el infinito. En este caso, el sensor de imagen debe situarse en el plano focal, a una distancia F de la lente. En la figura 5.10 se muestra un diagrama en el que se relaciona la distancia focal, con el ángulo de visión y el tamaño del sensor. Los objetos de la escena que serán capturados se encuentran en el ángulo definido por los rayos que pasan por el centro de la lente y que inciden en los extremos del sensor. La relación entre el ángulo de visión (α), la distancia focal (F) y el tamaño del sensor (D) puede obtenerse usando relaciones trigonométricas:

$$\frac{D/2}{F} = \tan(\alpha/2) \Rightarrow \alpha = 2 \cdot \arctan\left(\frac{D}{2 \cdot F}\right) \quad (5.7)$$

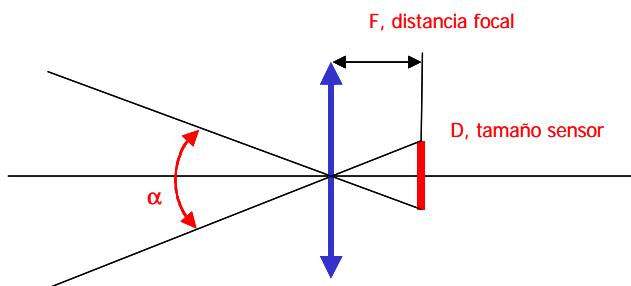


Fig. 5.10 Cálculo del ángulo de visión

Evidentemente, como el sensor no es cuadrado, puede definirse un ángulo de visión en el sentido horizontal y otro en el sentido vertical. La dependencia del ángulo de visión con el tamaño del sensor es la razón por la que los objetivos usados en las cámaras de vídeo son mucho más pequeños que los objetivos utilizados en fotografía convencional. En efecto, en la ecuación 5.7 resulta evidente que si deseamos mantener el ángulo de visión constante la relación entre el tamaño del sensor y la distancia focal también debe mantenerse constante. Por lo tanto, un aumento del tamaño del sensor supone un aumento proporcional de la distancia focal del objetivo.

Consideremos como ejemplo el cálculo de los ángulos de visión que se obtienen con un objetivo de 50mm de distancia focal en cámaras de fotografía en miniatura de 35 mm. En este tipo de fotografía se utiliza una película que tiene 35 mm de altura, pero de la que sólo se aprovechan 24 mm, pues el resto se utilizan para las guías de arrastre de la película. La anchura efectiva del fotograma es, tal y como se indica en la figura 5.11, de 36 mm. Debe observarse que en este tipo de fotografía la relación de aspecto es de 3:2, es decir, algo más alargada que en televisión convencional (4:3) y ligeramente inferior a la que se utiliza en los formatos panorámicos (16:9). Los ángulos de visión vertical y horizontal (suponemos que la toma se realiza en la forma apaisada) que se obtienen para un objetivo de 50 mm son:

$$\alpha_V = 26,99^\circ; \quad \alpha_H = 39,59^\circ \quad (5.8)$$

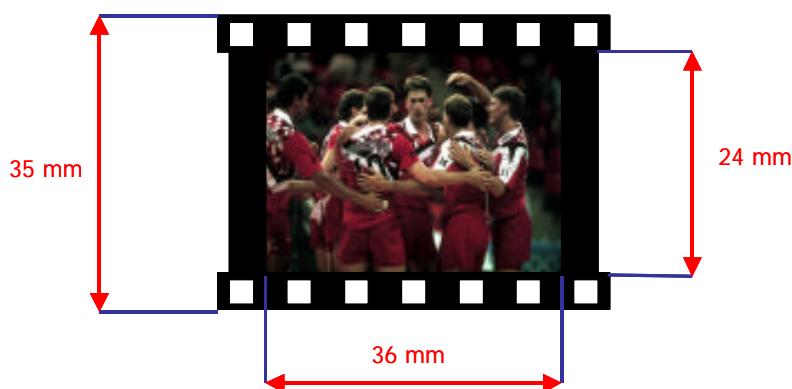


Fig. 5.11 Tamaño del sensor utilizado en fotografía en miniatura de 35 mm

La perspectiva que proporciona con un objetivo de 50 mm en este tipo de fotografía es parecido al que tiene el sistema visual humano. Las fotografías obtenidas con estos objetivos resultan por tanto muy naturales. Los objetivos con una distancia focal menor producen una mayor sensación de perspectiva en la escena. Las distancias focales de 35 mm suelen utilizarse para paisajes, entre 24 y 28 mm tenemos los grandes angulares, que ya introducen una distorsión apreciable de la imagen. Los ojos de pez de 16 o 18 mm permiten capturar escenas con un ángulo de visión horizontal de más de 90°. Por el contrario, si se utilizan distancias focales superiores a 50 mm, se reduce la sensación de perspectiva. La distancia focal de 105 mm, muy utilizada en retratos de personas, produce unos ángulos de visión de 13,04° (vertical) y 19,45° (horizontal). Un teleobjetivo de 300 mm reduce notablemente la perspectiva y tiene unos ángulos de visión de 4,58° (vertical) y 6,86° (horizontal).

El sistema de fotografía APS (*Advanced Photo System*) fue creado conjuntamente por varios fabricantes (*Kodak, Fuji, Canon, Nikon y Minolta*) como una alternativa universal al clásico formato de 35 mm. El objetivo fundamental del lanzamiento de este nuevo sistema era reactivar el mercado de cámaras fotográficas y ha tenido cierto éxito en la gama de cámaras compactas. El APS introduce diversas prestaciones adicionales entre las que destaca la presencia de una banda magnética en los laterales de la parte posterior de la película y en la que puede incorporarse información digital diversa

sobre las tomas: fecha, hora, diafragma, velocidad de obturación, tipo de encuadre y otros datos digitales. El envase del carrete también contiene información sobre la sensibilidad de la película y el número de exposiciones realizado, que se intercambian con la cámara. El usuario puede seleccionar el formato con el que se realiza la toma existiendo los modos normal, angular y panorámico en los que se modifica el tamaño o zona de la película en la que se realiza el registro de la imagen. El sistema APS dispone de una completa gama de productos adicionales para la presentación de imágenes en una pantalla de televisión, equipos para realizar escaneado de los negativos, etc. Se utiliza una película fotográfica de 24 mm de anchura que, en el modo de encuadre normal, tiene un tamaño efectivo de 16,7x30,2 mm. Este tamaño de sensor representa una superficie un 40 % menor que la de un fotograma en el formato de 35 mm. No obstante, los fabricantes aseguran que las copias tienen una calidad parecida en ambos formatos debido a las mejoras introducidas en las emulsiones sensibles.

Al modificar la relación de aspecto del sensor, la relación entre ángulos de visión y la sensación de perspectiva no se mantiene. Así, un objetivo de 42 mm para el formato APS cubre el mismo ángulo de visión horizontal que un objetivo de 50 mm en el formato de 35 mm. En efecto, utilizando el ángulo de visión horizontal que hemos obtenido en la ecuación 5.8, podemos calcular la nueva distancia focal mediante:

$$F = \frac{D}{2 \cdot \tan(\alpha/2)} = 42 \text{ mm} \quad (5.9)$$

No obstante, si utilizamos esta misma distancia focal para calcular, mediante la ecuación 5.7, el ángulo de visión vertical que obtenemos es de 22,48°, algo inferior al que se obtiene en el formato de 35 mm. Una alternativa que se utiliza con cierta frecuencia para obtener las distancias focales equivalentes en formatos que tienen relaciones de aspecto distintas consiste en igualar los ángulos de visión que se obtienen en la diagonal del sensor. Con ello, tal y como se representa en la figura 5.12, se modifican los dos campos de visión ajustándose a la relación de aspecto de modo que las imágenes tendrán una sensación de perspectiva parecida.

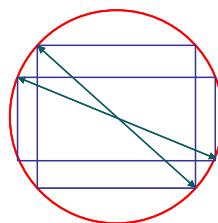


Fig. 5.12 Ángulos de visión equivalentes en imágenes con distinta relación de aspecto

En cámaras de vídeo el sensor tiene casi siempre una relación de aspecto de 4:3, o de 16:9 si se trata de cámaras que trabajan en el formato panorámico. El tamaño del sensor depende de la tecnología y de las prestaciones de la cámara. En cámaras domésticas suelen usarse sensores de 1/4 " ó 1/3". En cámaras profesionales pueden usarse sensores de mayor tamaño para mejorar la resolución (2/3"), aunque la tendencia es ir reduciendo el tamaño del sensor a medida que el avance tecnológico permite mejorar la resolución de los sensores de estado sólido. En la actualidad pueden encontrarse cámaras en

miniatura que incorporan sensores de 1/6". La terminología que expresa el tamaño del sensor en pulgadas procede de las antiguas cámaras que usaban tubos de imagen y no tiene mucho que ver con las dimensiones físicas reales del área de imagen. El número de pulgadas indicaba el diámetro de todo el tubo de imagen, del que sólo la parte central se utilizaba para la lectura de la imagen. Las primeras CCD indicaban el tipo de tubo de imagen con el que eran compatibles. Así, una CCD de 2/3" indicaba que tenía un área de imagen útil que podía sustituir a un tubo de imagen de 2/3" sin necesidad de modificar el grupo óptico. Una CCD de 2/3" tiene una zona de imagen útil de 11 mm de diagonal. En una CCD de 1/4 " la zona de imagen útil tiene una diagonal de aproximadamente 4 mm. Las dimensiones reales de la zona de imagen suelen depender del fabricante, sobre todo en sensores de reducido tamaño, y del modelo de cámara en el que se incorporan, ya que algunos sistemas de tratamiento digital reducen la zona de imagen útil para poder aplicar algoritmos de estabilización del movimiento de cámara. De todos modos, resulta evidente que la reducción del tamaño del sensor en cámaras de vídeo es notable y por tanto, los objetivos requieren distancias focales mucho menores. Consideremos como ejemplo un sensor CCD de 1/4 " que tiene un área de imagen de 2,1 mm x 2,8 mm. Este sensor requiere una óptica con una distancia focal de sólo 3,9 mm para obtener el mismo ángulo de visión horizontal que se obtiene con una óptica de 50 mm en fotografía de 35 mm. Esto indica que los grupos ópticos para cámaras de vídeo suelen ser menos aparatosos (menor distancia focal) que los utilizados en fotografía. La comparación resultaría aún más exagerada si hubiéramos considerado formatos de fotografía de estudio o de cine de 70 mm en los que el tamaño del sensor es todavía mayor.

5.3.5 Regulación de la cantidad de luz: diafragma y obturador

La cantidad de luz que incide sobre el sensor se controla mediante dos dispositivos que pueden actuar directamente sobre la cantidad de luz (*diafragma*) o sobre el tiempo de exposición del sensor (*obturador*). El diafragma está situado en el interior del grupo óptico, entre las lentes, y puede abrirse o cerrarse para dejar pasar más o menos luz por el objetivo. Su función es parecida a la del iris y la pupila en el sistema visual humano y actúa como regulador del margen dinámico en el que puede trabajar la cámara. En condiciones de elevada iluminación, el diafragma se cierra dejando pasar únicamente una fracción de la luz incidente, mientras que cuando la luz es insuficiente permanece abierto dejando que toda la luz que pasa por el objetivo incida sobre el sensor. El obturador controla el tiempo que el sensor es expuesto a la luz en cada fotograma (tiempo de exposición). También puede actuar como regulador del margen dinámico aumentando o disminuyendo este tiempo en función de las condiciones de luz. La acción conjunta de ambos elementos de control permite que pueda operarse en condiciones de luz muy dispares sin necesidad de modificar la sensibilidad del dispositivo de captura de imagen.

El diafragma (también denominado iris en cámaras de vídeo) y el obturador tienen efectos distintos sobre la imagen. El operador de cámara o fotógrafo deberá dar prioridad a uno u a otro en función de las características de la escena y los efectos de imagen que pretenda conseguir. Examinaremos con algún detalle estos efectos más adelante.

El diafragma tiene forma circular y está formado por un sistema de finas láminas metálicas superpuestas que pueden abrirse o cerrarse en función de la apertura deseada. En la figura 5.13 se muestran tres posibles posiciones de un diafragma. En cámaras de fotografía del tipo *reflex* suele estar

siempre abierto mientras no se realizan fotografías para permitir que toda la luz que entra por el objetivo fluya hacia el visor. Cuando se realiza la fotografía, el diafragma se cierra momentáneamente a la apertura seleccionada. En cámaras de fotografía que tienen un visor independiente y en cámaras de vídeo, el diafragma está permanentemente ajustado a la apertura seleccionada por el usuario o por el modo de exposición automático.

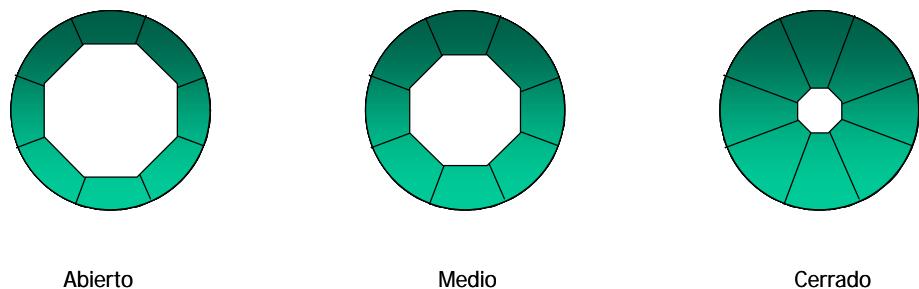


Fig. 5.13 Distintas aperturas del diafragma

En la figura 5.14 se muestra el efecto regulador del diafragma desde el punto de vista de la lente ideal. La posición del diafragma en el grupo óptico debe estar perfectamente ajustada para que su centro coincida con el eje visual de la lente y para que el diafragma equivalente esté situado exactamente en la misma posición que la lente equivalente. El diagrama de la figura 5.14 indica claramente que el uso del diafragma limita la energía total de luz incidente sobre el sensor. Esta representación del efecto del diafragma en la formación de imágenes será de gran trascendencia cuando consideremos sus efectos sobre la profundidad de campo.

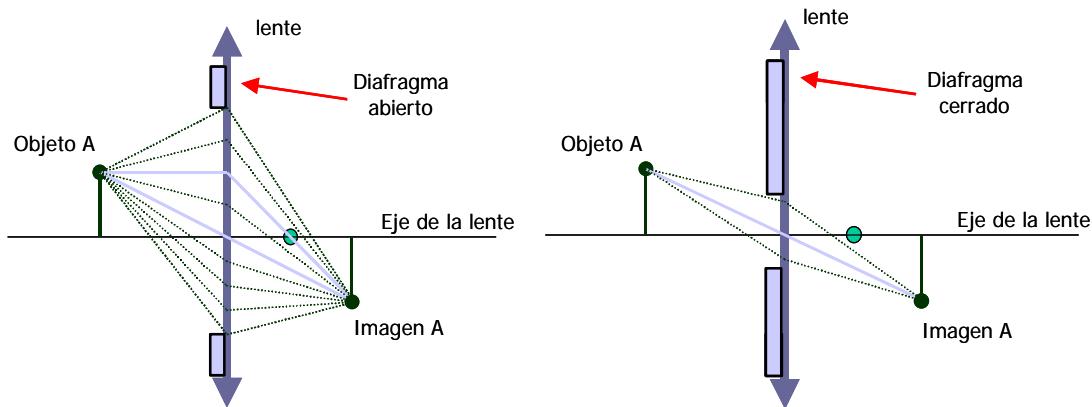


Fig. 5.14 Control de la cantidad de luz incidente en el sensor mediante el uso del diafragma.

La apertura del diafragma se indica mediante la apertura numérica (N.A) o los *números f*. Estos se obtienen como el cociente entre el diámetro del diafragma y la distancia focal. Así:

$$\text{número } f = \frac{F}{D_{\text{diafragma}}} \quad (5.10)$$

De este modo, un número *f* igual a 4 significa que el diámetro del diafragma es 4 veces menor que la distancia focal. Esta relación se indica de forma esquemática en la figura 5.15.

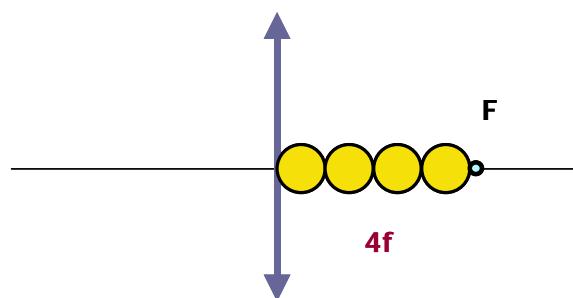


Fig. 5.15 Relación entre números *f*, apertura del diafragma y distancia focal

Los números *f* no pueden variarse de forma continua, sino mediante una secuencia discreta que toma los valores 1,4; 2; 2,8; 4; 5,6; 8; 11; 16; 22; 32; etc. El número más pequeño indica la apertura máxima de la lente y es variable en función de la calidad del objetivo. Pueden encontrarse objetivos que empiezan la secuencia en 1,8; 2,8; 4; etc., u otros con menor apertura que siguen la secuencia 3,5; 5,6; 8; etc. La relación entre un número *f* y el siguiente es de aproximadamente la raíz de 2. Esto significa que el diámetro del diafragma se reduce en un factor de 1,41 entre pasos sucesivos, lo que equivale a que la superficie de la apertura del diafragma se reduzca en un factor 2 (la apertura es circular por lo que viene dada por $\pi(D/2)^2$). Según esta relación, la cantidad de luz que pasa a través del objetivo para dos pasos sucesivos del diafragma se reduce en la mitad. Esto puede parecer en principio exagerado, pero está de acuerdo con las características del sistema visual humano que observará estos cambios exponenciales como si fueran lineales. Algunas cámaras pueden modificar los pasos del diafragma en pasos de 1/3 cuando trabajan en modo automático. En modo manual, muchos fotógrafos dejan la anilla selectora entre dos pasos para ajustar la cantidad de luz de forma más precisa.

El coste de un objetivo está muy relacionado con su apertura máxima, sobre todo cuando la distancia focal es grande. La razón de este aumento de coste se debe a que para conseguir aperturas grandes debe aumentarse el diámetro de las lentes, resultando más costoso reducir las aberraciones que pueden aparecer en la periferia.

El obturador también realiza el control del tiempo de exposición en pasos discretos, que varían en factores de 2. Como en el caso del diafragma, esto significa que entre dos pasos consecutivos la cantidad de luz incidente en el sensor se reduce a la mitad. El efecto sobre la imagen es, no obstante,

lineal y gradual. Las velocidades de obturación se indican en fracciones de segundos o directamente en segundos, dependiendo de si el tiempo de exposición es inferior o superior al segundo. Así, una velocidad de 250 significa que el tiempo de exposición es de 1/250.

La secuencia de velocidades de obturación utilizada en cámaras de fotografía es: 8000, 4000, 2000, 1000, 500, 250, 125, 60, 30, 15, 8, 4, 2, 1", 2", 4", 8", ..., *bulb*. Las velocidades de 1/8000 o 1/4000 sólo se consiguen con las cámaras de fotografía más avanzadas. El modo *bulb* se utiliza para tiempos de exposición muy prolongados y consiste en mantener abierto el obturador durante todo el tiempo que se mantenga pulsado el disparador.

Resulta indiscutible que si el tiempo de exposición es muy prolongado y existen objetos móviles en la escena este movimiento quedará registrado en el sensor. La película fotográfica o el sensor CCD se comportan como un integrador de luz, de modo que registran todos los movimientos que ocurren en la escena durante el tiempo en que el obturador permanece abierto. Por lo tanto, es importante evaluar el movimiento de los objetos de la imagen si queremos tener la garantía que la toma será correcta. Para tiempos de exposición de 1/30 o mayores, empieza a apreciarse el propio movimiento de la mano y es aconsejable utilizar un trípode para mantener la cámara estacionaria. En la fotografía de escenas deportivas, donde existe un movimiento considerable resulta aconsejable aumentar en lo posible la velocidad del obturador. La pérdida de luz puede compensarse abriendo el diafragma.

Pueden realizarse exposiciones intencionadamente lentas para conseguir efectos artísticos. Entre los más conocidos destacan las fotografías de saltos de agua, que si se registran con velocidades lentes producen un efecto espumoso, mientras que si se capturan con velocidades altas podrán apreciarse con nitidez las gotas de agua. En la fotografía de pruebas de motociclismo o automovilismo pueden usarse velocidades deliberadamente lentes y seguir con la cámara el movimiento del objeto mientras se realiza la fotografía. El efecto es que el sujeto móvil aparece nítido, mientras el fondo registra el movimiento de la cámara, produciendo una agradable sensación de movimiento. Otros efectos típicos obtenidos con velocidades lentes son las fotografías de fuegos artificiales o las fotografías nocturnas de calles, donde quedan registradas las luces de los automóviles durante la circulación.

Las velocidades de 1/60 y 1/30 pueden utilizarse para capturar imágenes de un televisor en aquellos países que utilizan el sistema NTSC (frecuencia de campo de 60 Hz). Con estas velocidades el haz del receptor de televisión realizará uno o dos barridos completos de la pantalla, de modo que las imágenes resultantes quedarán registradas en la película. Es obvio que si la cámara no está sincronizada con el receptor, los dos barridos pueden corresponder a fotogramas distintos dentro de la secuencia de vídeo, observándose el posible desplazamiento entre los dos fotogramas. Si utilizamos una velocidad de obturación de 1/60 en el sistema de televisión PAL (con 50 Hz de frecuencia de campo), el haz no tendrá tiempo de explorar toda la pantalla, con lo que observaremos una parte de la pantalla del televisor en negro. Si se usa la velocidad de 1/30, el haz no habrá tenido tiempo de realizar dos barridos completos, por lo que la zona de la pantalla donde se han producido dos exploraciones aparecerá más brillante que la zona donde el haz sólo ha pasado una vez.

No obstante, es posible realizar fotografías de calidad de imágenes estáticas en la pantalla de un ordenador (p.ej. diapositivas de una presentación). Para ello deberemos seleccionar una velocidad de exposición lenta (alrededor de 1" o 1/2"), con lo que conseguiremos que el haz pase varias mientras se registra la imagen. Si la cámara permanece inmóvil durante la exposición, la imagen quedará nítida.

Para poder realizar las fotografías con estos tiempos de exposición tan lentos será necesario reducir el brillo de la pantalla, cerrar el diafragma y posiblemente trabajar con película de baja sensibilidad. Por el contrario, si se reduce el tiempo de exposición a unos pocos cuadros, suelen aparecer problemas de diferencias de iluminación, debido a que una zona de la imagen ha sido explorada una vez menos que la otra.

El obturador en cámaras *reflex* es un elemento opaco situado por delante de la película que se retira al realizar la fotografía. En el momento de pulsar el disparador se levanta el espejo del visor, se cierra el diafragma a la apertura seleccionada y, finalmente, se abre el obturador durante el tiempo de exposición. Cuando este último se ha cerrado, el espejo y el diafragma vuelven a sus condiciones de reposo. El mecanismo de apertura del obturador debe tener una gran precisión para garantizar que los tiempos de exposición son los correctos. Además, debe tratarse de un sistema fiable y que pueda usarse prolongadamente sin ningún deterioro. Las altas velocidades de 1/8000 o 1/4000 exigen unos movimientos extremadamente rápidos que sólo pueden obtenerse a través de artilugios mecánicos de gran precisión.

La tecnología más utilizada para realizar el control del tiempo de exposición es el obturador de doble cortinilla. Su principio de funcionamiento se ilustra en la figura 5.16, donde se indica la presencia de dos cortinillas opacas que circulan desde arriba hacia abajo (existen modelos donde las cortinillas van desde la izquierda hacia la derecha) a una velocidad constante, pero no excesivamente rápida, para poder garantizar que el mecanismo tenga una vida prolongada. La primera cortinilla destapa la película, mientras que la segunda, que se dispara con un retardo igual al tiempo de exposición, vuelve a tapar la película. Para tiempos de exposición moderados o lentos, la primera cortinilla llega a destapar toda la película antes de que se dispare la segunda. Con ello, todo el fotograma es expuesto a la luz de forma prácticamente simultánea. No obstante, para velocidades altas, la segunda ventanilla se dispara mientras la primera aún está destapando una parte de la película. El resultado es que entre las dos cortinas existe una rendija a través de la cual se expone a la luz el negativo. Esta rendija se desplaza de arriba hacia abajo a la misma velocidad que las cortinillas y su distancia (D) depende del tiempo de exposición. Para tiempos de exposición muy pequeños, la distancia entre las cortinillas es extremadamente reducida. En consecuencia, todas las partes de la película quedan expuestas durante el tiempo deseado con un mecanismo sencillo y de elevada fiabilidad.

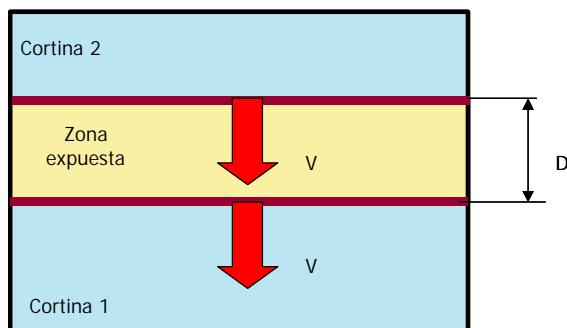


Fig. 5.16 Mecanismo de doble cortinilla para el control del tiempo de exposición

El sistema de doble cortinilla supone que cuando trabajamos con tiempos de exposición de muy pequeños no todas las partes del negativo son expuestas a la luz simultáneamente, sino que éste se va descubriendo a la misma velocidad que las cortinas. No obstante, este efecto resulta imposible de apreciar en la práctica. La doble cortinilla permite obtener velocidades de obturación de hasta 1/1000 o 1/2000. Las velocidades superiores suelen obtenerse con sistemas más sofisticados de múltiples cortinillas.

En cámaras de vídeo o fotografía digital, donde se utiliza un sensor de estado sólido, el tiempo de exposición suele regularse de forma electrónica, sin necesidad de interponer pantallas opacas delante del sensor. La obturación electrónica consiste en mantener las condiciones de adquisición de la CCD sólo durante el periodo de tiempo en que se desea realizar la exposición. Con esta tecnología se simplifican notablemente los aspectos mecánicos de la cámara y pueden obtenerse tiempos de exposición extremadamente cortos, del orden de 1/50.000.

Las cámaras de vídeo no permiten, como resulta obvio, utilizar tiempos de exposición mayores al periodo de campo. Normalmente, la adquisición de cada uno de los campos se realiza de forma secuencial, es decir, realmente cada campo corresponde a una imagen distinta, tomada a doble velocidad y se van alternando las líneas. Esto restringe que el valor máximo del tiempo de exposición coincida con el periodo de campo. En las cámaras del sistema PAL se sustituyen las velocidades de 1/60 por la de 1/50 y, habitualmente, también se incluye la velocidad de 1/100. Algunas cámaras de vídeo digital permiten que puedan utilizarse velocidades de obturación por debajo de la frecuencia de campo, utilizando tiempos bastante más bajos, para poder operar en condiciones de muy baja iluminación. En este caso se realiza un tratamiento digital de la señal de vídeo en el que se reduce la frecuencia de muestreo de las imágenes. La señal PAL que se proporciona en la salida sigue teniendo 25 imágenes por segundo, aunque algunas de ellas son obtenidas como repetición de las anteriores.

5.3.6 Profundidad de campo

La profundidad de campo es una medida del margen de distancias, alrededor del objeto de interés, donde puede considerarse que los objetos aún permanecen enfocados. Este concepto se representa esquemáticamente en la figura 5.17, donde suponemos que la lente enfoca al objeto situado en el punto A. Existirá un margen de distancias por delante y por detrás de este objeto en las que podemos considerar que se mantiene un enfoque más o menos correcto de la imagen.

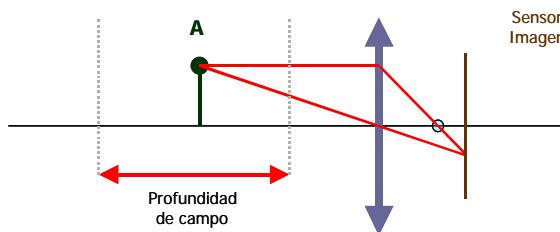


Fig. 5.17 La profundidad de campo indica el rango de distancias en el que los objetos permanecen enfocados

El concepto de la profundidad de campo resulta difuso y difícil de cuantificar. De hecho, desde un punto de vista teórico, todos los objetos que no están exactamente sobre el plano del objeto están desenfocados y los límites para indicar el grado de desenfoque dependen, como veremos, de muchos factores. En la figura 5.18 se muestra un ejemplo de cómo se forma la imagen de un objeto que está fuera de enfoque. El sensor está situado sobre el plano de imagen del objeto A. El objeto B, que está situado por delante de A, tendrá una imagen que se situará por detrás del plano de imagen. La imagen que captaremos en el sensor es un círculo cuyo diámetro será tanto más pequeño cuanto más próximos estén los dos objetos. Este círculo se denomina círculo de ambigüedad o de confusión y su diámetro es una medida del grado de desenfoque de la imagen. La profundidad de campo está directamente relacionada con el grado de definición o nitidez de las imágenes. Si una escena se registra con poca profundidad de campo sólo queda nítido el motivo principal, mientras que todos los demás objetos permanecerán borrosos. En principio es deseable que la profundidad de campo sea lo mayor posible para que aparezcan el máximo de objetos bien definidos. No obstante, es habitual que el fotógrafo fuerce la toma de imágenes con poca profundidad de campo para centrar la atención en el sujeto principal.

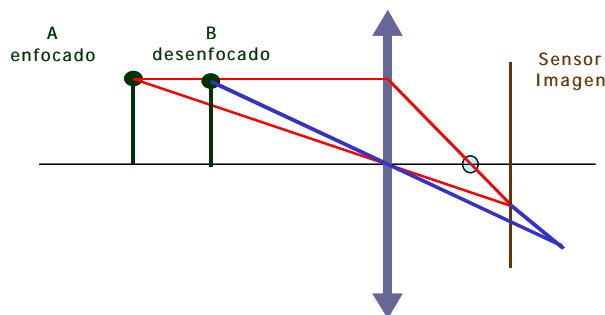


Fig. 5.18 Formación de imágenes de puntos desenfocados y círculo de confusión

El grado de desenfoque depende de varios factores que vamos a analizar con cierto detalle. En primer lugar debe considerarse el tipo de sensor utilizado. En efecto, en películas de gran sensibilidad el grano del material fotosensible es considerablemente mayor que en películas de baja sensibilidad. Es razonable, por tanto, que aparezcan nítidos los objetos que, aún no estando correctamente enfocados, tengan un círculo de confusión del mismo orden que el grano del negativo (casi tan nítidos como los que están enfocados). El tamaño del sensor también desempeña un papel importante, ya que las copias deberán ampliarse más o menos, con lo que el círculo de confusión tendrá mayor o menor importancia. En sensores de estado sólido, es evidente que el círculo de confusión no tendrá ningún efecto sobre la imagen, mientras no supere la distancia entre píxeles consecutivos.

El poder de resolución del grupo óptico también incide sobre la profundidad de campo. En un grupo óptico de baja calidad, la profundidad de campo es aparentemente superior debido a la propia falta de nitidez de los objetos que están correctamente enfocados. Podemos concluir, por tanto, que el círculo

de confusión no tendrá mayor importancia si tiene un diámetro inferior al poder de resolución del grupo óptico.

La profundidad de campo también depende de la distancia a la que se encuentre el objeto al que enfocamos. Si el objeto está próximo a la lente, cualquier pequeña variación de su posición provocará que el plano de imagen varíe considerablemente su posición. En la figura 5.19 se representan los círculos de ambigüedad de un objeto B, que está siempre a la misma distancia del objeto A. En el primer caso, el objeto A está situado próximo a la lente, mientras que en el segundo está alejado.

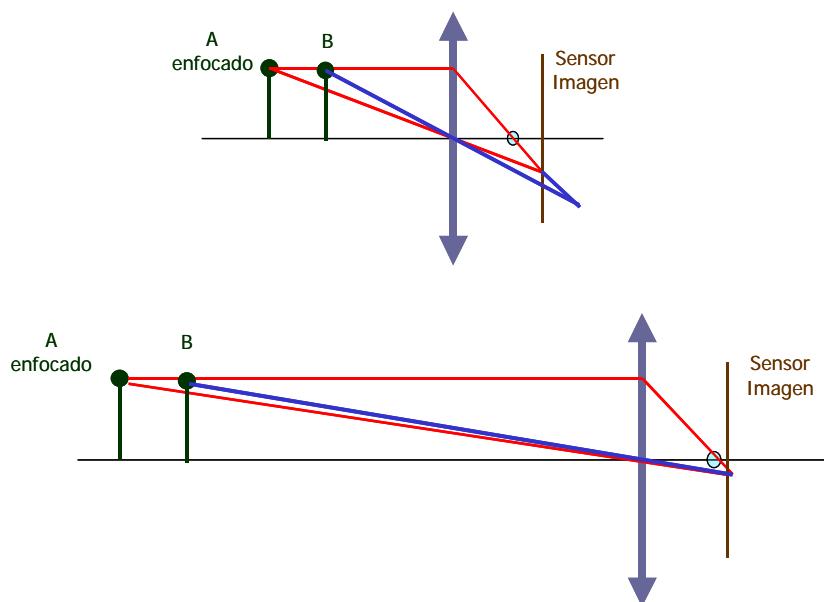


Fig. 5.19 Efecto de la distancia de los objetos en la profundidad de campo

Otro de los factores que inciden directamente sobre la profundidad de campo es la distancia focal de la lente. En la figura 5.20 se representa esquemáticamente esta dependencia para dos objetos situados a una distancia invariable de la lente y donde únicamente se modifica la distancia focal. La profundidad de campo es mucho mayor en lentes con una distancia focal reducida.

La apertura del diafragma también tiene un efecto directo sobre la profundidad de campo. Comprender la relación entre ambos factores es sumamente importante en fotografía y vídeo, puesto que a menudo suelen combinarse los ajustes de tiempo de exposición y diafragma para conseguir los efectos de profundidad de campo y movimiento de objetos deseados. La dependencia de la profundidad de campo con la apertura del diafragma se ilustra en la figura 5.21. En este caso los dos objetos están situados a una misma distancia de la lente y se utiliza una misma distancia focal. Cuando se cierra el diafragma se reduce el volumen de rayos que intervienen en la formación de punto imagen, por lo que el círculo de ambigüedad se reduce.

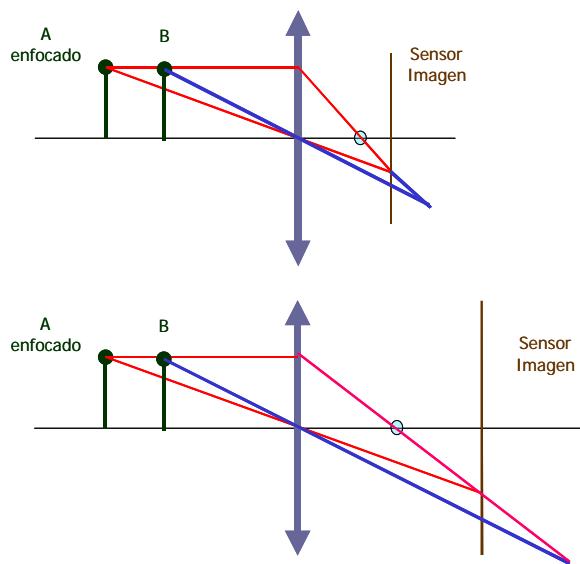


Fig. 5.20 Dependencia de la profundidad de campo con la distancia focal

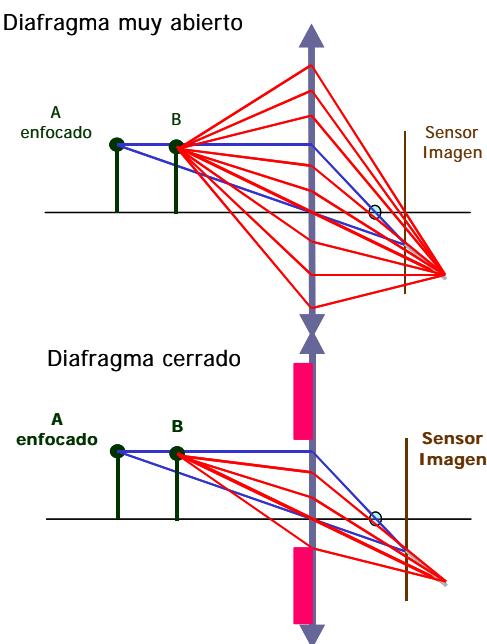


Fig. 5.21 Relación entre diafragma y profundidad de campo

A modo de síntesis sobre los efectos del diafragma y el obturador en la calidad final de la imagen, podemos concluir que la pérdida de nitidez puede ser debida, bien a que el objeto no está correctamente enfocado (fuera de campo), o a que el propio movimiento del objeto produce un desenfoque por tiempo de integración. Ambas pérdidas de definición son contrapuestas, ya que si queremos reducir el desenfoque por integración debe disminuirse el tiempo de exposición, lo que, para mantener la luz total que incide sobre el sensor, exige aumentar la apertura del diafragma, con la consiguiente pérdida en profundidad de campo. Así, si damos prioridad a la velocidad del obturador para garantizar que no aparecen imágenes borrosas originadas por el movimiento de los objetos o de la cámara, deberemos aceptar una reducción en la profundidad de campo. Por el contrario, si deseamos obtener una gran profundidad de campo, deberemos cerrar el diafragma, por lo que es posible que los objetos que tienen un movimiento elevado aparezcan borrosos.

Una alternativa para mejorar la profundidad de campo y reducir el enfoque por movimiento consiste en aumentar la sensibilidad del sensor. Con ello se requiere menor cantidad de luz para una exposición correcta, con lo que puede trabajarse con aperturas y tiempos de exposición más pequeños. No obstante, la reducción de la sensibilidad de la película supone un aumento del grano de la película (pérdida de resolución espacial) y una pérdida en colores puros. En sensores CCD puede aumentarse la ganancia de la señal electrónicamente, pero esto produce un aumento del ruido en la imagen.

Si la sensibilidad del sensor es fija, las condiciones de iluminación son las que determinan la fracción de luz que debemos dejar pasar hacia el sensor. Por lo tanto, en condiciones de iluminación diurna resultará más fácil obtener fotografías nítidas que en condiciones de baja iluminación.

Las cámaras económicas que no disponen de la capacidad de enfocar las imágenes suelen utilizar aperturas de diafragma muy pequeñas y ópticas de gran angular. Ambos factores tienden a aumentar la profundidad de campo del fotograma, por lo que, si el sensor está situado sobre el plano focal, todas las imágenes que estén a una cierta distancia de la cámara quedarán correctamente enfocadas. No obstante, el problema inherente a este tipo de cámaras es que sólo pueden trabajar en condiciones de fuerte iluminación.

Es posible cuantificar las relaciones entre todos estos parámetros y su incidencia sobre la profundidad de campo. Las ecuaciones más utilizadas para estimar la profundidad de campo vienen dadas en función del parámetro p que se define como:

$$p = \frac{4f \cdot \tan^2(\alpha_H / 2)}{N \cdot \left(\frac{W}{H}\right) \cdot b} \quad (5.11)$$

donde el número f representa la apertura del diafragma, α_H el ángulo de visión horizontal, W/H la relación de aspecto, b la anchura del sensor y N la resolución vertical del sensor (número de líneas que pueden representarse).

La profundidad de campo disminuye cuando p tiende a cero y aumenta si p crece. La ecuación 5.11 relaciona como incide cada uno de los parámetros que hemos mencionado sobre la profundidad de campo. Así, un aumento del número f o del ángulo de visión (disminución de la distancia focal) produce un incremento de la profundidad de campo. Análogamente, un aumento de la sensibilidad del

sensor representa una disminución de la profundidad de campo. El único parámetro que parece contradictorio con los comentarios que hemos realizado es que un aumento del tamaño del sensor (b) parece conducir a una disminución de la profundidad de campo. Para resolver esta aparente contradicción es conveniente expresar el ángulo de visión en función del tamaño del sensor y de la distancia focal:

$$\tan(\alpha_H / 2) = \frac{b}{2 \cdot F} \Rightarrow p = \frac{f \cdot b}{N \cdot F^2 \left(\frac{W}{H} \right)} \quad (5.12)$$

donde ahora resulta evidente que un aumento del tamaño del sensor conduce a una mejora de la profundidad de campo.

El parámetro p puede relacionarse con la distancia del objeto más próximo y más lejano que permanecerán enfocados. De acuerdo con el esquema de la figura 5.22, las distancias entre la lente y los límites de la profundidad de campo vienen dados por:

$$a_v = \frac{a}{1 + p \cdot a} \quad a_F = \frac{a}{1 - p \cdot a} \quad (5.13)$$

lo que resulta en una profundidad de campo total:

$$a_{campo} = a_F - a_v = \frac{p \cdot a^2}{1 - p^2 \cdot a^2} \quad (5.14)$$

que nos relaciona la profundidad de campo con la distancia del objeto y el parámetro p . Es usual que los objetivos de las cámaras incorporen en el propio anillo de diafragmas y rueda de enfoque una tabla simplificada que indica la profundidad de campo en metros y en pies.

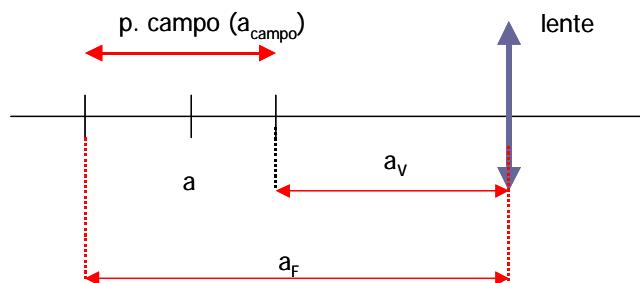


Fig. 5.22 Cálculo de los límites de la profundidad de campo

Evidentemente, las ecuaciones 5.13 y 5.14 sólo tienen sentido para denominadores positivos, puesto que en caso contrario no se producen imágenes reales sobre el sensor. Por otra parte el parámetro p es

generalmente pequeño y menor que la unidad, lo que conduce a que la profundidad de campo aumente cuando p crece. En la figura 5.23 se muestra la relación entre a_{campo} y p en la región de validez de este último.

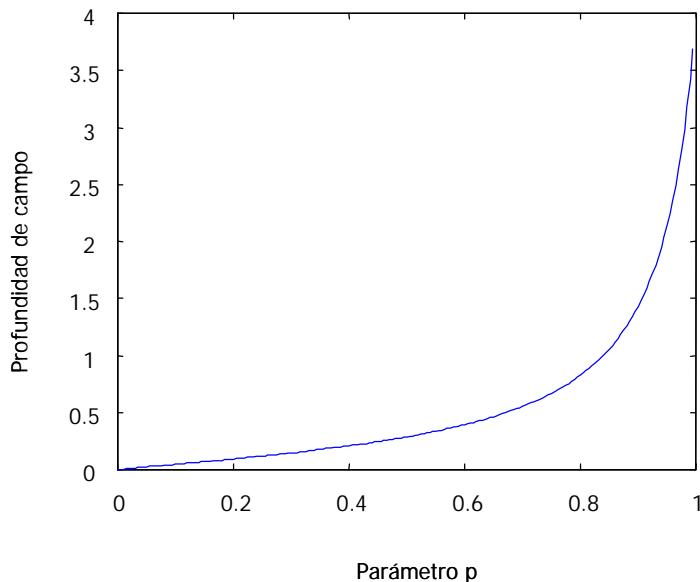


Fig. 5.23 Relación entre profundidad de campo y parámetro p

5.4 Fotometría

La medida de la cantidad de luz en una escena es fundamental para determinar el tiempo de exposición y la apertura del diafragma, con el que conviene realizar la toma de imágenes. Prácticamente todas las cámaras actuales incorporan un fotómetro o exposímetro que realiza los cálculos automáticos de la apertura y diafragma con el que se recomienda tomar la escena. Estos exposímetros permiten seleccionar el tipo de medición que se desea realizar: global, para toda la imagen que se observa en el visor, o de una zona o región pequeña de la imagen, o incluso de un único punto de la escena. Estos exposímetros también pueden trabajar con prioridad de velocidad o de diafragma. En el primer caso, el fotógrafo especifica la velocidad de obturación con la que desea realizar la toma y el sistema calcula automáticamente, en función de la sensibilidad del sensor, el diafragma adecuado. En el modo de prioridad de diafragma se determina automáticamente la velocidad de exposición en función de la apertura seleccionada.

Las unidades fotométricas también se utilizan para especificar la sensibilidad de los sensores de imagen basados en tubos de imagen o células CCD, que generalmente se proporcionan en el mínimo número de *lux* con el que pueden trabajar. En el caso de película fotosensible la sensibilidad se suele expresar en unidades ASA (*American Standard Association*), que actualmente ya son de carácter universal (ISO). Una sensibilidad de 100 ASA permite realizar tomas en condiciones de luz ambiental

diurna. Doblar el número de ASA de la película significa doblar su sensibilidad, lo que permite aumentar en un paso o la apertura del diafragma o la velocidad de obturación. El aumento de la sensibilidad de la película supone, por lo general, una reducción de la resolución espacial (menor detalle) y una pérdida en matices y pureza de color. Casi todas las cámaras modernas incorporan un lector de sensibilidad de la película y realizan los cálculos de exposición y apertura en función de ésta.

En este apartado se exponen las unidades básicas que se utilizan para medir la intensidad o energía de la luz. El objetivo es comprender las relaciones básicas entre las unidades más utilizadas para medir la intensidad luminosa (*candela*), el brillo, el flujo luminoso (*lumen*), y la iluminancia de los objetos (*lux*). Estas unidades se utilizan para especificar las características de iluminación de focos de luz, monitores, iluminación incidente en los objetos de una escena e iluminación recibida. Su uso y sus relaciones resultan a menudo confusos.

5.4.1 Intensidad luminosa: la candela

La intensidad luminosa proporcionada por una fuente puntual en cada dirección del espacio se mide en *candelas*. La definición formal de esta unidad establece que una *candela* (en la dirección θ) se corresponde con una fuente de luz monocromática de 540×10^{12} Hz que tiene una intensidad radiante en esa dirección de 1/683 Vatios (Watts) por estereoradián. Esta definición es bastante confusa y vamos a intentar esclarecer su significado examinando con detalle cada uno de los parámetros que intervienen y su interpretación en casos prácticos.

Históricamente, la definición de candela se basaba en la luz que emitía una llama. Posteriormente se redefinió en función del brillo emitido por 1 cm^2 de platino fundido solidificándose. Actualmente se define como la potencia radiada en Vatios por una fuente de luz monocromática a la frecuencia de 540×10^{12} Hz. Esta frecuencia se corresponde con la longitud de onda de 555 nm, en la que el sistema visual humano medio presenta una mayor sensibilidad en condiciones de luz diurna. El uso de la frecuencia en vez de la longitud de onda pretende que la definición de la unidad sea independiente del medio de transmisión (la frecuencia no varía con el medio de transmisión pero la longitud de onda sí).

El uso de 1/683 Watts para definir una unidad de candela se debe a la intención de la nueva definición de candela coincidiera, aproximadamente, con la antigua (intensidad emitida por 1 cm^2 de platino fundido).

La intensidad luminosa se mide en función de la potencia transmitida por ángulo sólido. Con ello se pretende que la medida caracterice la fuente de luz independientemente de la distancia a la que es observada. La unidad de ángulo sólido es el estereoradián, que se define como el ángulo subtendido por un cono que parte de la fuente de luz e intersecta a la esfera de radio unidad en una superficie de 1 m^2 . En la figura 5.24 se representa esquemáticamente esta unidad. El ángulo sólido subtendido por cualquier otra porción de superficie situada sobre la circunferencia unidad puede determinarse directamente mediante proporcionalidad. Así, una superficie de $0,25 \text{ m}^2$ sobre la esfera unidad equivale a un ángulo subtendido de $0,25$ estereoradianes. Según esta definición, es evidente que la esfera de radio unidad completa, que tiene un área de 4π , subtende un ángulo sólido de 4π estereoradianes. Los ángulos subtendidos por superficies que no se encuentran sobre la esfera unidad pueden determinarse proyectando la superficie de interés sobre la fuente de luz y calculando el área

resultante cuando esta proyección intersecta sobre la esfera unidad. Los detalles asociados a este tipo de cálculos se ilustran en la figura 5.25, donde se considera la proyección de un diferencial de superficie situado a una distancia R del foco de luz y que forma un ángulo θ con el vector unitario en la dirección radial. El diferencial de ángulo sólido subtendido por esta superficie viene dado por:

$$d\Omega = 4 \cdot \pi \cdot \frac{dS \cdot \cos \theta}{4 \cdot \pi \cdot R^2} = \frac{dS \cdot \cos \theta}{R^2} \quad (5.15)$$

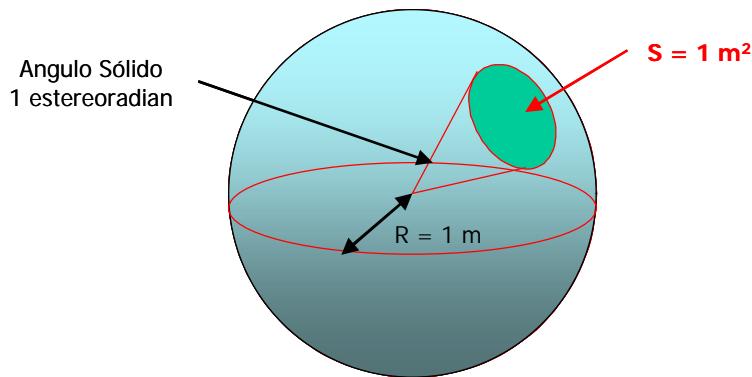


Fig. 5.24 Definición de estereorradián

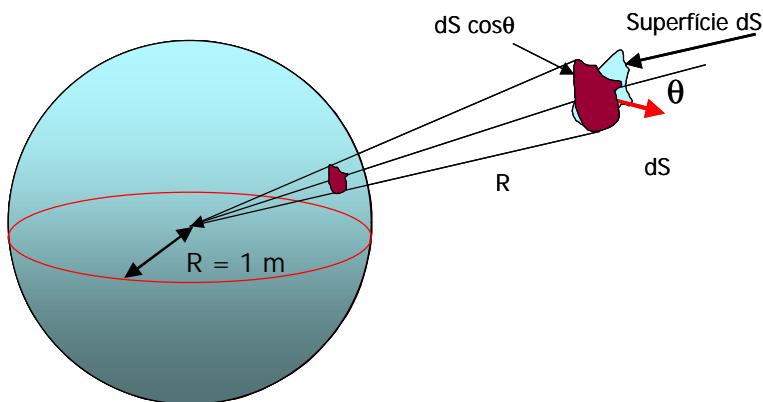


Fig. 5.25 Cálculo del ángulo subtendido por un diferencial de superficie situado a una distancia R

La candela es una unidad que se utiliza para definir las características de distribución direccional de la luz emitida por una fuente puntual. Los fabricantes de lámparas, focos o reflectores suelen proporcionar un diagrama direccional que indica cómo se distribuye la energía luminosa en el espacio.

En la figura 5.26 se representa uno de estos diagramas, cuya interpretación resulta bastante simple. En este ejemplo, la intensidad luminosa en la dirección de 60° es de aproximadamente 300 cd. Esto significa que si nos situamos a una cierta distancia R de la fuente de luz en esta dirección y medimos la potencia que se recibe en una superficie dS , el resultado que obtendremos es:

$$dP = 300 \text{ (cd)} \cdot d\Omega = 300 \cdot \left(\frac{1}{683} \text{ (W/sr)} \right) \cdot \frac{dS \cdot \cos \theta}{R^2} \text{ (sr)} = \frac{300 \cdot \cos \theta \cdot dS}{683 \cdot R^2} \text{ (W)} \quad (5.16)$$

La correcta interpretación de estos cálculos proporciona mucha información sobre el significado de la candela como unidad de medida de la intensidad lumínica. Debe tenerse en cuenta que en esta ecuación el ángulo θ representa el ángulo existente entre la perpendicular a la superficie en la que se realiza la medida y la dirección radial. Si suponemos que la superficie está orientada hacia la fuente la ecuación puede simplificarse a:

$$dP(\phi) = \frac{C(\phi) \cdot dS}{683 \cdot R^2} \text{ (W)} \quad (5.17)$$

donde hemos supuesto que $C(j)$ representa el diagrama de intensidad luminosa expresado en candelas. El factor 683 que aparece en el denominador realiza la conversión a Vatios de la potencia lumínica. De acuerdo con esta ecuación, la potencia recibida en la superficie de medida disminuye a medida que nos alejamos de la fuente. Esto es debido a que el ángulo sólido subtendido por la superficie también disminuye. Si aumentáramos la superficie a medida que nos alejamos de la fuente, de manera que el ángulo sólido se mantuviera constante, la potencia recibida en la superficie también se mantendría constante. Esto indica que la radiación luminosa se dispersa a medida que nos alejamos del punto emisor de luz. $C(j)$ caracteriza por tanto la intensidad luminosa que radia la fuente sin tener en cuenta los efectos de la distancia ni las características de la superficie sobre la que incide la luz.

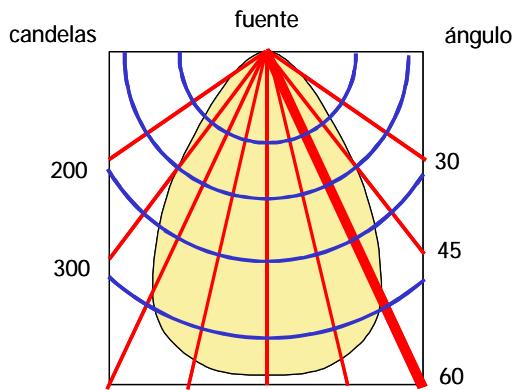


Fig. 5.26 Representación del diagrama de intensidad luminosa en candelas de una fuente de luz

Análogamente, podemos determinar $C(j)$ a partir de las medidas que obtengamos en el sensor de potencia lumínica simplemente despejando esta función en la ecuación (5.17). Si el sensor tiene una

superficie total equivalente S_{sensor} y mide una potencia P expresada en Vatios, la intensidad luminosa $C(\mathbf{j})$ que obtendremos cuando realizamos la medida desde una distancia R , con el sensor correctamente orientado hacia la fuente es:

$$C(\phi) = \frac{683 \cdot P \cdot R^2}{S_{sensor}} \quad (5.18)$$

Hemos visto que la candela está definida para radiaciones monocromáticas en la región de máxima sensibilidad del sistema visual. No obstante, la mayoría de fuentes luminosas son espectralmente complejas, por lo que es necesario tener en cuenta todas las componentes espectrales radiadas para que la medida sea indicativa de la iluminación real que produce la fuente. La forma de tener en cuenta las diferentes componentes espectrales en que se descompone una fuente de luz real es ponderar su energía mediante la respuesta de sensibilidad relativa del ojo humano medio. Esta curva está especificada y tabulada por el CIE (*International Commission of Illumination*) y tiene una forma parecida a la de la figura 5.27. Según esta curva, una fuente que radie a una frecuencia monocromática de 510 nm (verde) tiene un rendimiento visual del 50% respecto a la frecuencia de 555 nm. Esto significa que deberá radiar el doble de potencia en Vatios para producir las mismas candelas. Todas las medidas fotométricas tienen en cuenta la respuesta psicofísica del sistema visual humano, por lo que no proporcionan una medida absoluta de la energía lumínica, sino de su percepción. Según lo expuesto, podríamos modificar la ecuación (5.18) para tener en cuenta las fuentes no monocromáticas sino que rodian un espectro de potencia $P(\lambda)$. La ecuación resultante sería:

$$C(\phi) = \frac{683 \cdot R^2}{S_{sensor}} \cdot \int_{visible} P(\lambda) \cdot F(\lambda) \cdot d\lambda \quad (5.19)$$

donde $F(\lambda)$ representa la curva de sensibilidad relativa del sistema visual humano representada en la figura 5.27.

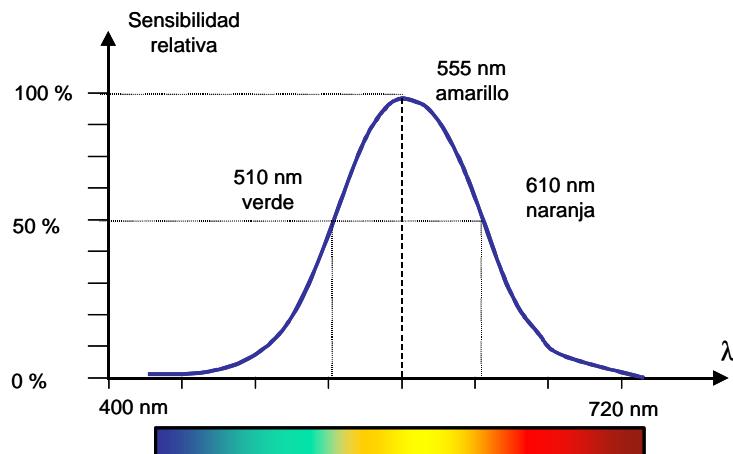


Fig. 5.27 Respuesta de sensibilidad relativa del sistema visual humano

La luz es una radiación electromagnética y, como tal, su potencia y energía puede ser medida con unidades de radiación convencionales. Esta forma de medir la intensidad, potencia y energía de las fuentes luminosas se conoce con el nombre de medidas radiométricas y, en contraposición con las medidas fotométricas que estamos presentando, no tienen en cuenta la respuesta del sistema visual.

La definición de la intensidad luminosa como candela supone que la fuente de luz es puntual o de reducido tamaño. El flujo de energía se define en función del ángulo sólido y, como hemos visto, es independiente de la distancia desde la que se realiza la medida. Si la fuente es puntual, la sensación de brillo (cuantificada como luminancia) que produce al observarla desde una cierta distancia puede ser muy fuerte, aunque su intensidad luminosa sea sólo de una candela. En cambio, si el tamaño de la fuente es algo mayor, aunque su energía total sea la misma, producirá una sensación de brillo más moderada. El brillo o luminancia de una fuente se mide en candelas por metro cuadrado (cd/m^2), de este modo, se tiene en cuenta la superficie real de la fuente de luz. Debe observarse que, en general, cualquier objeto visible puede interpretarse como una fuente de luz, independientemente de que realmente actúen como fuentes de luz o que reflejen la luz procedente de otro foco. De acuerdo con ello, el brillo o luminancia de un objeto depende de la iluminación a la que esté expuesto y de sus características para absorber y reflejar la luz.

5.4.2 El flujo luminoso: lumen

Hemos visto que la candela, al ser independiente de la distancia, puede interpretarse como una propiedad de la fuente de luz. Por tanto, se requiere una unidad distinta para reflejar la luz que fluye en el espacio y que finalmente incide sobre los objetos o los sensores. Esta unidad es el *lumen* (lm) y de hecho ya la hemos estado utilizando cuando definíamos la relación entre la candela y la potencia recibida por un sensor de luz situado a una cierta distancia de la fuente.

El flujo luminoso de una fuente cuya intensidad son $C(\phi)$ candelas, en un ángulo sólido $d\Omega$, viene dado por:

$$dF = C(\phi) \cdot d\Omega \quad (5.20)$$

la unidad de medida es el *lumen* que resulta equivalente a una candela por estereorradián (cd sr).

Consideremos como ejemplo las dos fuentes de luz representadas en la figura 5.28. La primera fuente radia una intensidad de una candela hacia todo el espacio situado frente a ella (la mitad de una esfera, es decir, 2π estereoradianes). La segunda fuente también radia una intensidad de una candela, pero distribuida únicamente sobre un ángulo sólido frontal de 1 estereorradián. Si determinamos el flujo total de la fuente radiado al espacio mediante la ecuación (5.20), obtendremos que la primera fuente emite un total de 2π lm mientras que la segunda emite únicamente 1 lm.

Una medida muy utilizada en catálogos de fuentes de luz son los lumen por Vatio (lm/W). Para una fuente que radia en distintas longitudes de onda, esta medida proporciona una idea de la eficiencia luminosa del dispositivo teniendo en cuenta la respuesta del sistema visual humano. Así, las lámparas de tungsteno, que se caracterizan por su luz cálida (mucha componente roja), tienen una eficiencia reducida (mucho consumo para la luminosidad que producen), debido a que la sensibilidad del sistema

visual humano a esta componente espectral es baja. En cambio, los fluorescentes radian mucha energía en las componentes del amarillo, verde y rojo, por lo que tienen una elevada eficiencia en lumen por Vatio.

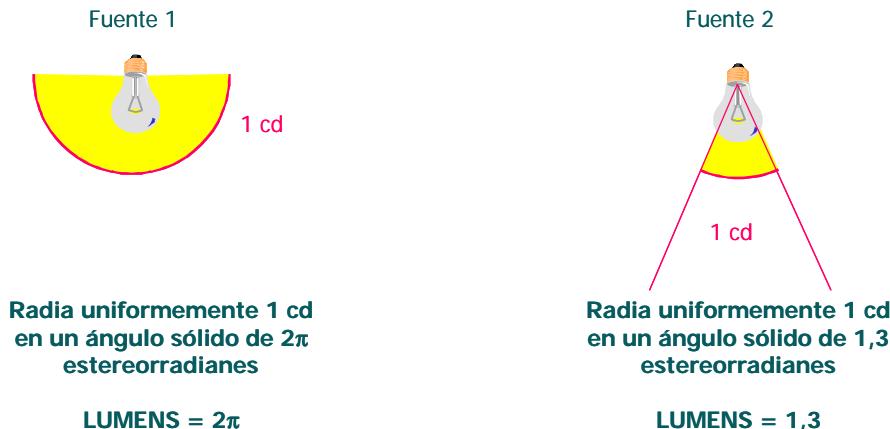


Fig. 5.28 Relación entre candela y lumen para fuentes directivas

5.4.3 La iluminancia: lux

La iluminancia se utiliza para determinar la densidad de energía lumínica que incide sobre un objeto y se obtiene como el número de lumens por unidad de área. La unidad de medida es el *lux* que es equivalente a lm/m^2 . Puede calcularse directamente a partir del diagrama de la fuente y depende de la posición relativa entre la fuente y el objeto y su distancia. Cuando un objeto es iluminado por varias fuentes, la iluminación total se obtiene mediante la superposición de todas las iluminaciones individuales (suponemos que las fuentes de luz son incoherentes). Existen paquetes de software que permiten determinar la iluminación que reciben los objetos dentro de un recinto a partir de la distribución y las características de las fuentes de luz.

Consideremos como ejemplo el cálculo de la iluminancia que reciben dos objetos A y B de una fuente de luz como la representada en la figura 5.29. En este ejemplo, es evidente que el objeto B, al estar situado fuera del diagrama de radiación de la fuente, no recibirá ninguna iluminación. Para determinar la iluminancia que recibe el objeto A deberemos tener en cuenta su distancia a la fuente y el área del objeto. El ángulo equivalente subtendido por el objeto es:

$$\Omega_{eq} = 4\pi \cdot \frac{A}{4\pi R^2} = \frac{A}{R^2} \quad (5.21)$$

Para determinarlo hemos tenido en cuenta la proporción entre el área del objeto A y el área de una esfera situada a la misma distancia, de la que sabemos que tiene un ángulo subtendido equivalente de 4π estereoradianes. Una vez determinado el ángulo equivalente resulta trivial realizar el cálculo del flujo luminoso que incide sobre el objeto:

$$F = C(\phi) \cdot \Omega_{eq} = \frac{A}{R^2} \quad (\text{lm}) \quad (5.22)$$

donde no ha sido necesario realizar la integración debido a que $C(\phi)$ es constante en la región de interés. Finalmente, podemos determinar la iluminancia teniendo en cuenta el área del objeto.

$$I = \frac{F}{A} = \frac{1}{R^2} \quad (\text{lux}) \quad (5.23)$$

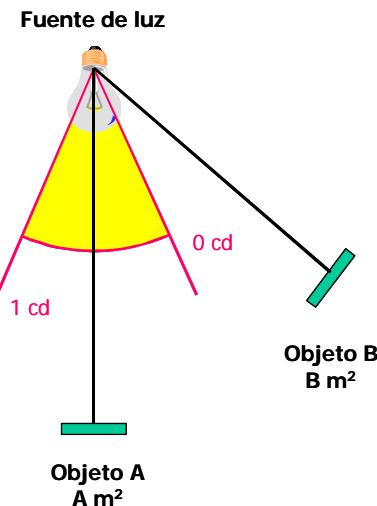


Fig.5.29 Cálculo de la iluminancia que reciben los objetos A y B

Es importante distinguir entre la iluminancia y la luminancia de un objeto. La primera está relacionada con la cantidad de luz que incide sobre el objeto, mientras que la segunda es una medida de la sensación de brillo, por lo que dependerá tanto de la luz incidente como de las propiedades de absorción y reflexión de la luz del objeto. Para medir la iluminancia se utilizan sensores electrónicos que evalúan la cantidad de luz incidente sobre el objeto. Deben situarse por tanto en una posición próxima y orientarse paralelos al objeto. La luminancia se mide usando el mismo equipo, pero con la superficie sensora encarada hacia el objeto. Entre ambas medidas hay un factor de escala y un cambio de unidades.

Para medir la luminancia y la iluminancia de forma aproximada, puede utilizarse el exposímetro de una cámara *reflex* automática. Los lux pueden medirse colocando una hoja de papel o pantalla blanca en el punto en que deseamos medir la iluminancia, para que refleje toda la luz. Dejando la cámara en modo automático podemos utilizar la siguiente ecuación:

$$\text{Lux} = 50 \cdot f^2 / (T_{\text{exposición}} \cdot ASA_{\text{película}}) \quad (5.24)$$

donde tanto el número focal como el tiempo de exposición se obtienen de los valores que ha ajustado automáticamente la cámara. La medida suele tener un error considerable debido a que los pasos de números f y tiempos de exposición son logarítmicos, pero proporciona una primera aproximación a la medida de la iluminancia. La medida de la luminancia en cd/m² puede obtenerse con una fórmula parecida:

$$\text{Luminancia} = 12,4 \cdot f^2 / (T_{\text{exposición}} \cdot \text{ASA}_{\text{película}}) \quad (5.25)$$

En este caso, no es necesario utilizar la superficie blanca, ya que lo que deseamos estimar es la sensación de brillo que producen los objetos. Si se dispone de una cámara que pueda realizar mediciones puntuales de los ajustes del exposímetro, resulta muy fácil estimar a distancia la luminancia de los distintos objetos de la escena.

5.5 Sensores de imagen de estado sólido

Actualmente, casi todos los tipos de cámaras de vídeo o de fotografía digital utilizan sensores de imagen de estado sólido, también conocidos como dispositivos de acoplamiento de carga (*CCD – Charge Coupled Device*). Estos dispositivos empezaron a desarrollarse a principios de la década de los 70 y aunque en sus inicios sólo permitían obtener resoluciones bajas, actualmente se han convertido en los sensores de imagen más populares y casi han desplazado completamente otros tipos de tecnologías. El principio de funcionamiento de los sensores CCD se basa en la estructura de los condensadores MOSFET que se representa en la figura 5.30. El electrodo de polisilicio tiene propiedades metálicas, pero, a diferencia de los condensadores MOSFET convencionales, es transparente permitiendo el paso de la luz y la por tanto de generación eléctrica a partir de los fotones incidentes sobre el semiconductor. El electrodo metálico está aislado del semiconductor tipo P mediante una película de dióxido de silicio. Si se aplica una tensión positiva al electrodo, gran parte de los electrones generados a partir de la descomposición de un fotón en un par electrón-hueco se concentrarán, por atracción eléctrica, bajo el electrodo positivo. Si no existiera esta tensión positiva los electrones y los huecos se recombinarían dando lugar a una carga neta nula. En consecuencia, la carga almacenada debajo del contacto de polisilicio es proporcional a la cantidad de luz incidente sobre la superficie del electrodo.

La estructura básica de una célula CCD consiste en una cadena de condensadores MOSFET, con un substrato tipo N común a todos ellos y situados suficientemente próximos como para que pueda realizarse una interacción entre las cargas a partir del control electrónico de las tensiones aplicadas a los terminales. Es importante observar que la carga almacenada es un valor analógico que de alguna manera deberemos extraer de la estructura CCD para obtener su valor. Las células CCD se usan en multitud de aplicaciones, ya sea como dispositivos de almacenamiento de datos analógicos, memorias digitales de alta capacidad, unidades de retardo analógicas o elementos sensores de imagen. Únicamente en este último caso, los electrodos metálicos deben ser transparentes para permitir el paso de la luz.

El control electrónico de un sensor de imagen del tipo CCD se realiza en dos fases: adquisición de imagen y lectura de datos. Las tensiones aplicadas a los electrodos durante la fase de adquisición de imagen se representa en la figura 5.31 para una célula CCD de tres fases (las más populares). Nótese

que en esta configuración se aplica tensión positiva a uno de cada tres terminales, de modo que las cargas acumuladas quedan registradas bajo los contactos de estos electrodos. El tiempo durante el que se aplican las condiciones de adquisición de la imagen es equivalente al tiempo de exposición del sensor, ya que la carga total almacenada será proporcional al valor medio de la luz incidente durante este periodo de tiempo.

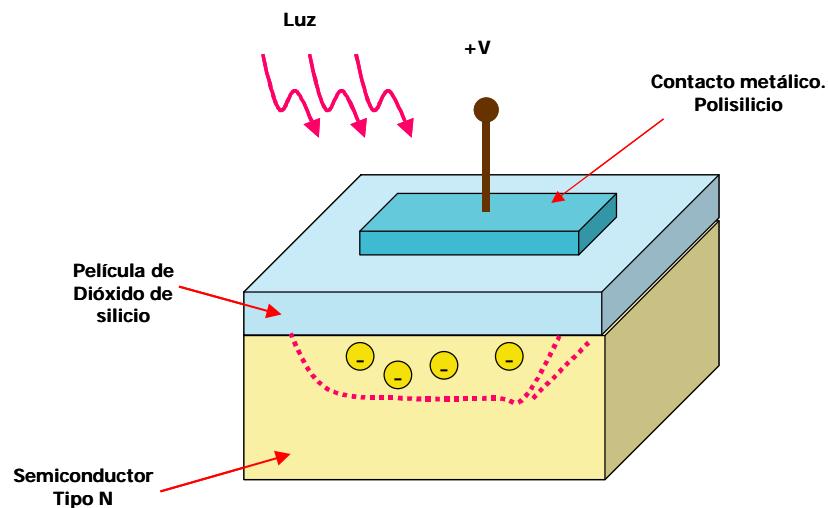


Fig. 5.30 Diagrama básico de un condensador MOSFET usado en células CCD

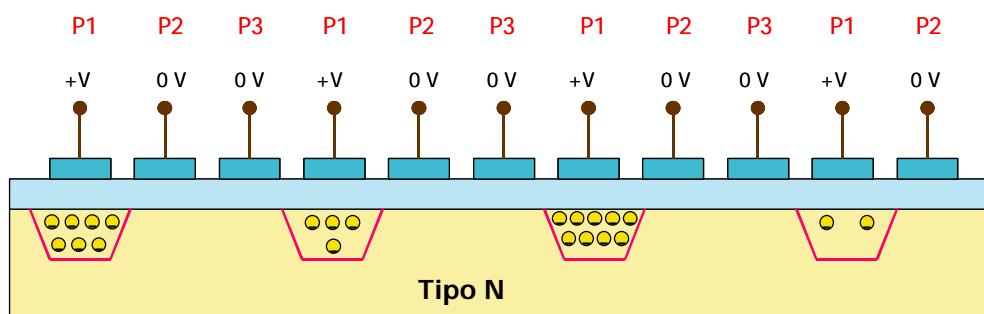


Fig. 5.31 Estructura CCD durante la fase de adquisición de imagen

Una vez registrada la luz incidente bajo la estructura CCD debe procederse a su lectura. El proceso de lectura se realiza modificando las tensiones aplicadas sobre los terminales P1, P2 y P3, de forma que externamente se provoca una transferencia de las cargas entre los condensadores adyacentes. Para que

la lectura sea factible es necesario incorporar en uno de los extremos de la CCD un elemento que pueda sensar la cantidad de carga desplazada. La forma en que se realiza esta lectura de la carga se representa también en la figura 5.32, donde se incluye un acabamiento con un dopado tipo P del substrato de silicio. Esta terminación resulta equivalente a un transistor MOSFET donde, ahora, la tensión del penúltimo electrodo actúa como señal de puerta, permitiendo la lectura de la carga y su posterior amplificación.

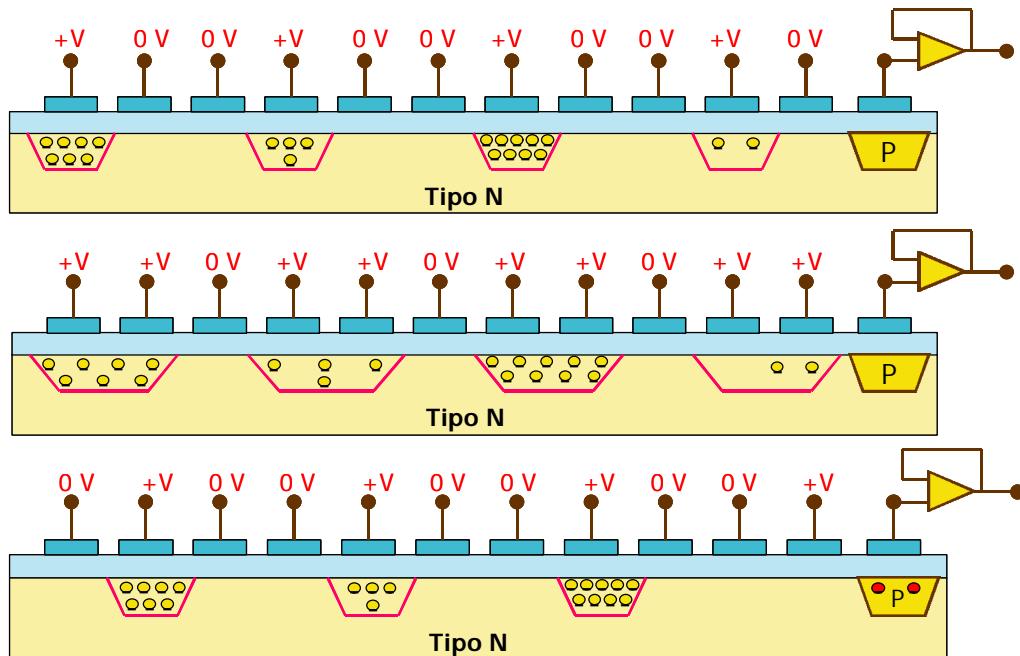


Fig. 5.32 Desplazamiento de las cargas para la lectura de la CCD

El desplazamiento de la carga se obtiene aplicando señales de reloj externas a los electrodos de la estructura. En la fase de adquisición la carga se encuentra bajo los contactos P1, que están a una tensión positiva. Para desplazar la carga al contacto P2 se aplica una tensión positiva a este electrodo, manteniendo constante la tensión del electrodo P1. Bajo estas condiciones, la carga almacenada bajo el contacto P1 se reparte ahora entre los contactos P1 y P2. Si ahora bajamos la tensión del contacto P1, toda la carga se situará bajo el terminal P2. De modo análogo, es posible desplazar la carga de P2 a P3 y de P3 a P1, desplazándose todas las cargas hacia la derecha hasta que son leídas en la célula de terminación de la estructura.

Normalmente, las señales de reloj para la lectura de la CCD se proporcionan desde un circuito integrado externo. Las cámaras actuales suelen integrar las funciones de relojes, drivers, control y acondicionamiento de señales en circuitos integrados externos al sensor CCD. La generación de las señales de reloj para la lectura no es trivial y requiere un circuito integrado especializado para proporcionar las formas de onda y niveles de tensión correctos. La mayoría de cámaras operan con baterías externas por lo que requieren diversos reguladores de tensión para generar internamente las

tensiones que requiere el desplazamiento de las cargas. En la figura 5.33 se representa un diagrama básico de las formas de onda que deben aplicarse a los terminales P1, P2 y P3 para el desplazamiento de las cargas. El caso que hemos considerado corresponde a las CCD de tres fases (por las tres señales de reloj independientes que son necesarias). Evidentemente, el desplazamiento de las cargas puede realizarse hacia la derecha o hacia la izquierda sin más que modificar la relación entre las señales de control. Existen configuraciones de CCD que pueden funcionar con dos fases de reloj, aunque su uso se restringe a aplicaciones muy específicas.

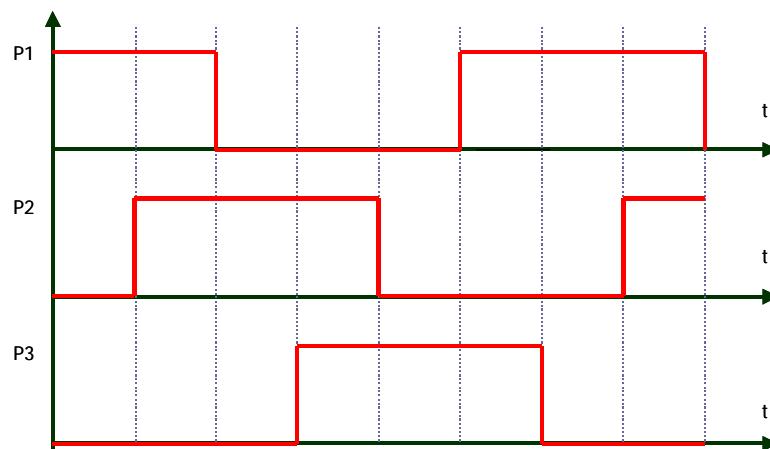


Fig 5.33 Diagrama de las señales de lectura en una CCD de tres fases

Las aplicaciones de los dispositivos CCD para memorias analógicas o células de retardo tienen una estructura análoga a la utilizada para los sensores de imagen. En este caso, no se permite que la luz incida en el interior de la estructura y las cargas se introducen externamente mediante una configuración de transistor MOSFET similar a la que se utiliza para la lectura. Las memorias suelen tener una configuración FIFO (primer dato en entrar – primero en salir), ya que los elementos de entrada y salida de cargas se sitúan en extremos opuestos. En las células de retardo el tiempo de retardo puede controlarse a partir de la frecuencia de las señales de desplazamiento de cargas. Si se aplican frecuencias altas, las cargas fluyen rápidamente entre el extremo de entrada y el de salida de la CCD, dando lugar a un retardo pequeño entre entrada y salida.

La configuración de CCD que hemos considerado hasta este momento se conoce como CCD de línea debido a que los sensores están situados uno al lado del otro formando una línea recta. Este tipo de estructuras se utilizan en *scanners*, fotocopiadoras, faxes y las denominadas cámaras lineales, que se utilizan en algunos sistemas de visión industrial por computador. Para obtener una imagen con este tipo de dispositivos es necesario que exista un desplazamiento relativo entre el objeto y el sensor CCD. La imagen se va explorando línea a línea y se reconstruye a partir del movimiento relativo existente entre el sensor y el objeto. En la figura 5.34 se representa esquemáticamente una CCD de línea donde cada cuadro representa un sensor de imagen y la terminación representa el elemento de lectura de la carga. Normalmente, se omite la representación de las señales de reloj para la lectura.

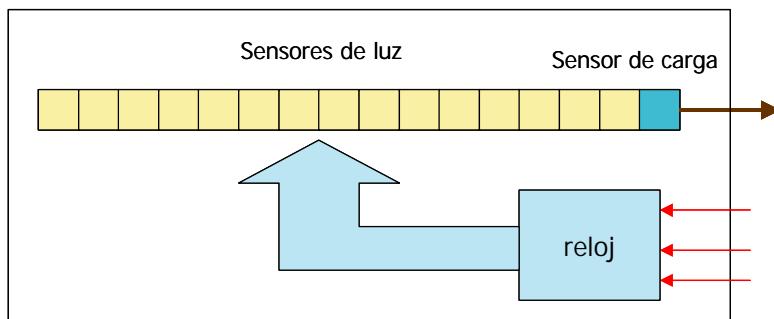


Fig.5.34 Representación de una CCD de línea.

Uno de los problemas inherentes al procedimiento de lectura de la CCD es que si la estructura se mantiene expuesta a la luz mientras se van desplazando las cargas, se va añadiendo una carga remanente a medida que éstas se desplazan hacia el terminal de lectura. El efecto es particularmente notorio cuando existe un punto de luz de gran intensidad. En este caso, todas las elementos situados a la izquierda de este punto de luz (suponemos que las cargas se desplazan hacia la derecha) adquirirán una carga adicional cuando pasen por debajo del contacto expuesto a una gran intensidad lumínica. El efecto se conoce con el nombre de *smear* y visualmente se observa como que el punto más brillante deja una estela de luz en todos los elementos de imagen situados a su izquierda y en su misma línea.

En las cámaras de vídeo actuales los sensores CCD tienen una estructura matricial, de modo que puede capturarse toda la imagen de forma simultánea sin necesidad de desplazar el sensor o el objeto. En los apartados siguientes se analizan los tipos de CCD matriciales más habituales. También se detallan los principios de algunas variantes tecnológicas de los sensores CCD que están empleándose para la captura de imágenes.

5.5.1 Sensores CCD matriciales

Un sensor matricial está formado por una agrupación de elementos CCD lineales dispuestos en forma de una matriz, tal y como se ilustra en la figura 5.35. El grupo óptico se encarga de formar la imagen sobre la superficie del sensor de modo que todos los elementos de la imagen son adquiridos simultáneamente. Una vez capturada la imagen, se efectúa la lectura de la información desplazando las cargas de cada CCD de línea hacia la salida. Normalmente, la lectura debe realizarse de forma coordinada con la señal de vídeo de salida, es decir, la lectura de una línea debe realizarse en el tiempo de línea de la señal de vídeo y la adquisición de una nueva imagen debe sincronizarse con el periodo de borrado de campo. Esta restricción impone una fuerte limitación en cámaras de vídeo que deben proporcionar una señal en tiempo real, ya que el sensor únicamente puede exponerse a la luz cuando la imagen ha sido completamente leída, con lo que no podrá iniciarse la adquisición de un nuevo campo hasta que se haya finalizado completamente la lectura del campo actual. Esta filosofía de lectura exige que el tiempo de exposición de la CCD sea inferior al tiempo total de borrado de campo, de esta forma las imágenes pueden fluir a la misma velocidad que la señal de vídeo. Por lo tanto, para cámaras de tipo PAL, el tiempo de exposición máximo coincide con el periodo de borrado

de campo, es decir, $(25/312,5)*40\text{ms} = 3,2 \text{ ms}$. Además, debe incorporarse un obturador mecánico que impida que la luz incidente modifique parte de la información de la imagen mientras se está realizando la lectura. El uso del obturador mecánico impide que aparezca el efecto de *smear* en la imagen que estaría presente si los contactos no estuvieran protegidos de la luz.

Este tipo de sensores se utiliza a menudo en aplicaciones de captura de imagen en las que no se requiere proporcionar una señal de vídeo en tiempo real, sino un único fotograma. Las aplicaciones más comunes son la obtención de imágenes de astronomía, donde pueden requerirse tiempos de exposición elevados, o aplicaciones industriales de visión por ordenador, en las que el reconocimiento de los objetos de interés puede realizarse a partir de un único fotograma.

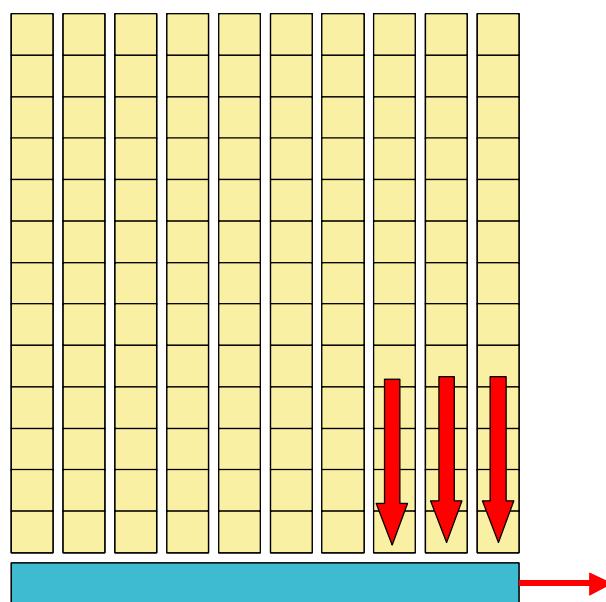


Fig. 5.35 Sensor CCD matricial

5.5.2 Sensores CCD de transferencia de cuadro (*Frame Transfer CCD*)

Las CCD de transferencia de cuadro son probablemente las más utilizadas en cámaras de vídeo. Conceptualmente son parecidas a las CCD matriciales, con la diferencia de que sólo la mitad de la estructura CCD está expuesta a la luz, mientras que la otra mitad está protegida. La idea básica de su funcionamiento consiste en adquirir la imagen utilizando la superficie expuesta a la luz y, una vez adquirida, desplazar las cargas rápidamente y en paralelo hacia la zona protegida. Con ello, es posible realizar los procesos de adquisición y lectura de los datos simultánea e independientemente. El tiempo máximo de exposición puede llegar a ser igual al periodo de campo, ya que una vez transferida toda la carga a la zona protegida puede iniciarse la captura de una nueva imagen, sin necesidad de esperar a que los datos hayan sido proporcionados a los sistemas exteriores. Además, no es necesario utilizar un obturador mecánico, ya que la imagen adquirida está protegida de la luz cuando se realiza la

conversión a señal de vídeo. Por otra parte, el *smear* depende del tiempo requerido para transferir la carga desde la región activa a la región de almacenamiento. Cuanto menor sea este tiempo menores son sus efectos sobre la imagen. En la figura 5.36 se representa la estructura de una CCD de transferencia de cuadro.

Es importante advertir que estamos asumiendo que las imágenes se están capturando a una velocidad igual a la frecuencia de campo, en vez de la frecuencia de imagen o cuadro. Esto significa que los dos campos de la señal de vídeo corresponden a instantes de muestreo distintos, lo que resulta ser habitual en muchos modelos de cámara. Este problema, y su incidencia en la resolución real de la imagen, fue considerado con cierto detalle en el primer capítulo. No obstante, en algunos modelos de cámara es posible trabajar a la frecuencia de imagen real (25 imágenes por segundo). En este caso, los dos campos pueden obtenerse a partir de las muestras de la CCD, realizando promedios entre líneas consecutivas. Así, en el campo par se promedian las líneas 1 y 2, 3 y 4, 5 y 6, etc, mientras que en el campo impar se promedian las líneas 2 y 3, 4 y 5, 6 y 7, ect.

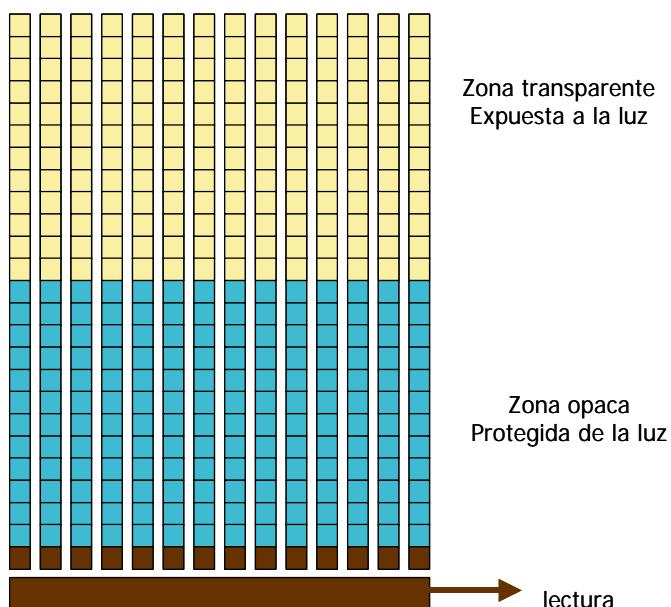


Fig. 5.36 Estructura de una CCD de transferencia de cuadro

5.5.3 Sensores CCD de transferencia de cuadro divididos (*Split Frame Transfer CCD*)

Esta arquitectura es similar a la de transferencia de cuadro convencional, con la salvedad de que la parte de almacenamiento está dividida en dos secciones (superior e inferior). La parte central de la CCD es la que queda expuesta a la luz. Una vez capturada la imagen, su transferencia a las regiones de almacenamiento se realiza simultáneamente, la mitad superior hacia la mitad superior y la inferior hacia la inferior. Esta estrategia reduce el *smear* del sensor, ya que el tiempo de desplazamiento se

reduce a la mitad. No obstante, deben incorporarse dos registros de desplazamiento y dos sensores de carga en la CCD. El esquema simplificado de esta configuración se representa en la figura 5.37.

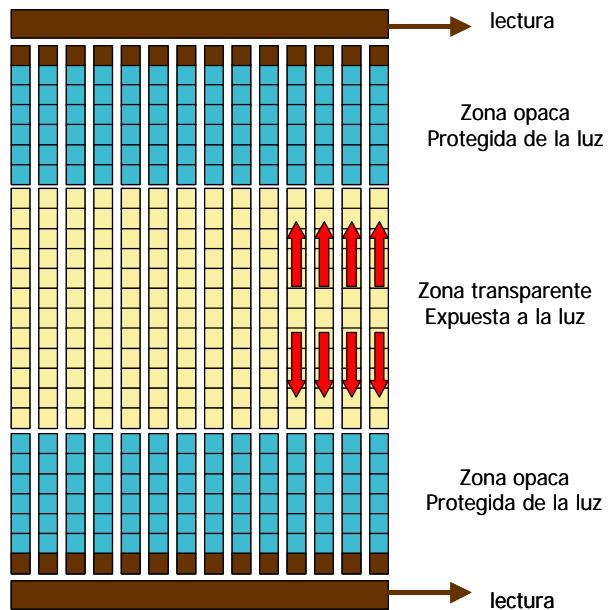


Fig. 5.37 Estructura de una CCD de transferencia de cuadro dividido

5.5.4 Sensores CCD de transferencia interlínea (Interline Transfer)

La matriz de sensores está formada, tal y como se ilustra en la figura 5.38, por estructuras CCD lineales en las que se intercalan las filas expuestas a la luz y las protegidas. Esta configuración permite que el desplazamiento de la carga a la zona protegida pueda realizarse en un único paso, desapareciendo completamente el efecto de *smear*. Una vez protegida la información de la fuente de luz, puede procederse a su lectura de forma sincronizada con la señal de vídeo. El tiempo de exposición puede controlarse directamente a partir de las señales de control de adquisición de imagen, por lo que no es necesario ningún obturador mecánico.

El problema principal de esta configuración, al compararla con las CCD de transferencia de cuadro, es que al intercalar líneas visibles con protegidas puede obtenerse una menor densidad de elementos sensibles y, por tanto, una menor resolución de imagen. Si se desea aumentar la resolución puede reducirse el área de cada elemento sensor, pero en este caso, también se disminuirá la sensibilidad de la CCD.

Las CCD de transferencia de cuadro y las de transferencia de línea son las más utilizadas en cámaras profesionales y domésticas. El uso de una u otra tecnología depende del fabricante y de las características con que se deseé dotar la cámara: alta sensibilidad o *smear* reducido.

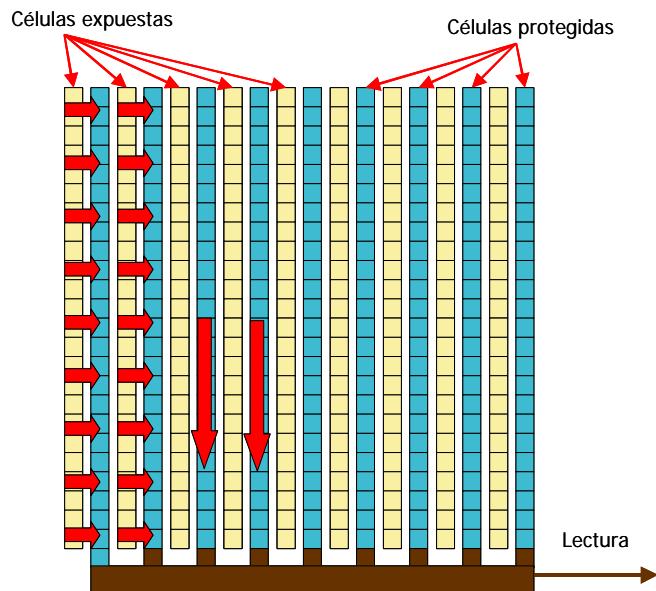


Fig. 5.38 Configuración de una CCD de transferencia interlínea

En general, todas las estructuras CCD descritas presentan una excelente fiabilidad y robustez, lo que las hace ideales para aplicaciones de vídeo y de captura de imágenes en cualquier condición. No presentan distorsión geométrica y su respuesta es muy lineal con la luz incidente.

A parte de su aplicación en sistemas de vídeo se utilizan para la adquisición de imágenes en astronomía y microscopía de fluorescencia. Estas aplicaciones requieren tiempos de exposición muy elevados (desde varios segundos hasta horas) que pueden obtenerse manteniendo las condiciones de adquisición de imagen durante tiempos prolongados. Los dispositivos CCD utilizados para estas aplicaciones requieren que el sustrato semiconductor sea enfriado mediante sistemas externos para reducir la generación espontánea de pares electrón-hueco debida a la temperatura. Esta generación espontánea se denomina *corriente de oscuridad* y representa la carga que se almacenará en los electrodos en ausencia de luz y, por lo tanto, debida únicamente a la generación térmica de carga eléctrica. Resulta obvio que si deseamos capturar imágenes con muy poca iluminación, será necesario que la generación térmica de carga esté por debajo de la generación de carga a partir de la iluminación. Así, la temperatura del sustrato deberá reducirse más o menos en función de las condiciones de iluminación. Debido a las imperfecciones en la fabricación del dispositivo, la corriente de oscuridad depende del elemento de imagen, pudiendo obtener puntos de la estructura CCD donde ésta es relativamente elevada. La carga adquirida en condiciones de oscuridad total produce una imagen parecida a un campo de estrellas (donde los elementos de imagen con mayor corriente de oscuridad producen los puntos luminosos). Esta imagen es estacionaria y es característica de cada CCD, por lo que puede sustraerse de la imagen final para reducir la incidencia de la corriente de oscuridad sobre la imagen adquirida.

Los defectos en la estructura y fabricación del CCD pueden producir algunos píxeles muertos (*dead pixels*), que se caracterizan por ser contactos metálicos bajo los que no se almacena carga. La presencia de estos elementos defectuosos puede afectar a todos los elementos de la misma columna si se produce una pérdida de carga en estos terminales durante el proceso de lectura.

Los esquemas de CCD que hemos presentado son muy simplificados, pues no se han tenido en cuenta células y configuraciones de condensadores MOSFET que se utilizan para la eliminar parte de la carga indeseada o para iniciar la captura de una nueva imagen. Uno de estos circuitos son los sistemas de protección frente al *blooming*. Este fenómeno se produce cuando se sobreexpone la CCD o existe algún punto luminoso de gran intensidad. En estos casos, la carga generada bajo el contacto metálico próximo al punto luminoso es muy elevada y puede desbordar hacia las células adyacentes. Para evitar este problema suelen ponerse células que descargan a masa los excedentes de carga.

En toda esta exposición hemos supuesto que el sustrato de silicio es del tipo N, por lo que los portadores mayoritarios son los electrones. En la práctica, la mayoría de CCD se fabrican con un sustrato de silicio del tipo P. Esto significa que los portadores mayoritarios son en realidad los huecos y que las tensiones que realmente deben aplicarse a los contactos metálicos son negativas.

5.5.5 Dispositivos de inyección de carga (*Charge Injection Devices, CID*)

Los principios físicos para la captura de imagen de estos dispositivos son los mismos que en los CCD. La diferencia entre ambos estriba en la forma en la que se realiza la lectura de la carga. En este caso, los elementos de imagen pueden direccionarse de forma individual, especificando la fila y la columna de la matriz de sensores que se desea leer, proporcionando un acceso aleatorio a los píxeles de la imagen.

Cada uno de los elementos de imagen contiene dos condensadores MOS acoplados, uno dedicado a la captura de imagen y otro dedicado al almacenamiento de la información (protegido de la luz) mediante una configuración como la representada en la figura 5.39. La transferencia de carga entre el elemento expuesto a la luz y el elemento de almacenamiento se realiza seleccionando primero la fila y luego la columna en la que se desea realizar la lectura. Nótese que gracias a la configuración matricial de los sensores, sólo la carga asociada al elemento seleccionado será transferida. Por otra parte, la lectura de la información se realiza de modo no destructivo, mediante sensores de carga conectados a los buses de direccionamiento. Esto permite que la carga pueda volver a restablecerse al elemento sensor una vez leída.

Las aplicaciones de este tipo de sensores se concentran en utilidades donde el tiempo de exposición es un factor muy crítico y es conveniente controlarlo adaptativamente. Obsérvese que es posible utilizar tiempos de exposición diferentes para distintas partes de la imagen e incluso continuar exponiendo un elemento de imagen después de su lectura. Otra ventaja inherente a esta estructura es que no existe ninguna transferencia de carga entre sensores adyacentes, por lo que no es posible la aparición de efectos de *blooming* o de *smear*. Asimismo, la presencia de elementos defectuosos no tiene ninguna incidencia en los píxeles adyacentes.

El inconveniente principal de estas células es que, debido a su procedimiento de lectura, su nivel de ruido es bastante superior al de los CCD convencionales. La carga obtenida bajo el elemento sensor es

traspasada al bus de lectura mediante un divisor de capacidades. La propia resistencia del bus introduce un nivel de ruido significativamente mayor que en las CCD y reduce el ámbito de aplicación de estos elementos.

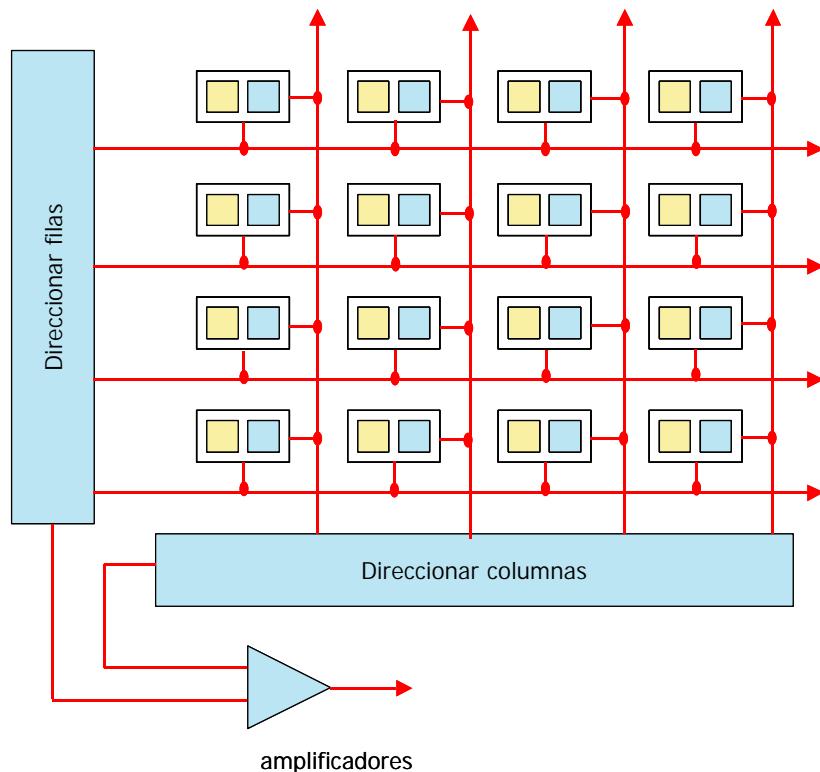


Fig. 5.39 Estructura de los sensores de inyección de carga

5.5.6 Sensores de píxel activo (Active Pixel Sensors)

Los *sensores de píxel activo* también pueden seleccionar los elementos de imagen de modo individual y no requieren que se realice una transferencia de las cargas entre los sensores. Se fabrican utilizando tecnología CMOS y utilizan un amplificador activo para cada elemento de imagen, eliminando los problemas de la capacidad y resistencia de los buses en los dispositivos CID.

Con la tecnología actual, los transistores que amplifican la carga de los sensores cubren aproximadamente un 75 % del área de imagen, lo que reduce el número de elementos activos del sensor. Debido a esta restricción, prácticamente sólo se utilizan en estructuras lineales. Aunque la tecnología debe mejorarse apreciablemente para poder competir con las densidades actuales de los dispositivos CCD, se usan con cierta frecuencia en juguetes o equipos de gama baja. Su principal ventaja es que al estar basados en tecnología CMOS pueden producirse en líneas convencionales, con

un coste reducido y con la posibilidad de integrar otras funciones lógicas o de conversión analógico digital en el mismo circuito integrado. El avance de esta tecnología permitiría introducir funciones avanzadas en los sensores de imagen sin necesidad de utilizar circuitos exteriores.

5.5.7 Captura de imágenes en color con un único sensor

En las cámaras profesionales o semiprofesionales suele utilizarse un sensor CCD para cada componente de color. La luz procedente del grupo óptico se descompone en tres componentes R, G y B mediante el uso de un prisma dicroico cuyos principios de funcionamiento ya han sido analizados en el primer capítulo. No obstante, en cámaras de gama baja, suele utilizarse un único sensor CCD para realizar la captura de la imagen en color. Con ello se reduce el coste de los sensores y, sobre todo, el coste del prisma de separación de las componentes de color.

Para realizar la captura de la imagen mediante un único sensor es necesario interponer filtros de color delante de los elementos de imagen del sensor de modo que sólo se permita el paso de las componentes de luz que interesen. La idea básica es distribuir sobre la CCD filtros de color de modo que las cargas de los elementos correspondan en unos casos al rojo, en otros al verde y en otros al azul. Si conocemos qué tipo de filtro se ha utilizado en cada elemento de imagen podremos reconstruir las tres componentes de color. El coste de esta alternativa es, evidentemente, una pérdida significativa de resolución.

Podría pensarse en una alternativa que interpusiera un filtro de color delante de toda la CCD, que fueran cambiando con el tiempo y que permitiera la captura de las tres componentes de color multiplexadas en el tiempo. No obstante, esta alternativa supondría tener que incorporar sistemas mecánicos complejos para el cambio de filtro. Además reduciría el tiempo máximo de exposición en un factor 3 y significaría (esto es lo más importante) que las tres componentes de color asociadas a una imagen no se corresponden con el mismo instante de tiempo, dando lugar a distorsiones de color significativas en los objetos móviles de la escena.

En la práctica, durante el proceso de fabricación de las CCD que deben utilizarse en cámaras de un único sensor se aplica un revestimiento de filtros de colores que afectan a cada uno de los elementos de imagen. No obstante, los filtros que se utilizan no corresponden a los tres colores básicos, ya que obtendríamos un bajo rendimiento lumínico (y por tanto, mucha sensibilidad al ruido) en algunas componentes (rojo y azul). Los filtros que se suelen utilizar son el verde (G), amarillo (Ye), magenta (Mg) y ciano (Cy) y su disposición geométrica en la matriz de sensores es como la que se indica en la figura 5.40. En esta figura también se indica cómo se realiza la lectura de la imagen en dos campos. En el campo A se combinan los elementos de imagen obtenidos en las líneas A1, A2, etc. de la CCD mientras que en el campo B se combinan los elementos de imagen B1, B2, etc.

A partir de las señales obtenidas con estos filtros se determinan las componentes de luminancia y diferencia de color utilizando únicamente sumas y restas de píxeles. Los pares de elementos correspondientes a la misma columna son siempre sumados. Así, la información que podemos tratar correspondiente a los pares de línea A1 es (Cy+G) y (Ye+Mg). En los pares de línea del tipo A2 disponemos de la información (Cy+Mg) y (Ye+G). Esto significa que tanto las componentes de luminancia como de color sólo serán obtenidas de modo aproximado. La luminancia se determina para

todos los pares de líneas (A1, A2) mientras que las señales diferencia de color sólo se determinan en pares alternados (en A1 se estima R-Y y en A2 la componente B-Y).

La luminancia en el par de líneas A1 se obtiene a partir de la suma de (G+Cy) y (Mg+Ye), de acuerdo con la siguiente expresión:

$$Y \approx \{(G + Cy) + (Mg + Ye)\} \times \frac{1}{2} = \frac{1}{2}(2B + 3G + 2R) \quad (5.26)$$

La componente de croma se aproxima restando las combinaciones entre dos elementos de imagen sucesivos:

$$R - Y \approx \{(Mg + Ye) - (G + Cy)\} = (2R - G) \quad (5.27)$$

En la línea A2 la luminancia se obtiene sumando las dos componentes:

$$Y \approx \{(Mg + Cy) + (G + Ye)\} \times \frac{1}{2} = \frac{1}{2}(2B + 3G + 2R) \quad (5.28)$$

y la señal (B-Y) se obtiene mediante la diferencia de las dos componentes

$$-(B - Y) \approx \{(G + Ye) - (Mg + Cy)\} = -(2B - G) \quad (5.29)$$

Los resultados que se obtienen para el campo B son parecidos. Tal y como se desprende de estos resultados, los valores de las componentes Y, (R-Y) y (B-Y) que realmente se utilizan en este tipo de cámaras son sólo valores aproximados a los colores reales.

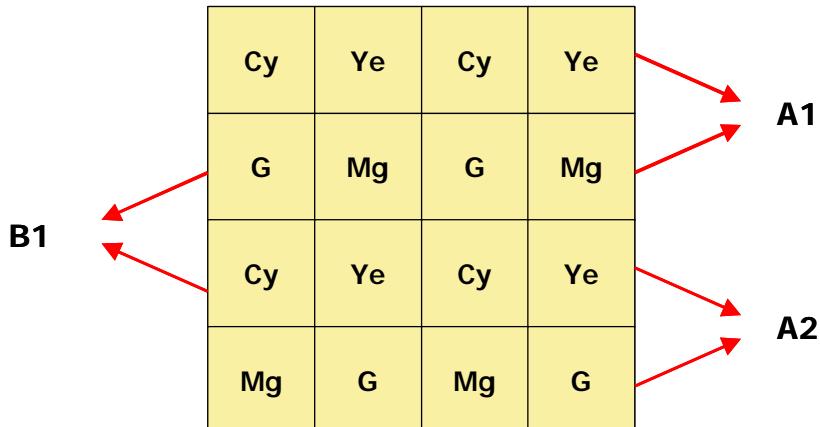


Fig. 5.40 Configuración de filtros de color en sistemas con un único sensor CCD

5.6 Sistemas automáticos y control de la señal de vídeo

En esta sección veremos los principios de funcionamiento de algunos subsistemas que usualmente se incluyen en cámaras de vídeo o de fotografía. Aunque veremos algunos controles que se utilizan sólo en cámaras profesionales, nos concentraremos principalmente en las cámaras domésticas, puesto que al estar orientadas a usuarios no expertos son las que incorporan un mayor número de sistemas de control automático.

5.6.1 Zoom óptico y zoom digital

El *zoom* o ángulo de visión de la cámara se controla directamente por el usuario que puede modificarlo accionando un botón o deslizador. En cámaras profesionales este control actúa directamente sobre el motor de *zoom* del grupo óptico, variando la distancia focal. El *factor de zoom* o número de aumentos del grupo óptico se define como la relación entre la distancia focal más grande y la más pequeña que pueden obtenerse. Son habituales factores de *zoom* de 20x en cámaras domésticas y de 10x en profesionales. Algunas cámaras incorporan también el llamado *zoom digital*, que consiste en utilizar sólo la parte central de la CCD para obtener la imagen final, interpolando el resto de muestras. El tipo de filtro interpolador que se utiliza puede ser tan simple como la repetición de la muestra anterior o la interpolación lineal entre elementos de imagen adyacentes. En general, si se utiliza un factor de *zoom* digital excesivamente grande, se observa una pérdida de calidad de imagen considerable que aparece como una imagen con un grado elevado de pixelado. Existen cámaras domésticas que con la acción conjunta del *zoom* óptico y el *zoom* digital pueden llegar a factores de aumento superiores a 200. La toma de imágenes con este factor de aumento exige, evidentemente, el uso de trípodes muy estables.

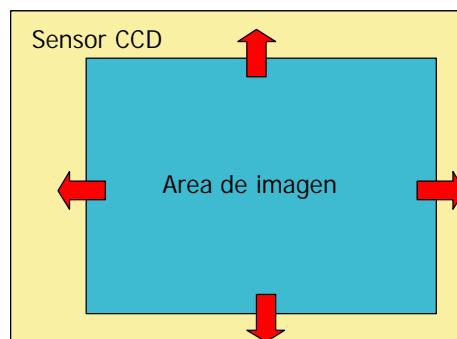


Fig. 5.41 Estabilización de imagen mediante sistemas electrónicos

5.6.2 Mecanismos de estabilización de imagen

Se utilizan básicamente en cámaras domésticas y en algunas cámaras de campo y tienen como objetivo reducir el movimiento involuntario de la cámara por parte del operador que produce imágenes poco estables. Existen dos alternativas para obtener la estabilización de la imagen. La

primera consiste en fijar la superficie donde se forma la imagen actuando directamente sobre el grupo óptico mediante sistemas mecánicos. Se trata de un sistema complejo y caro, pero con el que se obtienen excelentes resultados. La segunda alternativa es puramente electrónica y se basa en actuar sobre la imagen inestable obtenida en el sensor CCD. En este tipo de sistemas, la superficie de imagen del sensor CCD es mayor que la superficie de imagen útil, tal y como se ilustra en la figura 5.41. Un sistema electrónico se encarga de determinar el índice de la fila y la columna de la imagen capturada por la CCD que debe presentarse como primera fila y primera columna de la imagen útil. Los algoritmos de actualización de estos valores son específicos de cada fabricante y se basan en cálculos simples de correlación entre imágenes y predicciones de movimiento.

5.6.3 Sistemas de autoenfoque

Los sistemas de autoenfoque actúan sobre el motor de enfoque del grupo óptico con el objetivo de mantener el sujeto principal de la acción permanentemente enfocado, sin necesidad de que el operador actúe manualmente sobre el mando de enfoque. Pueden ser activos o pasivos en función de si utilizan una señal auxiliar para determinar la distancia a la que se encuentran los objetos o realizan el control de enfoque directamente a partir de las imágenes recibidas. Los sistemas activos más populares son los basados en ultrasonidos e infrarrojos.

En la figura 5.42 se ilustra el principio de funcionamiento de un sistema basado en ultrasonidos. La cámara transmite pulsos de corta duración y se mide el tiempo que transcurre entre la emisión y la recepción del pulso, enfocando el grupo óptico a esta distancia equivalente. Si no se detecta la recepción del pulso el sistema enfoca automáticamente hacia el infinito. Se trata de un sistema muy económico que se utiliza en cámaras de gama baja (normalmente cámaras miniatura de fotografía). Uno de los principales inconvenientes de los sistemas basados en ultrasonidos es que no pueden enfocar a través de un cristal (queda enfocado el cristal)

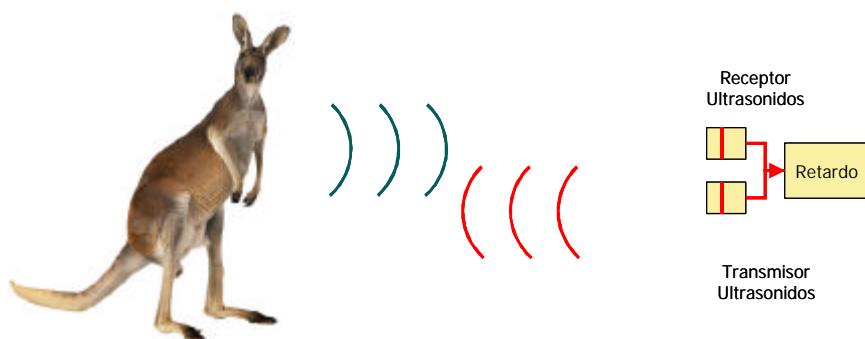


Fig. 5.42 Sistema de autoenfoque por ultrasonidos

Los sistemas basados en infrarrojos transmiten una señal continua en esta banda de frecuencia. La distancia del objeto principal se obtiene por trigonometría a partir del diagrama de la figura 5.43. La distancia entre el emisor y el receptor de infrarrojos (D) es conocida. El sistema mide la distancia d,

respecto al centro del array de fotodiodos, a la que se recibe el rayo transmitido. La distancia real del objeto puede determinarse a partir de estos dos parámetros. En general, los sistemas de autoenfoque basados en infrarrojos suelen proporcionar mejores resultados que los sistemas basados en ultrasonidos.

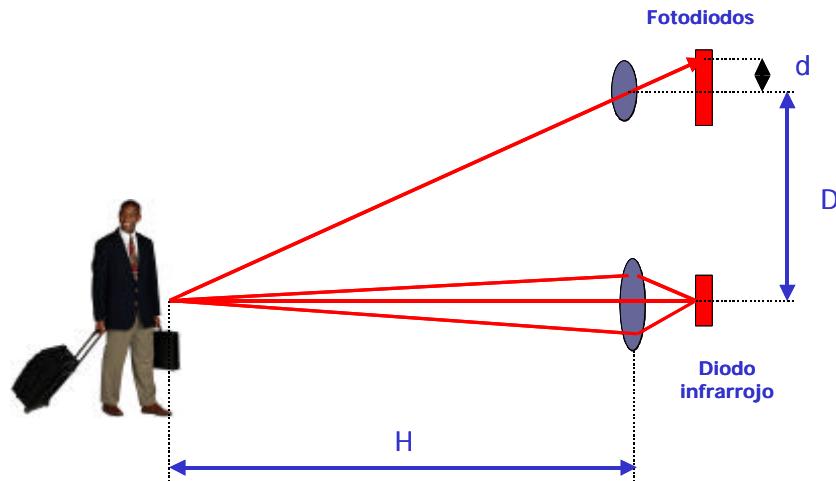


Fig. 5.43 Sistema de autoenfoque basado en infrarrojos

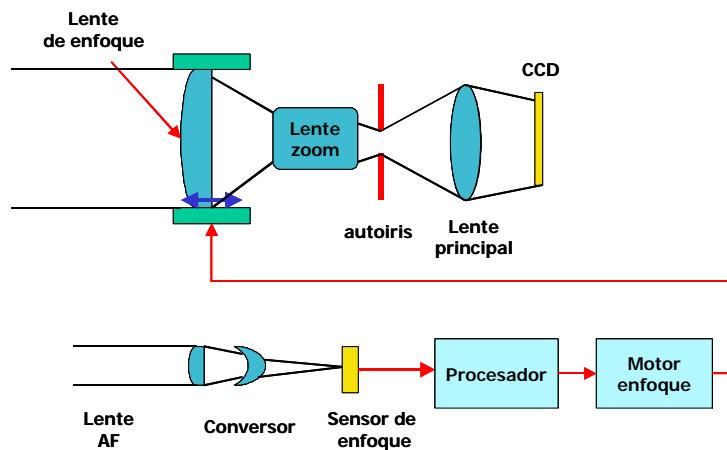


Fig. 5.44 Enfoque pasivo mediante sistema óptico auxiliar

Los sistemas pasivos suelen proporcionar una mayor precisión en el enfoque de las imágenes. Existen dos posibles variantes en función de la imagen que se utiliza para calcular el punto de enfoque. En las figuras 5.44 y 5.45 se representan estas dos alternativas. En el primer caso, la imagen utilizada para determinar el enfoque se obtiene mediante un sistema óptico auxiliar, independiente del grupo óptico que se utiliza para registrar la escena. La alternativa a este sistema consiste en integrar un prisma

separador de luz, que permite utilizar la misma imagen para el cálculo del foco. En cámaras reflex de 35 mm el propio espejo reflector del visor realiza esta función.

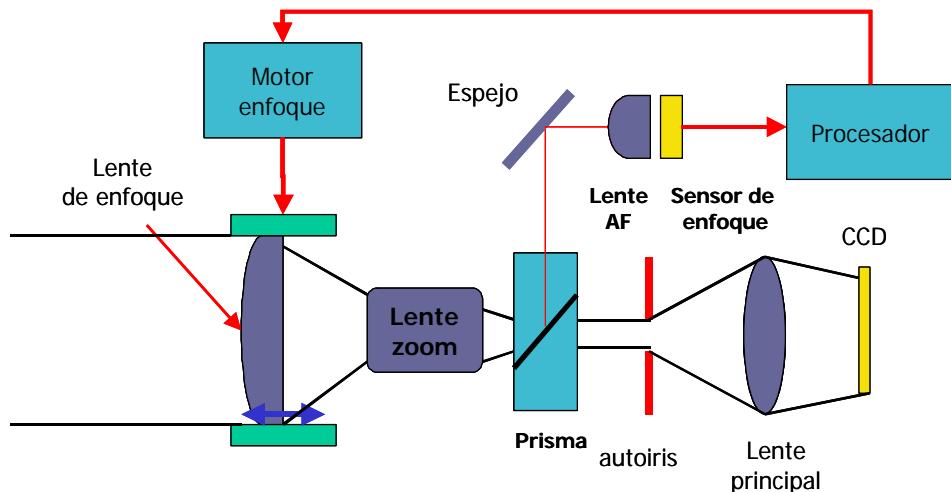


Fig. 5.45 Enfoque pasivo integrado dentro del propio grupo óptico

Para determinar el enfoque pueden usarse divisores ópticos de imagen parecidos a los que se utilizan en las cámaras de fotografía para el enfoque manual. Estos sistemas dividen la imagen en el sentido horizontal en dos partes, de forma que, tal y como se ilustra en la parte derecha de la figura 5.46, al modificar la posición del foco las imágenes resultantes se desplazan hacia la izquierda o hacia la derecha. La imagen sólo permanece correctamente enfocada cuando las líneas verticales de ambas subimágenes coinciden. Para realizar el ajuste automático de foco mediante este sistema, es necesario utilizar lentes lenticulares y pares de fotodiodos situados a ambos extremos de la imagen dividida. El sistema de control actúa sobre el motor de foco hasta que la señales recibidas por los pares de fotosensores coincidan.

Una alternativa a este sistema consiste en utilizar un sensor CCD lineal, que se sitúa en el centro de la imagen, tal y como se muestra en la figura 5.46. Incluso, para cámaras basadas en CCD matriciales, es posible aprovechar la información obtenida en los sensores centrales. El control de enfoque se realiza a partir del análisis de la energía de señal recibida en estos sensores, moviendo el motor hasta que se obtiene una señal en la que aparecen rápidas transiciones entre el fondo y los objetos. El principio de funcionamiento de este sistema se ilustra en la figura 5.47, donde se representa la señal obtenida en el sensor CCD para una imagen enfocada y una imagen desenfocada. Para determinar el foco correcto suele estimarse la derivada de la luminancia recogida en el sensor. Esta derivada se determina mediante diferencias entre los niveles de luminancia entre píxeles consecutivos y es una medida de lo abruptos que son los contornos y, por tanto, de la nitidez de la imagen. Nótese que en la figura 5.47 se obtiene una derivada con una energía mucho mayor para la imagen enfocada que para la imagen desenfocada. En general, la energía de la derivada es una buena medida del enfoque de la imagen.

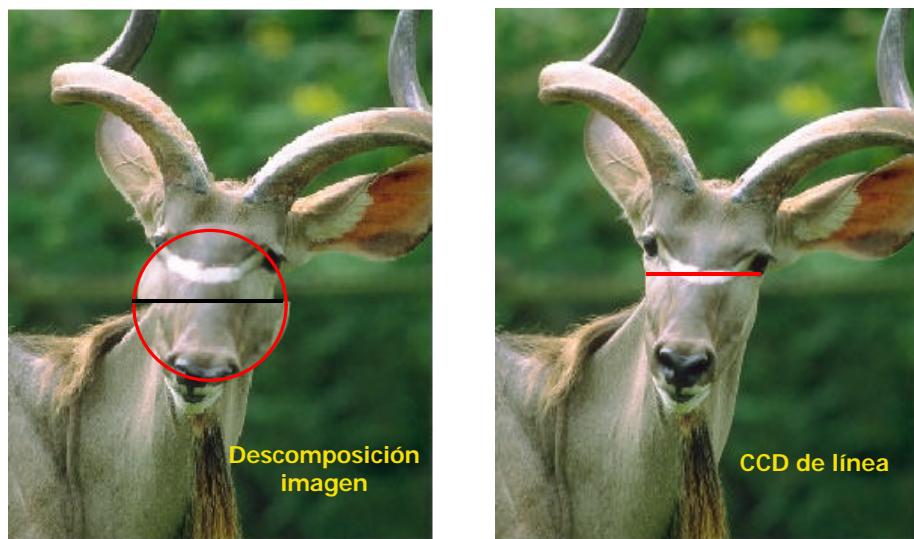


Fig. 5.46 Enfoque pasivo mediante descomposición de imagen y CCD de línea

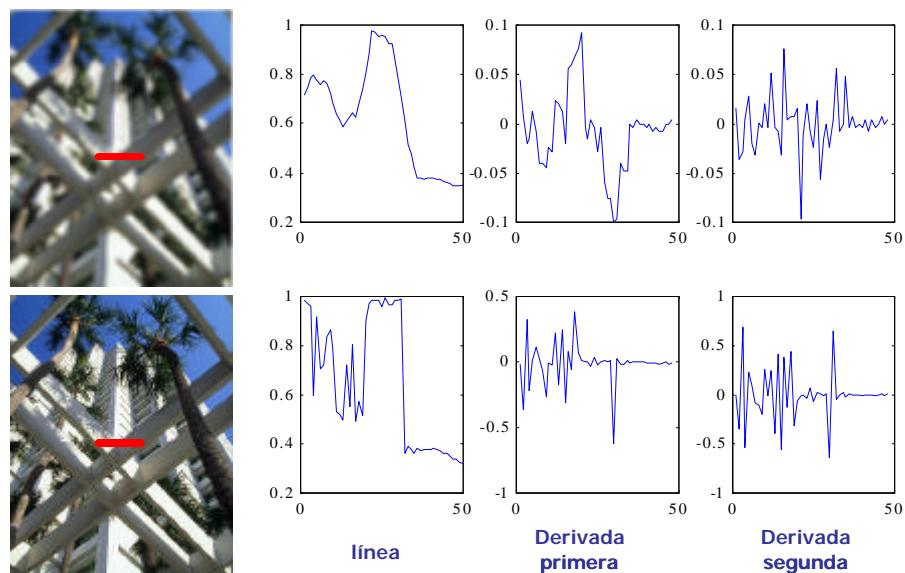


Fig. 5.47 Gradiente de la señal obtenida en el sensor CCD

Los principales problemas de los sistemas de enfoque pasivo es que sólo pueden enfocar en la zona central de la imagen y que es necesario que existan contornos verticales en esta zona para poder realizar un enfoque correcto. En cámaras de fotografía estos problemas se solucionan mediante un

botón que mantiene el ajuste de enfoque. Si desea realizarse una fotografía de un motivo que no tiene contornos en la parte central, deberemos enfocar primero sus contornos y posteriormente desplazar la cámara, con el enfoque fijado, hasta obtener el ángulo deseado. Si se requiere realizar fotografías de objetos que tienen sólo contornos horizontales se recomienda girar la cámara 90° (con lo que situamos el sensor perpendicular a los contornos), ajustar y fijar el enfoque y posteriormente realizar la toma. En cámaras de vídeo pueden utilizarse los mismos trucos, siempre que no se requiera registrar constantemente la escena, en cuyo caso, se recomienda utilizar el enfoque manual.

Algunos modelos de cámaras avanzadas disponen de varios sensores CCD con los que el usuario puede seleccionar la zona de la imagen en la que se realizará el cálculo de enfoque. Un ejemplo de estos sistemas es el MultiCam 1300AF de Nikon, cuya distribución de sensores CCD se representa en la figura 5.48. El subsistema de autoenfoque dispone de un total de 14 sensores CCD del tipo lineal. El usuario puede seleccionar cinco regiones distintas para el enfoque de la imagen (el centro y las cuatro direcciones principales). En las posiciones central, derecha e izquierda, los sensores están dispuestos en forma de cruz para poder detectar tanto los contornos horizontales como los verticales. Además de las CCD convencionales se integran unas CCD de mayor área, y por tanto mayor sensibilidad a la luz, que permiten realizar enfoques de gran precisión en condiciones de baja iluminación.

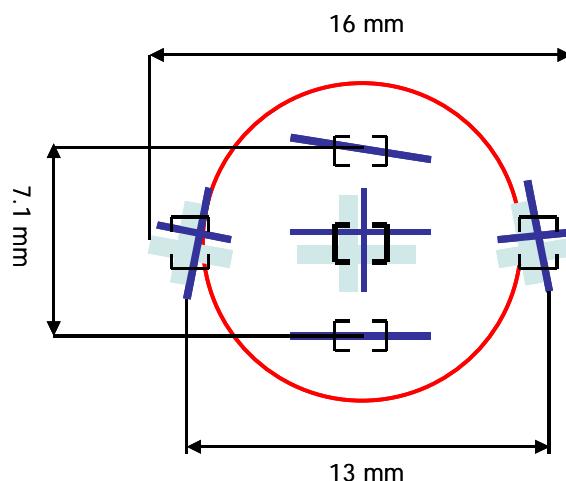


Fig. 5.48 Sistema de autoenfoque pasivo con múltiples sensores CCD

5.6.4 Ajuste del balance de blancos

Los colores capturados por la cámara dependen, como es evidente, de la iluminación. En condiciones de luz natural diurna la energía lumínica está distribuida de forma aproximadamente igual en las tres componentes de color. Sin embargo, con iluminación artificial es muy probable que una de las componentes de color sea más importante que las otras. Así, para iluminación de tungsteno predomina la componente roja, lo que resulta apropiado en escenarios cálidos en los que predominan los tonos rojizos. Los efectos de la iluminación en la imagen pueden en principio compensarse actuando sobre

la ganancia de cada una de las componentes de color en la cámara. Este ajuste se conoce como el balance de blancos y es conveniente realizarlo antes de empezar un registro de imagen para adaptarnos a las condiciones de iluminación. El ajuste del balance blancos en las cámaras actuales se ha simplificado notablemente y basta con enfocar la cámara hacia una hoja de papel u objeto blanco y pulsar el botón de calibración de blancos. La ganancia de las tres componentes de color se ajusta automáticamente para que den el mismo nivel de señal en estas condiciones de iluminación. El ajuste de blancos se incorpora en todas las cámaras profesionales y en los modelos domésticos de gama alta.

5.6.5 Modos de exposición y efectos

La mayor parte de cámaras de vídeo domésticas incorporan programas que controlan la exposición y el diafragma de forma automática en función de las condiciones de iluminación y tipo de escenas que se estén registrando. Los modos de exposición más habituales son:

Deportes: Se utilizan velocidades de obturación elevadas (mayores que 1/400) para que los sujetos no queden borrosos debido al movimiento. Esto significa que se proporciona prioridad al tiempo de exposición, de modo que en condiciones de baja iluminación el diafragma deberá abrirse resultando poca profundidad de campo.

Retratos: En este modo se pretende resaltar el sujeto principal dejando el fondo borroso. Por ello, se fuerza que el diagrama esté muy abierto para reducir la profundidad de campo.

Playa y Nieve: Se trata de un modo pensado para condiciones de iluminación intensa en los que es habitual que se produzca sobre-exposición si se utiliza un modo automático convencional.

Puestas de Sol y Escenas Nocturnas: En condiciones de baja iluminación los modos automáticos tienden a aumentar la ganancia de la señal proporcionada por el sensor reproduciendo escenas poco naturales.

Baja Iluminación: Este modo sólo se incorpora en cámaras digitales en las que se puede aumentar el tiempo de exposición del fotograma por debajo de la frecuencia de campo. En efecto, los fotogramas, tomados a velocidades lentes, se almacenan en memoria y se repiten durante varias imágenes, pudiendo capturar escenas en condiciones de baja iluminación. Evidentemente, la frecuencia de imágenes reales se reduce, por lo que se aprecia discontinuidad en el movimiento.

Paisajes: En este caso se fuerza que la cámara enfoque al infinito, por lo que pueden capturarse imágenes del paisaje a través de cristales de un automóvil o con la presencia de obstáculos en primer plano.

La mayoría de cámaras portátiles también integran efectos de imagen simples que proporcionan al usuario cierta flexibilidad para producir vídeos domésticos con buenos acabados sin necesidad de realizar ediciones posteriores. Entre estos efectos son especialmente importantes los fundidos entre escenas, el registro de imágenes en blanco y negro, sepia o muy baja resolución (efecto mosaico), la

mezcla de imágenes en memoria con imágenes de vídeo real, etc. Algunos modelos incorporan incluso utilidades de titulación.

5.7 Tipos de cámaras

En este apartado revisaremos, de forma muy breve, los diferentes tipos de cámaras de vídeo que pueden encontrarse en el mercado, así como sus características y sus aplicaciones. El objetivo es proporcionar una idea general de los criterios que deben utilizarse para seleccionar un determinado tipo de cámara en función de la aplicación.

5.7.1 Cámaras de estudio

Estas cámaras están orientadas a estudios de grabación y suelen ser de un tamaño y peso considerable, por lo que, generalmente, están montadas sobre soportes estables o carreles móviles. Existen algunos modelos portátiles que pueden ser transportados por el operador para realizar tomas móviles. Proporcionan como señal de salida las tres componentes de color independientes, aunque el formato de estas señales puede variar en función de las características de la cámara y los equipos del estudio. Existen modelos en que las señales se transmiten en banda base por cables independientes, modelos en los que se realiza una multiplexación temporal; otros emplean transmisiones bidireccionales mediante cables del tipo Triax o incluso transmisiones analógicas o digitales vía fibra óptica.

La característica principal de estas cámaras es que están conectadas a una unidad de control de cámara externa (*CCU-Camera Control Unit*), que es donde generalmente se realiza la compensación de la matriz de color, el ajuste de ganancia de las componentes de color, la mejora de imagen, la corrección de gamma y las conversiones a señal de luminancia y componentes diferencia de color. El cable que une el cabezal de cámara con la unidad de control de cámara se utiliza también para proporcionar diversas señales de retorno al operador de cámara, como por ejemplo la señal de audio del ayudante de realización, la señal de vídeo que se observa por el visor, etc. También se proporcionan señales de control del diafragma y foco así como la alimentación de la cámara. Debe tenerse en cuenta que, generalmente, el operador de cámara sólo controla el enfoque y el *zoom* de la cámara. El diafragma o iris, los tiempos de exposición, la ganancia de las componentes de color, etc., se controlan desde paneles de control independientes. Al centralizar el control de todas las cámaras en un único punto, permite obtener un equilibrado perfecto de todas las cámaras que intervienen en la producción.

Los formatos de señal de vídeo proporcionados por las cámaras de estudio son muy diversos y dependen de las características de la cámara. Así, en una cámara de alta definición los tiempos de línea o de campo dependen del número de líneas de la cámara y del número de imágenes por segundo. Análogamente, en modelos pensados para la captura de escenas en cámara lenta, el formato de salida suele depender del número de fotogramas por segundo. En general, la conversión a formatos estándar puede realizarse en la unidad de control de cámara.

5.7.2 Cámaras de campo (*Electronic News Gathering – ENG*)

Estas cámaras también están orientadas al mercado profesional. Se trata de cámaras portátiles pensadas para la realización de reportajes en exteriores y se caracterizan por incorporar un subsistema de registro de la secuencia de vídeo sobre soporte magnético (magnetoscopios). Existen modelos en los que el cuerpo principal de la cámara (óptica y captura de señales) es independiente del sistema de registro de la señal, que puede intercambiarse en función de las necesidades. Generalmente, disponen de múltiples formatos de salida de la señal de vídeo (vídeo compuesto, componentes, formatos digitales, etc).

5.7.3 Cámaras de circuito cerrado de TV o de video vigilancia

Generalmente son cámaras en miniatura que proporcionan una señal de vídeo compuesto PAL o NTSC estándar. Existen modelos en blanco y negro y en color y tienen un coste relativamente bajo. En algunos modelos es posible controlar el foco y el *zoom* del grupo óptico a distancia. La mayoría disponen de un sistema de autoiris que regula la cantidad de luz incidente en función de las condiciones de iluminación, que, en algunos modelos, también puede controlarse manualmente. En la mayoría de instalaciones las cámaras se distribuyen en distintos puntos del edificio y las señales se dirigen a un centro de control, donde un multiplexor de entrada permite seleccionar al operador la cámara que se visualiza. Existe paquetes de *software* específicos para video vigilancia que gestionan la inspección automática de las cámaras y el registro de las imágenes en disco duro o en magnetoscopios convencionales en función de las alarmas que se producen en el edificio. En edificios de dimensiones considerables es probable que deban incorporarse moduladores de vídeo para cable, para que las imágenes puedan ser recibidas con calidad desde el centro de control. Algunos modelos de cámaras para video vigilancia son de tamaño muy reducido, por lo que pueden disimularse en el entorno y pasar fácilmente desapercibidas. Este tipo de cámaras se utilizan con frecuencia en centros comerciales.

5.7.4 Cámaras industriales y de visión

Se engloban dentro de este apartado todas las cámaras que se utilizan en aplicaciones de reconocimiento automático de formas y visión por computador. Las características de las cámaras dependen del tipo de aplicación y pueden encontrarse tipos muy distintos. Normalmente la cámara está conectada a una tarjeta de digitalización de imagen que puede realizar la adquisición de la secuencia de vídeo en tiempo real o adquirir un único fotograma cada vez que se produzcan las condiciones de la imagen que debe reconocerse. Este segundo caso es el más extendido en la mayoría de aplicaciones, utilizando tarjetas de digitalización que se denominan capturadoras de imagen (*frame grabber*), que generalmente aceptan distintos formatos de señal de cámara y que disponen de puertos para controlar algunos de los parámetros de la cámara.

En las aplicaciones más simples pueden utilizarse cámaras de video vigilancia que proporcionan una señal PAL estándar en color o, en la mayoría de los casos, en blanco y negro. La adquisición de la imagen se realiza cuando se producen las condiciones en las que se requiere realizar un reconocimiento de la imagen. Generalmente estas condiciones son controladas por circuitos externos que producen una señal de inicio de captura. Una vez la imagen ha sido digitalizada, se realiza la

identificación o reconocimiento de los objetos produciendo señales externas de control y quedando el sistema en condiciones de capturar una nueva imagen. Así, en una aplicación de reconocimiento de matrículas en un *parking*, la señal de inicio de captura la produciría un sensor que detecta la presencia de un automóvil en la barrera de entrada al *parking*. Una vez adquirida la imagen y procesada se procede a actuar sobre el sistema proporcionando el *ticket* de entrada y abriendo la barrera del *parking*. Es evidente que, en función de las características del problema puede resultar necesario que las características de la señal PAL no proporcionen suficiente resolución de imagen para resolver el problema de reconocimiento de formas. En estos casos, se suelen utilizar cámaras de alta definición que proporcionan señales de vídeo no estándar pero cuyo formato debe ser interpretado correctamente por la tarjeta digitalizadora.

Aparte de las cámaras convencionales, las más utilizadas son las progresivas, las lineales y las denominadas cámaras inteligentes. En cualquier caso, es posible que la señal proporcionada sea analógica o directamente digital. Las señales digitales suelen entregarse mediante un cable específico en el que todos los bits del elemento de imagen se suministran en paralelo. También existen versiones que utilizan protocolos serie estándar a través de un puerto RS-232 o RS-422. La lectura de la información de los datos proporcionados por la cámara requiere el uso de programas adaptados a estos formatos.

Las cámaras progresivas proporcionan todas las líneas que constituyen la imagen en un único campo. Con ello, aumenta la calidad y resolución de la imagen, sobre todo cuando se realiza la captura de objetos móviles. Existen modelos de alta definición que pueden entregar imágenes de hasta 2000 líneas. Generalmente, estas cámaras pueden controlarse mediante una interfaz digital en la que es posible variar el tiempo de exposición, la apertura del diafragma, el enfoque y el ángulo de visión. Como generalmente no es necesario que la cámara proporcione señales de vídeo en tiempo real, también es posible controlar el instante en el que se realiza la captura de la imagen en el sensor CCD. Asimismo, el tiempo de exposición puede superar los 1/50 segundos de las cámaras convencionales, ya que la frecuencia de imagen puede controlarse externamente e incluso, trabajar basándose exclusivamente en fotogramas. En estos modelos, el tiempo mínimo entre dos fotogramas viene determinado por el tiempo de transferencia de la imagen desde la cámara a la tarjeta digitalizadora.

Las cámaras lineales están basadas en una CCD de línea en vez de una CCD matricial. La ventaja principal de este tipo de cámaras es que pueden obtener un gran número de elementos de imagen por línea (hasta 4000 píxeles). No obstante, la captura de la imagen debe realizarse línea a línea, por lo que es preciso que exista un desplazamiento relativo entre la cámara y el objeto que se está capturando. La velocidad de desplazamiento entre ellos determina la resolución de la imagen en el sentido ortogonal a la línea de CCD. Es importante que esta velocidad sea lineal para reducir las posibles distorsiones geométricas de imagen que obtendríamos si el movimiento relativo no es uniforme. Generalmente, el movimiento del objeto se realiza con sistemas de posicionamiento mecánicos de gran precisión que son controlados directamente por el sistema encargado del reconocimiento de imagen. Los formatos de señal que se obtienen en la salida de la cámara pueden ser analógicos o digitales, aunque estos últimos son los que más predominan en los modelos actuales. La tasa de transmisión de bits entre la cámara y la tarjeta de digitalización depende del modelo de cámara, el número de píxeles, si es en color o en blanco y negro, etc. Las tarjetas de digitalización suelen poder configurarse para adaptarse a las características de las cámaras.

Algunas cámaras incorporan la posibilidad de realizar tratamiento de datos interno, por lo que se denominan cámaras inteligentes. Entre las operaciones que pueden realizar estas cámaras destaca la estimación de gradiente, el ajuste digital de contraste o la transformada de Fourier de la imagen. La ventaja de estos sistemas es que proporcionan una información previamente tratada, por lo que simplifican la tarea del procesador durante el reconocimiento de formas. En algunos casos el tratamiento puede realizarse directamente por métodos ópticos (transformada de Fourier) utilizando lentes especialmente diseñadas para que proporcionen este tipo de información. En general se trata de cámaras de precio muy elevado cuyo uso sólo se ha popularizado en aplicaciones muy concretas.

5.7.5 Cámaras de vídeo domésticas

También se denominan cámaras portátiles y se caracterizan por ser equipos de tamaño reducido en los que se integra el cabezal de la cámara y un sistema de registro de señal en cinta magnética (*camcorders*). Al orientarse al mercado doméstico suelen incorporar sistemas electrónicos de ajuste automático de las componentes de color, ganancia de la cámara, control de iris, selección de modos de exposición en función del tipo de escena, etc., con objeto de simplificar el control de la cámara a usuarios no expertos. Los sistemas de registro de la señal de vídeo más utilizados actualmente son el Hi-8 y el Video8, en formatos analógicos, y el miniDV o el Digital8 en digital. Existen algunos modelos, denominados semiprofesionales, de altas prestaciones, que se utilizan en ocasiones para la captura de imágenes en exteriores en sustitución de las cámaras ENG. Su calidad es bastante aceptable y su reducido coste las hace más rentable para estudios de TV local o productoras pequeñas.

5.7.6 Cámaras para aplicaciones científicas

En este apartado englobamos un gran número de sistemas de adquisición de imagen que cubre los campos de la medicina, la microscopía, la astronomía, etc. Las características de estas cámaras son muy diversas y quedan fuera del alcance de este texto. En la mayoría de aplicaciones se utilizan cámaras convencionales o de alta definición en las que se incorporan sistemas ópticos que adaptan la imagen que se desea adquirir. Así, es habitual en medicina disponer de sofisticadas sondas de inspección controladas electrónicamente que proporcionan una imagen que posteriormente es capturada por una CCD externa. En microscopía suelen utilizarse adaptadores de la cámara de vídeo al microscopio. En astronomía se utilizan CCD de grandes dimensiones (para tener una excelente resolución) y de alta sensibilidad a la luz.

5.7.7 WebCams

Reciben este nombre las cámaras que se utilizan en aplicaciones de videoconferencia por internet. En principio, puede utilizarse cualquier cámara de vídeo convencional conectada a una tarjeta de digitalización de vídeo siempre que se utilice un software de codificación de la señal de vídeo y audio adecuado al ancho de banda que proporciona el sistema. No obstante, la característica fundamental de una *WebCam* es su inmediata integración al ordenador personal. Normalmente se conectan directamente al puerto serie RS-232 o al puerto USB sin necesidad de disponer de tarjetas de digitalización genéricas, por lo que facilitan notablemente las tareas de configuración del sistema. La cámara suele ser de reducidas dimensiones, con una óptica del tipo gran angular, y pueden fijarse

sobre el monitor o la mesa de trabajo con facilidad. Algunos modelos disponen de una memoria interna que permite utilizarlas como cámaras de fotografía digital cuando se desconectan del ordenador.