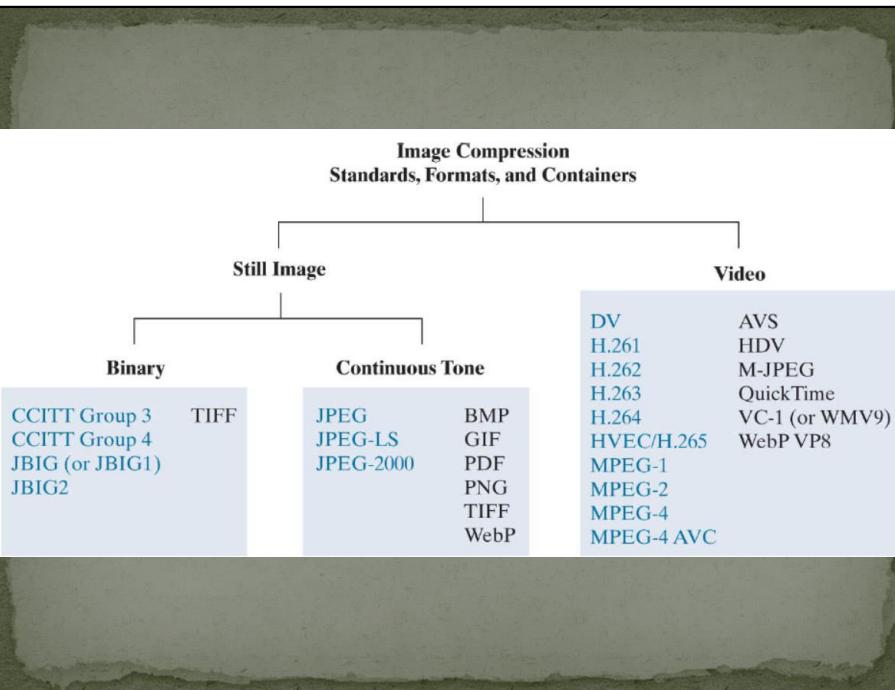


3.9. Nén ảnh màu (Ch. 8 Image Compression and Watermarking)

- ❖ Vì sao cần phải nén hình ảnh và video?
- ❖ Bản chất của quá trình nén tín hiệu?
- ❖ Phân loại các phương pháp nén?



❖ Tổng quan về kỹ thuật nén ảnh

Phân loại theo nguyên lý nén:

- Nén không tổn hao (lossless data reduction)
- Nén có tổn hao (loss data reduction)

Phân loại theo cách thực hiện nén:

- Phương pháp không gian (Spatial Data Compression): các phương pháp nén bằng cách tác động trực tiếp lên điểm ảnh.
- Phương pháp sử dụng biến đổi (Transform Coding): phương pháp nén sử dụng các phép biến đổi không gian, quá trình nén được thực hiện bằng cách tác động lên ảnh biến đổi.

Chất lượng ảnh nén có thể thay đổi tùy theo đặc điểm của hình ảnh nguồn và nội dung ảnh.

Có thể đánh giá **hiệu quả nén** theo số bit cho một điểm trong ảnh nén (N_b). N_b được xác định bằng tổng số bit dùng để mô tả ảnh nén chia cho tổng số điểm ảnh:

$$N_b = \text{Số bit nén/Số điểm}$$

Nguyên lý nén: là loại bỏ dư thừa có trong tín hiệu.

- Dư thừa mã (Coding Redundancy)

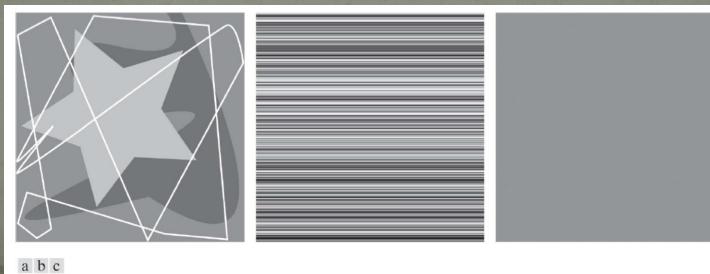
Nếu các mức của tín hiệu video được mã hóa bằng các symbol nhiều hơn cần thiết thì tín hiệu nhận được sẽ có độ dư thừa mã.

- Dư thừa trong pixel (Interpixel Redundancy)

Mức độ tương quan giữa hai điểm ảnh gần nhau là tương đối lớn, do đó trong ảnh số tồn tại lượng dư thừa pixel (còn gọi là dư thừa không gian) tương đối lớn.

- Dư thừa tâm sinh lý

Ảnh có thể được tách ra thành hai phần: một phần chứa các tin tức quan trọng cho người xem, phần khác là các thông tin mà người xem hầu như không cảm nhận được.



Đánh giá mức độ sai số giữa hai ảnh:

➤ Sai số tuyệt đối $e(x,y) = \hat{f}(x,y) - f(x,y)$

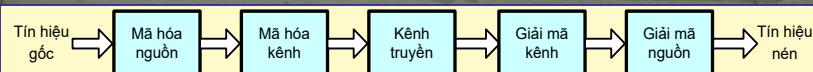
➤ Sai số trung bình bình phương - RMS (Root Mean Square).

$$e_{RMS} = \left[\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (\hat{f}(x,y) - f(x,y))^2 \right]^{1/2}$$

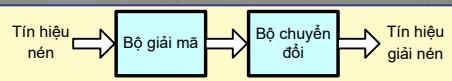
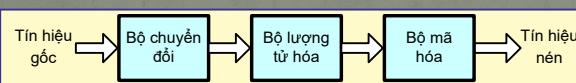
➤ Tỷ lệ tín hiệu/nhiều SNR (Signal to Noise Ratio) :

$$SNR = \left[\frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \hat{f}(x,y)^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (\hat{f}(x,y) - f(x,y))^2} \right]^{1/2}$$

Chất lượng hình ảnh phụ thuộc rất nhiều vào cảm nhận tâm sinh lý của con người. Chính vì vậy, trong kỹ thuật truyền hình thường sử dụng phương pháp đánh giá chất lượng ảnh dựa trên **cảm nhận chủ quan** của người quan sát.



Sơ đồ khái niệm khái quát



Sơ đồ khái niệm khái quát

- **Bộ chuyển đổi:** dùng các phép biến đổi không gian để chuyển ảnh trong không gian thực sang một không gian khác, nơi các hệ số chuyển đổi có mức độ tương quan thấp hơn.

- **Bộ lượng tử hóa:** sử dụng phương pháp lượng tử không đồng đều nhằm triệt tiêu các hệ số biến đổi có năng lượng thấp hoặc đóng vai trò không quan trọng khi khôi phục ảnh. Quá trình lượng tử không có tính thuận nghịch.

- **Bộ mã hóa:** gắn một từ mã cho một mức lượng tử.

➤ Các phương pháp nén không tổn hao (lossless data reduction)

- Phương pháp mã hóa Huffman

Phương pháp mã hóa Huffman là phương pháp giảm dư thừa thống kê. Các ký tự có tần suất cao nhất sẽ được mã hóa bằng từ mã ngắn nhất và ngược lại. Như vậy chúng ta làm giảm chiều dài trung bình của từ mã được dùng để mã hóa dữ liệu.

- Phương pháp mã LZW (Lempel-Ziv-Welch)

Quá trình mã hóa được thực hiện bằng cách thay chuỗi dữ liệu gốc có độ dài khác nhau bằng từ mã có độ dài cố định. Chuỗi dữ liệu được đưa vào tra cứu trong "từ điển" để tìm ra mã tương ứng với nó.

"Từ điển" 9 bits với 512 từ mã được khởi tạo.

Ban đầu chỉ có các từ mã từ 0-255

25	25	90	90
25	25	90	90
25	25	90	90
25	25	90	90

Vị trí trong "từ điển" (từ mã)	Nội dung (chuỗi ký tự)
0	0
1	1
...	...
255	255
256	-
...	...
511	-

25	25	90	90
25	25	90	90
25	25	90	90
25	25	90	90

Quá trình mã hóa được thực hiện từ trái sang phải, từ trên xuống dưới. Mức chói của các điểm ảnh liên tiếp được nối lại thành một chuỗi để tiến hành nhận dạng.

Chuỗi ký tự đã được nhận dạng	Các ký tự đưa vào coder	Từ mã ở đầu ra coder	Vị trí trong "từ điển"	Nội dung chuỗi ký tự
25				25
25	25	25	256	25-25
25	90	25	257	25-90
90	90	90	258	90-90
90	25	90	259	90-25
25	25			
25-25	90	256	260	25-25-90
90	90			
90-90	25	258	261	90-90-25
25	25			
25-25	90			
25-25-90	90	260	262	25-25-90-90
90	25			
90-25	25	259	263	90-25-25
25	90			
25-90	90	257	264	25-90-90
90		90		

Kết quả mã hóa theo kỹ thuật LZW trong ví dụ trên cho thấy: ảnh bao gồm 16 pixel, mỗi pixel mã hóa bằng 8 bits được thay bằng chuỗi mã bao gồm 10 từ mã 9 bits, hệ số nén sẽ là $(16 \times 8) / (10 \times 9) = 1.42$.

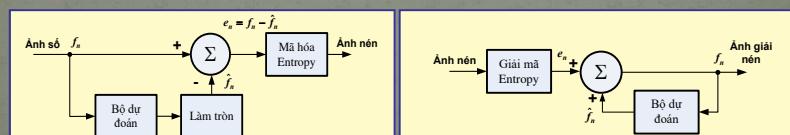
"từ điển" sử dụng khi mã hóa không cần phải chuyển sang phía giải mã, trong quá trình giải mã LZW, "từ điển" này sẽ được khôi phục tự động.

- Phương pháp mã hóa loạt dài (RLC - Run Length Coding)

Kỹ thuật nén này dựa trên nguyên lý tìm trong tín hiệu số nhị phân chuỗi bit "1" hoặc "0" liên tiếp, sau đó mã hóa chuỗi đó bằng hai thông số: giá trị bit lặp và số lượng bit lặp (chiều dài chuỗi). Đôi khi, để nâng cao hiệu quả nén, giá trị chiều dài các chuỗi có thể được mã hóa bằng mã có chiều dài thay đổi (ví dụ mã Huffman).

- Phương pháp mã dự đoán không tổn thất

Phương pháp mã hóa này dựa trên tính tương quan giữa hai điểm ảnh nằm kề nhau và kỹ thuật tách và truyền thông tin "mới" chứa trong mỗi pixel ảnh. Thông tin "mới" trong điểm ảnh được xác định bằng hiệu giữa điểm ảnh thực và điểm ảnh dự đoán. Phương pháp mã dự đoán còn gọi là điều xung mã vi sai (DPCM – Differential Pulse Code Modulation). Hệ thống bao gồm bộ mã hóa (hình 3.3.6a) và bộ giải mã (hình 3.3.6b).

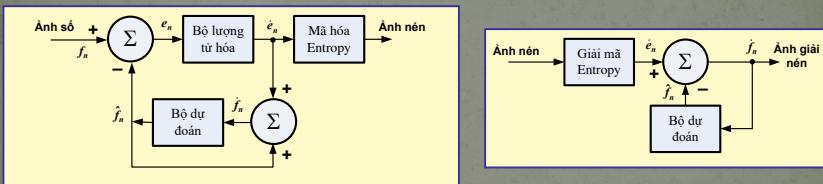


Bộ mã hóa DPCM và giải mã DPCM

➤ Các phương pháp nén tổn hao (loss data reduction)

- Phương pháp mã dự đoán có tổn hao

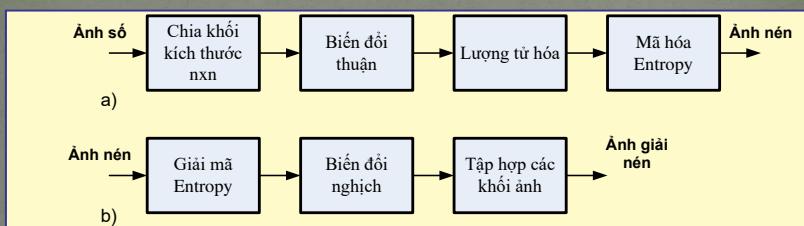
Bộ mã hóa dự đoán có thêm bộ lượng tử hóa. Quá trình lượng tử hóa sẽ làm tròn giá trị e_n tới mức lượng tử gần nhất là \hat{e}_n .



- Các phương pháp nén sử dụng phép biến đổi không gian tín hiệu

Nguyên lý chung của các phương pháp nén này là sử dụng các phép biến đổi tuyển tính (như biến đổi Fourier, biến đổi Cosin, biến đổi DCT - Discrete Cosine Transform v.v.) để ánh xạ ảnh số sang không gian khác. Tập hệ số kết quả của các biến đổi tuyển tính sau đó được lượng tử hóa và mã hóa.

Mã hóa (nén) sử dụng biến đổi



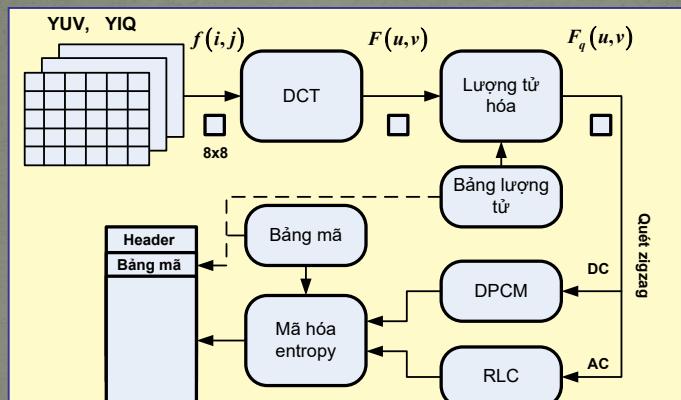
Ưu điểm:

- Các hệ số khai triển thường có mức tương quan nhỏ hơn so với mức độ tương quan giữa các điểm ảnh trong không gian thực.
- Phần lớn năng lượng của tín hiệu tập trung tại các hệ số nằm trong miền tần số thấp.

Quá trình nén dữ liệu có thể được thực hiện trong giai đoạn **lượng tử hóa**: khi lựa chọn mức lượng tử tương đối lớn, các hệ số biến đổi có năng lượng nhỏ (thường nằm ở miền tần số cao) sẽ bị loại bỏ bớt.

Phương pháp nén ảnh tĩnh theo chuẩn JPEG

JPEG (Joint Photographic Expert Group) là tên của một tổ chức nghiên cứu về các chuẩn nén ảnh (trước đây là ISO) được thành lập vào năm 1982.



Sơ đồ khái bô mã hóa ảnh tĩnh JPEG

- Biến đổi DCT (Discrete Cosine Transform)

DCT biến đổi dữ liệu từ miền không gian sang miền tần số. **DCT** có đặc tính "gói" năng lượng tốt, biến đổi **DCT** cho kết quả là các số thực, ngoài ra có các thuật toán nhanh để thực hiện biến đổi này.

DCT hai chiều:

Biến đổi DCT hai chiều (2-D) được dùng cho các khối ảnh có kích thước 8×8 . Quá trình biến đổi thuận DCT (Forward DCT) được định nghĩa như sau:

$$F(u,v) = \frac{C(u)C(v)}{4} \sum_{j=0}^7 \sum_{k=0}^7 f(j,k) \cos \frac{(2j+1)u\pi}{16} \cos \frac{(2k+1)v\pi}{16}$$

$f(j,k)$ - các mẫu của ảnh gốc trong block 8×8 pixel.

F(u,v)-các hệ số của khối DCT 8x8.

$$C(u), C(v) = \begin{cases} 1/\sqrt{2} & u, v = 0 \\ 1 & u, v \neq 0 \end{cases}$$

Trong ma trận hệ số DCT hai chiều, **hệ số thứ nhất** $F(0,0)$ bằng giá trị trung bình của các điểm ảnh trong block 8x8:

$$F(0,0) = \frac{1}{8} \sum_{j=0}^7 \sum_{k=0}^7 f(j,k)$$

Phép biến đổi DCT hai chiều là biến đổi đối xứng và **biến đổi nghịch** cho phép tái tạo lại các giá trị mẫu $f(j,k)$ trên cơ sở các hệ số $F(u,v)$ theo công thức sau:

$$f(j,k) = \sum_{u=0}^7 \sum_{v=0}^7 \frac{C(u)C(v)}{4} F(u,v) \cos \frac{(2j+1)u\pi}{16} \cos \frac{(2k+1)v\pi}{16}$$

Phép biến đổi DCT **không nén được dữ liệu**, từ 64 mẫu ta nhận được 64 hệ số. Trong các hệ số DCT, thành phần DC thường có giá trị lớn nhất, các hệ số nằm kề nó ứng với tần số thấp có giá trị nhỏ hơn, các hệ số còn lại ứng với tần số cao thường có giá trị rất nhỏ.

• Lượng tử hóa (không đồng đều)

Để thực hiện quá trình nén dữ liệu, ma trận các hệ số khai triển sau DCT phải được chia cho bảng trọng số $Q(u,v)$ để loại bỏ một phần các hệ số DCT có biên độ nhỏ (thường là các thành phần cao tần).

$$F_q(u,v) = \text{round} \left[\frac{F(u,v)}{Q(u,v)} \right]$$

0	u	7
v		
7		
16	11	10
12	12	14
14	13	16
14	17	22
18	22	37
24	35	55
49	64	78
72	92	95
11	16	16
19	26	24
40	58	40
51	60	51
55	56	56
57	69	57
80	62	62
109	103	77
113	92	77
104	121	92
120	101	92
103	112	98
100	112	98
103	99	99
99	99	99

Bảng trọng số (theo chuẩn JPEG cho mẫu tín hiệu chói)

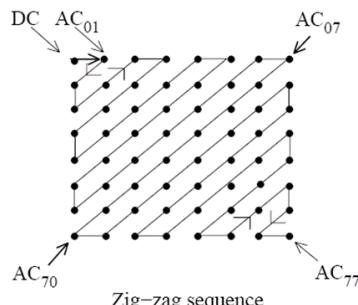
0	u	7
v		
7		
17	18	24
18	21	26
24	26	56
47	66	99
99	99	99
99	99	99
99	99	99
99	99	99
99	99	99
99	99	99
99	99	99
99	99	99

Bảng trọng số (theo chuẩn JPEG cho mẫu tín hiệu màu)

• Quét zig-zag

JPEG sử dụng phương pháp đọc các hệ số theo đường zig-zag (quét zig-zag). Mục đích làm tăng tối đa độ dài của chuỗi các giá trị 0 liên tiếp. Như vậy hiệu quả nén khi dùng mã RLC sẽ tăng.

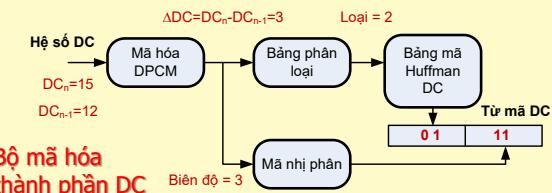
240	0	-10	0	0	0	0	0	0
-24	-12	0	0	0	0	0	0	0
-14	-13	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0



• Mã hóa thành phần DC

Theo JPEG, các **hệ số DC** được mã hóa theo phương pháp **DPCM**. Để tăng hiệu suất nén, kết quả nhận được sau đó được mã hóa tiếp bằng **mã Huffman**.

Dài động của các giá trị nhận được sau khi mã hóa DPCM lấp gấp đôi dài động của các hệ số DCT, có nghĩa là thành phần này sẽ có giá trị nằm trong khoảng -2^{11} đến $2^{11} - 1$. Số lượng bits cần thiết để mã hóa thành phần DC có thể lên tới 11.



Các hệ số DC sai lệch	Phân loại	Tùy mã (chối)
-255...-128; 128...255	8	1111 110
-127...-64; 64...127	7	1111 10
-63...-32; 32...63	6	1111 0
-31...-16; 16...31	5	1110
-15...-8; 8...15	4	110
-7...4; 4...7	3	101
-3; 2; 3	2	01
-1; 1	1	00
0	0	100

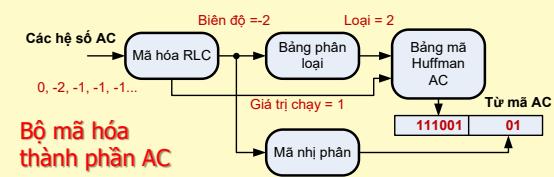
Bảng phân loại các hệ số AC và DC

Ví dụ: Thành phần DC có giá trị $DC_n=15$. Giả sử thành phần DC của block trước đó là $DC_{n-1}=12$. Mã DPCM: $\Delta DC = DC_n - DC_{n-1} = 3$. Tra bảng tìm ra $\Delta DC = 3$ thuộc "loại" 2.

Tra bảng mã Huffman, từ mã ứng với "lợi" 2 là (01) (2 chính là độ dài của mã). Giá trị $ADC = 3$ được biểu diễn bằng chuỗi nhị phân "11". Như vậy từ **mã DC sẽ là 0111**.

15	0	-1	0	0	0	0	0
-2	-1	0	0	0	0	0	0
-1	-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

Block hệ số khai triển DCT



Mã hóa RLC cho các hệ số AC ta có: 1) Giá trị chạy – là số lượng bit "0" đứng trước hệ số khác "0" đang được mã hóa; 2) biên độ của hệ số khác "0" nói trên.

Từ mã Huffman ứng với cặp giá trị trên được tìm ra trong bảng phân loại và bảng mã Huffman cho thành phần AC. Từ mã AC sẽ bao gồm từ mã Huffman và giá trị biên độ (nhi phân) của hệ số AC.

Ví dụ: Sau quá trình quét zig-zag, ta nhận được chuỗi hệ số AC sau:

$$0, -2, -1, -1, -1, 0, 0, -1, 0, 0, \dots$$

Chuỗi bít nhân được sau bô mã RLC là:

(1,-2) (0, -1) (0, -1) (0, -1) (2,-1) (EOB)

Tín hiệu được đưa vào mã hóa Huffman có cấu trúc sau:

$$(1,2)(-2), (0,1)(-1), (0,1)(-1), (0,1)(-1), (2,1)(-1), (0,0)$$

(1,2)(-2), (0,1)(-1), (0,1)(-1), (0,1)(-1), (2,1)(-1), (0,0)

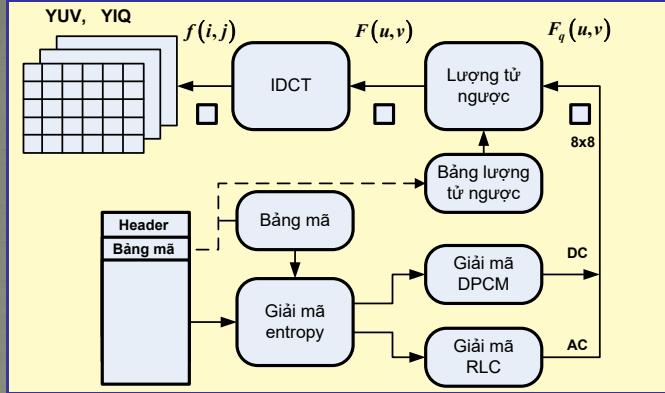
O chay	loai	Đô dài mã	Tử mã
0	1	2	00
0	2	2	01
0	3	3	100
0	4	4	1011
0	5	5	11010
0	6	6	111000
0	7	7	1111000
.	.	.	
1	1	4	1100
1	2	6	111001
1	3	7	1111001
1	4	9	111110110
.	.	.	
11	1	9	111111010
.	.	.	
EOB	.	4	1010

Chỉ cần 35 bits để truyền đi block 64 điểm ảnh, như vậy hiệu quả nén của phương pháp JPEG trong trường hợp này là 0.5 bit/điểm ảnh.

Bảng mã Huffman cho thành phần AC

Kết quả mã hóa các thành phần AC và DC được tập hợp lại thành chuỗi bit có dạng như sau:

0111 11000101 0000 0000 0000 1101100 1010
code Huffman



Sơ đồ khỗi bộ giải mã ảnh tĩnh JPEG

Quá trình **lượng tử ngược** được thực hiện theo công thức:

$$F(u,v) = F_q(u,v) \cdot Q(u,v)$$

Nén JPEG lũy tiến

Có hai phương pháp nén JPEG: 1) nén **JPEG căn bản** (Baseline Lossy JPEG) hay JPEG tuần tự. 2) Nén **JPEG lũy tiến**.

Phương pháp nén **JPEG lũy tiến** thực hiện ghi dữ liệu theo cách đặc biệt để decoder có thể giải mã toàn bộ ảnh ở "mức thô" ngay mà không cần chờ đến khi quá trình truyền ảnh kết thúc.

Nén JPEG lũy tiến có thể thực hiện theo 3 **giải thuật sau**:

- 1) Giải thuật lựa chọn phổ (progressive spectral selection algorithm).
- 2) Giải thuật xấp xỉ thành công (progressive successive approximation).
- 3) Giải thuật kết hợp (combined progressive algorithm).

Giải thuật **lựa chọn phổ**: truyền các hệ số trong miền tần số thấp trước, sau đó truyền các hệ số còn lại.

Giải thuật **xấp xỉ**: tất cả các hệ số DCT được gửi đi trước có độ chính xác thấp hơn, thí dụ: bước 1 truyền 4 bits, bước hai truyền thêm 4 bits, bước ba truyền nốt 2 bits còn lại của số nhị phân biểu diễn biên độ hệ số AC.



Giải mã ảnh tĩnh
JPEG tuần tự



Giải mã ảnh tĩnh JPEG lũy tiến

Hai khôi ảnh (hình dưới) được đưa vào bộ nén ảnh JPEG.

1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3
1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3
1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3
1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3
1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3

Thực hiện quá trình nén và giải nén JPEG cho khôi ảnh thứ hai. (Bảng phân loại và bảng mã tra trong tài liệu). Đánh giá kết quả và tìm tỷ lệ nén.

Các tham số tiêu chuẩn của phương pháp nén JPEG

Tham số	Đặc điểm
Tín hiệu mã hóa	RGB hoặc Y và C _R , C _B
Cấu trúc lấy mẫu	4:4:4, 4:2:2 và 4:2:0
Kích thước ảnh tối đa (điểm ảnh)	65 536 x 65 536
Biểu diễn mẫu	8 bit cho hệ thống cơ bản 8-12 bit cho hệ thống mở rộng DCT
Độ chính xác của quá trình lượng tử hóa và biến đổi DCT	9 bit
Phương pháp lượng tử hóa hệ số DC	DPCM
Cấu trúc khôi trong quá trình lượng tử hóa thích nghi	16x16 bit
Độ chính xác cực đại của hệ số DC	11 bit
Bảng lượng tử	Sai lệch giữa các giá trị Y và C _R , C _B
Biến đổi RLC	Mã Huffman
Hệ số cân bằng các khôi	Có thể biến đổi
Bù chuyển động	Không
Quét	Tuần tự hay xen kẽ
Kênh truyền	Được quản lý lỗi

➤ Các phương pháp nén ảnh động

➤ Phương pháp nén ảnh động theo chuẩn M-JPEG

M – JPEG là chuẩn mở rộng của JPEG. Chuỗi các ảnh tĩnh của dòng video có thể được nén **trong ảnh** bằng phương pháp JPEG.

Hệ số nén đạt được khi nén ảnh động theo phương pháp M – JPEG phụ thuộc vào chất lượng ảnh nén: tín hiệu video nén M-JPEG chất lượng cao (dùng trong studio) có tốc độ dòng bits trên **24 Mbit/s**. Vì **các ảnh được mã hóa độc lập** với nhau nên quá trình ghép cảnh khi dàn dựng trong studio sẽ có độ chính xác tới từng ảnh.

Các thiết bị sử dụng định dạng nén M – JPEG có **nhược điểm** sau:

- Các thiết bị sử dụng phương pháp nén theo định dạng M – JPEG không thể sử dụng cho truyền dẫn, phát sóng vì **tốc độ dòng bit sau khi nén còn cao**.
- Kỹ thuật nén M-JPEG chưa được chuẩn hóa như các định dạng nén khác (JPEG, MPEG-2 v.v.), nên thiết bị nén của các hãng khác nhau thường **không có tính tương thích cao**, do đó khó có thể trao đổi trực tiếp số liệu cho nhau.
- Các phần mềm giải mã video hỗ trợ định dạng M-JPEG còn chưa phổ biến.

➤ Dư thừa “trong pixel”

139	144	149	153	155	155	155	155
144	151	153	156	159	156	156	156
150	155	160	163	158	156	156	156
159	161	162	160	160	159	159	159
159	160	161	162	162	155	155	155
161	161	161	161	160	157	157	157
162	162	161	163	162	157	157	157
162	162	161	161	163	158	158	158

235.6	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	1.3	16	11
-22.6	-17.5	-6.2	-3.2	-2.9	-0.1	0.4	-1.2	12	12
-10.9	-9.3	-1.6	1.5	0.2	-0.9	-0.6	-0.1	14	13
-7.1	-1.9	0.2	1.5	0.9	-0.1	0.0	0.3	14	17
-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	1.3	18	22
1.8	-0.2	1.6	-0.3	-0.8	1.5	1.0	-1.0	24	35
-1.3	-0.4	-0.3	-1.5	-0.5	1.7	1.1	-0.8	49	64
-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4	72	92

a) Block ảnh gốc 8x8

b) Block các hệ số DCT

15	0	-1	0	0	0	0	0
-2	-1	0	0	0	0	0	0
-1	-1	0	0	0	0	0	0

240	0	-10	0	0	0	0	0	144	146
-24	-12	0	0	0	0	0	0	148	150
-14	-13	0	0	0	0	0	0	155	156

Ảnh n

Ảnh n+1

➤ Phương pháp nén ảnh động theo chuẩn MPEG

Nén tín hiệu video theo **chuẩn MPEG** (Moving Picture Experts Group) là phương pháp nén ảnh động theo nguyên tắc **giảm dư thừa trong không gian** và **giảm dư thừa giữa các khung ảnh** (frame).

Các kỹ thuật nén được chuẩn hóa bởi MPEG bao gồm MPEG-1, MPEG-2, MPEG-4, MPEG-7. Song song với các định dạng nén cho tín hiệu video, trong các chuẩn MPEG còn xác định các phương pháp nén cho tín hiệu audio. Ví dụ, chuẩn MPEG-1 định nghĩa 3 lớp nén audio như sau:

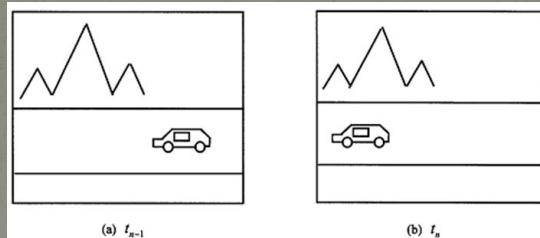
- [MP1](#) or MPEG-1 Part 3 Layer 1 ([MPEG-1 Audio Layer I](#))
- [MP2](#) or MPEG-1 Part 3 Layer 2 ([MPEG-1 Audio Layer II](#))
- [MP3](#) or MPEG-1 Part 3 Layer 3 ([MPEG-1 Audio Layer III](#))

Level	Profile			
	Simple 4:2:0	Main 4:2:0	SNR scalable 4:2:0	Spatially scalable 4:2:0 High 4:2:0 or 4:2:2
High 1920 × 1152 (60 frames/s)		62.7 Ms/s 80 Mbit/s		100 Mbit/s for 3 layers
High-1440 1440 × 1152 (60 frames/s)		47 Ms/s 60 Mbit/s	47 Ms/s 60 Mbit/s for 3 layers	80 Mbit/s for 3 layers
Main 720 × 576 (30 frames/s)	10.4 Ms/s 15 Mbit/s	10.4 Ms/s 15 Mbit/s	10.4 Ms/s 15 Mbit/s for 2 layers	20 Mbit/s for 3 layers
Low 352 × 288 (30 frames/s)		3.04 Ms/s 4 Mbit/s	3.04 Ms/s 4 Mbit/s for 2 layers	

➤ Kỹ thuật bù chuyển động trong MPEG

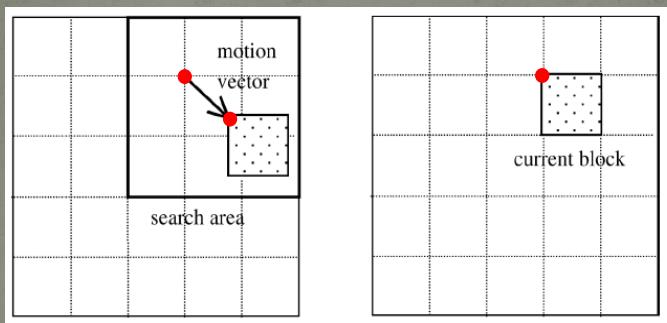
Nguyên lý **bù chuyển động** là tìm ra các vùng ảnh giống nhau hoặc gần giống nhau giữa hai ảnh kế tiếp (gọi là **ảnh tham khảo** và **ảnh dự đoán**). Thông tin về ảnh dự đoán sẽ được truyền đi dưới dạng **vector chuyển động** tức vector xác định vị trí của vùng ảnh thuộc ảnh tham khảo trong ảnh dự đoán.

Quá trình tìm ra vector chuyển động gọi là quá trình **xấp xỉ chuyển động** (**motion estimation**), quá trình khôi phục ảnh dựa trên vector này gọi là quá trình **bù chuyển động** (**motion compensation**).



Kỹ thuật so khớp khối (block matching)

Ảnh được chia thành các khối hình chữ nhật không chồng lên nhau có kích thước 16×16 . Tất cả các điểm ảnh trong khối có cùng vector chuyển động. Với mỗi khối, chúng ta sẽ tìm được khối tương xứng nhất với nó trong ảnh trước đó qua sự tương quan. Cụ thể, khối nào trong ảnh trước đó cho giá trị tương quan lớn nhất thì vị trí tương đối của hai khối này sẽ cho xác định vector chuyển động.



Ảnh tham khảo (ảnh trước)

Ảnh dự đoán (ảnh tức thời)

Đánh giá mức độ sai số giữa hai ảnh:

- Sai số tuyệt đối $e(x,y) = \hat{f}(x,y) - f(x,y)$
- Sai số trung bình bình phương - RMS (Root Mean Square).

$$e_{RMS} = \left[\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (\hat{f}(x,y) - f(x,y))^2 \right]^{1/2}$$

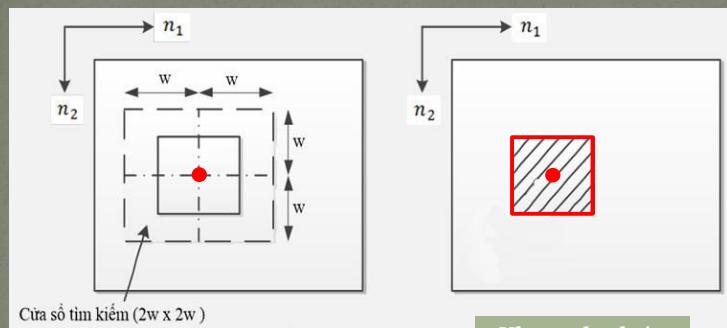
- Tỷ lệ tín hiệu/nhiễu SNR (Signal to Noise Ratio) :

$$SNR = \left[\frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \hat{f}(x,y)^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (\hat{f}(x,y) - f(x,y))^2} \right]^{1/2}$$

Sum of absolute differences (SAD)

$$SAD(i,j) = \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} |s(n_1, n_2, k) - s(n_1 + i, n_2 + j, k - l)|.$$

Kỹ thuật so khớp khôi



➤ Giải thuật vét cạn (Exhaust Search – ES)

- Độ chính xác
- Chi phí tính toán -> vùng tìm kiếm.
- Mức độ phức tạp không đổi xứng
- Không chuẩn hóa phương pháp ước lượng

Các phép so sánh khôi theo tiêu chí SAD yêu cầu 3 phép toán (trừ, giá trị tuyệt đối, cộng). Vì vậy, chi phí để tìm vector chuyển động cho một khôi ảnh là: $(2p+1).(2p+1).N^2.3$ phép tính.
 $P=2W$ – kích thước vùng tìm kiếm, N – kích thước macroblock

Nếu ảnh FHD có $1920 \times 1080 / (N.N)$ khôi ảnh thì tổng phép toán cần thực hiện trên giây là:
 $(2p+1).(2p+1).N^2.3 * 1920 \times 1080 / (N.N)$ phép tính.

Tìm hệ số nén tối đa cho video FHD sử dụng MV?



Các vector chuyển động trong ảnh bù chuyển động

Hiệp hội viễn thông quốc tế (ITU) và tổ chức tiêu chuẩn quốc tế/ Uỷ ban kỹ thuật điện tử quốc tế (ISO/IEC) là hai tổ chức phát triển các tiêu chuẩn mã hoá Video.

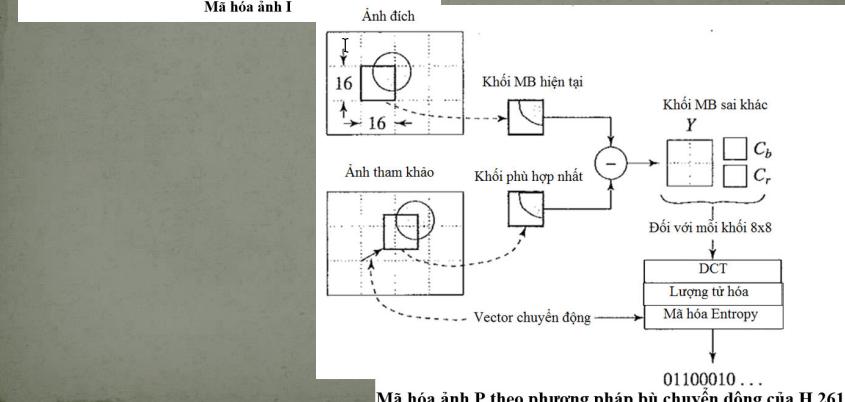
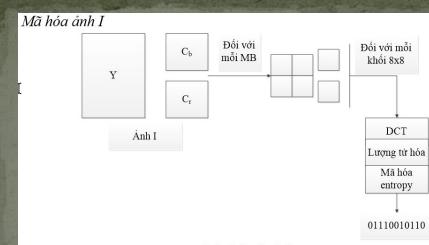
Những khuyến nghị của ITU được thiết kế dành cho các ứng dụng truyền thông Video thời gian thực như Video Conferencing hay điện thoại truyền hình.

Các chuẩn mã hóa video MPEG của ISO nhằm vào việc lưu trữ và phân phối video cho lĩnh vực truyền hình số.

Chuẩn mã hóa video đầu tiên của ITU-T là H.261 ưng dụng cho hội nghị truyền hình và kênh truyền là ISDN.

Chuẩn H.263 cung cấp hiệu suất nén tốt có 4 chế độ chất lượng tùy chọn.

H.264 (kết hợp giữa ITU-T và ISO/IEC MPEG) cải thiện hiệu suất nén lên khoảng 50% ở các tốc độ bit thấp. Ngoài ra, H.264, được thông qua như là một phần của chuẩn MPEG-4, thường được gọi là **MPEG-4 AVC (Advanced Video Coding)**



Các loại ảnh nén trong MPEG

1- Ảnh loại I (Intra-picture)

Là ảnh được mã hóa trong ảnh bằng kỹ thuật nén giống JPEG. Ảnh I là ảnh tham khảo để dự đoán các ảnh B,P.

2- Ảnh loại P (Predicted-picture)

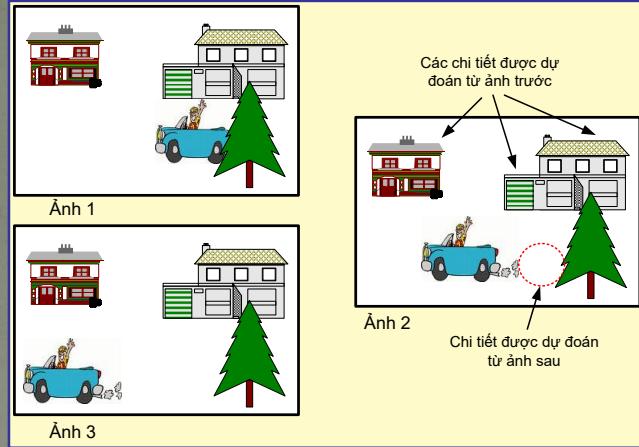
Là ảnh được mã hóa với vector chuyển động, được xấp xỉ từ ảnh I hay P trước nó.

3- Ảnh loại B (Bi-directional predicted picture)

Là ảnh được mã hóa với vector chuyển động được xấp xỉ dựa trên các ảnh I hoặc P ở phía trước và ở phía sau nó.

4- Ảnh loại D (DC-coded picture)

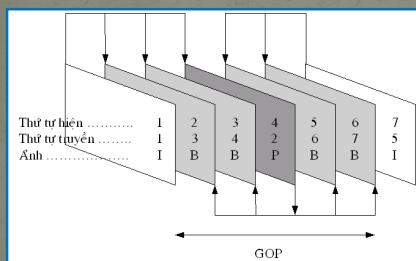
Ảnh D được mã hóa giống như ảnh I, tuy nhiên chỉ có thành phần một chiều (DC) được lưu lại.



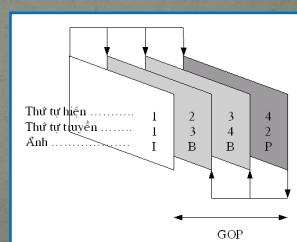
Quá trình bù chuyển động trong ảnh B

➤ Nhóm ảnh GOP (Group of Picture)

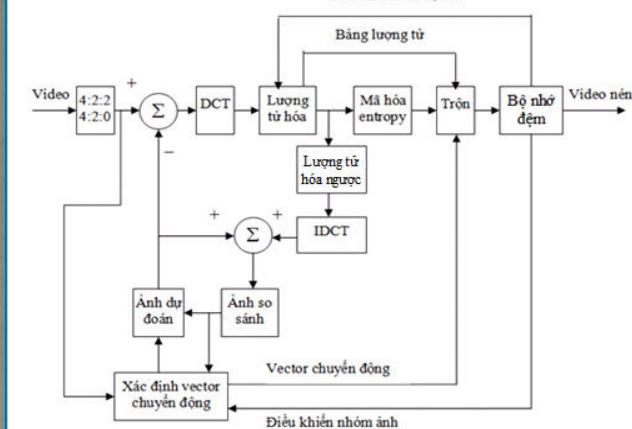
Nhóm ảnh mở luôn bắt đầu từ một ảnh I và kết thúc bằng một ảnh B trước ảnh I của nhóm sau.



Thay đổi trình tự ảnh khi
truyền tín hiệu MPEG



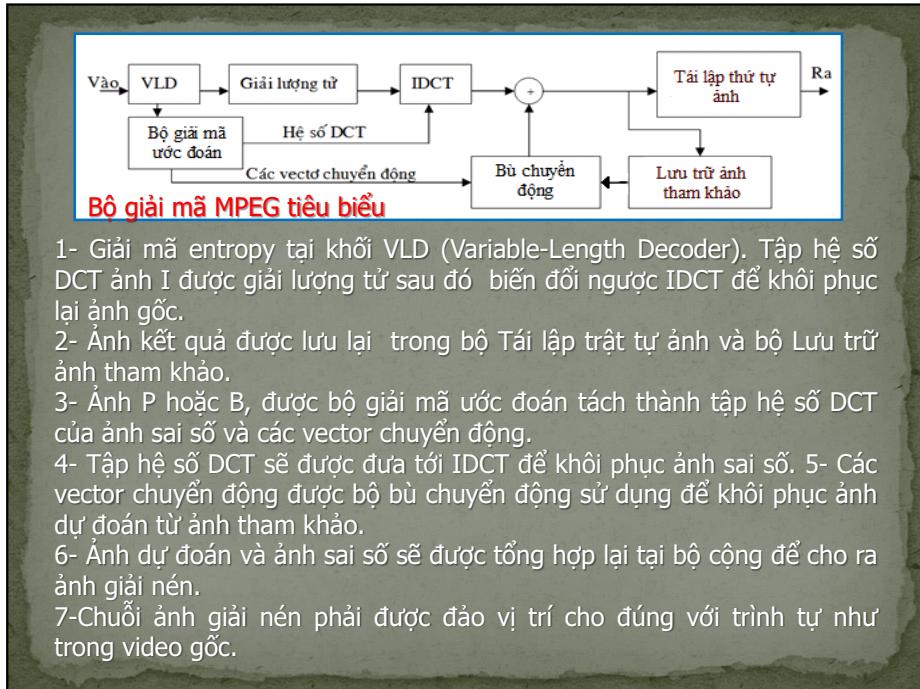
GOP có **cấu trúc khép kín** (đóng),
bắt đầu bằng I và kết thúc bằng I
hoặc P. Do đó dự đoán ảnh không
sử dụng thông tin của GOP khác



Bộ mã hóa MPEG tiêu biểu

Hai cơ chế điều khiển tốc độ luồng video nén trong MPEG-2:

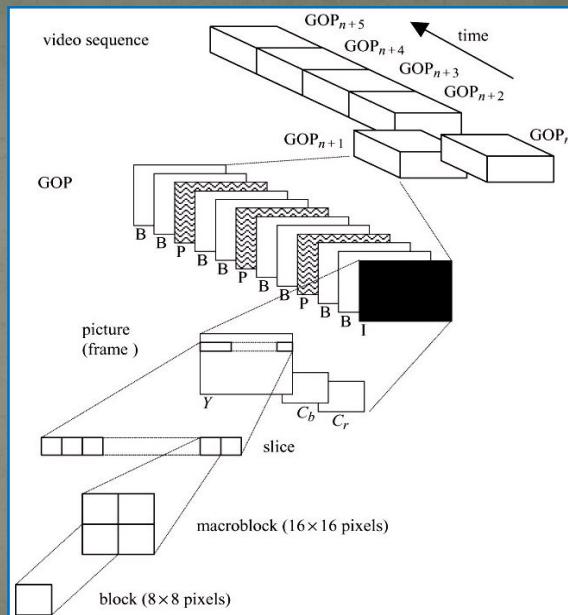
- Dựa trên nguyên tắc thay đổi cấu trúc GOB.
- Thay đổi ma trận lượng tử



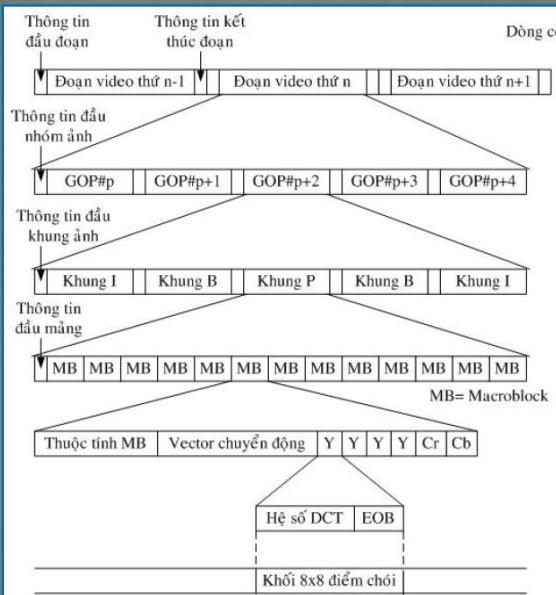
➤ Các thành phần cơ bản trong ảnh nén MPEG

1. **Lớp khối:** là block 8×8 điểm ảnh của tín hiệu chói hoặc tín hiệu màu .
2. **Lớp macroblock:** Macroblock có kích thước 16×16 . Trong một macroblock có thông tin về 4 block Y, 1 block Cr và 1 block Cb (theo chuẩn lấy mẫu 4:2:0). Macroblock sau khi mã hóa sẽ chứa thông tin về vector chuyển động và sai số dự đoán.
3. **Mảng (Slice):** là một chuỗi macroblock kề nhau. Kích thước lớn nhất của mảng có thể bao gồm toàn bộ bức ảnh và kích thước nhỏ nhất của mảng là một macroblock. **Slice header** chứa đựng vị trí của mảng trong toàn bộ ảnh, và hệ số lượng tử dùng để xác định ma trận lượng tử trong quá trình giải mã slice.
4. **Ảnh (Picture):** có 3 loại ảnh là ảnh P, I hay ảnh B. Picture header chứa thông tin về: thứ tự ảnh trong nhóm ảnh (thông tin này dùng để sắp xếp các ảnh thứ tự ảnh ở decoder), loại ảnh, kích thước vùng tìm kiếm vector chuyển động.
5. **Nhóm ảnh (GOP):** nhóm ảnh là tổ hợp của nhiều ảnh I, P và B. Mỗi một nhóm ảnh bắt đầu bằng một khung I. GOP header chứa mã xác định thời gian của ảnh đầu tiên trong nhóm.
6. **Đoạn ảnh (Sequence of pictures):** Đoạn ảnh bắt đầu bằng sequence header, sau đó là một hoặc nhiều GOP, cuối cùng là từ mã "end-of-sequence". Sequence header chứa đựng các thông số như: kích thước của ảnh, khổ ảnh, tần số ảnh, tốc độ bit của dòng video số, tần số ảnh và kích thước bộ nhớ đệm.

Cấu trúc luồng video nén theo chuẩn MPEG-1



Kiến trúc dòng dữ liệu MPEG



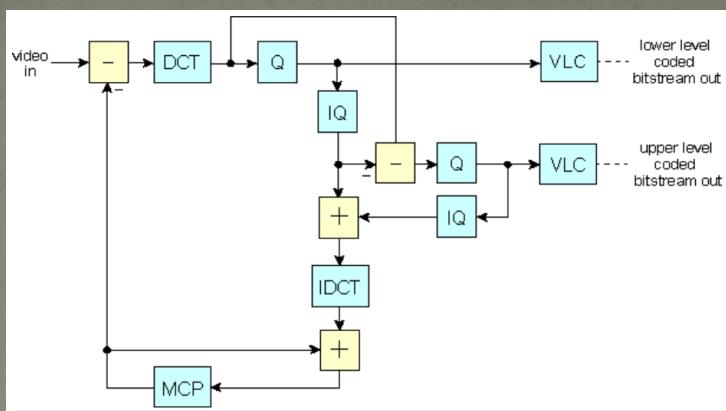
MPEG-2 là giải pháp mã hóa tín hiệu có **tính co dãn (scalability)** và **tính tương hợp (compatibility)**.

Tính "co dãn" (Scalable Coding) trong MPEG-2 dựa trên nguyên tắc nén tín hiệu video ở mức cơ bản (chất lượng trung bình) và các mức nâng cao (chất lượng cao).

Tính tương hợp cho phép decoder MPEG-1 giải mã tín hiệu MPEG-2 và ngược lại.

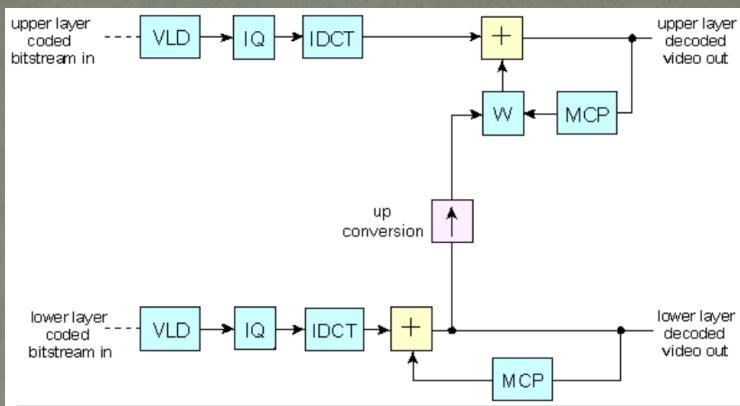
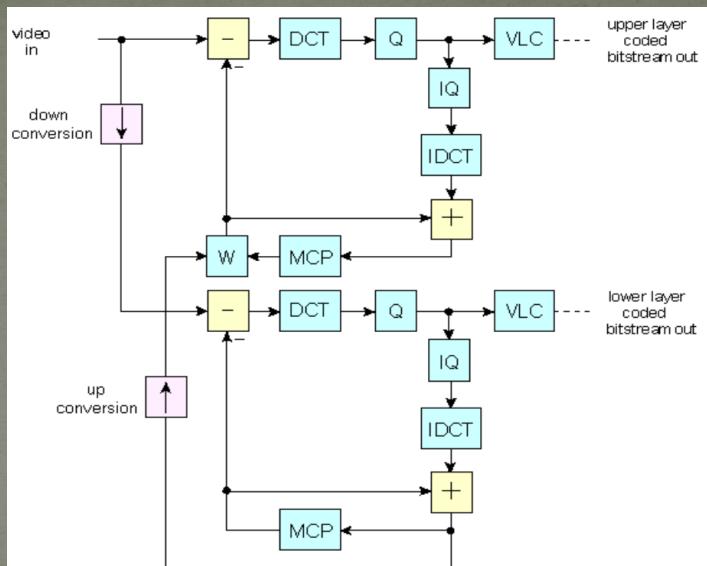
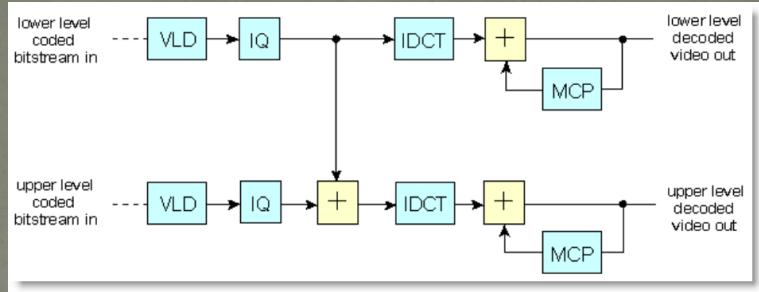
MPEG-2 hỗ trợ 3 dạng scalable coding:

- **Phân cấp chất lượng theo tỷ lệ SNR:** sử dụng các bảng lượng tử thô đồng thời truyền thêm ảnh sai số ở mức nâng cao.
- **Phân cấp chất lượng theo độ phân giải trong không gian (spatial scalability):** Tại mức cơ bản, tín hiệu được đưa vào mã hóa được lấy mẫu với tốc độ thấp. Tại mức nâng cao, người ta mã hóa và truyền đi ảnh sai số giữa ảnh có độ phân giải thấp và ảnh có độ phân giải cao (được lấy mẫu với tốc độ cao hơn).
- **Phân cấp chất lượng theo thời gian (temporal scalability):** Tại mức cơ bản, tín hiệu video đưa vào mã hóa có tần số ảnh thấp. Tại mức nâng cao, thông tin về các vector chuyển động sẽ được truyền sang phía thu, decoder (dựa trên các ảnh được truyền trong mức cơ bản) thực hiện quá trình nội suy ảnh, làm tăng tần số ảnh của tín hiệu giải nén.



MPC - Motion-compensated prediction

Bộ mã hóa MPEG-2 phân cấp chất lượng theo SNR



MPEG-2 Profiles					
Abbr.	Name	Frames	chroma format	Streams	Comment
SP	Simple Profile	P, I	4:2:0	1	no interlacing
MP	Main Profile	P, I, B	4:2:0	1	
422P	4:2:2 Profile	P, I, B	4:2:2	1	
SNR	SNR Profile	P, I, B	4:2:0	1-2	SNR: Signal to Noise Ratio
SP	Spatial Profile	P, I, B	4:2:0	1-3	
HP	High Profile	P, I, B	4:2:2	1-3	low, normal and high quality decoding

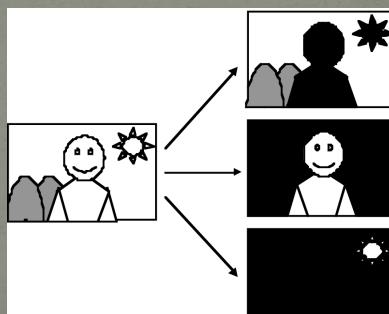
MPEG-2 Levels					
Abbr.	Name	Pixel/line	Lines	Framerate (Hz)	Bitrate (Mbit/s)
LL	Low Level	352	288	30	4
ML	Main Level	720	576	30	15
H-14	High 1440	1440	1152	30	60
HL	High Level	1920	1152	30	80

MPEG có 6 profiles và 4 mức nén với các độ phân giải khác nhau: tạo ra 24 tiêu chuẩn chất lượng

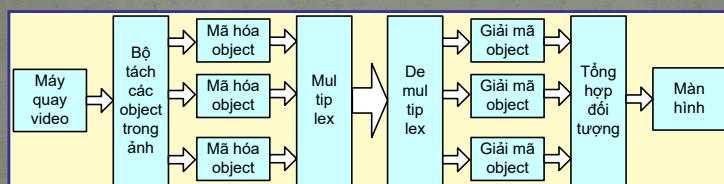
➤ Phương pháp nén ảnh động theo chuẩn MPEG-4

MPEG-4 được dùng để mã hóa video và audio với tốc độ rất thấp (<64 kbps), cũng như nén video chất lượng cao và có thể được sử dụng trong hệ thống HDTV. Những profile và level khác nhau trong MPEG-4 cho phép sử dụng tốc độ bit lên đến 1.2Gbps. MPEG-4 là chuẩn quốc tế đầu tiên cho phép mã hóa các **đối tượng (object) video**. Kỹ thuật mã hóa đối tượng làm tăng hiệu quả nén và tính linh động của MPEG-4.

Ví dụ các Object trong MPEG 4

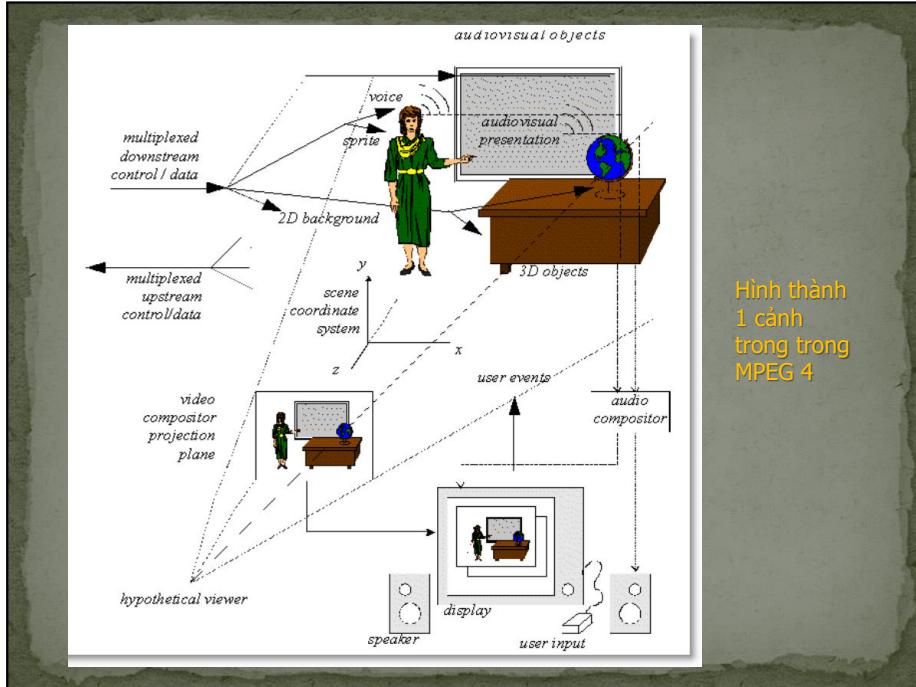


Trong MPEG-4, các **đối tượng** của ảnh được tách rời và mã hóa riêng rẽ sau đó truyền tới bộ giải mã. Việc tách rời các đối tượng video như vậy làm tăng **tính mềm dẻo** khi thực hiện mã hóa thích nghi và đồng thời làm tăng hiệu quả nén tín hiệu. Các đối tượng khác nhau như **video object** và **audio object** và được kết hợp tại bộ giải mã. Các loại object khác nhau sẽ được mã hóa với những kỹ thuật khác nhau và với các công cụ phù hợp nhất.



Chuẩn MPEG 4 cho phép mô tả cảnh trong đoạn video theo cách sau:

- Xác định vị trí của các thành phần video (VO) trong không gian cho trước
- Thực hiện biến đổi hình dạng cũng như âm thanh kèm theo của các VO
- Mô tả một VO dựa trên các thành phần ảnh cơ bản (primitive)
- Thay đổi góc nhìn cũng như góc nghe của người xem để tạo ra cảm giác tồn tại trong các điểm khác nhau của cảnh quay



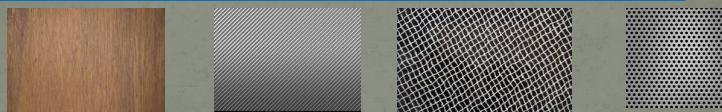
Hình thành
1 cảnh
trong trong
MPEG 4

Ba đặc tính rất quan trọng của MPEG-4 là:

- 1- Nhiều object có thể được mã hóa với các kỹ thuật khác nhau và kết hợp lại ở bộ giải mã.
- 2- Các object có thể là các cảnh tự nhiên có được từ camera hay các cảnh tạo ra trên máy tính như text hay cảnh hoạt hình 2 hoặc 3-D.
- 3- Từ luồng bit MPEG-4, bộ giải mã có thể tách ra các thông tin khác nhau tùy theo lựa chọn người xem chẳng hạn như ngôn ngữ thuyết minh hay thành phần video được lọc bỏ các thông tin không cần thiết (dành riêng cho trẻ em) v.v.

Một cảnh tiêu biểu trong MPEG-4 bao gồm bao gồm **ảnh nền** (background) và một hoặc nhiều **đối tượng cận ảnh** (foreground) chẳng hạn như đồ vật, một hoặc nhiều người, một vài phần tử đồ họa.

Video object được mô tả bởi hai phần tử là ảnh trong đối tượng gọi là "texture" và **key signal** hay alpha channel được xem là hình dạng của đối tượng.



Kết cấu dạng vân gỗ, sọc chéo, dạng lưới, dạng ô tròn v.v.

Mỗi object trong MPE-4 được lấy mẫu để tạo ra chuỗi ảnh gọi là **video object plane** (VOP). Object tĩnh có thể mô tả bằng một VOP. Dữ liệu được truyền đi bao gồm thông tin về **hình dạng (shape)** và **texture** của VOP. VOP cũng giống như các frame trong MPEG-2, chúng có thể được mã hóa trong ảnh hoặc mã hóa liên ảnh sử dụng phương pháp bù chuyển động.

MODELS	CODED INFORMATION	EXAMPLES
Pixels	Color of pixels	PCM
Statistically dependent pixels	Prediction error or transform coeffs	Predictive Coding Transform Coding
Moving blocks	Motion vectors and prediction error	Block-based coding H.261/263, MPEG-1/2
Moving regions	Shapes, motion, and colors of regions	Region-based coding H.263+, MPEG-4
Moving objects	Shapes, motion, and colors of objects	Model-based coding MPEG-4
Facial models	Action units	MPEG-4
A/V objects	Descriptive languages	MPEG-7

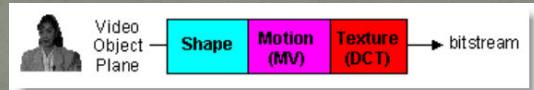
Cấp độ của video MPEG-4

Mỗi object trong MPEG-4 được lấy mẫu để tạo ra chuỗi ảnh gọi là **video object plane** (VOP). Object tính có thể mô tả bằng một VOP. Dữ liệu được truyền đi bao gồm thông tin về **hình dạng (shape)** và **texture** của VOP. VOP cũng giống như các frame trong MPEG-2, chúng có thể được mã hóa trong ảnh hoặc mã hóa liên ảnh sử dụng phương pháp bù chuyển động.

Lớp trên của VOP là **GOV** (**Group of video object planes**). GOV tương tự như GOP (group of pictures) trong MPEG-2.

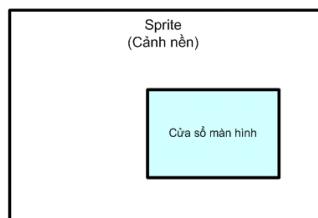
Lớp trên GOV là **VOL** (**Video object layer**) bao gồm nhiều GOV

Mức **video object** (**VO**) bao gồm các VOL dùng để dùng để mô tả đối tượng video. Video session (VS) là mức cao nhất của cảnh MPEG-4 bao gồm tất cả đối tượng video cả tự nhiên và tự tạo trong một cảnh.

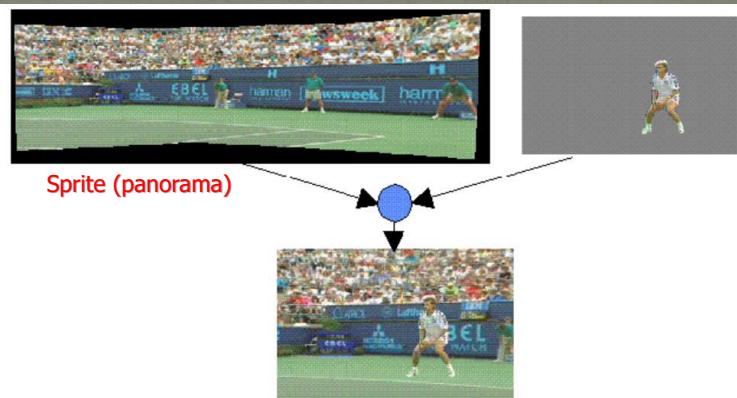


Sprites

MPEG-4 có một loại đối tượng đặc biệt được dùng làm cảnh nền gọi là **sprite**. Sprite là đối tượng video có kích thước lớn hơn màn hình hiển thị. Sprite là đối tượng được sử dụng liên tục trong một cảnh. MPEG-4 cho phép truyền toàn bộ cảnh nền như sprite và ảnh nền tức thời được đọc ra từ sprite khi cần. Trong video game, các thành phần của một sprite có thể được sử dụng nhiều lần vì thế lượng dữ liệu cần truyền sẽ giảm đáng kể.



Sprite được mã hóa như tín hiệu chói với hai thành phần màu như trong MPEG-2 và luôn được mã Intra bởi vì bản chất của sprite là ảnh tĩnh.



Hình thành 1 cảnh trong trong MPEG 4 với thành phần sprite

Ứng dụng của các tiêu chuẩn nén.

Ứng dụng	Tiêu chuẩn mã hóa	Độ phân giải cực đại	Tốc độ bit cực đại
Videophone	H.261	176x144	64 ÷ 128 Kbps
Hội nghị truyền hình	H.261	352x288	0.384 ÷ 1.554 Mbps
Truyền hình cáp	MPEG-2	720x576	4...9 Mbps
Truyền hình	MPEG-2	720x576	< 50 Mbps
Studio / sản xuất	MJPEG	720x576	< 50 Mbps
	MPEG-2(@4:2:2)		
HDTV / sản xuất	MPEG-2	1920x1280	100 Mbps
HDTV / truyền dẫn	MPEG-2	1920x1280	20 Mbps

EBU System	Nomenclature and abbreviation [samples horiz. x active lines/ Scanning/ frame rate]	Luma or R'G'B' Samples per active line (S/AL)	Active lines per frame (picture) (AL/F)	Frame rate, Hz	Luma or R'G'B' sampling ³ frequency (fs), MHz	Luma sample periods per total line (S/TL)	Total lines per frame	Net image Bit Rate (4:2:2, 10 bit) [Mbit/s]	Corresponding SMPTE system nomenclature
S1	1280x720/P/50 (abbreviated: 720/P/50)	1280	720	50	74.25	1980	750	921.6	Corresponds to SMPTE 296M System 3
S2	1920 x 1080/I/25 (abbreviated: 1080/I/25)	1920	1080	25 (50Hz field rate)	74.25	2640	1125	1036.8	Corresponds to SMPTE 274 System 6
S3	1920 x 1080/P/25 (abbreviated: 1080/P/25)	1920	1080	25	74.25	2640	1125	1036.8	Corresponds to SMPTE 274 System 9
S4	1920 x 1080/P/50 (abbreviated: 1080/P/50)	1920	1080	50	148.5	2640	1125	2073.6	Corresponds to SMPTE 274 System 3

HDTV Systems 1 to 4