

Visión por Computador

Domingo Mery

Departamento de Ciencia de la Computación
Universidad Católica de Chile
Santiago de Chile
17 de agosto de 2004

**Departamento de Ciencia de la Computación
Universidad Católica de Chile**

Domingo Mery:

Av. Vicuña Mackenna 4860(143)

eMail: dmery@ieee.org

<http://www.ing.puc.cl/~dmery>

Prefacio

En este documento se pretende resumir el curso Visión por Computador. Este curso debe tomarse como una introducción a la teoría desarrollada en las dos últimas décadas por la comunidad de *Computer Vision*. Con esta teoría es posible analizar en detalle la información que puede ser obtenida a partir de un conjunto de vistas de una escena.

El curso comienza con una breve introducción al tema siguiendo con una descripción de la la teoría que mejor modela la formación de las imágenes: La Geometría Proyectiva. A continuación se aprenderán los distintos modelos existentes para establecer la función de transferencia $3D \rightarrow 2D$ en un Sistema de Visión por Computador y se analizarán los algoritmos más conocidos para llevar a cabo la calibración del sistema. En el siguiente capítulo del curso se estudiará la Visión Estéreo en el que se establecerán las relaciones existentes entre dos, tres y cuatro vistas de una misma escena. Seguidamente, se mostrarán los principales algoritmos para hacer una reconstrucción 3D de una escena a partir de sus vistas. En el siguiente capítulo, Matching y Tracking, se estudiará cómo se puede encontrar la correspondencia entre las distintas imágenes de un objeto. Finalmente, se mostrarán algunas aplicaciones de la Visión por Computador en distintas áreas de Ciencias e Ingeniería.

Un agradecimiento muy sincero a mis alumnos y alumnas del curso de Visión Artificial dictado en el Magister de Ingeniería Informática de la Universidad de Santiago de Chile y en la carrera de Ingeniería Eléctrica de la Pontificia Universidad Católica de Chile. En especial quiero agradecer (en orden alfabético) a Daniela Cerón, Rolando Dünner, Sebastián Fingerhuth, Roberto Mir, Egor Montecinos y Felipe Ochoa, quienes me ayudaron muchísimo a corregir el manuscrito original.

Santiago, 2001-2004

Índice general

Prefacio	III
Indice General	VII
1. Introducción	1
1.1. ¿Qué es la Visión por Computador?	1
1.2. Ejemplos	8
1.3. Historia	10
1.4. Ejercicios	15
2. Geometría Proyectiva	17
2.1. Planos, puntos y líneas rectas	18
2.2. Transformaciones Proyectivas 2D	21
2.3. Categorización de las Transformaciones Proyectivas 2D	25
2.3.1. Transformación Isométrica (Euclídea)	25
2.3.2. Transformación de Similitud	27
2.3.3. Transformación Afín	27
2.3.4. Transformación Proyectiva General	28
2.3.5. Resumen de Transformaciones Proyectivas 2D	29
2.4. Transformaciones Proyectivas 3D	29
2.5. Ejercicios	34

3. Modelación Geométrica de un Sistema de Visión por Computador	41
3.1. Descripción de un Sistema de Visión por Computador	41
3.1.1. El Manipulador	41
3.1.2. Fuente de Energía	42
3.1.3. Sensor de Imagen	43
3.1.4. Conversor Análogo-Digital	43
3.1.5. Computador	43
3.2. La Cámara <i>Pinhole</i>	43
3.3. Cámara CCD	48
3.4. Distorsión del lente	51
3.5. Modelación de un manipulador	53
3.6. Calibración de un sistema de visión por computador	54
3.7. Ejercicios	55
4. Visión Estéreo	61
4.1. Análisis Bifocal	61
4.1.1. Análisis geométrico de dos vistas	66
4.1.2. Propiedades de la Matriz Fundamental	69
4.1.3. Análisis algebraico de dos vistas	70
4.1.4. Restricción bifocal práctica	73
4.2. Análisis Trifocal	74
4.2.1. Análisis algebraico de la geometría trifocal	75
4.2.2. Deducción alternativa de los tensores trifocales	79
4.2.3. Interpretación geométrica de las trilinearidades	81
4.2.4. Propiedades de las trilinearidades	83
4.2.5. Relación entre la geometría bifocal y trifocal	84
4.3. Correspondencia en cuatro vistas	85
4.4. Ejercicios	88

5. Reconstrucción 3D	91
5.1. Método de reconstrucción lineal para dos vistas	92
5.2. Reconstrucción 3D para dos o más vistas	95
6. Matching en dos Vistas	97
6.1. Introducción	97
6.2. Matching	97
6.2.1. Correspondencia entre puntos	98
6.2.2. Correspondencia entre líneas	100
6.2.3. Correspondencia entre regiones	101
6.2.4. Características geométricas	102
6.2.5. Características de color	106
6.2.6. Criterios de correspondencia entre regiones	111
7. Tracking en una Secuencia	115
7.1. Definiciones	116
7.2. Matching en dos imágenes	118
7.3. Seguimiento en más imágenes	120
7.3.1. Seguimiento en tres imágenes	121
7.3.2. Seguimiento en cuatro imágenes	124
7.3.3. Resumen de trayectorias	126
7.4. Verificación	126
Indice de Figuras	133
Bibliografía	138

Capítulo 1

Introducción

1.1. ¿Qué es la Visión por Computador?

El término Visión por Computador ha sido muy utilizado en los últimos años y tiende a ser confundido con otros conceptos. Con el fin de aclarar éste y otros términos se presenta en esta sección algunas definiciones.

Procesamiento Digital de Imágenes: Proceso mediante el cual se toma una imagen y se produce una versión modificada de esta imagen.

En las Figuras 1.1 y 1.2 se ilustran dos ejemplos. En el primero se observa la *segmentación* en la que se trata de identificar en una imagen cuáles son

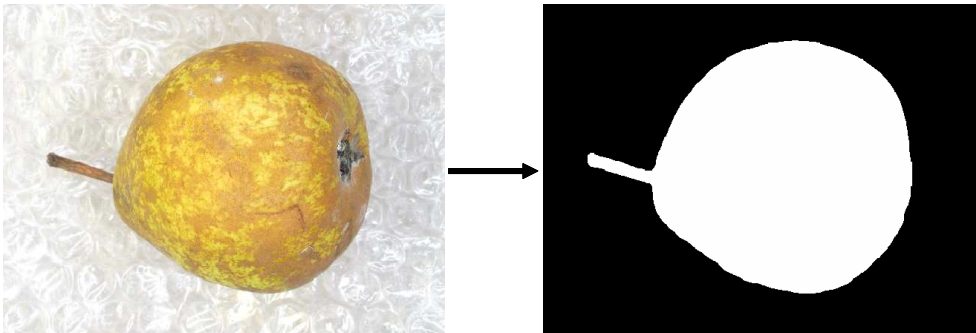


Figura 1.1: Procesamiento de imágenes– segmentación: Se pretende separar el objeto de estudio del fondo de la imagen [40].

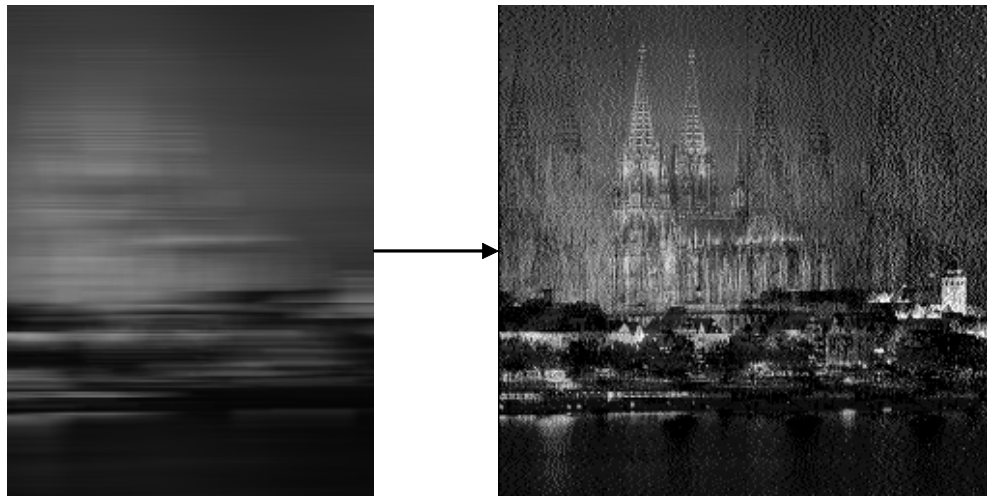


Figura 1.2: Procesamiento de imágenes– restauración: Se pretende eliminar el movimiento de la cámara al tomar la foto [36].

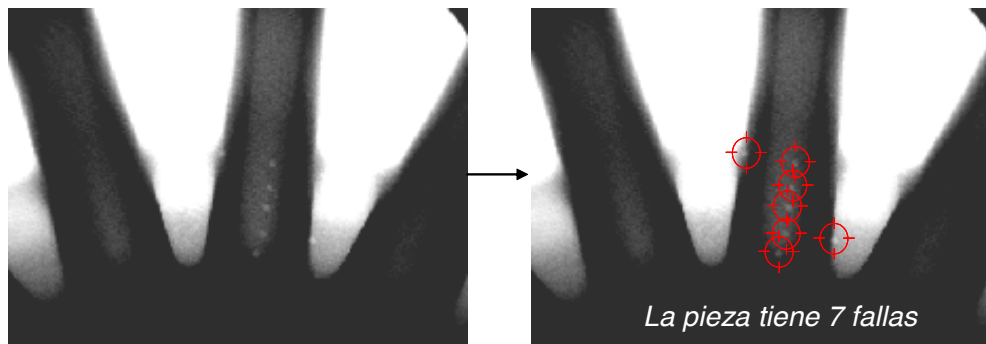


Figura 1.3: En el análisis de imágenes se pretende obtener cierta información a partir de una imagen. A través del análisis de la imagen radioscópica de la pieza mostrada se determina que la pieza no pasa el control de calidad por encontrarse defectuosa.

los píxeles que perteneces a un objeto. En este caso la salida es una imagen binaria constituidas por píxeles blancos o negros que significan ‘objeto’ o ‘no-objeto’. El segundo ejemplo es el de *restauración de imágenes*, en este caso una imagen borrosa se hace más nítida.

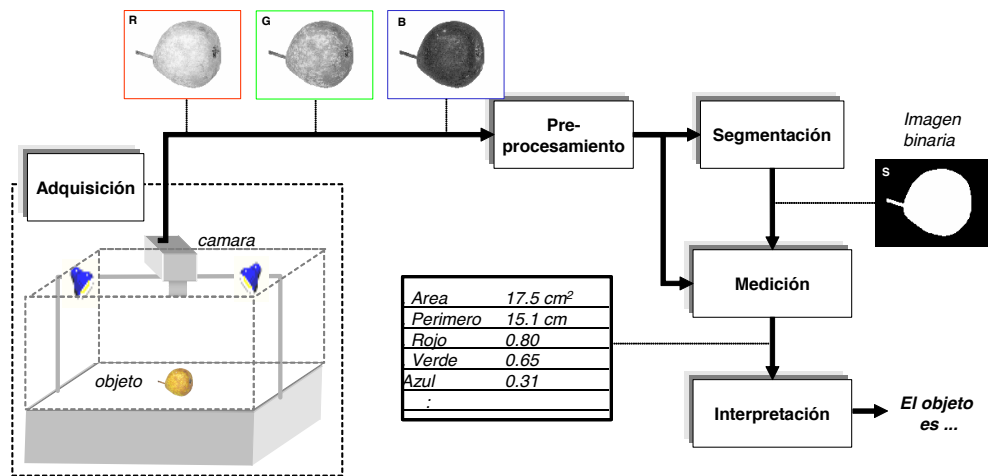


Figura 1.4: Esquema de un proceso de análisis de imágenes: adquisición de la imagen, preprocesamiento, segmentación, medición (o extracción de características), interpretación o clasificación [40].

Análisis de Imágenes: Proceso mediante el cual a partir de una imagen se obtiene una medición, interpretación o decisión.

Ejemplos de análisis de imágenes abundan en los procesos de control de inspección visual automática [41], en los que a partir de una imagen se pretende saber si un producto tiene fallas, tal como se aprecia en la Figura 1.3. Otro ejemplo es en el área de la industria de alimentos, en donde se analizan imágenes a color con el fin de detectar grados de calidad o anomalías en los alimentos. El esquema típico del análisis de imágenes se ilustra en la Figura 1.4 para el análisis de un alimento. El análisis consiste en cinco etapas [15]:

- **Adquisición de la imagen:** se obtiene la imagen adecuada del objeto de estudio. Dependiendo de la aplicación la imagen puede ser una fotografía, radiografía, termografía, etc.
- **Preprocesamiento:** con el fin de mejorar la calidad de la imagen obtenida se emplean ciertos filtros digitales que eliminan el ruido en la imagen o bien aumentan el contraste.
- **Segmentación:** como ya se explicó anteriormente se identifica el objeto de estudio en la imagen.

- **Medición** (extracción de características): se realiza una medición objetiva de ciertos atributos de interés del objeto de estudio.
- **Interpretación** (clasificación): de acuerdo a los valores obtenidos en las mediciones se lleva a cabo una interpretación del objeto.

El modelo de la Figura 1.4 puede usarse por ejemplo para la determinación del grado de maduración de la fruta. En este caso se mide el color y a partir de esta información se determina si la fruta cumple con las normas para salir al mercado.

Reconocimiento de Patrones: Asignación de objetos (patrones) a diferentes clases a partir de mediciones de los objetos.

El análisis de imágenes está estrechamente relacionado con el reconocimiento de patrones ya que en muchas aplicaciones el universo de interpretaciones es un conjunto discreto determinado por clases. En el ejemplo del análisis de la fruta mostrado en la Figura 1.4 podrían existir tres clases posibles: ‘poco madura’, ‘madura’ y ‘demasiado madura’. Mediante el estudio del color es posible clasificar la fruta en una de estas tres clases. El diseño del clasificador se determina en un estudio de casos conocidos [7]. Para casos sencillos basta con determinar correctamente ciertos umbrales de decisión. Para otros casos es necesario emplear clasificaciones más sofisticadas como las redes neuronales. En este estudio es necesario también determinar cuáles atributos son los que se van a medir, es decir, es necesario conocer en cuáles atributos está la información relevante para poder realizar una clasificación adecuada.

Es necesario resaltar que reconocimiento de patrones no implica necesariamente análisis de imágenes. Existen varias aplicaciones en las que los atributos no se obtienen de imágenes. A manera de ejemplo se muestra en Figura 1.5 un reconocimiento de patrones para detectar situaciones anómalas en deportistas. Si se observan los cinco pasos del análisis de imágenes se llega a la conclusión que sólo los últimos dos tienen relación con el reconocimiento de patrones.

Computación Gráfica: Generación computacional de imágenes a partir de modelos.

En la computación gráfica se generan imágenes artificiales a partir de modelos geométricos y físicos. La computación gráfica ha tenido un gran auge con

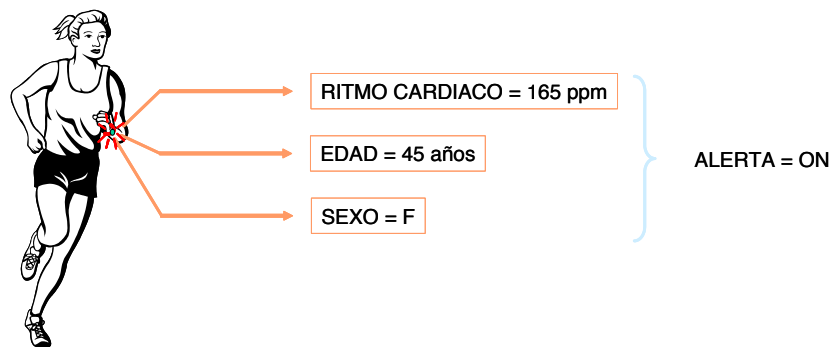


Figura 1.5: Reconocimiento de patrones: a partir de mediciones se determina una clase específica. En este ejemplo se ha detectado una situación de alerta y es recomendable que la deportista baje su ritmo cardiaco.

el desarrollo de la industria cinematográfica y la industria de video-juegos. Generalmente se construye la superficie de objetos tridimensionales en base a mallas poligonales (por ejemplo triángulos) [13]. Cada uno de estos polígonos es tratado en forma individual y dependiendo de su posición en el espacio 3D y de su textura es posible simular la proyección de una vista bidimensional. Un ejemplo de mallas poligonales se ilustra en la Figura 1.6.

En esta introducción se han presentado las definiciones de Procesamiento de Imágenes, Análisis de Imágenes, Reconocimiento de Patrones y Computación Gráfica... pero ¿qué es realmente Visión por Computador?. Para aclarar esta pregunta se presenta a continuación una serie de definiciones dadas por reconocidos investigadores en el área:

- Ciencia que desarrolla las bases teóricas y algorítmicas para obtener información sobre el mundo real a partir de una o varias imágenes (Haralick, 1992) [18].
- Disciplina que desarrolla sistemas capaces de interpretar el contenido de escenas naturales (Castleman, 1996) [4].
- La Visión por Computador, que ha emergido como una disciplina propia basada principalmente en las matemáticas y ciencias de la computación, consiste en hacer que un computador vea. Esto, sin embargo, es todavía un problema no resuelto... (Faugeras, 2001) [23].

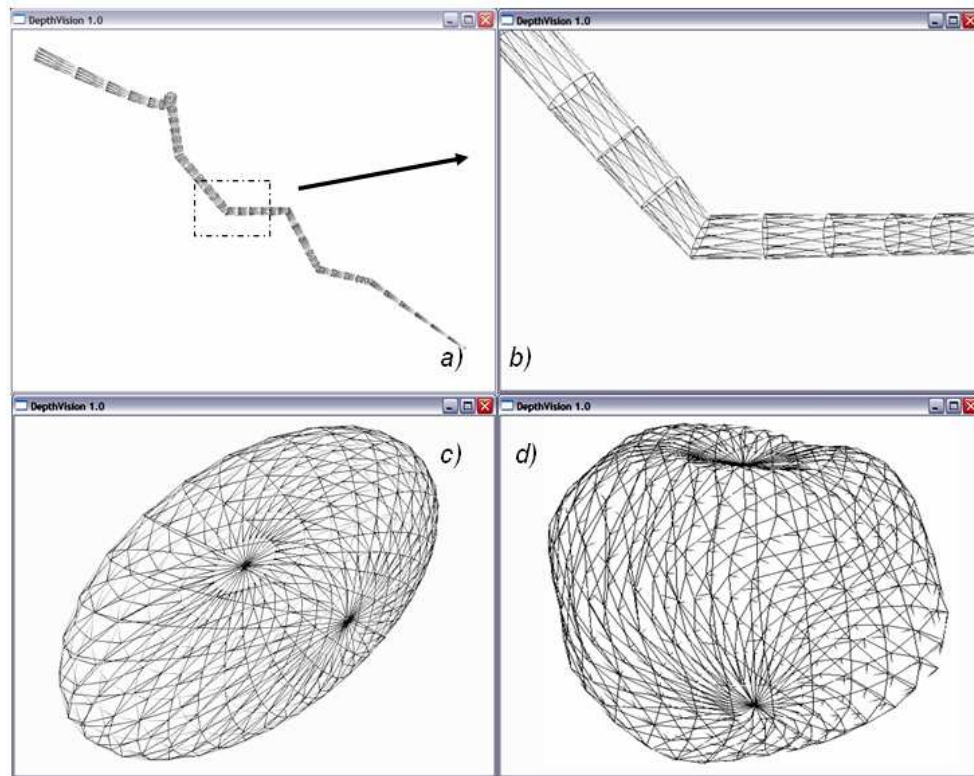


Figura 1.6: Computación Gráfica: cuatro ejemplos de construcción de superficies tridimensionales a partir de mallas poligonales [39].

Se observa que estas definiciones son muy amplias y que abarcan muchos temas. Incluso algunas definiciones se pueden confundir con la ciencia ficción. En los últimos años sin embargo se han dejado de lado términos rimbombantes como Visión Artificial y se ha empezado a utilizar con más fuerza Visión por Computador. Los términos en inglés *machine vision*, *computational vision*, *image understanding*, *robot vision*, *scenes analysis* y hasta *image analysis* están siendo desplazados por *computer vision* [6].

Para ahondar aún más en lo que es en la práctica Visión por Computador, se presenta en la Figura 1.7 una enumeración de los tópicos tratados en la *European Conference on Computer Vision (ECCV04)*, que tuvo lugar recientemente en Praga. Se observa la amplia gama de temas tratados.

- Illumination and Reflectance Modelling
- Image-based Modelling
- Image Formation, Acquisition Devices and Sensors
- Matching, Correspondence and Flow
- Low-level Vision, Image Features and Detectors
- Object Recognition, Retrieval and Indexing
- Segmentation and Grouping
- Visual Learning and Cognitive Vision
- Texture
- Active and Robot Vision
- Tracking and Motion
- Biologically-inspired Vision
- Video Analysis and Event Recognition
- Vision Systems Engineering & Evaluation
- Multiple View Geometry
- Industrial and Medical Applications
- 3D Reconstruction and Shape (from X)

Figura 1.7: Tópicos de Visión por Computador tratados en la última conferencia europea ECCV04 realizada en Praga.

Los temas tratados en este libro sin embargo no abarcan todos los tópicos mostrados en la Figura 1.7, ya que esto sería un proyecto demasiado ambicioso para un solo curso. En este libro no se verán los aspectos que están más relacionados con el procesamiento de imágenes, análisis de imágenes y reconocimiento de patrones, ya que estos temas han sido ampliamente cubiertos por otros libros y cursos. En este libro nos vamos a limitar a estudiar aquello que las mencionadas teorías no incorporan necesariamente en su análisis: la información del mundo tridimensional existente en las imágenes bidimensionales.

En este libro se concibe entonces la Visión por Computador como una gran herramienta para establecer la relación entre el mundo tridimensional y sus vistas bidimensionales tomadas de él. Por medio de esta teoría se puede hacer,

por una parte, una reconstrucción del espacio tridimensional a partir de sus vistas y, por otra parte, llevar a cabo una simulación de una proyección de una escena tridimensional en la posición deseada a un plano bidimensional.

El material de este libro ha sido desarrollado como libro de texto para un curso de un semestre de Ciencias de la Computación a nivel de postgrado.

1.2. Ejemplos

En esta sección se muestran algunos ejemplos en los que se puede apreciar el campo de aplicaciones de la Visión por Computador.

■ Fotogrametría

En la fotogrametría se persigue realizar mediciones del espacio 3D a partir de fotografías tomadas de él. De esta manera es posible medir superficies, construcciones, objetos, etc. Asimismo se puede llevar a cabo una topología de un terreno.

■ Rectificación Métrica

Mediante esta técnica es posible hacer correcciones de perspectiva (ver Figuras 1.8 y 2.4) y correcciones de distorsión de lente (ver Figura 1.9).

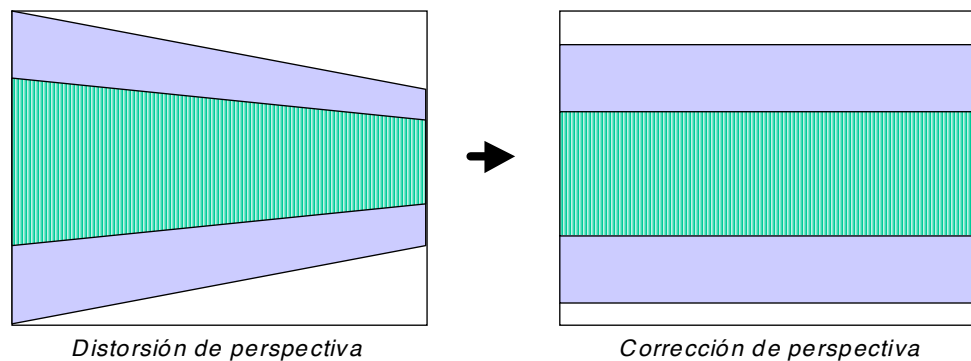


Figura 1.8: Ejemplo de rectificación de perspectiva.

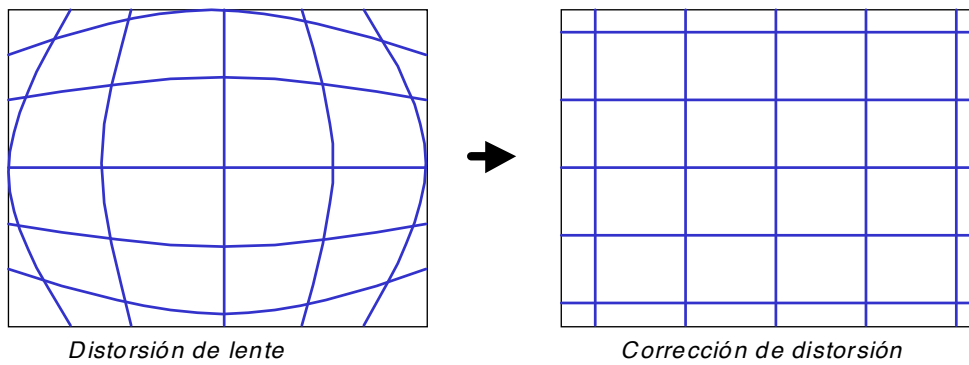


Figura 1.9: Ejemplo de rectificación de distorsión de lente.

■ Reconstrucción 3D

A partir de las vistas, mediante la técnica de triangulación, es posible obtener un modelo 3D del objeto proyectado en las vistas. El principio

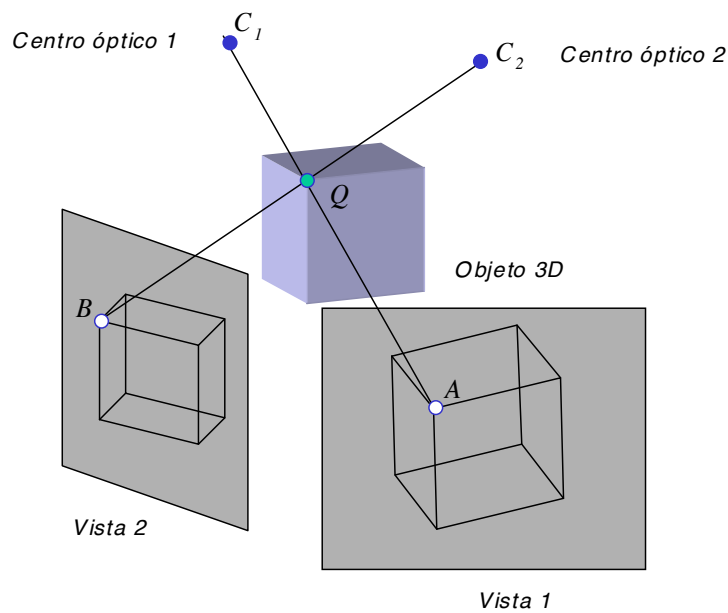


Figura 1.10: Triangulación: estimación de Q a partir de A y B .

de triangulación es mostrado en la Figura 1.10: sabiendo que los puntos A y B son proyecciones de un mismo punto tridimensional Q , es decir A y B son correspondientes, y conociendo los centros ópticos de la proyección C_1 y C_2 , se puede encontrar el punto Q a partir de la intersección entre las dos rectas $\langle C_1, A \rangle$ y $\langle C_2, B \rangle$.

■ Matching y Tracking

Por medio del Matching y Tracking es posible encontrar la correspondencia entre puntos de varias imágenes. Los puntos correspondientes son aquellos que representan una proyección del mismo punto físico en el espacio 3D. En la Figura 1.11 se puede apreciar tres vistas de una taza tomadas por una cámara fija mediante la rotación del eje central de la taza. Se puede observar que los puntos m_1 , m_2 y m_3 en las imágenes 1, 2 y 3 respectivamente, son correspondientes entre sí porque son proyecciones del mismo punto m de la taza. Mediante la teoría de Visión por Computador podemos responder las siguientes preguntas: *i)* Conociendo el punto m_1 en la imagen 1, ¿dónde está su punto correspondiente en las imágenes 2 y 3? *ii)* Conociendo los puntos m_1 y m_2 y sabiendo que son correspondientes, ¿dónde se encuentra el punto correspondiente en la tercera imagen? Estas preguntas serán respondidas a lo largo de este curso.

■ Computación Gráfica

Si se tiene un modelo de la formación de la imagen $f : 3D \rightarrow 2D$, es posible entonces simular gráficamente las vistas bidimensionales que se obtendrían de un objeto tridimensional. Las aplicaciones de realidad virtual emplean esta teoría.

■ Estimación de Movimiento

Mediante una cámara que toma imágenes de un objeto en movimiento es posible estimar el movimiento del objeto a partir de los puntos de correspondencia en la secuencia de imágenes.

1.3. Historia

Una cámara produce imágenes planas de un mundo físico percibido como tridimensional. Antes de la invención de la fotografía existía un gran interés

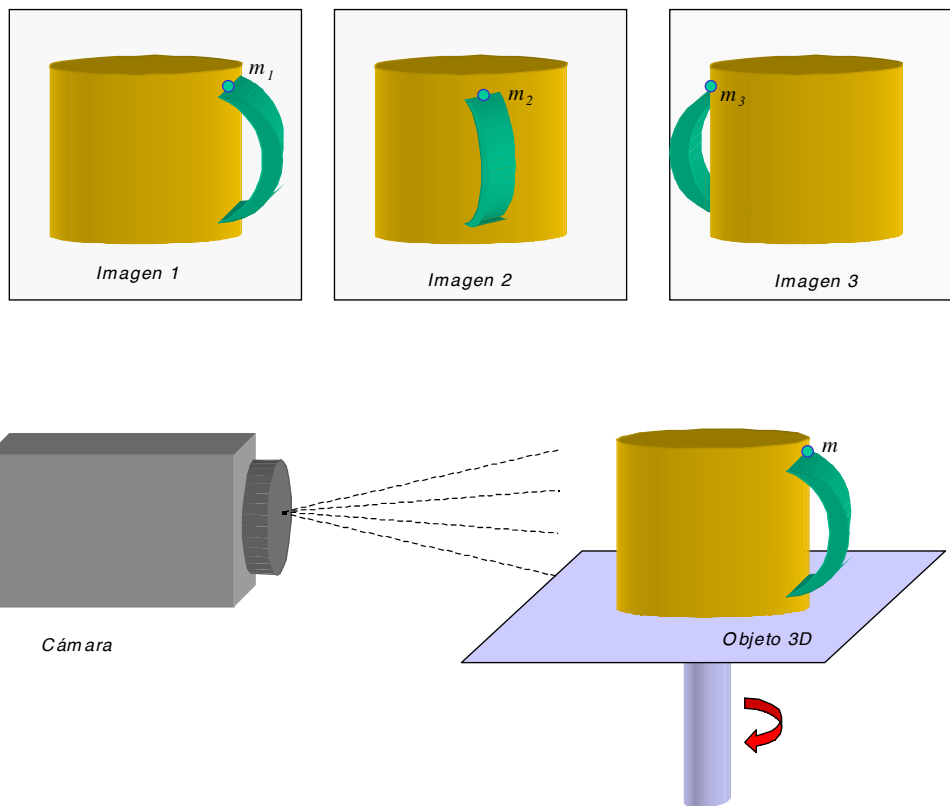


Figura 1.11: Correspondencia en tres puntos.

en representar este mundo 3D en imágenes planas 2D, como es el caso de la pintura¹.

Los griegos llegaron a conocer muchas de las propiedades geométricas de la proyección. Como es el caso de Tales de Mileto (640 AC - 548? AC) que con sus conocimientos de la Geometría pudo predecir un eclipse solar y también pudo medir la altura de una pirámide a partir de su sombra proyectada. Sin embargo, los griegos pensaban que la visión era activa, es decir que los ojos emitían partículas al mundo 3D en vez de considerar a los ojos como dispositivos pasivos receptores de luz.

Cabe mencionar dentro de los matemáticos griegos a Euclides, quien en

¹Gran parte de esta sección fue obtenida de la sección 4.1 de [9].

el siglo IV AC ideó la geometría plana. Para Euclides la geometría era concebida como un conjunto de líneas y puntos, independientes de un sistema de coordenadas.

Posteriormente, los pintores italianos del Renacimiento fueron los primeros en entender la formación de las imágenes y fueron los primeros en estudiar la Geometría para reproducir correctamente los efectos de la perspectiva en las imágenes del mundo que observaban. La pintura anterior a esta época era *plana*, es decir no mostraba la diferencia de profundidad en los objetos representados, como se muestra en la Figura 1.12.

La perspectiva fue inventada por Filippo Brunelleschi (1377-1446) alrededor de 1413. Brunelleschi fue un gran arquitecto del Renacimiento Temprano. Fue también escultor y pintor. Sus principales obras se encuentran en Florencia, como por ejemplo la Catedral Santa María de Fiore, cuya cúpula es la más grande del mundo con más de 50m de diámetro.

Artistas como Piero della Francesca (1415-1492), Leonardo da Vinci (1452-1519) y Albrecht Dürer (1471-1528), los dos primeros italianos y el tercero alemán que viaja a Italia para posteriormente llevar el Renacimiento a Alemania, realizan serios estudios geométricos que se usan hasta el día de hoy. A partir de esta época se empieza a considerar el *punto de fuga*, en el que líneas paralelas que se alejan del observador convergen en un punto. A manera



Figura 1.12: Pintura pre-renacentista y renacentista. Izquierda: Jesús entrando a Jerusalén. Derecha: Iglesia del Espíritu Santo, Bruneleschi.



Figura 1.13: Máquina de perspectiva por Albrecht Dürer [8, 9].

de ejemplo, en la Figura 1.12 se muestran dos pinturas: una pre-renacentista y otra renacentista². En la primera se puede observar la *polidimensionalidad*, en la cual los puntos de vista de los objetos representados no son únicos. Asimismo, el tamaño de los objetos está relacionado más con la importancia dentro de la obra que con la ubicación espacial. En la segunda pintura se aprecia claramente la profundidad producida por las líneas que convergen en un punto de fuga. De esta manera se le hace creer al observador que está frente a una escena tridimensional.

En el siglo XVI se desarrolla la teoría de la perspectiva. Se introducen las *Máquinas de Perspectiva* para ayudar a los pintores a reproducir exactamente la perspectiva sin tener que hacer cálculos matemáticos. Una de estas máquinas es representada en la Figura 1.13 por Albrecht Dürer. En esta figura, el ojo del dibujante es mantenido fijo y un dispositivo es utilizado para materializar la intersección de cada rayo visual con el plano de la imagen.

Las máquinas de la perspectiva pueden ser consideradas como el primer intento de una cámara. Ellas utilizan un plano R (plano de la imagen, ver rejilla en Figura 1.13) donde se forma la imagen y un punto C (centro óptico, ver ojo del dibujante en Figura 1.13) que no pertenece a R en el que se intersectan todos los rayos que forman la imagen.

En el año 1545 el astrónomo Germina Frisius publica un estudio donde

²La primera pintura fue tomada de Dutch, Steven: *Perspective in Art - Slides* (<http://weba.uwgb.edu/dutchs/190outln/perslide.htm>). La segunda pintura fue tomada de Dauben, Joseph W.: *The Art of Renaissance Science* (<http://www.crs4.it/Ars/arshtml/arstoc.html>).

presenta la *cámara oscura*. En la Figura 1.14 se representa un esquema de la cámara oscura. Mediante un orificio muy pequeño C en una pared se deja entrar la luz externa que es proyectada en una pared interior de la cámara oscura. El resultado es una imagen invertida del mundo exterior. La cámara oscura sirvió a algunos pintores como a Vermeer (1632-1675) para representar de la manera más precisa posible la realidad.

A partir de la teoría del plano cartesiano introducida por Descartes (1596-1650) se empieza a concebir la geometría desde un punto de vista algebraico. Así, las entidades geométricas son descritas como coordenadas y entidades algebraicas. Se cuenta que Descartes desarrolló esta teoría acostado en su cama al ver volar una mosca en la habitación. El movimiento de la mosca con respecto a los tres ejes conformados por la intersección de dos paredes y el techo, hizo que Descartes pensara en que con tres coordenadas (X, Y, Z) se podía definir la posición de la mosca en cualquier instante [16].

En el año 1826 el químico francés Niepce (1765-1833) llevó a cabo la primera fotografía, colocando una superficie fotosensible dentro de una cámara oscura para fijar la imagen. Posteriormente, en 1838 el químico francés Daguerre (1787-1851) hizo el primer proceso fotográfico práctico. Daguerre utilizó una placa fotográfica que era revelada con vapor de mercurio y fijada con trisulfato de sodio.

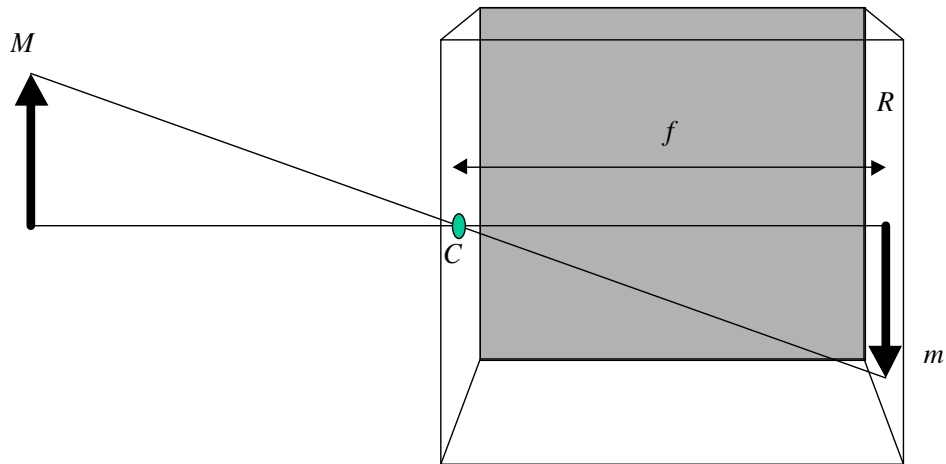


Figura 1.14: Cámara oscura.

En la actualidad se utilizan cámaras reflex y CCD que emplean lentes para incrementar la potencia de la luz y mejorar el enfoque de la imagen. Estas cámaras serán estudiadas posteriormente en el curso.

1.4. Ejercicios

Ejercicio 1.1 *Simular una máquina de perspectiva utilizando una transparencia. Colocar una transparencia entre el ojo de el/la observador/a y una escena tridimensional que contenga objetos del mismo tamaño dispuestos en distintas profundidades. Dibujar lo que se ve en la transparencia y observar el punto de fuga así como la reducción del tamaño de los objetos a medida que se van alejando de la transparencia.*

Capítulo 2

Geometría Projectiva

En este capítulo se introducen las principales ideas geométricas y la notación que se utilizará en el resto del curso. Se empezará con la geometría plana de dos dimensiones que es más fácil de visualizar y comprender para luego pasar a la geometría del espacio tridimensional¹.

Se estudiarán las *transformaciones proyectivas* de un plano. Mediante estas transformaciones es posible modelar la distorsión geométrica que le ocurre a un plano que es visto por una cámara perspectiva. Con estas transformaciones algunas propiedades se conservan como la colinearidad (líneas rectas son vistas como líneas rectas) mientras que otras propiedades no (las líneas paralelas por lo general no son vistas como líneas paralelas).

En un nivel elemental la geometría es el estudio de puntos y líneas, y sus relaciones. A lo largo de la historia la geometría ha sido concebida inicialmente como una disciplina netamente *geométrica*, en la que las líneas y puntos se estudian sin considerar un sistema de coordenadas. Posteriormente, mediante la introducción de un sistema de coordenadas cartesiano se logra *algebraizar* a la geometría. De esta manera, las entidades geométricas pueden ser descritas como coordenadas y entidades algebraicas. Por medio de las relaciones algebraicas se obtiene una representación matemática apropiada para implementar algoritmos y programar métodos computacionales. A lo largo del curso se utilizará tanto la concepción geométrica como la algebraica. En algunos casos la geometría logra visualizar mejor un problema dado, en

¹La teoría de este capítulo puede ser encontrada en los capítulos 1 y 2 de [23], en el capítulo 2 de [9] así como en el capítulo 2 de [8].

otros el álgebra puede representarlo y resolverlo más fácilmente.

2.1. Planos, puntos y líneas rectas

Un punto en un plano se representa como un par de coordenadas (x, y) en \mathcal{R}^2 . Comúnmente, \mathcal{R}^2 se identifica con un plano. Se puede considerar entonces \mathcal{R}^2 como un espacio vectorial en el que (x, y) es un vector. Se asocia así, un punto a un vector. En este curso los vectores serán representados en letra negrita y sus elementos serán dispuestos de manera vertical: $\mathbf{m} = (x, y)^\top$.

Una línea recta en el plano (x, y) está representada por la ecuación:

$$ax + by + c = 0. \quad (2.1)$$

De esta manera, una línea recta puede ser representada por un vector $\mathbf{l} = (a, b, c)^\top$.

La correspondencia entre líneas rectas y vectores no es uno a uno, ya que $(a, b, c)^\top$ y $(ka, kb, kc)^\top$ representan exactamente la misma línea recta para $k \neq 0$, sin embargo son vectores distintos. Estos vectores son considerados como equivalentes y se definen como *vectores homogéneos*.

Un punto (x, y) está en la recta $\mathbf{l} = (a, b, c)^\top$ si y sólo si la ecuación (2.1) es verdadera. Esta ecuación puede escribirse utilizando el producto punto entre los vectores $\mathbf{l} = [a \ b \ c]^\top$ y $\mathbf{m} = [x \ y \ 1]^\top$ como:

$$\mathbf{l} \bullet \mathbf{m} = \mathbf{l}^\top \mathbf{m} = [a \ b \ c][x \ y \ 1]^\top = ax + by + c = 0. \quad (2.2)$$

Así, el punto (x, y) es representado por un vector $(x, y, 1)^\top$. Sin embargo, los vectores (kx, ky, k) , para $k \neq 0$, pueden considerarse también como representaciones del mismo punto (x, y) ya que satisfacen la ecuación $[a \ b \ c]^\top [kx \ ky \ k] = k(ax + by + c) = 0$. Se dice entonces, que los puntos en un plano bidimensional pueden ser representados por vectores homogéneos de tres dimensiones cuyos dos primeros elementos son las coordenadas del punto en el plano y el tercer elemento es 1.

En términos generales, si se tiene un vector homogéneo de tres dimensiones dado por $(x_1, x_2, x_3)^\top$ que representa un punto en un plano, las coordenadas de este punto en el plano están definidas como $(x, y) = (x_1/x_3, x_2/x_3)$.

Ejemplo 2.1 *Intersección de dos líneas rectas:*

Dadas dos rectas $\mathbf{l} = (a, b, c)^T$ y $\mathbf{l}' = (a', b', c')^T$ se desea encontrar el punto de intersección de ambas rectas.

Del álgebra vectorial se sabe que el producto cruz de dos vectores produce un vector que es perpendicular a ellos. También se sabe que el producto punto entre dos vectores perpendiculares es igual a cero. Por lo tanto se puede afirmar que $\mathbf{l} \bullet (\mathbf{l} \times \mathbf{l}') = \mathbf{l}' \bullet (\mathbf{l} \times \mathbf{l}') = 0$. Si se define el punto \mathbf{m} como $\mathbf{l} \times \mathbf{l}'$ obtenemos la representación homogénea que satisface las ecuaciones $\mathbf{l}^T \mathbf{m} = 0$ y $\mathbf{l}'^T \mathbf{m} = 0$, lo que quiere decir que el punto \mathbf{m} es la intersección de las rectas ya que pertenece a ellas:

$$\mathbf{m} = \mathbf{l} \times \mathbf{l}' \quad (2.3)$$

Ejemplo 2.2 *Línea recta que pasa por dos puntos:*

Dados dos puntos $\mathbf{m} = (x, y, 1)^T$ y $\mathbf{m}' = (x', y', 1)^T$ se desea encontrar la línea recta que pasa por ambos puntos.

Mediante el mismo razonamiento del ejemplo anterior se puede afirmar que la recta está definida por:

$$\mathbf{l} = \mathbf{m} \times \mathbf{m}' \quad (2.4)$$

ya que $\mathbf{l}^T \mathbf{m} = 0$ y $\mathbf{l}^T \mathbf{m}' = 0$, lo cual quiere decir que tanto \mathbf{m} como \mathbf{m}' pertenecen a la misma recta \mathbf{l} .

Ejemplo 2.3 *Intersección de dos rectas paralelas:*

Dadas dos rectas paralelas $\mathbf{l} = (a, b, c)^T$ y $\mathbf{l}' = (a, b, c')^T$ se desea encontrar el punto de intersección.

Según (2.3) la representación homogénea del punto de intersección está dada por $\mathbf{m} = \mathbf{l} \times \mathbf{l}' = (c' - c)[b \ -a \ 0]^T$. La representación no homogénea de este punto presenta una singularidad debido a la división por cero. El punto obtenido de la división por cero puede ser interpretado como un punto en el infinito. Los puntos que tienen una representación homogénea de este tipo, $\mathbf{m} = [x_1 \ x_2 \ x_3]^T$ con $x_3 = 0$, son conocidos como *puntos ideales*. Todos los puntos ideales pertenecen a la línea en el infinito $\mathbf{l}_\infty = [0 \ 0 \ 1]^T$, ya que $[x_1 \ x_2 \ 0]^T [0 \ 0 \ 1] = 0$. De esta manera una recta $\mathbf{l} = (a, b, c)^T$ intersecta a \mathbf{l}_∞ en $\mathbf{m} = \mathbf{l} \times \mathbf{l}_\infty = [b \ -a \ 0]^T$. Una línea \mathbf{l}' paralela a \mathbf{l} intersecta a \mathbf{l}_∞ en el mismo punto. Es interesante observar que en la Geometría Proyectiva dos

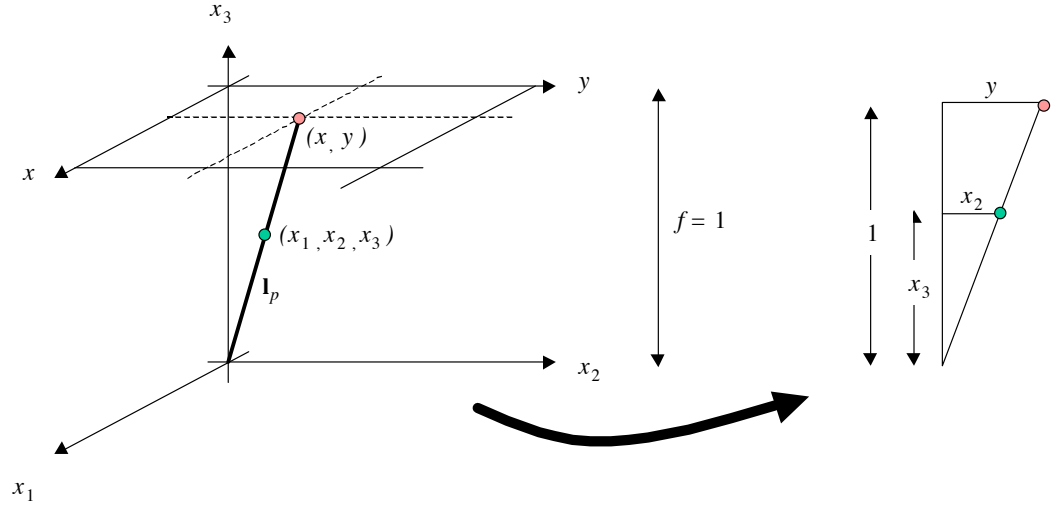


Figura 2.1: Proyección $(x_1, x_2, x_3) \rightarrow (x, y)$.

líneas rectas paralelas se intersectan en un punto (ideal), mientras que en la Geometría Euclídea la intersección de dos rectas paralelas no está definida.

Como ya se ha visto anteriormente, un punto (x, y) en un plano tiene una representación homogénea en \mathcal{R}^3 dada por $\mathbf{m} = [x_1 \ x_2 \ x_3]^T$ con $x = x_1/x_3$ y $y = x_2/x_3$. El plano proyectivo \mathcal{P}^2 en el que están representados todos los puntos no homogéneos $x = x_1/x_3$ y $y = x_2/x_3$ puede ser interpretado utilizando la representación de la Figura 2.1. En este esquema la proyección de (x_1, x_2, x_3) en un plano (x, y) paralelo al plano $(x_1, x_2, 0)$ ubicado en $x_3 = 1$, está dada por el punto (x, y) el cual puede ser calculado aplicando el Teorema de Thales como $x : x_1 = y : x_2 = 1 : x_3$. Se obtiene entonces $x = x_1/x_3$ y $y = x_2/x_3$. Se puede observar que cualquier punto 3D sobre la línea de proyección (ver \mathbf{l}_p en Figura 2.1) produce el mismo punto proyectado en el plano (x, y) . Esto es lo mismo que decir que la proyección de $k(x_1, x_2, x_3)$ es igual a la proyección de (x_1, x_2, x_3) , para $k \neq 0$, y que esta proyección está dada por $x = x_1/x_3$ y $y = x_2/x_3$. En otras palabras, la clase de vectores homogéneos $k(x_1, x_2, x_3)$ representan el mismo punto en \mathcal{P}^2 .

2.2. Transformaciones Projectivas 2D

La geometría projectiva 2D es el estudio de las propiedades del plano projectivo \mathcal{P}^2 que son invariantes bajo un grupo de transformaciones conocidas como *projectividades*.

Una projectividad es una transformación invertible dada por $h : \mathcal{P}^2 \rightarrow \mathcal{P}^2$ de manera tal que una línea recta es transformada como una línea recta. La projectividad está definida como:

$$h(\mathbf{m}) = \mathbf{m}' = \mathbf{H}\mathbf{m}, \quad (2.5)$$

donde \mathbf{H} es una matriz 3×3 no singular. Se dice entonces que \mathbf{m}' es la transformación lineal \mathbf{H} de \mathbf{m} . Esta transformación es biunívoca entre dos planos 2D, cuyos puntos son representados homogéneamente por \mathbf{m} y \mathbf{m}' . Es decir, un punto en un plano 2D tiene una única correspondencia en un punto de otro plano 2D, y cada punto en un plano tiene un solo punto correspondiente en el otro plano.

La condición invariante (una recta es transformada en una recta) puede comprobarse de la siguiente manera: si \mathbf{m}_1 , \mathbf{m}_2 y \mathbf{m}_3 están en la misma línea recta \mathbf{l} , entonces $\mathbf{l}^T \mathbf{m}_i = 0$ para $i = 1, 2, 3$. Esta ecuación puede ser escrita como $\mathbf{l}^T \mathbf{H}^{-1} \mathbf{H} \mathbf{m}_i = 0$, ya que $\mathbf{H}^{-1} \mathbf{H} = \mathbf{I}$. Como los tres puntos transformados por h quedan definidos como $\mathbf{m}'_i = \mathbf{H} \mathbf{m}_i$, se puede escribir entonces $\mathbf{l}'^T \mathbf{m}'_i = 0$ con $\mathbf{l}'^T = \mathbf{l}^T \mathbf{H}^{-1}$. Esto quiere decir que los puntos \mathbf{m}'_i pertenecen a una recta \mathbf{l}' .

La ecuación (2.5) puede escribirse de manera explícita como

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \quad (2.6)$$

Se observa que los resultados en la transformación de coordenadas no son afectados si se cambia \mathbf{H} por $k\mathbf{H}$, para $k \neq 0$, lo cual quiere decir que \mathbf{H} es una matriz homogénea.

Ejemplo 2.4 Calcular la matriz projectiva entre los planos paralelos mostrados en la Figura 2.2:

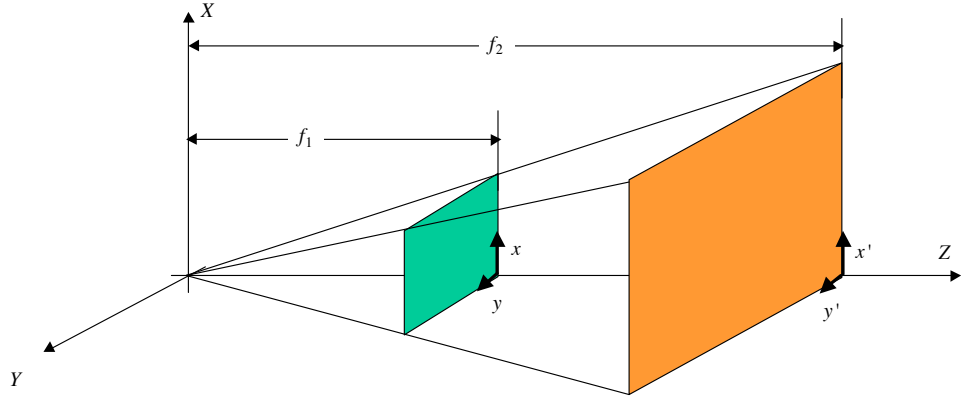


Figura 2.2: Proyección en dos planos paralelos.

Utilizando el Teorema de Thales, se tienen las siguientes relaciones:

$$\frac{x'}{f_2} = \frac{x}{f_1} \quad ; \quad \frac{y'}{f_2} = \frac{y}{f_1}$$

Estas ecuaciones se pueden escribir usando la forma matricial $\mathbf{m}' = \mathbf{H}\mathbf{m}$, con $\mathbf{m} = [x \ y \ 1]^T$, $\mathbf{m}' = [x' \ y' \ 1]^T$ y

$$\mathbf{H} = \begin{bmatrix} f_2/f_1 & 0 & 0 \\ 0 & f_2/f_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Para el caso de dos planos no paralelos como se muestra en la Figura 2.3 se pueden establecer dos características de la transformación proyectiva: *i)* hay una correspondencia biunívoca entre los puntos pertenecientes a ambos planos y *ii)* una línea recta en un plano corresponde a una línea recta en el otro plano. De esta manera se puede afirmar que la relación proyectiva entre ambos planos esta dada por la ecuación general (2.6). El próximo ejemplo ilustra una aplicación de esta transformación.

Ejemplo 2.5 *Rectificación de distorsión proyectiva:*

En este ejemplo se busca la matriz \mathbf{H} que logre transformar una imagen que presente una distorsión proyectiva, en una imagen libre de distorsión.

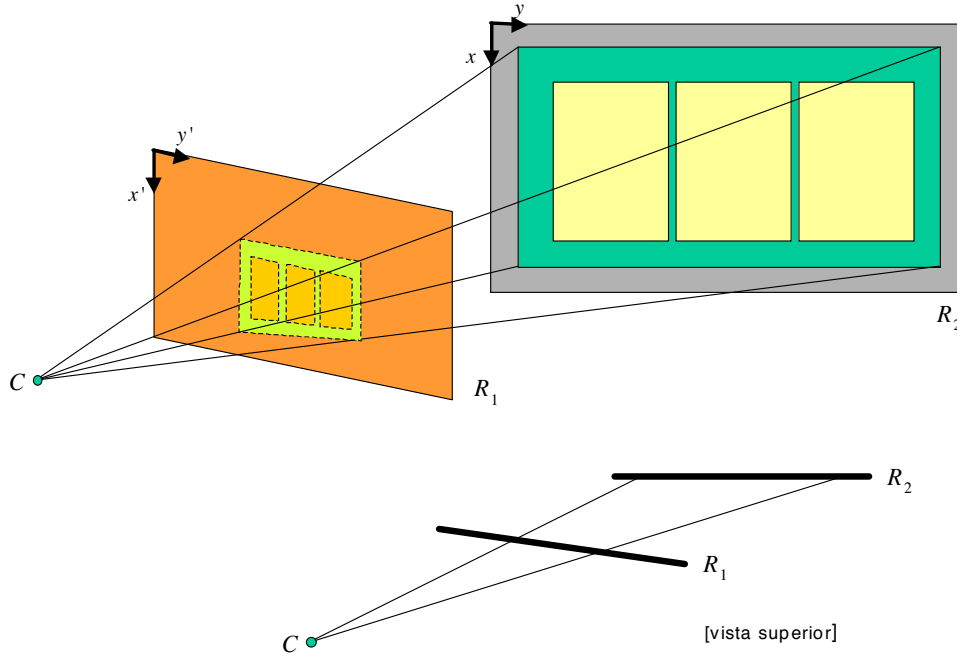


Figura 2.3: Proyección en dos planos no paralelos.

La operación de rectificación de distorsión se muestra en la Figura 2.3. Los datos de entrada en este problema son las coordenadas de n puntos (x'_i, y'_i) , $i = 1, \dots, n$, en la imagen original y las coordenadas deseadas de estos puntos (x_i, y_i) en la imagen rectificada. Por lo general se escogen puntos (x_i, y_i) que pertenezcan a un rectángulo. A partir de (2.6) se obtiene para cada punto (x', y') y su correspondiente (x, y) las ecuaciones:

$$x' = \frac{x'_1}{x'_3} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}$$

$$y' = \frac{x'_2}{x'_3} = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}$$

En el problema interesa encontrar los 9 elementos de \mathbf{H} . Sin embargo, como la matriz \mathbf{H} es homogénea, $k\mathbf{H}$ también sería solución al problema. De esta manera es posible dividir cada elemento de \mathbf{H} por h_{33} para obtener una

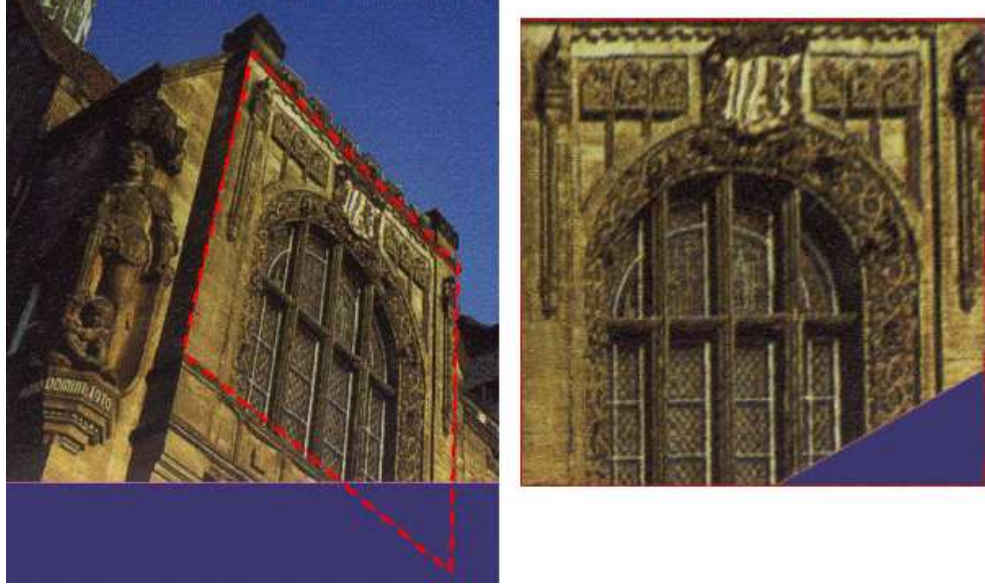


Figura 2.4: Rectificación de distorsión proyectiva.

matriz \mathbf{H} con sólo 8 elementos desconocidos, ya que el último elemento sería igual a uno.

Entonces, con $h_{33} = 1$ las dos últimas ecuaciones pueden ser escritas de manera matricial como:

$$\begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -x'x & -x'y \\ 0 & 0 & 0 & x & y & 1 & -y'x & -y'y \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix} \quad (2.7)$$

o bien $\mathbf{A}\mathbf{h} = \mathbf{b}$. Se observa que para cada correspondencia de puntos se obtienen dos ecuaciones. Suponiendo n pares de puntos correspondientes se

puede establecer el siguiente sistema de $2n$ ecuaciones y 8 incógnitas.

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_n \end{bmatrix} \mathbf{h} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}$$

donde \mathbf{A}_i y \mathbf{b}_i son la matriz \mathbf{A} y el vector \mathbf{b} obtenidas en (2.7) para el punto i . El sistema de ecuaciones anterior puede ser expresado como $\tilde{\mathbf{A}}\mathbf{h} = \mathbf{b}$. Para $n = 4$ existe una solución directa dada por $\mathbf{h} = \tilde{\mathbf{A}}^{-1}\mathbf{b}$. Si $n > 4$ sin embargo el sistema queda sobredeterminado. En este caso se utiliza el método de los mínimos cuadrados en el que se encuentra un vector \mathbf{h} tal que minimice $\|\tilde{\mathbf{A}}\mathbf{h} - \mathbf{b}\|$. La solución entonces estaría dada por $\mathbf{h} = [\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}}]^{-1} \tilde{\mathbf{A}}^\top \mathbf{b}$.

Cabe señalar que el método aquí expuesto funciona sólo si $h_{33} \neq 0$. Para el caso en que $h_{33} = 0$ se recomienda un método alternativo en el que la restricción para \mathbf{H} no es $h_{33} = 1$ sino $\|\tilde{\mathbf{h}}\| = 1$ donde $\tilde{\mathbf{h}} = [h_{11} \dots h_{33}]^\top$. Para mayores detalles ver [23].

En la Figura 2.4 se puede apreciar un ejemplo práctico en el que se rectifica la distorsión proyectiva presente en la fotografía de la fachada de una iglesia.

2.3. Categorización de las Transformaciones Projectivas 2D

A continuación se presentan 4 categorías existentes en las transformaciones proyectivas $\mathcal{R}^2 \rightarrow \mathcal{R}^2$.

2.3.1. Transformación Isométrica (Euclídea)

En la transformación isométrica se conserva la distancia Euclídea, es decir la distancia entre dos puntos es igual a la distancia entre los puntos transformados. La transformación proyectiva isométrica es ilustrada en la Figura 2.5 y corresponde a la transformación de coordenadas $(x', y') \leftrightarrow (x, y)$. La transformación $(x', y') \rightarrow (x, y)$ es (ver Ejercicio 2.10):

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (2.8)$$

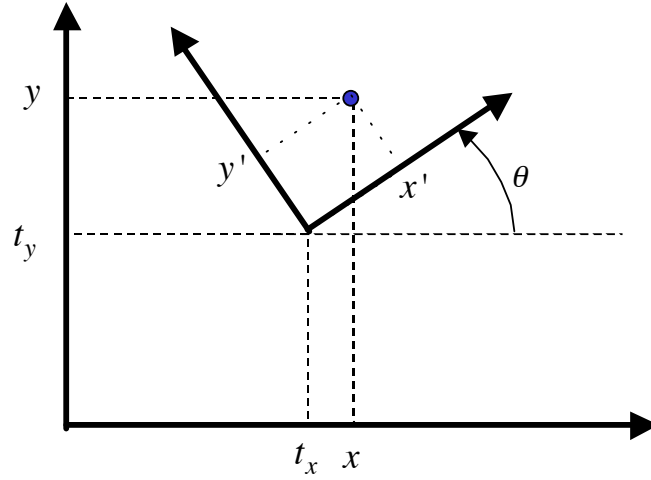


Figura 2.5: Transformación 2D isométrica (Euclídea).

donde θ es el ángulo de la rotación entre los ejes y (t_x, t_y) es el desplazamiento del origen. Esta ecuación puede ser escrita como:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \mathbf{R}' \begin{bmatrix} x' \\ y' \end{bmatrix} + \mathbf{t}' \quad (2.9)$$

o bien

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (2.10)$$

con $\mathbf{0}^\top = [0 \ 0]$. La transformación inversa $(x, y) \rightarrow (x', y')$ se obtiene de (2.9):

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = [\mathbf{R}']^{-1} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \mathbf{t}' \right). \quad (2.11)$$

Como la matriz \mathbf{R}' es ortonormal, es decir que $\mathbf{R}'[\mathbf{R}']^\top = \mathbf{I}_{2 \times 2}$ se sabe entonces que la inversa de \mathbf{R}' es su transpuesta. Definiendo $\mathbf{R} = [\mathbf{R}']^\top = [\mathbf{R}']^{-1}$ y $\mathbf{t} = -\mathbf{R}\mathbf{t}'$ se obtiene

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (2.12)$$

Otra propiedad de las matrices \mathbf{R} y \mathbf{R}' se obtiene observando que ambas son funciones del ángulo de rotación θ . Si se define $\mathbf{R} = \mathbf{R}(\theta)$ es fácil comprobar que

$$\mathbf{R}' = \mathbf{R}(-\theta). \quad (2.13)$$

Se puede observar que la matriz 3×3 de la transformación isométrica (2.12) tiene la siguiente forma

$$\mathbf{H}_E = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}. \quad (2.14)$$

Las *invariantes* de esta transformada, es decir las propiedades que se mantienen después de aplicar la transformación isométrica son: *i)* longitud entre puntos, *ii)* ángulo entre dos rectas y *iii)* área.

2.3.2. Transformación de Similitud

En la transformación de similitud se conserva la forma de los objetos. Sin embargo, en este caso la distancia entre dos puntos ya no es igual a la distancia entre los puntos transformados. La transformación proyectiva de similitud, definida por medio de la matriz \mathbf{H}_S , es:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}}_{\mathbf{H}_S} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (2.15)$$

Las variables \mathbf{t} , $\mathbf{0}^\top$ y \mathbf{R} están definidas en la Sección 2.3.1. A través del parámetro s se obtiene la ampliación ($s > 1$) o reducción de los objetos ($0 < s < 1$).

Las invariantes de esta transformada son: *i)* ángulos entre rectas, *ii)* líneas paralelas, *iii)* razón entre dos distancias y *iv)* razón entre dos áreas.

2.3.3. Transformación Afín

En la transformación afín se distorsiona la forma de los objetos introduciendo una matriz 2×2 \mathbf{A} no ortonormal en lugar de \mathbf{R} . La transformación afín,

definida por medio de la matriz \mathbf{H}_A , se expresa como:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}}_{\mathbf{H}_A} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.16)$$

Los vectores \mathbf{t} y $\mathbf{0}^\top$ están definidos en la Sección 2.3.1. Las invariantes de esta transformada son: *i)* líneas paralelas y *ii)* razón entre dos áreas.

2.3.4. Transformación Proyectiva General

La transformación proyectiva es la generalización de las transformaciones lineales $\mathcal{R}^2 \rightarrow \mathcal{R}^2$ en la que las líneas paralelas no son transformadas necesariamente como tales. La transformación proyectiva ya se expresó en coordenadas homogéneas en (2.5). En este caso la matriz 3×3 de la transformación se denota como \mathbf{H}_P .

La invariante de esta transformada es la razón de cruz.

Transformación	Matriz \mathbf{H}	Invariantes
Euclídea	$\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$	longitud entre puntos.
Similitud	$\begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$	ángulos entre rectas, razón entre dos distancias.
Afín	$\begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$	líneas paralelas, razón entre dos áreas.
General	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$	razón de cruz [23].

Tabla 2.1: Resumen de Transformaciones Proyectivas 2D.

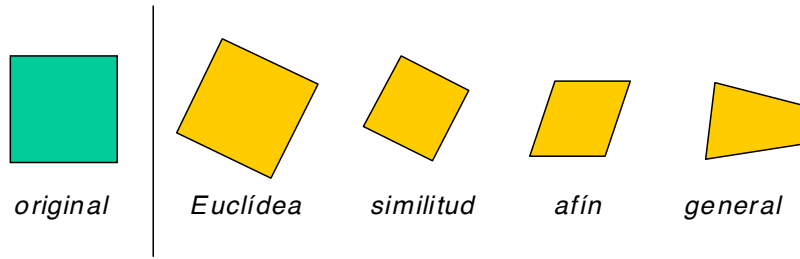


Figura 2.6: Transformaciones proyectivas 2D.

2.3.5. Resumen de Transformaciones Proyectivas 2D

Un resumen de las transformaciones proyectivas 2D explicadas en esta sección se muestra en la Tabla 2.1. Es necesario señalar que las invariantes de un grupo inferior son heredadas por su grupo superior mas no en sentido inverso, es decir la conservación de las líneas paralelas es invariante de las tres primeras transformaciones, y la longitud entre dos puntos sólo es invariante de la transformación Ecuclídea. Adicionalmente, en la Figura 2.6 se muestran las distintas transformaciones de un cuadrado que se pueden realizar con las cuatro transformaciones explicadas.

2.4. Transformaciones Proyectivas 3D

Un punto M que se encuentra en el espacio 3D se representa en coordenadas homogéneas como un vector de cuatro elementos. Si el punto 3D tiene coordenadas (no homogéneas) $(X, Y, Z)^T$ se expresará entonces como $\mathbf{M} = [X_1 \ X_2 \ X_3 \ X_4]^T$ donde $X = X_1/X_4$, $Y = X_2/X_4$, y $Z = X_3/X_4$. Una forma sencilla de pasar de coordenadas no homogéneas a homogéneas es agregando un uno al final del vector, es decir $\mathbf{M} = [X \ Y \ Z \ 1]^T$.

Un plano se define como:

$$p_1X + p_2Y + p_3Z + p_4 = 0. \quad (2.17)$$

Esta ecuación también puede ser escrita como $\mathbf{p}^T \mathbf{M} = 0$ con $\mathbf{p} = [p_1 \ p_2 \ p_3 \ p_4]^T$ y $\mathbf{M} = [X \ Y \ Z \ 1]^T$. Se observa entonces que al igual que la recta en \mathcal{R}^2 un

Transformación	Matriz \mathbf{H}	Invariantes
Euclídea	$\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$	longitud entre puntos, volumen.
Similitud	$\begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$	ángulos entre planos, forma.
Afín	$\begin{bmatrix} a_{11} & a_{12} & a_{13} & t_x \\ a_{21} & a_{22} & a_{23} & t_y \\ a_{31} & a_{32} & a_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$	líneas paralelas.
General	$\begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{bmatrix}$	intersección y tangentes de superficies en contacto.

Tabla 2.2: Resumen de Transformaciones Projectivas 3D.

plano tiene una representación homogénea, ya que $k\mathbf{p}$, para $k \neq 0$, representa el mismo plano definido en (2.17).

Una transformación proyectiva se define como:

$$\mathbf{M}' = \mathbf{H}\mathbf{M} \quad (2.18)$$

donde \mathbf{H} es una matriz 4×4 invertible. Al igual que en el caso de la transformación proyectiva bidimensional en la que las rectas son transformadas como rectas, en el caso de la transformación proyectiva 3D un plano transformado con (2.18) sigue siendo un plano. La representación homogénea del plano transformado esta dada por $\mathbf{p}'^\top = \mathbf{p}^\top \mathbf{H}^{-1}$.

En esta sección las transformaciones proyectivas 3D no serán discutidas tan en detalle como las transformaciones proyectivas 2D. Un resumen de ellas se puede encontrar en la Tabla 2.2. La Figura 2.7 presenta las distintas transformaciones aplicadas a un cubo.

A lo largo de este documento sólo se utilizará la transformación 3D Euclídea, ya que ella representa los cambios de coordenadas que pueden sufrir los objetos rígidos al pasar de un sistema de coordenadas a otro. Deformaciones de objetos 3D no serán contempladas en este documento.

Dado un sistema de coordenadas 3D (X, Y, Z) que ha sufrido una rotación

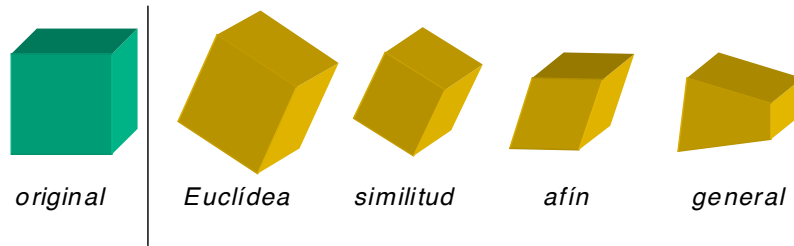


Figura 2.7: Transformaciones proyectivas 3D.

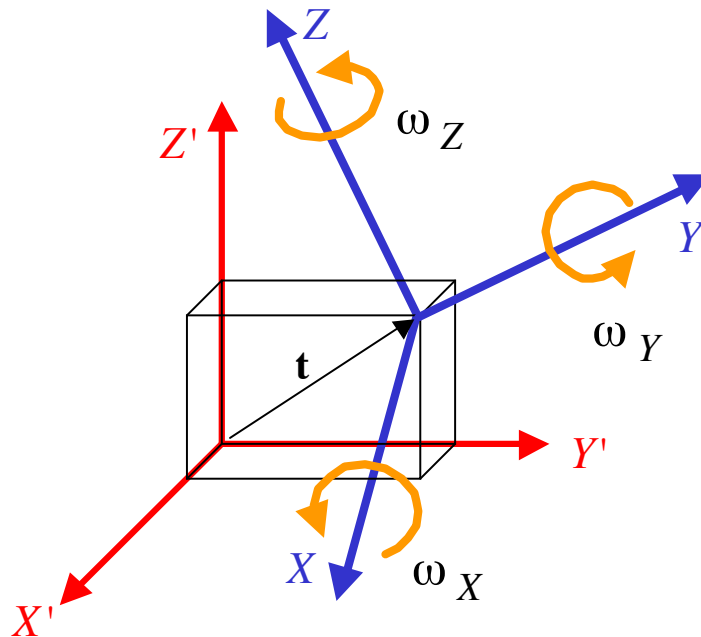


Figura 2.8: Transformación 3D Euclídea.

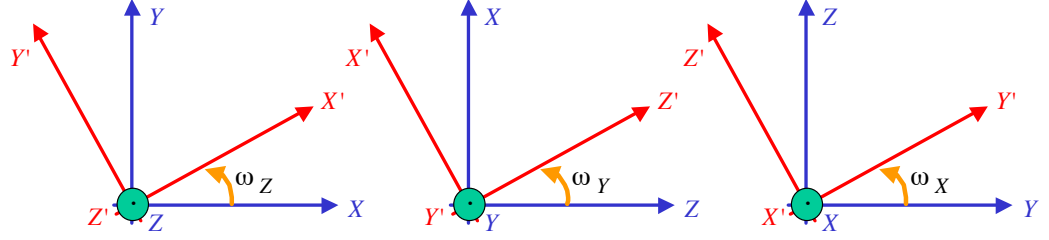


Figura 2.9: Rotación de los ejes Z , Y , y X .

y una traslación como se aprecia en la Figura 2.8, el espacio 3D en el nuevo sistema de coordenadas (X', Y', Z') queda definido por una transformación 3D Euclídea definida por:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \mathbf{R} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \mathbf{t} \quad (2.19)$$

o en coordenadas homogéneas:

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.20)$$

donde \mathbf{R} es una matriz 3×3 ortonormal y \mathbf{t} es un vector 3×1 que definen la rotación y traslación del sistema de coordenadas respectivamente.

Al igual que en la transformación Euclídea 2D explicada en la Sección 2.3.1, se puede expresar (X, Y, Z) en función de (X', Y', Z') de la siguiente manera:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} \quad (2.21)$$

donde $\mathbf{R}' = \mathbf{R}^\top$ y $\mathbf{t}' = -\mathbf{R}'\mathbf{t}$.

A continuación se definirá la matriz ortonormal \mathbf{R} presente en la transformación Euclídea $\mathcal{R}^3 \rightarrow \mathcal{R}^3$. Una rotación de los ejes de coordenadas puede

Rotación	Matriz de rotación
Eje Z	$\mathbf{R}_Z = \begin{bmatrix} \cos(\omega_Z) & \sin(\omega_Z) & 0 \\ -\sin(\omega_Z) & \cos(\omega_Z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Eje Y	$\mathbf{R}_Y = \begin{bmatrix} \cos(\omega_Y) & 0 & -\sin(\omega_Y) \\ 0 & 1 & 0 \\ \sin(\omega_Y) & 0 & \cos(\omega_Y) \end{bmatrix}$
Eje X	$\mathbf{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega_X) & \sin(\omega_X) \\ 0 & -\sin(\omega_X) & \cos(\omega_X) \end{bmatrix}$

Tabla 2.3: Matriz de rotación de los ejes Z , Y y X .

ser descompuesto en rotaciones de cada uno de los ejes tal como se muestra en la Figura 2.9. Las transformaciones de cada una de estas rotaciones están dadas por \mathbf{R}_Z , \mathbf{R}_Y y \mathbf{R}_X en la Tabla 2.3. A manera de ejemplo si el único movimiento existente es la rotación del eje X , la ecuación que transforma las coordenadas sería:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega_X) & \sin(\omega_X) \\ 0 & -\sin(\omega_X) & \cos(\omega_X) \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

La rotación total se puede definir entonces como primero una rotación del eje Z , luego del eje Y y luego del eje X , eso se puede expresar matemáticamente como una multiplicación de las tres matrices de rotación en el siguiente orden:

$$\mathbf{R}(\omega_X, \omega_Y, \omega_Z) = \mathbf{R}_X(\omega_X) \mathbf{R}_Y(\omega_Y) \mathbf{R}_Z(\omega_Z) = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}, \quad (2.22)$$

donde los elementos R_{ij} pueden ser expresados como:

$$\begin{aligned}
R_{11} &= \cos(\omega_Y) \cos(\omega_Z) \\
R_{12} &= \cos(\omega_Y) \sin(\omega_Z) \\
R_{13} &= -\sin(\omega_Y) \\
R_{21} &= \sin(\omega_X) \sin(\omega_Y) \cos(\omega_Z) - \cos(\omega_X) \sin(\omega_Z) \\
R_{22} &= \sin(\omega_X) \sin(\omega_Y) \sin(\omega_Z) + \cos(\omega_X) \cos(\omega_Z) \\
R_{23} &= \sin(\omega_X) \cos(\omega_Y) \\
R_{31} &= \cos(\omega_X) \sin(\omega_Y) \cos(\omega_Z) + \sin(\omega_X) \sin(\omega_Z) \\
R_{32} &= \cos(\omega_X) \sin(\omega_Y) \sin(\omega_Z) - \sin(\omega_X) \cos(\omega_Z) \\
R_{33} &= \cos(\omega_X) \cos(\omega_Y)
\end{aligned}$$

Es necesario resaltar que la multiplicación de matrices no es conmutativa, esto quiere decir que $\mathbf{R} = \mathbf{R}_X(\omega_X) \mathbf{R}_Y(\omega_Y) \mathbf{R}_Z(\omega_Z) \neq \mathbf{R}_Z(\omega_Z) \mathbf{R}_Y(\omega_Y) \mathbf{R}_X(\omega_X)$, sin embargo es posible obtener el mismo resultado para \mathbf{R} cambiando el orden de las matrices siempre y cuando se consideren otros ángulos de rotación para cada eje, es decir: $\mathbf{R} = \mathbf{R}_X(\omega_X) \mathbf{R}_Y(\omega_Y) \mathbf{R}_Z(\omega_Z) = \mathbf{R}_Z(\omega'_Z) \mathbf{R}_Y(\omega'_Y) \mathbf{R}_X(\omega'_X)$. A manera de ejemplo se puede comprobar que si primero hay una rotación $\omega_Z = 90^\circ$ y luego una rotación $\omega_Y = -90^\circ$ el resultado es el mismo que con una primera rotación de $\omega'_X = 90^\circ$ y luego una de $\omega'_Z = 90^\circ$ (ver ejercicio 2.7).

La matriz $\mathbf{R}' = \mathbf{R}^\top$ definida en (2.21) para calcular la transformación del sistema de coordenadas (X', Y', Z') al sistema de coordenadas (X, Y, Z) sería para $\mathbf{R} = \mathbf{R}_X(\omega_X) \mathbf{R}_Y(\omega_Y) \mathbf{R}_Z(\omega_Z)$:

$$\begin{aligned}
\mathbf{R}' &= \mathbf{R}^\top = \mathbf{R}_Z^\top(\omega_Z) \mathbf{R}_Y^\top(\omega_Y) \mathbf{R}_X^\top(\omega_X) \\
&= \mathbf{R}_Z(-\omega_Z) \mathbf{R}_Y(-\omega_Y) \mathbf{R}_X(-\omega_X)
\end{aligned} \tag{2.23}$$

En la última igualdad se utilizó la propiedad (2.13).

2.5. Ejercicios

Ejercicio 2.1 Resolver el sistema de ecuaciones

$$\begin{cases} x + 2y = 8 \\ x - 2y = 0 \end{cases}$$

usando (2.3).

Solución: $\mathbf{l} = (1, 2, -8)^\top$, $\mathbf{l}' = (1, -2, 0)^\top$, $\mathbf{m} = \mathbf{l} \times \mathbf{l}' = (-16, -8, -4)^\top$. Esto implica que $x = -16/-4 = 4$, $y = -8/-4 = 2$.

Ejercicio 2.2 En un sistema de coordenadas (x, y) se definen los vértices de un cuadrado $(-1, 1)$, $(1, 1)$, $(1, -1)$ y $(-1, -1)$. Se desea conocer las coordenadas de estos vértices en un nuevo sistema de coordenadas (x', y') como el mostrado en la Figura 2.5 en donde $\theta = 30^\circ$, $t_x = 1$ y $t_y = 2$

Solución: Utilizando la transformación (2.12) se obtiene que los puntos en el nuevo sistema de coordenadas son $(-2,2321; 0,1340)$, $(-0,5; -0,8660)$, $(-1,5; -2,5981)$ y $(-3,2321; -1,5981)$ respectivamente.

Ejercicio 2.3 Transformar las coordenadas de los vértices de un cuadrado $(-1, 1)$, $(1, 1)$, $(1, -1)$ y $(-1, -1)$ a un nuevo sistema de coordenadas en el que ha ocurrido una rotación de 30° , un desplazamiento en $(1, 2)$ y un cambio de escala $1 : 2$.

Solución: Se tiene $\theta = 30^\circ$, $t_x = 1$, $t_y = 2$ y $s = 0,5$. Utilizando la transformación (2.15) se obtiene que los puntos en el nuevo sistema de coordenadas están dados por $(-2,0490; -0,5490)$, $(-1,1830; -1,0490)$, $(-1,6830; -1,9151)$ y $(-2,5490; -1,4151)$ respectivamente.

Ejercicio 2.4 Escribir un programa que implemente la transformada proyectiva \mathbf{H} entre una imagen original \mathbf{I}' una y rectificadas \mathbf{I} .

Solución: Los datos de entrada al programa son \mathbf{H} y la imagen original \mathbf{I}' cuyos elementos $I'(x', y')$ son los valores de gris en los píxeles denotados por (x', y') para $x' = 1, \dots, N'$ y $y' = 1, \dots, M'$. Se desea encontrar entonces la imagen rectificada \mathbf{I} donde $I(x, y)$ son los valores de gris en los píxeles (x, y) para $x = 1, \dots, N$ y $y = 1, \dots, M$. La idea de este programa es que para cada píxel (x, y) en la imagen rectificada se busque su píxel correspondiente (x', y') en la imagen original utilizando la transformación \mathbf{H} . Seguramente el valor de gris en (x', y') tendrá que ser interpolado de sus cuatro vecinos enteros, ya que al ser \mathbf{I}' una imagen digitalizada $I'(x', y')$ está definido sólo para valores enteros de (x', y') . El algoritmo se presenta a continuación:

Algoritmo:

1. Para $x = 1, \dots, N$ y para $y = 1, \dots, M$:
2. Definir $\mathbf{m} = [x \ y \ 1]^\top$.
3. Evaluar $\mathbf{m}' = \mathbf{H}\mathbf{m}$.
4. Evaluar $x' = m_1/m_3$ y $y' = m_2/m_3$.
5. Definir $\tilde{x} = \text{fix}(x')$, $\tilde{y} = \text{fix}(y')$, $\Delta x' = x' - \tilde{x}$ y $\Delta y' = y' - \tilde{y}$.
6. Evaluar $I(x, y) = [I'(\tilde{x}+1, \tilde{y}) - I'(\tilde{x}, \tilde{y})]\Delta x' + [I'(\tilde{x}, \tilde{y}+1) - I'(\tilde{x}, \tilde{y})]\Delta y' + [I'(\tilde{x}+1, \tilde{y}+1) + I'(\tilde{x}, \tilde{y}) - I'(\tilde{x}+1, \tilde{y}) - I'(\tilde{x}, \tilde{y}+1)]\Delta x'\Delta y' + I'(\tilde{x}, \tilde{y})$

En este algoritmo se ha utilizado la función ‘fix’ que extrae la parte entera de un número real. La interpolación tiene lugar en el punto 6 del algoritmo, para mayores detalles de esta interpolación bilineal consultar [4].

Ejercicio 2.5 *Encontrar la representación homogénea de un plano \mathbf{p} que contenga los puntos M_1 , M_2 y M_3 .*

Solución: Utilizando la representación homogénea de los 3 puntos \mathbf{M}_i , para $i = 1, 2, 3$ se puede escribir la ecuación del plano para cada punto $\mathbf{M}_i^\top \mathbf{p} = 0$, o bien:

$$\begin{bmatrix} \mathbf{M}_1^\top \\ \mathbf{M}_2^\top \\ \mathbf{M}_3^\top \end{bmatrix} \mathbf{p} = 0$$

La solución a este sistema de ecuaciones está dada por

$$\mathbf{p} = [D_{234} \ -D_{134} \ D_{124} \ -D_{123}]^\top$$

donde D_{jkl} es el determinante de las filas j , k y l de la matriz $[\mathbf{M}_1 \mathbf{M}_2 \mathbf{M}_3]$.

Ejercicio 2.6 *Encontrar la intersección de tres planos representados en vectores homogéneos como \mathbf{p}_1 , \mathbf{p}_2 y \mathbf{p}_3 .*

Solución: La intersección de tres planos es un punto. Utilizando la representación homogénea de este punto \mathbf{M} , y sabiendo que este punto pertenece a los tres planos, se puede escribir entonces $\mathbf{M}^T \mathbf{p}_i = 0$, o bien:

$$\begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \mathbf{p}_3^T \end{bmatrix} \mathbf{M} = 0$$

Al igual que el ejemplo anterior, la solución a este sistema de ecuaciones está dada por $\mathbf{M} = (D_{234}, -D_{134}, D_{124}, -D_{123})^T$ donde D_{jkl} es el determinante de las filas j , k y l de la matriz $[\mathbf{p}_1 \mathbf{p}_2 \mathbf{p}_3]$.

Ejercicio 2.7 Encontrar la matriz de rotación \mathbf{R} para las siguientes rotaciones de ejes *i)* primero $\omega_Z = 90^0$ y luego $\omega_Y = -90^0$. *ii)* primero $\omega'_X = 90^0$ y luego $\omega'_Z = 90^0$.

Solución: Para el caso *i)* la matriz de rotación queda $\mathbf{R} = \mathbf{R}_X(\omega_X = 0^0) \mathbf{R}_Y(\omega_Y = -90^0) \mathbf{R}_Z(\omega_Z = 90^0)$. Para el caso *ii)* la matriz de rotación queda $\mathbf{R} = \mathbf{R}_Z(\omega'_Z = 90^0) \mathbf{R}_Y(\omega'_Y = 0^0) \mathbf{R}_X(\omega'_X = 90^0)$. Para ambos casos el resultado es:

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}.$$

Ejercicio 2.8 En la Figura 2.10, encontrar la transformación Euclídea del sistema de coordenadas del objeto (X_4, Y_4, Z_4) al sistema de coordenadas de referencia (X_1, Y_1, Z_1) .

Ejercicio 2.9 Encontrar la transformación proyectiva $2D \rightarrow 2D$, definida por la matriz 3×3 \mathbf{H} que realiza la corrección de distorsión de perspectiva mostrada en la Figura 3.10 suponiendo que $(x'_1, y'_1) = (1, 1)$; $(x'_2, y'_2) = (4, 1)$; $(x'_3, y'_3) = (3, 4)$; $(x'_4, y'_4) = (2, 4)$; $(x_1, y_1) = (1, 1)$; $(x_2, y_2) = (4, 1)$; $(x_3, y_3) = (4, 4)$; $(x_4, y_4) = (1, 4)$. Para los cálculos suponer $h_{33}=1$.

Solución: A partir de (2.7) se obtienen dos ecuaciones por cada par de puntos correspondientes. Esto quiere decir que para los cuatro puntos

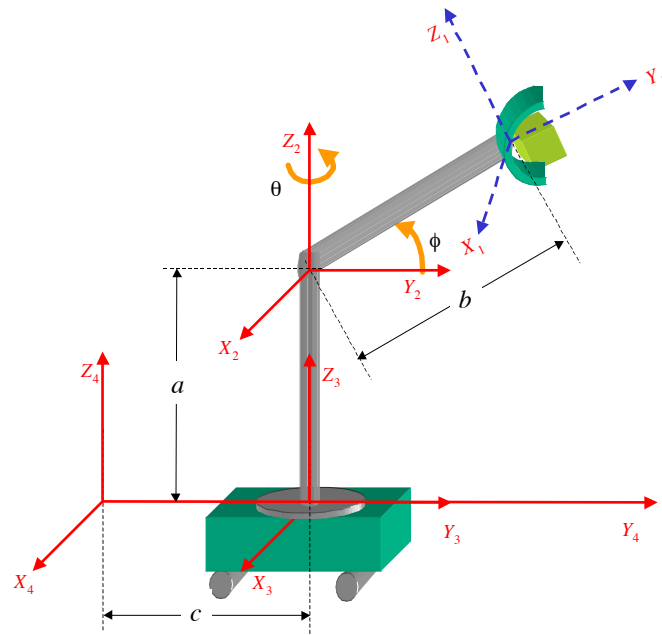


Figura 2.10: Figura del Ejercicio 2.8: Transformación $(X_4, Y_4, Z_4) \rightarrow (X_1, Y_1, Z_1)$.

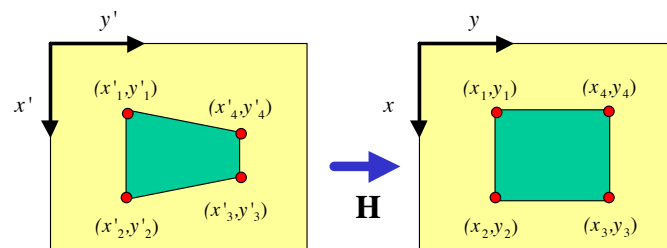


Figura 2.11: Distorsión de perspectiva (ver Ejercicio 2.9).

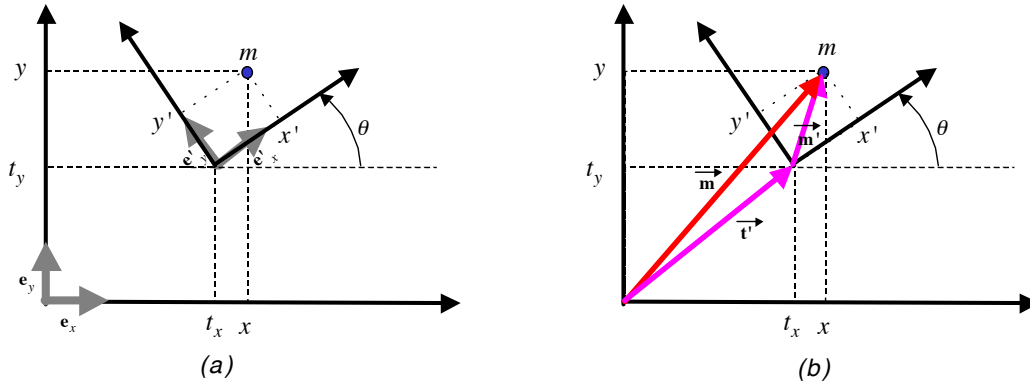


Figura 2.12: Transformación Euclídea 2D (Ejercicio 2.10).

correspondientes el sistema de ecuaciones quedaría:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 \\ 4 & 1 & 1 & 0 & 0 & 0 & -16 & -4 \\ 0 & 0 & 0 & 4 & 1 & 1 & -4 & -1 \\ 4 & 4 & 1 & 0 & 0 & 0 & -12 & -12 \\ 0 & 0 & 0 & 4 & 4 & 1 & -16 & -16 \\ 1 & 4 & 1 & 0 & 0 & 0 & -2 & -8 \\ 0 & 0 & 0 & 1 & 4 & 1 & -4 & -16 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 4 \\ 1 \\ 3 \\ 4 \\ 2 \\ 4 \end{bmatrix}$$

o bien $\tilde{\mathbf{A}}\mathbf{h} = \mathbf{b}$. Los elementos h_{11}, \dots, h_{32} se encuentran entonces a partir de $\mathbf{h} = [h_{11} \dots h_{32}]^T = [\tilde{\mathbf{A}}]^{-1}\mathbf{b} = (3, 5, -5, 0, 11, -8, 0, 2)^T$. Esto quiere decir que la matriz \mathbf{H} quedaría:

$$\mathbf{H} = \begin{bmatrix} 3 & 5 & -5 \\ 0 & 11 & -8 \\ 0 & 2 & 1 \end{bmatrix}.$$

Ejercicio 2.10 Demostrar que la relación entre las coordenadas (x', y') y (x, y) en la Figura 2.5 está dada por (2.8).

Solución: Se definen los vectores unitarios de los ejes x y y como \mathbf{e}_x y \mathbf{e}_y , y de los ejes x' y y' como \mathbf{e}'_x y \mathbf{e}'_y , tal como se aprecia en la Figura 2.12a. Se

sabe que la relación entre ellos es:

$$\begin{aligned} \mathbf{e}'_x &= \cos(\theta)\mathbf{e}_x + \sin(\theta)\mathbf{e}_y \\ \mathbf{e}'_y &= -\sin(\theta)\mathbf{e}_x + \cos(\theta)\mathbf{e}_y \end{aligned} \quad (2.24)$$

En la Figura 2.12b, el punto m puede ser representado vectorialmente como $\vec{\mathbf{m}}$ o bien como su descomposición $\vec{\mathbf{t}}' + \vec{\mathbf{m}}'$. Igualándolos se obtiene:

$$\underbrace{x\mathbf{e}_x + y\mathbf{e}_y}_{\vec{\mathbf{m}}} = \underbrace{t_x\mathbf{e}_x + t_y\mathbf{e}_y}_{\vec{\mathbf{t}}'} + \underbrace{x'\mathbf{e}'_x + y'\mathbf{e}'_y}_{\vec{\mathbf{m}}'}.$$

Reemplazando (2.24) en esta última ecuación se obtiene una expresión que depende de los vectores unitarios \mathbf{e}_x y \mathbf{e}_y :

$$x\mathbf{e}_x + y\mathbf{e}_y = t_x\mathbf{e}_x + t_y\mathbf{e}_y + x'\cos(\theta)\mathbf{e}_x + x'\sin(\theta)\mathbf{e}_y - y'\sin(\theta)\mathbf{e}_x + y'\cos(\theta)\mathbf{e}_y.$$

Igualando las componentes en la dirección x y las componentes en la dirección y de manera independiente se obtiene (2.8):

$$\begin{aligned} x &= x'\cos(\theta) - y'\sin(\theta) + t_x \\ y &= x'\sin(\theta) + y'\cos(\theta) + t_y \end{aligned}.$$

Capítulo 3

Modelación Geométrica de un Sistema de Visión por Computador

En este capítulo se describe cómo funciona un sistema de visión por computador y se presenta un modelo geométrico para poder calcular la función de transferencia del espacio 3D y la imagen 2D tomada de él.

3.1. Descripción de un Sistema de Visión por Computador

En esta sección se describirá cuáles son los principales componentes de un sistema de visión por computador. Tal como se aprecia en la Figura 3.1 estos componentes son: el manipulador, la fuente de energía, el sensor, el conversor análogo-digital y el computador.

3.1.1. El Manipulador

El manipulador es un aparato que mueve y ubica el objeto de estudio a una posición deseada sin ser tocado por el ser humano. Un manipulador posee *grados de libertad* que indican los posibles movimientos que puede hacer para

mover el objeto. Los grados de libertad pueden ser de traslación y/o rotación. Muchas veces el manipulador se acciona mediante *joysticks*, otras veces por medio de una interfaz con un PLC o computador.

El manipulador consta de elementos deslizantes y de giro con los que se lleva a cabo la traslación y rotación del objeto respectivamente.

Hay configuraciones en las que el manipulador no mueve al objeto sino a la(s) cámara(s), esto es muy ventajoso cuando se trata de analizar objetos muy pesados, ya que mover la(s) cámara(s) requiere de una mecánica más sencilla y económica.

3.1.2. Fuente de Energía

Dependiendo del tipo de análisis que se desea hacer del objeto de estudio se debe escoger la energía necesaria para poder tomar una imagen de él. Los tipos de energía utilizados son: luz (visible) para la fotografía, rayos X y rayos γ para la radiografía y también para tomografía, ultrasonido para la ecografía, campos magnéticos para la magneto-resonancia, calor para la termografía, etc.

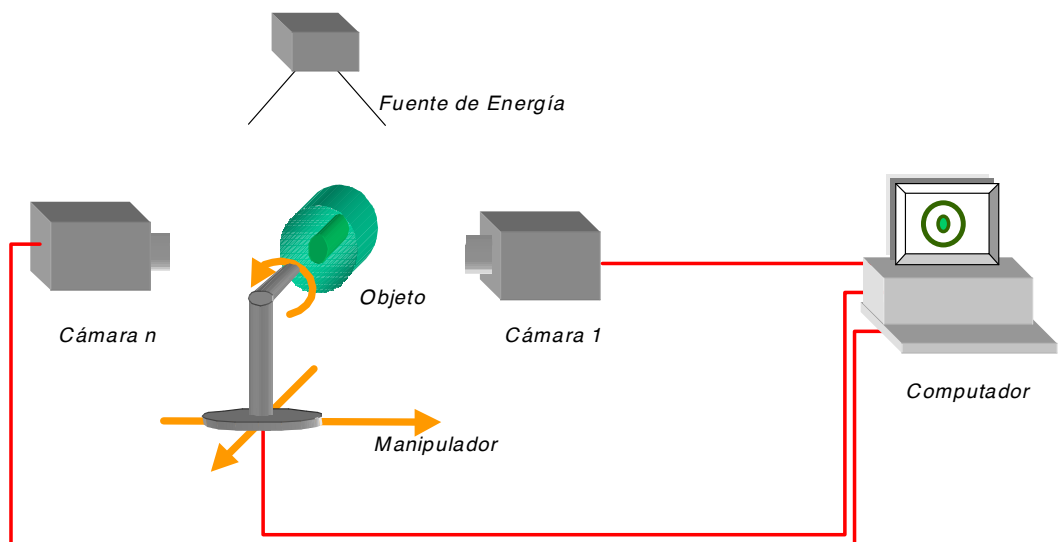


Figura 3.1: Sistema de visión por computador.

En la gran mayoría de casos se utilizan filtros para restringir el espectro de frecuencias de la energía. En el caso de iluminación es importante analizar si se prefiere luz difusa o directa y también su color (espectro).

3.1.3. Sensor de Imagen

El sensor debe ser sensible a la energía utilizada. Si es luz por ejemplo será necesario utilizar algún tipo de elemento fotosensible que transforme los fotones reflejados por el objeto de estudio a alguna señal eléctrica (generalmente voltaje). Para el caso de los rayos X estos elementos son muy poco sensibles a los fotones de este espectro por lo que se utiliza entre el objeto y el sensor fotosensible un amplificador de imagen que transforma los rayos X en luz visible¹

El sensor debe ser bidimensional (o unidimensional en movimiento) para poder captar las dos dimensiones de la imagen.

3.1.4. Conversor Análogo-Digital

El conversor A/D convierte la señal eléctrica a un código binario que puede ser interpretado por el computador para conformar una imagen digital del objeto de estudio.

3.1.5. Computador

El computador se encarga de procesar la información entregada por el conversor A/D. Las tareas típicas de un computador utilizado en un sistema de visión por computador son: *i)* mejoramiento de la imagen, *ii)* segmentación, *iii)* clasificación de patrones y *iv)* análisis espacial.

3.2. La Cámara *Pinhole*

El modelo básico de la cámara pinhole ya ha sido visto en la Sección 2. Un esquema de este modelo se presenta en la Figura 3.2. El modelo consiste en

¹Existen elementos de estado sólido sensibles a los rayos X, sin embargo el uso del amplificador de imagen resulta tres veces más económico.

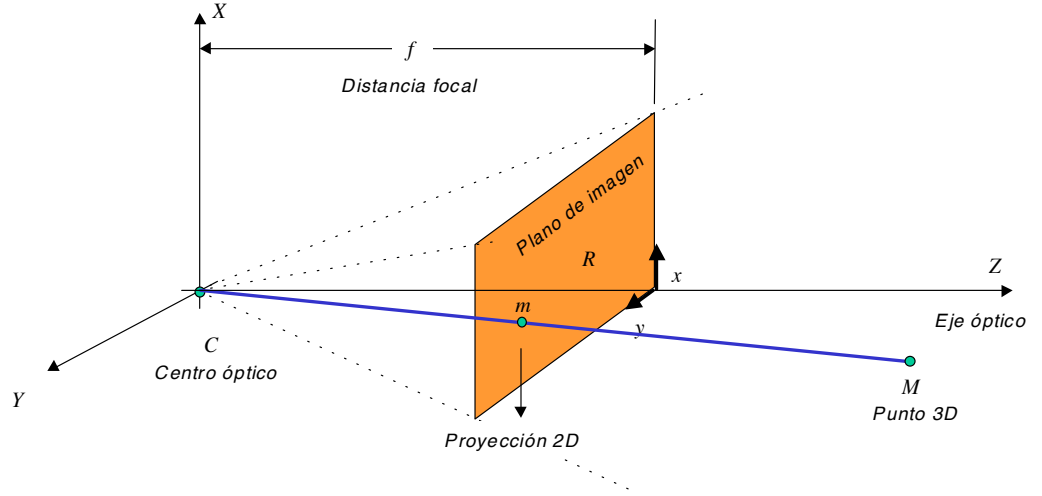


Figura 3.2: Modelo geométrico de cámara *pinhole*.

un centro óptico C , en donde convergen todos los rayos de la proyección, y un plano de imagen R en el cual la imagen es proyectada. El plano de imagen está ubicado a una distancia focal f del centro óptico y perpendicular al eje óptico Z .

Un punto 3D M es proyectado en el plano de imagen como m . El punto 2D m se define como la intersección de la recta $\langle C, M \rangle$ con el plano R , donde la notación $\langle A, B \rangle$ denota la línea recta que contiene los puntos A y B :

$$m = \langle C, M \rangle \cap R \quad (3.1)$$

Suponiendo que las coordenadas (inhomogéneas) de los puntos M y m son $(X, Y, Z)^T$ y $(x, y)^T$ respectivamente, se puede encontrar una relación entre ellas aplicando el teorema de Tales.

$$\begin{cases} Zx = fX \\ Zy = fY \end{cases} \quad (3.2)$$

o bien en coordenadas homogéneas:

$$Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.3)$$

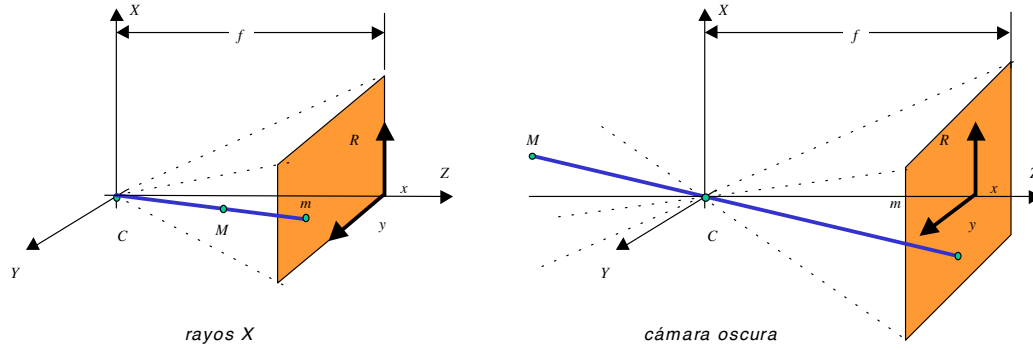


Figura 3.3: Modelo geométrico de proyección rayos X y cámara oscura.

que puede ser escrita en forma matricial como

$$\lambda \mathbf{m} = \mathbf{P} \mathbf{M} \quad (3.4)$$

siendo $\mathbf{M} = [X \ Y \ Z \ 1]^T$ y $\mathbf{m} = [x \ y \ 1]^T$ las coordenadas homogéneas de M y m respectivamente, y \mathbf{P} la matriz de 3×4 denominada *matriz de proyección perspectiva* de la cámara. El factor λ es un factor de escala para mantener la igualdad y es igual a Z . Se observa que la ecuación no lineal de la proyección en coordenadas inhomogéneas (3.2) se convierte en una ecuación lineal en coordenadas homogéneas, lo cual constituye una de las principales ventajas del uso de las coordenadas homogéneas en la geometría proyectiva.

Modelos geométricos similares presentan la proyección de rayos X y la cámara oscura (ver Figura 3.3). En los rayos X la fuente de emisión es modelada como un punto y coincide con el centro óptico C . El objeto es ubicado entre la fuente de rayos X y la placa fotográfica (plano de imagen R). En este caso, las ecuaciones que describen la proyección coinciden con las expresadas anteriormente. Para la cámara oscura, descrita en la Sección 1.3 e ilustrada en la Figura 1.14, el orificio de ingreso de luz corresponde al centro óptico C ya que por él pasan todos los haces de luz que conforman la imagen en el plano de imagen R . Sin embargo, debido a que el centro óptico se ubica entre el objeto y el plano de imagen ocurre una inversión de la imagen, es decir $X/Z = -x/f$ y $Y/Z = -y/f$. En la matriz de proyección perspectiva \mathbf{P} entonces es necesario cambiar f por $-f$:

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.5)$$

En el modelo de proyección $\lambda \mathbf{m} = \mathbf{P} \mathbf{M}$ se ha asumido que:

- i) el origen del sistema de coordenadas del espacio 3D coincide con el centro óptico C ,
- ii) el eje óptico coincide con el eje Z de este sistema de coordenadas, y
- iii) el origen del sistema de coordenadas del plano de la imagen coincide con la intersección de Z con R . Esta intersección es conocida con el punto principal c de la imagen.

¿Qué pasa si estos tres supuestos no se cumplen? Se considerará inicialmente un nuevo sistema de coordenadas (X', Y', Z') como se ilustra en la Figura 3.4. Este nuevo sistema de coordenadas no cumple las dos primeras condiciones anteriormente señaladas ya que su origen no coincide con el centro óptico y su eje Z' no coincide con el eje óptico de la proyección. Generalmente, este

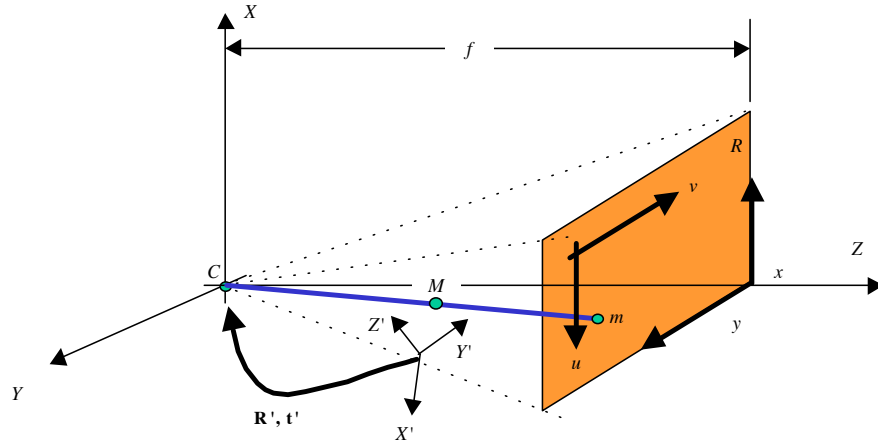


Figura 3.4: Modelo geométrico de proyección con rotación de ejes.

nuevo sistema de coordenadas está referido al objeto de estudio. Este cambio de coordenadas corresponde a una transformación 3D Euclídea (ver Figura 2.8) ya que sólo están involucradas la rotación y la traslación del objeto (y no la deformación o cambio de escala), como ya se analizó en la Sección 2.4. Considerando en este caso que la rotación de los ejes X' , Y' y Z' respecto a los ejes X , Y , y Z está definida por la matriz ortonormal 3×3 \mathbf{R}' y que el origen del nuevo sistema de coordenadas (X', Y', Z') se representa por el vector 3×1 \mathbf{t}' en el sistema de coordenadas (X, Y, Z) entonces se puede escribir:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix}, \quad (3.6)$$

o bien

$$\mathbf{M} = \mathbf{S}'\mathbf{M}', \quad (3.7)$$

con $\mathbf{M} = [X \ Y \ Z \ 1]^T$, $\mathbf{M}' = [X' \ Y' \ Z' \ 1]^T$ y \mathbf{S}' la matriz 4×4 que incluye a \mathbf{R}' y \mathbf{t}' tal como se muestra en (3.6)². Utilizando (3.7) y (3.4) la proyección de un punto \mathbf{M}' en la imagen queda definida por:

$$\lambda \mathbf{m} = \mathbf{P}\mathbf{S}'\mathbf{M}'. \quad (3.8)$$

En el plano de imagen, se considerará ahora un nuevo sistema de coordenadas dado por los ejes u , y v , como se ilustra en la Figura 3.4, de tal manera que la condición que el origen del sistema de coordenadas del plano de imagen no coincida con el punto principal de la proyección. Si entre los ejes x, y y los ejes u, v sólo existe una traslación y rotación entonces entre estos dos sistemas de coordenadas se da una transformación 2D Euclídea definida en la Sección 2.3.1. Es posible también que haya un cambio de escala entre ambos ejes como ocurre en las cámaras CCD, donde las unidades de los ejes están dadas en píxeles y no en milímetros por ejemplo. En términos generales se puede decir que entre ambos sistemas de coordenadas existe una transformación proyectiva 2D que puede ser expresada como:

$$\mathbf{w} = \mathbf{H}\mathbf{m} \quad (3.9)$$

con $\mathbf{w} = [u \ v \ 1]^T$, $\mathbf{m} = [x \ y \ 1]^T$ y \mathbf{H} la matriz 3×3 que define la transformación proyectiva $2D \rightarrow 2D$ tal como se explicó en la Sección 2.2.

²Se puede observar que la transformación inversa está definida en (2.20).

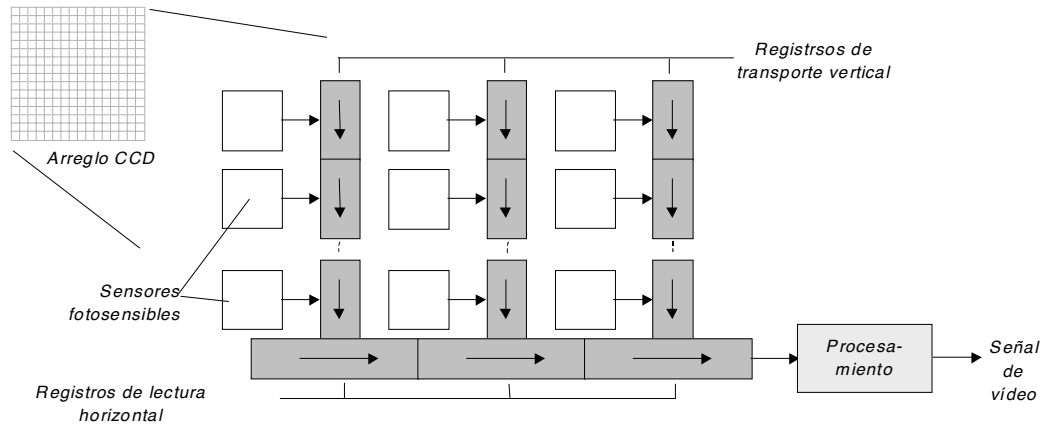


Figura 3.5: Construcción de un arreglo CCD.

La proyección de un punto 3D (X', Y', Z') (representado en coordenadas homogéneas por el vector \mathbf{M}') en un punto (u, v) en el plano de imagen (representado en coordenadas homogéneas por el vector \mathbf{w}) se puede expresar usando las ecuaciones (3.9) y (3.8) de la siguiente manera:

$$\lambda \mathbf{w} = \mathbf{HPS}'\mathbf{M}'. \quad (3.10)$$

3.3. Cámara CCD

La cámara CCD (*charge-coupled-device*) es un sensor de imagen que utiliza elementos semiconductores fotosensibles en forma de arreglos matriciales. Los receptores activos de este arreglo son distribuidos en píxeles individuales. En un sensor CCD se transforma la energía de la luz incidente en energía eléctrica. La carga eléctrica almacenada en la celda es posteriormente transportada utilizando un registro de desplazamiento (*shift register*) para conformar una señal de vídeo. Cabe señalar que en las cámaras CCD se discretiza la imagen en píxeles, sin embargo el valor de la carga eléctrica almacenado en cada celda no se digitaliza en el arreglo CCD sino en una conversión posterior realizada por un conversor análogo-digital.

Las cámaras CCD son consideradas como dispositivos que poseen una muy baja deformación geométrica de la imagen, una muy buena sensibilidad a la

luz y una muy buena resolución llegando típicamente a los 400.000 píxels³. El tiempo de captura de la imagen está típicamente en el rango de 1/60s y 1/10000s [4].

La formación geométrica de la imagen se considera como una transformación de las coordenadas x, y en un nuevo sistema de coordenadas u, v tal como se explicó a grandes rasgos en la sección anterior (ver Figura 3.4). Para esta transformación es necesario considerar los siguientes cuatro aspectos:

Cambio de escala: Las coordenadas de una imagen están comúnmente expresadas en otras unidades métricas que las empleadas para medir el espacio 3D. Por lo general la imagen se expresa en píxels y el espacio 3D en milímetros. Por esta razón, en la transformación de coordenadas $(x, y) \rightarrow (u, v)$ que se mencionó en la Sección 3.2 es necesario considerar un factor de escala. Adicionalmente es necesario tomar en cuenta que debido a que los píxels no son cuadrados, sino rectangulares, el factor de escala es distinto en cada eje de la imagen. Los factores de escala utilizados son α_x y α_y expresados en [pixel/mm].

Traslación del origen: Se introducen las variables (u_0, v_0) para denotar el punto principal de la imagen en el nuevo sistema de coordenadas, es decir $u = u_0, v = v_0$ corresponden al punto $x = 0, y = 0$.

Rotación de los ejes: Los ejes x, y y los ejes u, v no tienen la misma orientación. En la modelación de la cámara existen dos métodos: uno que considera un ángulo θ de rotación, y otro que asume que este ángulo es cero y que el ajuste debe hacerse en los ejes X, Y, Z del sistema de coordenadas del espacio 3D. En este segundo método el eje Z sigue siendo el eje óptico de la proyección y los ejes x, y siguen siendo paralelos a los ejes X, Y , sin embargo se hace coincidir, mediante una rotación del eje Z , la orientación de los ejes x, y con los ejes u, v .

Factor de torcimiento: Muchas veces los ejes u, v no son ortogonales debido a que los píxels en los arreglos CCD no son rectangulares. En este caso es necesario introducirle al modelo de la cámara un factor de torcimiento (*skew factor*) s . En la gran mayoría de cámaras s es cero.

Resumiendo los cuatro aspectos y considerando que la orientación de u, v y

³Hoy en día existen cámaras CCD para HDTV(*High Definition Television*) con alrededor de 2.200.000 píxels.

x, y es la misma, la transformación de coordenadas está definida por:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (3.11)$$

o bien

$$\mathbf{w} = \mathbf{K}\mathbf{m}, \quad (3.12)$$

con $\mathbf{w} = [u \ v \ 1]^T$, $\mathbf{m} = [x \ y \ 1]^T$ y \mathbf{K} la matriz 3×3 escrita en (3.11). La matriz \mathbf{K} es conocida como la matriz de calibración de la cámara.

La proyección $(X', Y', Z') \rightarrow (u, v)$ ($3D \rightarrow 2D$) puede ser escrito utilizando (3.10):

$$\lambda \mathbf{w} = \mathbf{KPS}'\mathbf{M}'. \quad (3.13)$$

En esta ecuación se sabe que la matriz \mathbf{K} depende de 5 parámetros: los dos factores de escala α_x y α_y , las coordenadas del punto principal (u_0, v_0) y el factor de torcimiento s ; la matriz de proyección perspectiva \mathbf{P} depende de la distancia focal f ; y la matriz de la transformación 3D Euclídea depende de 6 parámetros: 3 para la traslación tridimensional (t_X, t_Y, t_Z) y 3 para la rotación de los ejes: $(\omega_X, \omega_Y, \omega_Z)$. Esto quiere decir que el modelo consta de 12 parámetros, sin embargo al realizar la multiplicación \mathbf{KP} :

$$\mathbf{KP} = \begin{bmatrix} \alpha_x & s & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} f\alpha_x & fs & u_0 & 0 \\ 0 & f\alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3.14)$$

se observa que hay elementos que se resumen en la multiplicación. A manera de ejemplo, con este modelo sólo es posible conocer el producto $f\alpha_x$ y no f y α_x por separado. Por esta razón muchas veces el producto anterior se expresa como:

$$\mathbf{KP} = \begin{bmatrix} \alpha'_x & s' & u_0 \\ 0 & \alpha'_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3.15)$$

con $\alpha'_x = f\alpha_x$, $\alpha'_y = f\alpha_y$ y $s' = fs$. Este nuevo modelo puede ser interpretado como una proyección con longitud focal normalizada $f = 1$ y con factores de escala y torcimiento referidos a esta normalización. De esta manera el modelo de la cámara queda definido por sólo 11 parámetros, los que se descomponen en 5 intrínsecos a la cámara $(\alpha'_x, \alpha'_y, u_0, v_0, s')$ y 6 extrínsecos a la cámara $(t_X, t_Y, t_Z, \omega_X, \omega_Y, \omega_Z)$.

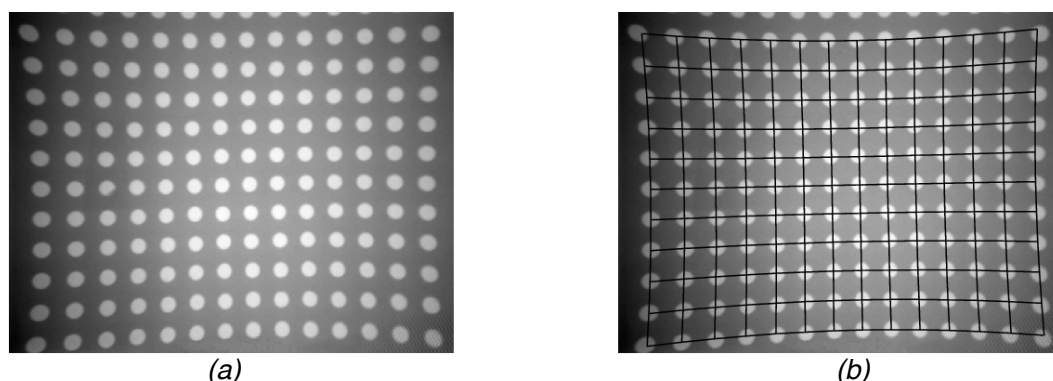


Figura 3.6: Ejemplo de Distorsión: a) Radiografía de un objeto de calibración, b) Modelación de la distorsión

3.4. Distorsión del lente

La curvatura del lente utilizado en las cámaras introduce una deformación en la imagen. Debido a esta distorsión las líneas que en el espacio 3D son rectas ya no son vistas en la proyección como líneas rectas sino como líneas curvas (ver Figura 1.9). Este efecto puede ser despreciable en el centro de la imagen, sin embargo es considerable en los extremos de la imagen, donde la normal de la superficie del lente no es paralela al eje óptico de la proyección. Un ejemplo real es mostrado en la Figura 3.6.

En los casos en que la distorsión de la imagen es grande, el modelo lineal de cámara CCD que se introdujo en la Sección 3.3 deberá ser modificado. En la literatura existe una gran variedad de modelos que incluyen esta distorsión. La idea general de estos modelos consiste en tener una imagen ideal con coordenadas (x, y) y una real con coordenadas (x', y') . Lógicamente, sólo la imagen real es vista. De esta manera el modelo de proyección total, en el que un punto $\mathbf{M}' = [X' \ Y' \ Z' \ 1]^T$ es proyectado en la imagen como $\mathbf{w} = [u \ v \ 1]^T$, consta de cuatro partes:

- i) **Transformación Euclídea 3D:** Con esta transformación las coordenadas de un punto \mathbf{M}' en coordenadas relativas al objeto de estudio son transformadas a las coordenadas del sistema de proyección como un punto \mathbf{M} . La transformación, definida en (3.7), es: $\mathbf{M} = \mathbf{S}'\mathbf{M}'$.

- ii) **Proyección en perspectiva:** Con esta proyección se obtiene a partir del punto \mathbf{M} un punto (x, y) en el plano de la imagen. Esta proyección está definida en (3.4): $\lambda \mathbf{m} = \mathbf{PM}$, con $\mathbf{m} = [x \ y \ 1]^T$. Esta imagen proyectada es la que se denomina imagen *ideal*.
- iii) **Modelación de distorsión:** Con una función de distorsión se obtiene la imagen *real* a partir de la imagen ideal. Es decir a partir de un punto (x, y) de la imagen ideal se obtiene un punto (x', y') de la imagen real. La función de distorsión se expresa como:

$$x' = x + \delta_x(x, y) \quad , \quad y' = y + \delta_y(x, y) \quad (3.16)$$

- iv) **Proyección en la cámara:** La formación de la imagen en la cámara CCD se hace mediante la matriz \mathbf{K} tal como se explicó en la Sección 3.3. Utilizando (3.11) para las nuevas coordenadas se obtiene:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}, \quad (3.17)$$

A continuación se explicará brevemente como se puede modelar la función de distorsión. Generalmente la distorsión es modelada como una componente *radial* δ_r y otra *tangencial* δ_ϕ [50]. Estas componentes son ilustradas en la Figura 3.7 en la que se observa un punto de distorsión nula (x_0, y_0) , en donde $x = x'$, y $y = y'$, es decir $\delta_x(x_0, y_0) = \delta_y(x_0, y_0) = 0$. Comúnmente este punto es el punto principal de la imagen, sin embargo esta condición no siempre se cumple porque el lente puede encontrarse un poco desplazado del eje óptico de la proyección. Descomponiendo (x, y) en coordenadas polares (r, ϕ) con centro en (x_0, y_0) se puede escribir la distorsión como la suma de ambas componentes:

$$\begin{aligned} \delta_x(x, y) &= \delta_r(r) \cos(\phi) - r\delta_\phi(\phi) \sin(\phi) \\ \delta_y(x, y) &= \delta_r(r) \sin(\phi) + r\delta_\phi(\phi) \cos(\phi) \end{aligned} \quad (3.18)$$

La componente de distorsión radial asume que un punto ideal (x, y) se proyecta en la imagen real sobre la línea radial que contiene los puntos (x_0, y_0) y (x, y) . Análogamente, la distorsión tangencial asume que existe un cambio de ángulo δ_ϕ (y no de radio). El efecto es ilustrado en la Figura 3.7. Así como la distorsión radial depende sólo del radio, la distorsión tangencial depende sólo del ángulo. Las distorsiones pueden modelarse como polinomios de orden mayor que uno.

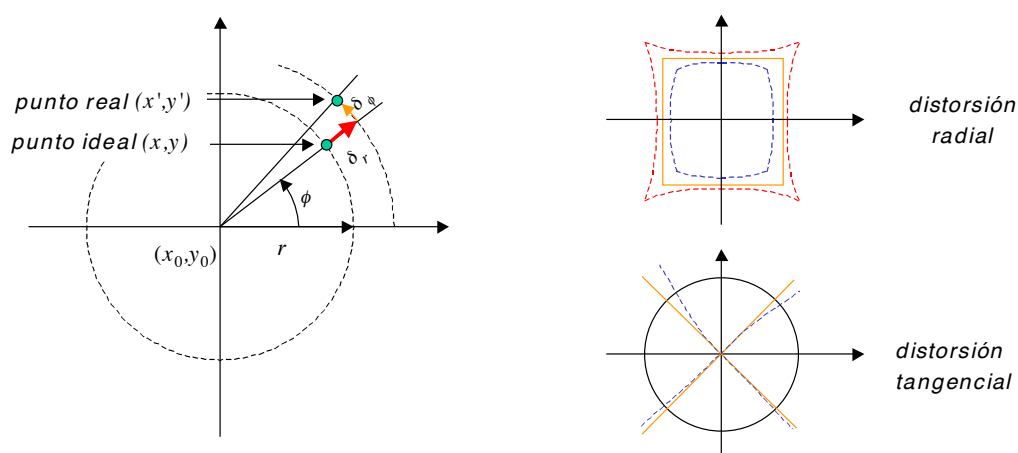


Figura 3.7: Modelación de distorsión de lente en componente radial y tangencial.

3.5. Modelación de un manipulador

La modelación de un manipulador consiste en encontrar las matrices proyectivas Euclídeas 3D que hacen la transformación de coordenadas de un sistema de coordenadas relativas al objeto de estudio al sistema de coordenadas del espacio 3D en el que tiene lugar la proyección perspectiva. En el modelo deben incluirse los grados de libertad que posee el manipulador. Comúnmente por cada centro de rotación se establece un sistema de coordenadas. Un ejemplo puede encontrarse en el Ejercicio 2.8.

En la modelación debe incluirse factores de cambio de escala, ya que por lo general las unidades en las que trabaja un manipulador son 'incrementos' y éstos deben convertirse a milímetros o radianes. La conversión de escala se modela en forma lineal. A manera de ejemplo, si la rotación de un eje es proporcionada por una variable de salida del manipulador denominada R , y ésta corresponde al ángulo de rotación ω_X de nuestra transformación Euclídea, deben encontrarse los parámetros K_{R0} y K_{R1} que establezcan la relación $\omega_X = K_{R0} + K_{R1}R$, de esta manera K_{R1} realiza la conversión de escala ('incrementos' \rightarrow milímetros) y K_{R0} denota el ángulo de rotación en radianes que corresponde a $R = 0$.

3.6. Calibración de un sistema de visión por computador

La calibración es el método mediante el cual se estiman los parámetros intrínsecos y extrínsecos de la cámara, así como los parámetros del manipulador. También es posible estimar los parámetros del modelo de distorsión del lente de la cámara.

Existen dos métodos comúnmente usados para la calibración: auto-calibración (*self-clibration*) y calibración fotogramétrica.

En la auto-calibración se toman varias imágenes de una misma escena y mediante la correspondencia entre puntos de distintas imágenes se puede encontrar los mejores parámetros del modelo que puedan otorgar esta correspondencia. La reconstrucción 3D realizada con el modelo encontrado está afectada sin embargo por un factor de escala ya que en este método no se puede saber cuál es el tamaño real de los objetos captados por las cámaras. Un objeto pequeño cerca del centro óptico puede tener la misma imagen que el mismo objeto agrandado más cerca del plano de imagen.

Si lo que se busca es una reconstrucción 3D precisa, como es el caso de muchas de las aplicaciones de la robótica, es recomendable utilizar la calibración fotogramétrica. Esta calibración utiliza un objeto 3D de referencia cuya geometría es conocida a la perfección. N puntos de interés son escogidos del objeto de referencia, obteniendo así las coordenadas $\mathbf{M}'_i = [X'_i \ Y'_i \ Z'_i \ 1]^T$, para $i = 1, \dots, N$. El objeto es a continuación captado por la cámara y sus puntos de interés son vistos como puntos 2D con coordenadas $\mathbf{w}_i = [u_i \ v_i \ 1]^T$. Teniendo un modelo de la proyección es posible obtener una estimación teórica de los puntos 3D. De esta manera se calculan los puntos:

$$\hat{\mathbf{w}}_i = \mathbf{f}(\mathbf{M}'_i) \quad (3.19)$$

donde \mathbf{f} es la función de proyección que involucra los parámetros de la cámara, del lente y del manipulador según sea el caso. Formalmente \mathbf{f} es una función no lineal que depende de un vector de parámetros Θ que agrupa los parámetros del modelo. Para el caso de una proyección sin distorsión se puede usar (3.13) como función de proyección \mathbf{f} , en este caso los once parámetros del modelo se agrupan como $\Theta = [t_X \ t_Y \ t_Z \ \omega_X \ \omega_Y \ \omega_Z \ \alpha'_x \ \alpha'_y \ u_0 \ v_0 \ s']^T$.

El problema de calibración se transforma en un problema de optimización mediante el cual una función objetivo que mide el error entre la proyección

estimada $\hat{\mathbf{w}}_i$ y la proyección medida \mathbf{w}_i debe ser minimizada. De esta manera, se debe encontrar los parámetros de la función de proyección \mathbf{f} de tal manera que se minimice la siguiente función objetivo:

$$J(\Theta) = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{w}}_i - \mathbf{w}_i\| \rightarrow \min \quad (3.20)$$

3.7. Ejercicios

Ejercicio 3.1 En la Figura 3.8 se presenta el sistema péndulo invertido, el cual consiste en un carrito que se desplaza en la dirección Z' y que sostiene una barra de longitud b la cual se debe mantener en equilibrio. La barra se encuentra a un ángulo θ del eje Z' . El control del péndulo invertido pretende mover el carrito de tal manera que $\theta = \pi/2$. El eje de rotación de la barra se encuentra a una distancia a del origen de un sistema de coordenadas (X', Y', Z') . En este sistema de coordenadas, la barra de longitud b se encuentra en el plano (X', Z') . Para estudiar el ángulo de la barra se cuenta con una cámara cuyo centro óptico C se encuentra en las coordenadas $(X' = X'_0, Y' = Y'_0, Z' = Z'_0)$. En este centro óptico se define un nuevo sistema de coordenadas (X, Y, Z) el cual presenta una rotación del eje X de $\pi/2 + \phi$. En $Z = f$ se encuentra el plano de imagen R de la cámara en el

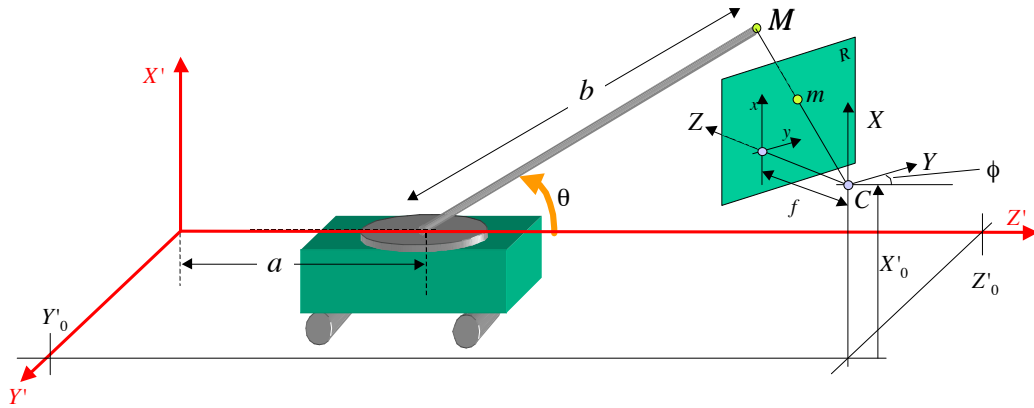


Figura 3.8: Péndulo invertido (ver Ejercicio 3.1).

cual se define un sistema de coordenadas bidimensional (x, y) , donde el eje x y el eje y son paralelos al eje X y al eje Y respectivamente. El origen de este sistema de coordenadas $(x = 0, y = 0)$ es la intersección del eje Z con el plano R . Encuentre las coordenadas (x, y) del punto m definido como la imagen del punto M (ubicado al extremo de la barra) en el plano R .

Solución: La solución a este problema se obtiene en tres etapas: *i)* cálculo de las coordenadas de M en el sistema de coordenadas (X', Y', Z') ; *ii)* transformación Euclídea 3D $(X', Y', Z') \rightarrow (X, Y, Z)$; y *iii)* proyección del punto (X, Y, Z) en el plano R .

i) Las coordenadas de M en el sistema de coordenadas (X', Y', Z') son: $X' = b \sin \theta$, $Y' = 0$ y $Z' = a + b \cos \theta$.

ii) El punto M en el sistema de coordenadas (X, Y, Z) se puede calcular utilizando el esquema de la Figura 2.8. En este caso el vector de traslación es $\mathbf{t} = [X'_0 \ Y'_0 \ Z'_0]^T$ y los ángulos de rotación son $\omega_X = -\phi - \pi/2$, $\omega_Y = 0$ y

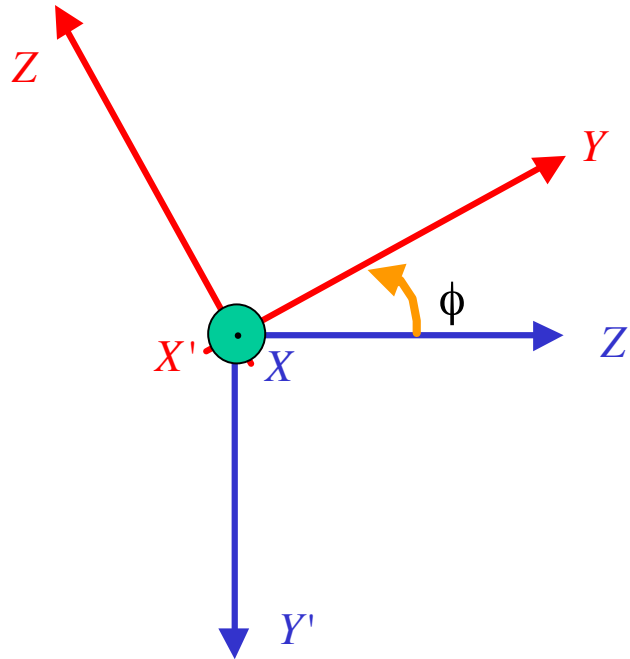


Figura 3.9: Rotación del eje X' en el péndulo invertido (ver Ejercicio 3.1).

$\omega_Z = 0$. El signo menos en ω_X se debe a que el ángulo se mide del eje Y al eje Y' como se aprecia en la Figura 2.9 (comparar con Figura 3.9). La matriz de rotación \mathbf{R} corresponde entonces a \mathbf{R}_X proporcionada por la Tabla 2.3. Según (2.20) la transformación entre (X', Y', Z') y (X, Y, Z) queda:

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & X'_0 \\ 0 & \cos(\omega_X) & -\sin(\omega_X) & Y'_0 \\ 0 & \sin(\omega_X) & \cos(\omega_X) & Z'_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

o bien

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_X & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

Sin embargo lo que interesa en este ejercicio es la transformación inversa $(X', Y', Z') \rightarrow (X, Y, Z)$, la que se obtiene de (2.21):

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_X^T & -\mathbf{R}_X^T \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix}$$

o bien en términos matriciales $\mathbf{M} = \mathbf{S}'\mathbf{M}'$, donde $\mathbf{M} = [X \ Y \ Z \ 1]^T$ y $\mathbf{M}' = [X' \ Y' \ Z' \ 1]^T$ y \mathbf{S}' la matriz 4×4 definida en la ecuación anterior.

iii) La proyección de \mathbf{M} en el plano R corresponde a la cámara *pinhole* mostrada en la Figura 3.2. Esta proyección se expresa matemáticamente en (3.4), es decir, $\lambda \mathbf{m} = \mathbf{P}\mathbf{M}$ con $\mathbf{m} = [x \ y \ 1]^T$, y \mathbf{P} la matriz de proyección perspectiva de la cámara:

$$\mathbf{P} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Finalmente se obtiene:

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{P}\mathbf{S}' \begin{bmatrix} b \sin \theta \\ 0 \\ a + b \cos \theta \\ 1 \end{bmatrix}.$$

Ejercicio 3.2 En un sistema óptico no lineal ocurre una distorsión como la que se muestra en la Figura 3.10. La relación entre puntos ideales (x, y) y reales (x', y') se muestra en la Tabla 3.1. a) ¿Existe distorsión tangencial? Justifique su respuesta. b) Proponga las funciones de distorsión $\delta_x(x, y)$ y $\delta_y(x, y)$ tal que cumplan:

$$x' = x + \delta_x(x, y) \quad , \quad y' = y + \delta_y(x, y)$$

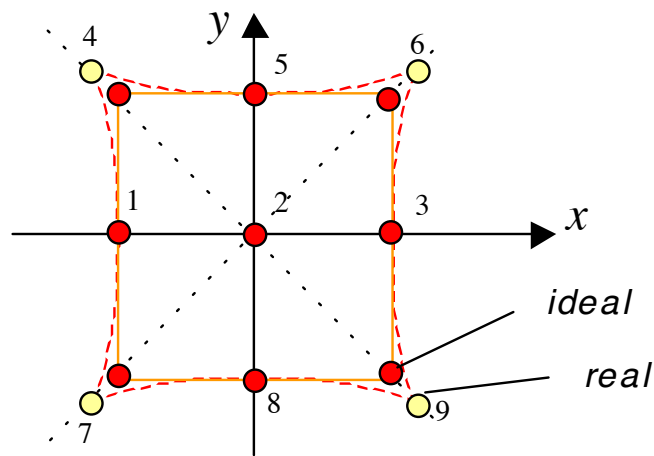


Figura 3.10: Distorsión no lineal (ver Ejercicio 3.2).

i	x	y	x'	y'
1	-1	0	-1	0
2	0	0	0	0
3	1	0	1	0
4	-1	1	-1.2	1.2
5	0	1	0	1
6	1	1	1.2	1.2
7	-1	-1	-1.2	-1.2
8	0	-1	0	-1
9	1	-1	1.2	-1.2

Tabla 3.1: Distorsión no lineal (ver Figura 3.10).

Solución: En los puntos presentados en la Figura 3.10 se observa que sólo hay distorsión radial (y no tangencial) ya que los puntos reales se obtienen desplazando los puntos ideales a lo largo de su radio.

La distorsión puede ser modelada de muchas maneras. Una de ellas se presenta a continuación en la que se asume solamente una distorsión radial $\delta_r(r)$. Utilizando (3.18) con $\delta_\phi(\phi) = 0$ se obtiene

$$\begin{aligned}\delta_x(x, y) &= \delta_r(r) \cos(\phi) \\ \delta_y(x, y) &= \delta_r(r) \sin(\phi)\end{aligned}\tag{3.21}$$

donde r, ϕ son las coordenadas polares de (x, y) , es decir $r = \sqrt{x^2 + y^2}$ y $\phi = \text{angle}(x, y)$. En la Tabla 3.1 se observa que no hay distorsión ($\delta_r(r) = 0$) para $r = 0$ (punto 2) y para $r = 1$ (puntos 1, 3, 5 y 8). La distorsión se presenta en los puntos 4, 6, 7 y 9, donde $r = \sqrt{2}$, en este caso $\delta_r(r)$ debe ser la distancia entre estos puntos ideales y sus respectivos puntos reales, es decir $(1, 2 - 1)\sqrt{2} = 0,2\sqrt{2}$. La función $\delta_r(r)$ puede tener la forma $\delta_r(r) = kr(r - 1)$, así se asegura que $\delta_r(r)$ sea cero para $r = 0$ y para $r = 1$. El factor k se obtiene evaluando $\delta_r(r)$ en $r = \sqrt{2}$, en este caso queda $\delta_r(\sqrt{2}) = k\sqrt{2}(\sqrt{2} - 1) = 0,2\sqrt{2}$. Despejando k se obtiene $k = 0,2/(\sqrt{2} - 1) = 0,4828$. La distorsión radial es entonces:

$$\delta_r(r) = 0,4828r(r - 1)$$

Sustituyendo esta función en (3.21) y considerando que $\cos(\phi) = x/r$ y $\sin(\phi) = y/r$, se obtiene

$$\begin{aligned}\delta_x(x, y) &= 0,4828x(\sqrt{x^2 + y^2} - 1) \\ \delta_y(x, y) &= 0,4828y(\sqrt{x^2 + y^2} - 1) \ .\end{aligned}$$

Capítulo 4

Visión Estéreo

El término *estéreo* en visión se utiliza cuando existe más de una vista de una escena. Estéreo, del griego $\sigma\tau\epsilon\rho\epsilon\omicron$, significa sólido, que en este caso se relaciona con la idea de tridimensionalidad. A través de varias imágenes de una escena, tomadas desde distintos puntos de vista, se puede tener la idea de las características tridimensionales de la escena en estudio.

En este capítulo se estudiarán las relaciones algebraicas y geométricas que existen cuando se ha tomado más de una vista de una escena. Se pondrá énfasis en el análisis de dos y tres vistas, geometría *bifocal* y *trifocal* respectivamente. Sin embargo, al final del capítulo se expondrá sólo a manera de introducción la geometría *quadrifocal* y de N vistas.

4.1. Análisis Bifocal

En el análisis bifocal se tiene un sistema de visión con dos cámaras, o bien una sola cámara que toma dos imágenes del objeto de estudio en dos tiempos distintos, suponiendo que en ese tiempo la cámara o el objeto se han movido. Para efectos de simplificación de la exposición del problema se estudiará la configuración de dos cámaras que toman al mismo tiempo una imagen del objeto de estudio. Sin embargo, con la teoría expuesta en este capítulo se puede deducir la solución al problema de dos vistas distintas con una sola cámara.

La geometría de dos vistas es conocida como la *Geometría Epipolar*. El

término epipolar viene del griego *epi* ($\epsilon\pi\iota$) que significa sobre, encima, y *polos* ($\pi\acute{o}\lambda\omicron\varsigma$) cuyo significado es punto de atracción o uno de los dos puntos de una esfera que son intersectados por su eje de rotación. La Geometría Epipolar lleva este nombre porque, como se verá más adelante, a cada una de las dos imágenes se le asocia un *epipolo*.

La geometría de dos vistas es presentada en la Figura 4.1. Un punto 3D M es visto en las dos imágenes como m_1 y m_2 (ver Figura 4.1a). Como se estudió en el capítulo anterior, la imagen es definida como la proyección del espacio 3D en un plano de imagen 2D por medio de un centro óptico. Los centros ópticos en este caso son C_1 y C_2 . A partir de m_1 solamente no se puede saber exactamente la ubicación exacta de M , ya que en el proceso de proyección se ha perdido la información de profundidad. Sin embargo, se

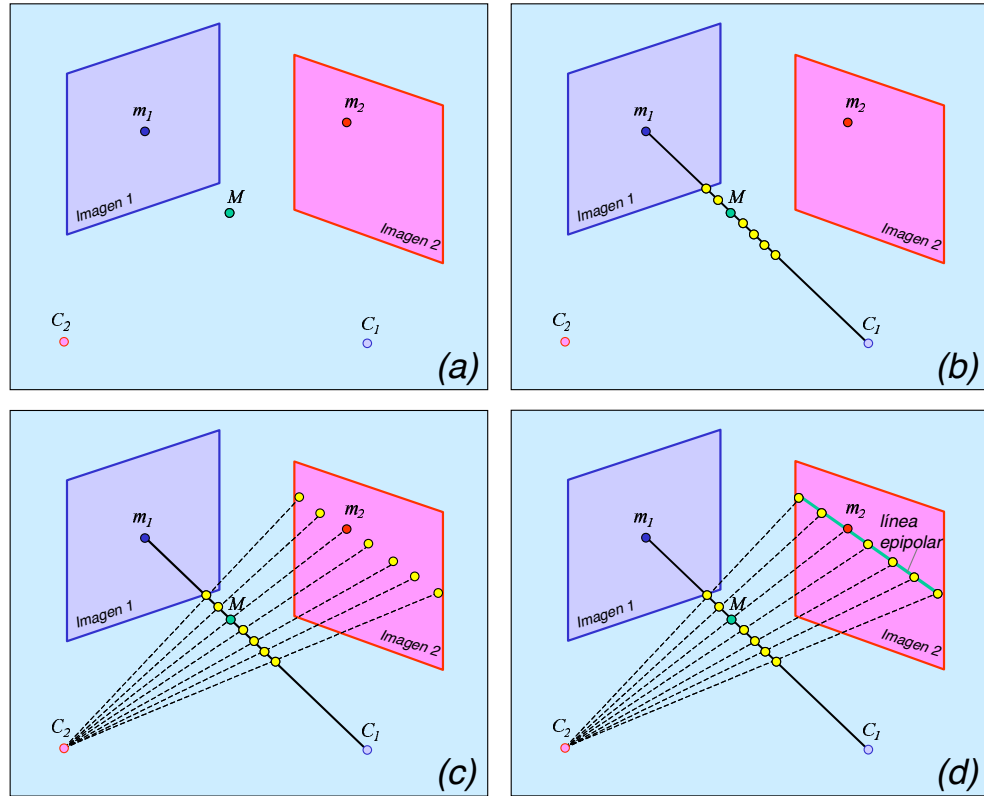


Figura 4.1: Geometría epipolar.

puede afirmar que M debe estar en el rayo que nace en el centro óptico C_1 para forma m_1 , es decir, M pertenece a la recta $\langle m_1, C_1 \rangle$. Esta situación es mostrada en la Figura 4.1b, donde varios puntos (M incluido) pertenecientes a la recta $\langle m_1, C_1 \rangle$ pueden ser los que forman el punto m_1 en la primera imagen. Si a partir de m_1 se desea conocer la ubicación de m_2 es necesario entonces proyectar en la imagen 2 los posibles puntos que pueden formar m_1 (ver Figura 4.1c). Se observa que m_2 es uno de estos puntos proyectados, sin embargo a partir de m_1 solamente no se puede saber la ubicación exacta de m_2 , sólo se puede afirmar que m_2 pertenece a la proyección de la recta $\langle m_1, C_1 \rangle$ realizada por el segundo centro óptico C_2 en la imagen 2. La proyección de esta recta, se denomina *línea epipolar* y se puede apreciar en la Figura 4.1d.

La *restricción epipolar* señala que para que m_1 y m_2 sean puntos correspondientes, el punto m_2 debe estar en la línea epipolar de m_1 . Esto no quiere decir que todos los puntos en la línea epipolar de m_1 son correspondientes a m_1 , ya que como bien se puede observar de la Figura 4.1 sólo un punto en la imagen 2 es correspondiente a m_1 , y en este caso es la proyección de M en la segunda imagen. La restricción epipolar es entonces una condición necesaria, mas no suficiente. A pesar de que no sea una condición suficiente, es de gran utilidad saber que el punto correspondiente a m_1 en la segunda imagen está sobre una línea y no está ubicado en cualquier parte de la imagen. Esto representa una reducción considerable en la dimensionalidad del problema de búsqueda de puntos correspondientes, ya que en vez de buscar en toda la imagen 2 (de dos dimensiones) se busca sólo a lo largo de una línea (una dimensión). A manera de ejemplo, si la segunda imagen tiene $N \times N$ píxels, la búsqueda de correspondencia se realiza sólo en N píxels de la imagen y no en N^2 píxels.

Una segunda representación de la Geometría Epipolar se aprecia en la Figura 4.2, en la que los planos de imagen están entre los centros ópticos y el punto 3D M . Al igual que en la representación anterior, las proyecciones de M son m_1 y m_2 en la primera y segunda imagen respectivamente. En esta configuración se observa también el mismo fenómeno: a partir de m_1 no se sabe exactamente dónde está ubicado el punto 3D M , sólo se sabe que se encuentra en algún punto de la recta que pasa por los puntos m_1 y C_1 . Los posibles puntos correspondientes a m_2 en la segunda imagen se obtienen entonces mediante la proyección de esta recta por el centro óptico C_2 en la segunda imagen. Esta recta en la imagen 2 es la línea epipolar l_2 .

De manera análoga, si se desea buscar los posibles puntos correspondientes

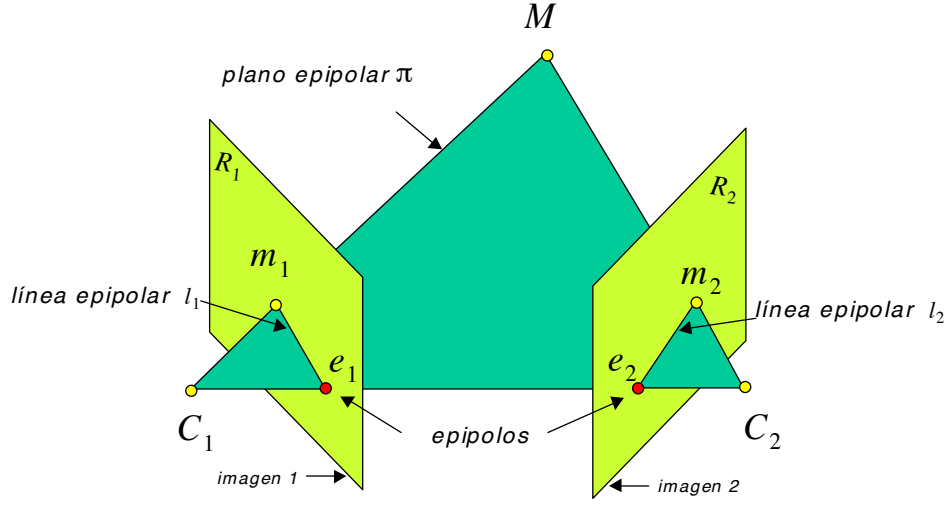


Figura 4.2: Líneas epipolares y epipolos.

a m_2 en la primera imagen se obtiene una recta epipolar l_1 definida como la proyección realizada por C_1 de la recta que contiene los puntos C_2 y m_2 en el plano de la primera imagen.

A continuación se define el *plano epipolar* π , como el plano que contiene los puntos C_1 , C_2 y M . Se observa que el plano epipolar contiene también los puntos m_1 y m_2 , y sus líneas epipolares l_1 y l_2 , las que se pueden definir entonces como las intersecciones del plano epipolar con los planos de imagen, es decir:

$$\begin{aligned} l_1 &= \pi \cap R_1 \\ l_2 &= \pi \cap R_2 \end{aligned} \quad (4.1)$$

Si se desea estudiar la Geometría Epipolar de un nuevo punto 3D M' (que no este en el plano π), se observa que en este sistema bifocal, en el que la ubicación de los planos de imagen (R_1 y R_2) y los centros ópticos (C_1 y C_2) no ha cambiado, existe un nuevo plano epipolar π' , como se muestra en la Figura 4.3. De acuerdo a la definición dada, π' contiene los puntos C_1 , C_2 y M' . Para este nuevo punto M' , existen las proyecciones m'_1 y m'_2 , definidas como las proyecciones de M' en las imágenes 1 y 2 respectivamente, y también existen sus líneas epipolares l'_1 y l'_2 , definidas como las intersecciones del plano epipolar π' con los planos de imagen R_1 y R_2 . Se observa que los planos π

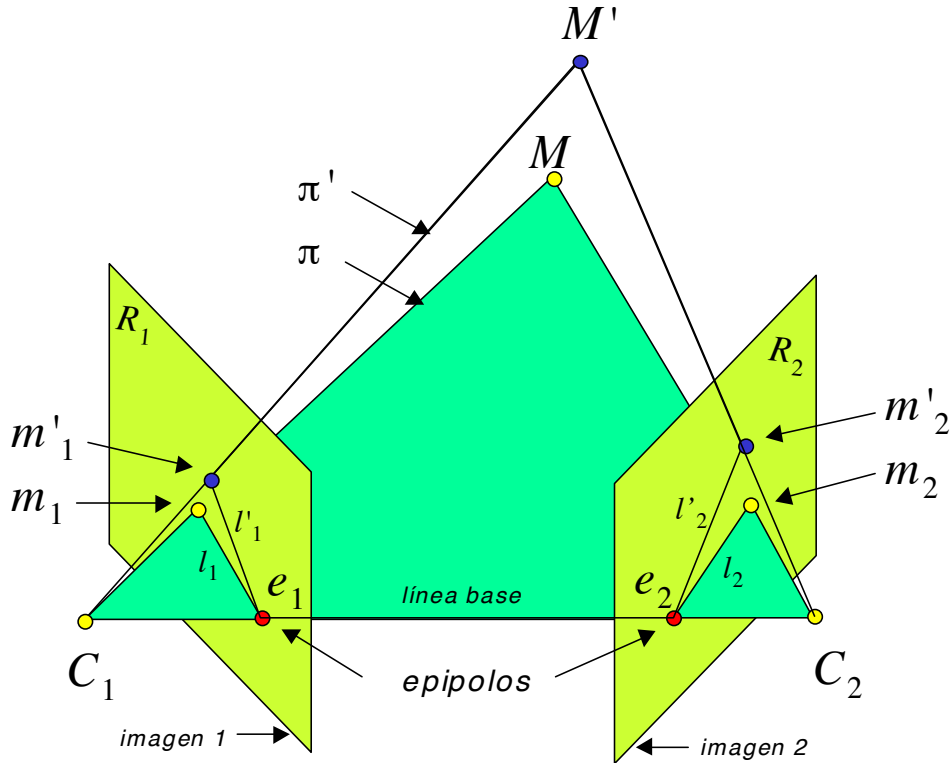


Figura 4.3: Planos epipolares.

y π' contienen no sólo los puntos C_1 y C_2 , sino que todos los puntos que están en la recta $\langle C_1, C_2 \rangle$, conocida como la *línea base*. De esta afirmación se puede deducir una propiedad muy importante de las líneas epipolares. Como las líneas epipolares se definen como la intersección de los planos epipolares con los planos de imagen, se obtiene entonces que todas las líneas epipolares en una imagen poseen un punto en común, conocido como el *epipolo*, definido como la intersección de la línea base con su plano de imagen:

$$\begin{aligned} e_1 &= \langle C_1, C_2 \rangle \cap R_1 \\ e_2 &= \langle C_1, C_2 \rangle \cap R_2 \end{aligned} \quad (4.2)$$

Ya que los epipolos son comunes a las líneas epipolares se deduce entonces

que

$$\begin{aligned} l_1 \cap l'_1 &= e_1 \\ l_2 \cap l'_2 &= e_2 \end{aligned}$$

4.1.1. Análisis geométrico de dos vistas

En el Capítulo 3 se logró establecer una transformación proyectiva de un punto 3D M a un punto 2D m . Este punto m se definió como la proyección M en el plano de imagen. Dependiendo del sistema de coordenadas en que están representados M y m se obtiene una ecuación como la presentada en (3.13). En términos generales, se puede afirmar que si la representación homogénea de M es $\mathbf{M} = [X \ Y \ Z \ 1]^T$ y de m es $\mathbf{m} = [x \ y \ 1]^T$ se puede escribir

$$\lambda \mathbf{m} = \mathbf{A} \mathbf{M} \quad (4.3)$$

donde \mathbf{A} , denominada la matriz de proyección general, es en una matriz de 3×4 elementos, encargada de convertir el punto 3D M en la proyección 2D m^1 .

Para dos vistas se tiene entonces el punto 3D M que es visto como m_1 y m_2 en la imagen 1 y 2 respectivamente. Como para cada imagen hay una matriz de proyección se obtiene el siguiente sistema de ecuaciones

$$\begin{cases} \lambda_1 \mathbf{m}_1 = \mathbf{A} \mathbf{M} \\ \lambda_2 \mathbf{m}_2 = \mathbf{B} \mathbf{M} \end{cases} \quad (4.4)$$

donde \mathbf{A} y \mathbf{B} son las matrices de proyección de las imágenes 1 y 2 respectivamente, y \mathbf{m}_1 y \mathbf{m}_2 son las representaciones homogéneas de m_1 y m_2 . Las coordenadas de \mathbf{M} en ambas ecuaciones están refeidas al mismo sistema de coordenadas.

A continuación se buscará una expresión matemática para l_2 , la línea epipolar de m_1 en la segunda imagen a partir de \mathbf{m}_1 , \mathbf{A} y \mathbf{B} . Como se mencionó en la introducción anterior, la línea epipolar l_2 es la proyección del rayo $\langle C_1, m_1 \rangle$ en la segunda imagen. Este rayo queda definido por dos puntos en el espacio 3D. El primero de ellos es C_1 cuyas representación homogénea utilizando las coordenadas en el sistema de coordenadas en que esta dado el punto 3D sería

¹Es necesario observar que en (3.13) la matriz \mathbf{A} equivale a \mathbf{KPS}' , sin embargo, los puntos 3D y 2D han sido representados como \mathbf{M}' y \mathbf{w} respectivamente.

\mathbf{C}_1 . Otro punto que está presente en el rayo $\langle C_1, m_1 \rangle$ es M sin embargo sus coordenadas son desconocidas. Las coordenadas de m_1 están dadas en un plano, no en el espacio 3D, sin embargo se puede calcular a partir de \mathbf{m}_1 un punto \mathbf{M}^+ que está en el rayo:

$$\mathbf{M}^+ = \mathbf{A}^+ \mathbf{m}_1 \quad (4.5)$$

donde \mathbf{A}^+ es la pseudo-inversa de \mathbf{A} . La pseudo-inversa de \mathbf{A} es una matriz que cumple con la siguiente propiedad:

$$\mathbf{A}\mathbf{A}^+ = \mathbf{I} \quad (4.6)$$

donde \mathbf{I} es una matriz identidad de 3×3 elementos. Debido a que \mathbf{A} es de 3×4 elementos la matriz \mathbf{A}^+ tiene que ser de 4×3 . Una expresión para la pseudo-inversa de \mathbf{A} es:

$$\mathbf{A}^+ = \mathbf{A}^T [\mathbf{A}^T \mathbf{A}]^{-1}. \quad \text{A}^+ \text{inv}(\mathbf{A}^* \mathbf{A}^*) \quad (4.7)$$

Es fácil comprobar que se cumple $\mathbf{A}\mathbf{A}^+ = \mathbf{I}$. Para demostrar que el punto \mathbf{M}^+ definido en (4.5) pertenece al rayo $\langle C_1, m_1 \rangle$ es necesario verificar si su proyección en la imagen 1 coincide con m_1 . Utilizando la primera ecuación de (4.4), la proyección de este punto sería:

$$\mathbf{A}\mathbf{M}^+ = \mathbf{A}\mathbf{A}^+ \mathbf{m}_1 = \mathbf{I}\mathbf{m}_1 = \mathbf{m}_1 \quad (4.8)$$

que como se observa coincide con la representación homogénea de m_1 . De esta manera se conocen dos puntos que pertenecen al rayo $\langle C_1, m_1 \rangle$: \mathbf{C}_1 y \mathbf{M}^+ . Por definición, la proyección de C_1 en la segunda imagen es e_2 , el epipolo de la imagen 2. La proyección de estos puntos en la segunda imagen serían entonces e_2 (proyección de C_1) y m_2^+ (proyección de M^+). Una representación homogénea de estos puntos se obtienen a partir de la segunda ecuación de (4.4):

$$\begin{cases} \mathbf{e}_2 &= \mathbf{B}\mathbf{C}_1 \\ \mathbf{m}_2^+ &= \mathbf{B}\mathbf{M}^+ \end{cases} \quad (4.9)$$

Si la recta epipolar l_2 contiene estos dos puntos, se puede decir entonces que su representación homogénea queda definida como:

$$\mathbf{l}_2 = \mathbf{e}_2 \times \mathbf{m}_2^+ = \mathbf{B}\mathbf{C}_1 \times \mathbf{B}\mathbf{M}^+ = \mathbf{B}\mathbf{C}_1 \times \mathbf{B}\mathbf{A}^+ \mathbf{m}_1 \quad (4.10)$$

A continuación se utilizará el concepto de *matriz antisimétrica* para encontrar una expresión más simple para \mathbf{l}_2 . Dados dos vectores \mathbf{u} y \mathbf{v} de tres elementos

cada uno, definiendo el vector \mathbf{w} como el producto cruz $\mathbf{u} \times \mathbf{v}$, se puede encontrar una matriz $[\mathbf{u}]_{\times}$, de 3×3 elementos, denominada la matriz antisimétrica de \mathbf{u} tal que:

$$\mathbf{w} = \mathbf{u} \times \mathbf{v} = [\mathbf{u}]_{\times} \mathbf{v} \quad (4.11)$$

Es fácil comprobar que si $\mathbf{u} = [u_1 \ u_2 \ u_3]^T$ la matriz antisimétrica de \mathbf{u} es:

$$[\mathbf{u}]_{\times} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \quad (4.12)$$

Utilizando la matriz antisimétrica de \mathbf{BC}_1 se obtiene una nueva expresión para \mathbf{l}_2 :

$$\mathbf{l}_2 = [\mathbf{BC}_1]_{\times} \mathbf{BA}^+ \mathbf{m}_1 \quad (4.13)$$

Definiendo la matriz \mathbf{F} de 3×3 elementos como:

$$\mathbf{F} = [\mathbf{BC}_1]_{\times} \mathbf{BA}^+ \quad (4.14)$$

se puede expresar la línea epipolar como

$$\mathbf{l}_2 = \mathbf{Fm}_1 \quad (4.15)$$

Si m_2 pertenece a esta recta entonces $\mathbf{m}_2^T \mathbf{l}_2 = 0$, o bien

$$\mathbf{m}_2^T \mathbf{Fm}_1 = 0 \quad (4.16)$$

La matriz \mathbf{F} es conocida como la *Matriz Fundamental* y es de gran importancia para el análisis de dos vistas, ya que \mathbf{F} es constante para una geometría bifocal dada, no depende de m_1 , m_2 ni M . La ecuación (4.16) es conocida como la *restricción epipolar* y señala que para que dos puntos m_1 y m_2 sean correspondientes, deben satisfacer (4.16).

Cabe mencionar que muchas veces las coordenadas de C_1 no se conocen, sin embargo a partir de la matriz de proyección \mathbf{A} es posible encontrar \mathbf{C}_1 . Se sabe que la proyección de C_1 en la imagen 1 no está definida, y que este es el único punto del espacio que no puede ser proyectado en el plano de imagen 1. Por lo tanto se puede señalar que el centro óptico debe satisfacer la siguiente ecuación $\mathbf{AC}_1 = [0 \ 0 \ 0]^T$, ya que el punto $[0 \ 0 \ 0]^T$ al tener su tercera componente igual a cero no está definido en el plano de imagen.

4.1.2. Propiedades de la Matriz Fundamental

La Matriz Fundamental \mathbf{F} tiene las siguientes propiedades

- i) Las representaciones homogéneas de las líneas epipolares l_1 y l_2 se definen como:

$$\begin{aligned} l_2 &= \mathbf{F}\mathbf{m}_1 \\ l_1 &= \mathbf{F}^\top \mathbf{m}_2 \end{aligned} \quad (4.17)$$

- ii) La restricción epipolar es

$$\mathbf{m}_2^\top \mathbf{F} \mathbf{m}_1 = 0 \quad (4.18)$$

- iii) La matriz \mathbf{F} es homogénea, ya que $k\mathbf{F}$ para $k \neq 0$ también puede ser utilizada en los cálculos anteriores.

- iv) El determinante de \mathbf{F} es cero, ya que

$$|\mathbf{F}| = |[\mathbf{e}_2]_\times \mathbf{B} \mathbf{A}^+| = |[\mathbf{e}_2]_\times| |\mathbf{B} \mathbf{A}^+| = 0 \quad (4.19)$$

La última igualdad se obtiene debido a que el determinante de una matriz antisimétrica es cero, como se puede deducir de (4.12).

- v) Como el determinante de \mathbf{F} es cero, y \mathbf{F} es homogénea se dice que \mathbf{F} tiene sólo siete grados de libertad, esto quiere decir que sólo siete (de los nueve) elementos de \mathbf{F} son linealmente independientes, los otros dos pueden ser calculados como función de los otros siete.
- vi) La matriz \mathbf{F} es constante para una geometría bifocal dada, no depende de m_1 , m_2 ni M , sólo depende de sus matrices de proyección \mathbf{A} y \mathbf{B} .
- vii) Los epipolos y la matriz Fundamental están relaciones de la siguiente manera:

$$\mathbf{F}\mathbf{e}_1 = \mathbf{0} \quad \text{y} \quad \mathbf{F}^\top \mathbf{e}_2 = \mathbf{0}, \quad (4.20)$$

siendo $\mathbf{0} = [0 \ 0 \ 0]^\top$. Estas ecuaciones sirven para calcular los epipolos, ya que se puede asumir que como \mathbf{e}_1 y \mathbf{e}_2 son representaciones homogéneas, su tercera componente es uno. La relación anterior se puede deducir a partir de la condición epipolar: si se tiene un punto m_1 cualquiera en la imagen 1, se sabe que su línea epipolar en la imagen 2 pasa por el epipolo e_2 , esto quiere decir que se cumple $\mathbf{e}_2^\top \mathbf{F} \mathbf{m}_1 = 0$.

Como esta condición se cumple siempre para cualquier m_1 entonces se puede afirmar que $\mathbf{e}_2^T \mathbf{F} = [0 \ 0 \ 0]$, o bien $\mathbf{F}^T \mathbf{e}_2 = \mathbf{0}$. El mismo razonamiento se puede hacer para el epipolo e_1 , con lo que se obtiene $\mathbf{F} \mathbf{e}_1 = \mathbf{0}$.

4.1.3. Análisis algebraico de dos vistas

El problema de correspondencia en dos vistas se puede resolver algebraicamente utilizando los *tensores bifocales* [20, 25]. A continuación se presenta detalladamente el análisis algebraico de dos vistas.

Las proyecciones de un punto 3D M en dos planos de imagen, imagen 1 e imagen 2, m_1 y m_2 respectivamente, tal como se aprecia en la Figura 4.1, se pueden calcular por medio de la ecuación general de proyección (4.3) utilizando la matriz de proyección \mathbf{A} para la imagen 1 y la matriz de proyección \mathbf{B} para la segunda:

$$\begin{cases} \lambda_1 \mathbf{m}_1 = \mathbf{A} \mathbf{M} \\ \lambda_2 \mathbf{m}_2 = \mathbf{B} \mathbf{M} \end{cases} \quad (4.21)$$

o bien haciendo una transformación de coordenadas para \mathbf{M}

$$\begin{cases} \lambda_1 \mathbf{m}_1 = [\mathbf{I} \mid \mathbf{0}] \tilde{\mathbf{M}} = \tilde{\mathbf{A}} \tilde{\mathbf{M}} \\ \lambda_2 \mathbf{m}_2 = \tilde{\mathbf{B}} \tilde{\mathbf{M}} \end{cases} \quad (4.22)$$

donde

$$\begin{aligned} \mathbf{M} &= \mathbf{H}^{-1} \tilde{\mathbf{M}} \\ \tilde{\mathbf{A}} &= \mathbf{A} \mathbf{H}^{-1} = [\mathbf{I} \mid \mathbf{0}] \\ \tilde{\mathbf{B}} &= \mathbf{B} \mathbf{H}^{-1} \end{aligned} \quad (4.23)$$

La matriz \mathbf{H} , de 4×4 elementos, es una matriz regular cuyas tres primeras filas corresponden a la matriz \mathbf{A} [19]. Suponiendo que la cuarta fila de \mathbf{H} es \mathbf{h} se obtiene:

$$\mathbf{H} \mathbf{H}^{-1} = \begin{bmatrix} \mathbf{A} \\ \mathbf{h} \end{bmatrix} \mathbf{H}^{-1} = \begin{bmatrix} \mathbf{A} \mathbf{H}^{-1} \\ \mathbf{h} \mathbf{H}^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

La última igualdad se obtiene sabiendo que \mathbf{H} es regular, entonces $\mathbf{H} \mathbf{H}^{-1}$ es una matriz identidad de 4×4 . Como $\mathbf{A} \mathbf{H}^{-1}$ corresponde a las primeras

tres filas de este resultado se dice que $\mathbf{A}\mathbf{H}^{-1} = [\mathbf{I} \mid \mathbf{0}]$, donde \mathbf{I} es una matriz identidad de 3×3 y $\mathbf{0} = [0 \ 0 \ 0]^T$.

Mediante la matriz \mathbf{H} se hace una transformación del sistema de coordenadas en el cual se había representado el punto M . Se trata de una transformación proyectiva 3D no Euclídea. En este nuevo sistema de coordenadas, las coordenadas de M ahora son representadas homogéneamente como $\tilde{\mathbf{M}}$. De esta manera se obtiene una matriz de proyección normalizada para la primera imagen del tipo $\tilde{\mathbf{A}} = [\mathbf{I} \mid \mathbf{0}]$.

Reformulando (4.22) se puede escribir el sistema de ecuaciones:

$$\underbrace{\begin{bmatrix} \tilde{\mathbf{a}}_1 & x_1 & 0 \\ \tilde{\mathbf{a}}_2 & y_1 & 0 \\ \tilde{\mathbf{a}}_3 & 1 & 0 \\ \tilde{\mathbf{b}}_1 & 0 & x_2 \\ \tilde{\mathbf{b}}_2 & 0 & y_2 \\ \tilde{\mathbf{b}}_3 & 0 & 1 \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \tilde{\mathbf{M}} \\ -\lambda_1 \\ -\lambda_2 \end{bmatrix}}_{\mathbf{v}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (4.24)$$

donde $\tilde{\mathbf{a}}_i$ y $\tilde{\mathbf{b}}_i$ corresponden a la fila i de la matriz $\tilde{\mathbf{A}}$ y $\tilde{\mathbf{B}}$ respectivamente, para $i = 1, 2, 3$.

Bajo la hipótesis de que m_1 y m_2 son puntos correspondientes, es decir que ambos son proyecciones de un mismo punto 3D, existe entonces un punto M único. En este caso el sistema de ecuaciones (4.24) tiene una solución no trivial para \mathbf{v} . Por lo tanto se puede afirmar que bajo esta hipótesis de correspondencia $\mathbf{v} \neq \mathbf{0}$. Se observa que la matriz \mathbf{G} es de 6×6 elementos, por lo tanto una condición necesaria y suficiente para la existencia de una solución no trivial de \mathbf{v} es que el rango de \mathbf{G} sea 5, o bien, que el determinante de \mathbf{G} sea igual a cero. Es decir

$$|\mathbf{G}| = 0. \quad (4.25)$$

El determinante de \mathbf{G} se puede obtener por medio de la fórmula de Laplace [3], en la que $|\mathbf{G}|$ se expande como una sumatoria de los elementos de \mathbf{G} de una fila o columna multiplicados por sus respectivos cofactores, lo cual resulta muy conveniente en matrices que tienen muchos elementos iguales a cero. Expandiendo $|\mathbf{G}|$ a través de la quinta columna en la que están los

elementos x_1 , y_1 y 1 (no hay que olvidar que los vectores fila $\tilde{\mathbf{a}}_i$ y $\tilde{\mathbf{b}}_i$ tienen cuatro elementos) se obtiene:

$$|\mathbf{G}| = x_1 \begin{vmatrix} \tilde{\mathbf{a}}_2 & 0 \\ \tilde{\mathbf{a}}_3 & 0 \\ \tilde{\mathbf{b}}_1 & x_2 \\ \tilde{\mathbf{b}}_2 & y_2 \\ \tilde{\mathbf{b}}_3 & 1 \end{vmatrix} - y_1 \begin{vmatrix} \tilde{\mathbf{a}}_1 & 0 \\ \tilde{\mathbf{a}}_3 & 0 \\ \tilde{\mathbf{b}}_1 & x_2 \\ \tilde{\mathbf{b}}_2 & y_2 \\ \tilde{\mathbf{b}}_3 & 1 \end{vmatrix} + \begin{vmatrix} \tilde{\mathbf{a}}_1 & 0 \\ \tilde{\mathbf{a}}_2 & 0 \\ \tilde{\mathbf{b}}_1 & x_2 \\ \tilde{\mathbf{b}}_2 & y_2 \\ \tilde{\mathbf{b}}_3 & 1 \end{vmatrix} = 0.$$

A continuación se emplea nuevamente la formula de Laplace para expandir cada uno de los tres determinantes de 5×5 elementos presentes en la ecuación anterior. En este caso se emplea la columna en la que están x_2 , y_2 y 1, obteniéndose así:

$$|\mathbf{G}| = x_1 x_2 \begin{vmatrix} \tilde{\mathbf{a}}_2 \\ \tilde{\mathbf{a}}_3 \\ \tilde{\mathbf{b}}_2 \\ \tilde{\mathbf{b}}_3 \end{vmatrix} - y_1 x_2 \begin{vmatrix} \tilde{\mathbf{a}}_1 \\ \tilde{\mathbf{a}}_3 \\ \tilde{\mathbf{b}}_2 \\ \tilde{\mathbf{b}}_3 \end{vmatrix} + x_2 \begin{vmatrix} \tilde{\mathbf{a}}_1 \\ \tilde{\mathbf{a}}_2 \\ \tilde{\mathbf{b}}_2 \\ \tilde{\mathbf{b}}_3 \end{vmatrix} +$$

$$-x_1 y_2 \begin{vmatrix} \tilde{\mathbf{a}}_2 \\ \tilde{\mathbf{a}}_3 \\ \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_3 \end{vmatrix} + y_1 y_2 \begin{vmatrix} \tilde{\mathbf{a}}_1 \\ \tilde{\mathbf{a}}_3 \\ \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_3 \end{vmatrix} - y_2 \begin{vmatrix} \tilde{\mathbf{a}}_1 \\ \tilde{\mathbf{a}}_2 \\ \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_3 \end{vmatrix} + x_1 \begin{vmatrix} \tilde{\mathbf{a}}_2 \\ \tilde{\mathbf{a}}_3 \\ \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_2 \end{vmatrix} - y_1 \begin{vmatrix} \tilde{\mathbf{a}}_1 \\ \tilde{\mathbf{a}}_3 \\ \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_2 \end{vmatrix} + \begin{vmatrix} \tilde{\mathbf{a}}_1 \\ \tilde{\mathbf{a}}_2 \\ \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_2 \end{vmatrix} = 0.$$

Esta ecuación puede ser escrita de la siguiente manera:

$$|\mathbf{G}| = [x_2 \ y_2 \ 1] \underbrace{\begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \mathbf{m}_2^T \mathbf{F} \mathbf{m}_1 = 0, \quad (4.26)$$

donde

$$F_{ij} = (-1)^{i+j} \begin{vmatrix} \sim \tilde{\mathbf{a}}_j \\ \sim \tilde{\mathbf{b}}_i \end{vmatrix} \quad \text{para } i, j = 1, 2, 3. \quad (4.27)$$

Los términos $\sim \tilde{\mathbf{a}}_j$ y $\sim \tilde{\mathbf{b}}_i$ significan respectivamente las matrices $\tilde{\mathbf{A}}$ sin la fila j y la matriz $\tilde{\mathbf{B}}$ sin la fila i .

La matriz \mathbf{F} , de 3×3 elementos, es la conocida *Matriz Fundamental* de la Geometría Epipolar. Los elementos F_{ij} se denominan los *Tensores Bifocales* [20].

La ecuación (4.26) se puede escribir de manera tensorial utilizando la convención de Einstein para la suma tensorial²:

$$|\mathbf{G}| = m_1^j m_2^i F_{ij} = 0, \quad (4.28)$$

donde $[m_1^1 \ m_1^2 \ m_1^3] = [x_1 \ y_1 \ 1] = \mathbf{m}_1^T$ y $[m_2^1 \ m_2^2 \ m_2^3] = [x_2 \ y_2 \ 1] = \mathbf{m}_2^T$.

Para la forma canónica de la matriz de proyección de la primera imagen, $\tilde{\mathbf{A}} = [\mathbf{I} \mid \mathbf{0}]$, los tensores bifocales calculados a partir de (4.27) son:

$$F_{ij} = \tilde{b}_{i \oplus 1, j} \tilde{b}_{i \oplus 2, 4} - \tilde{b}_{i \oplus 2, j} \tilde{b}_{i \oplus 1, 4}, \quad (4.29)$$

donde

$$i \oplus k = \begin{cases} i + k & \text{si } i + k \leq 3 \\ i + k - 3 & \text{en caso contrario} \end{cases}.$$

La ecuación (4.26), o bien la ecuación (4.28), expresan matemáticamente la ya mencionada restricción epipolar: para que m_1 y m_2 sean correspondientes, m_2 debe estar en la línea epipolar de m_1 en la segunda imagen. Se observa que la línea epipolar l está dada entonces por:

$$\mathbf{m}_2^T l = l_1 x_2 + l_2 y_2 + l_3 = 0 \quad (4.30)$$

donde $l = [l_1 \ l_2 \ l_3]^T = \mathbf{F} \mathbf{m}_1$, o bien en la forma tensorial:

$$l_i = m_1^j F_{ij}. \quad (4.31)$$

Es de suma importancia hacer notar que la Matriz Fundamental \mathbf{F} es independiente de las coordenadas de m_1 , m_2 y M . La Matriz Fundamental es definida como una función de las matrices de proyección A y B , esto quiere decir que \mathbf{F} es una función de la ubicación de los planos de proyección y de los centros ópticos de ambas imágenes.

4.1.4. Restricción bifocal práctica

En la práctica debido a errores en la medición y calibración, dos puntos correspondientes m_1 y m_2 satisfacen la condición epipolar con una probabilidad muy baja, ya que m_2 no está exactamente sobre la línea epipolar l ,

²La convención de Einstein para la suma de tensores indica que dos tensores que tienen el mismo índice deben desglosarse, multiplicándose y sumándose de la siguiente manera $\alpha^i \beta_i = \alpha_1 \beta_1 + \dots + \alpha_n \beta_n$, siendo n el número de elementos de cada tensor.

sino que está muy cerca. Por esta razón es necesario utilizar otro criterio de correspondencia. En la práctica se dice que m_1 y m_2 pueden ser puntos correspondientes si la distancia mínima de m_2 a l es menor que una distancia d_0 . Esta distancia se calcula a partir de una línea perpendicular a l que pase por m_2 (ver Ejercicio 4.2). De esta manera, se obtiene que la restricción epipolar *práctica* se expresa como [31]:

$$d = \frac{|\mathbf{m}_2^T \mathbf{F} \mathbf{m}_1|}{\sqrt{l_1^2 + l_2^2}} < d_0. \quad (4.32)$$

4.2. Análisis Trifocal

En el caso de tener tres vistas de una misma escena, se estudiará si los puntos de proyección m_1 , m_2 y m_3 en las imágenes 1, 2 y 3 respectivamente, son puntos correspondientes, es decir si los tres puntos son proyecciones de un mismo punto 3D M . Basándose en la geometría epipolar, se puede afirmar que si se calcula la línea epipolar de m_1 y la línea epipolar de m_2 en la tercera imagen, m_3 debe estar en la intersección de ambas líneas, ya que si m_1 y m_3 son correspondientes m_3 debe estar en la línea epipolar de m_1 en la tercera imagen y Como la misma deducción se puede hacer para m_2 ,

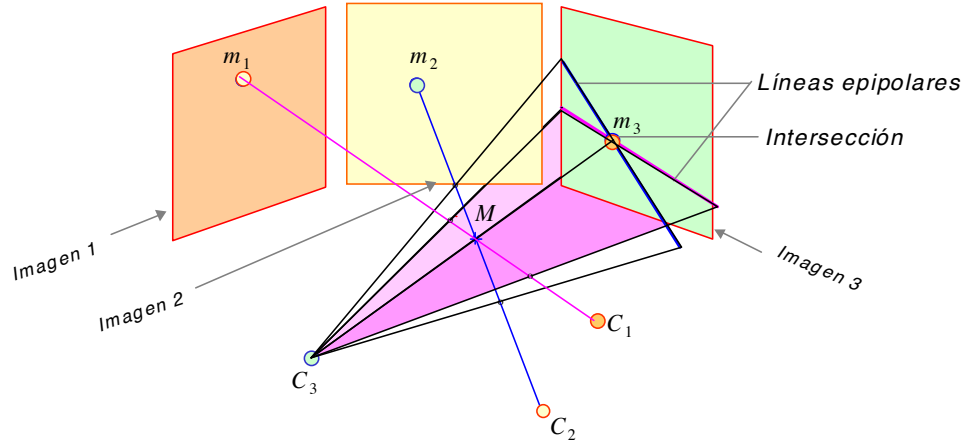


Figura 4.4: Geometría Epipolar para tres vistas.

entonces m_3 debe pertenecer a ambas líneas epipolares, es decir m_3 es el punto de intersección de las líneas, tal como se ilustra en la Figura 4.4.

La Geometría Epipolar en tres imágenes señala entonces que m_1 , m_2 y m_3 son puntos correspondientes si m_3 coincide con el punto de intersección de las líneas epipolares de m_1 y m_2 en la tercera imagen [12]. Esta es una condición necesaria y suficiente.

Sin embargo, el punto de intersección no está definido si ambas líneas epipolares son iguales. Lamentablemente esta situación no es poco común. Ambas líneas son iguales cuando los planos epipolares Π_{13} y Π_{23} , definidos como los planos que contienen M , C_1 y C_3 , y M , C_2 y C_3 respectivamente, son iguales. Esto sucede en dos ocasiones: *i*) cuando los tres centros ópticos C_1 , C_2 y C_3 son colineales; o bien *ii*) cuando los tres centros ópticos C_1 , C_2 y C_3 no son colineales y m_1 , m_2 y m_3 se encuentran sobre el plano definido por los tres centros ópticos [10, 44]. La primera de ellas ocurre en la práctica muy frecuentemente, ya que se obtiene al tomar tres imágenes con una misma cámara que se mueve en línea recta.

Además de las dos desventajas mencionadas para el uso de la Geometría Epipolar en tres vistas, hay que señalar que la Geometría Epipolar no proporciona un método directo para analizar la correspondencia de tres puntos, ya que es necesario calcular dos líneas epipolares y luego su intersección.

4.2.1. Análisis algebraico de la geometría trifocal

Una forma de estudiar el problema de la geometría trifocal es por medio de los *tensores trifocales* [21, 45], que serán presentados a continuación. Mediante los tensores trifocales se puede por una parte evitar las singularidades indicadas anteriormente y por otra parte obtener una solución directa para la correspondencia en tres vistas.

Las tres proyecciones de un punto 3D M en las imágenes 1, 2 y 3, pueden expresarse matemáticamente, como se hizo en la Sección 4.1.3, a partir de la ecuación general de proyección (4.3) utilizando las matrices de proyección \mathbf{A} , \mathbf{B} y \mathbf{C} . La forma canónica de estas ecuaciones es:

$$\begin{cases} \lambda_1 \mathbf{m}_1 = [\mathbf{I} \mid \mathbf{0}] \tilde{\mathbf{M}} = \tilde{\mathbf{A}} \tilde{\mathbf{M}} \\ \lambda_2 \mathbf{m}_2 = \tilde{\mathbf{B}} \tilde{\mathbf{M}} \\ \lambda_3 \mathbf{m}_3 = \tilde{\mathbf{C}} \tilde{\mathbf{M}} \end{cases}, \quad (4.33)$$

donde $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{H}^{-1}$. Las entidades $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, $\tilde{\mathbf{M}}$ fueron definidas en (4.23).

Se sabe que si m_1 , m_2 y m_3 son puntos correspondientes entonces debe existir una solución para M . Una solución conocida al problema de establecer la correspondencia se obtiene al reformular el sistema de ecuaciones (4.33) de la siguiente manera:

$$\underbrace{\begin{bmatrix} \tilde{\mathbf{a}}_1 & x_1 & 0 & 0 \\ \tilde{\mathbf{a}}_2 & y_1 & 0 & 0 \\ \tilde{\mathbf{a}}_3 & 1 & 0 & 0 \\ \tilde{\mathbf{b}}_1 & 0 & x_2 & 0 \\ \tilde{\mathbf{b}}_2 & 0 & y_2 & 0 \\ \tilde{\mathbf{b}}_3 & 0 & 1 & 0 \\ \tilde{\mathbf{c}}_1 & 0 & 0 & x_3 \\ \tilde{\mathbf{c}}_2 & 0 & 0 & y_3 \\ \tilde{\mathbf{c}}_3 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \tilde{\mathbf{M}} \\ -\lambda_1 \\ -\lambda_2 \\ -\lambda_3 \end{bmatrix}}_{\mathbf{v}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (4.34)$$

donde $\tilde{\mathbf{a}}_i$, $\tilde{\mathbf{b}}_i$ y $\tilde{\mathbf{c}}_i$ son respectivamente la fila i de las matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ y $\tilde{\mathbf{C}}$. Planteando la hipótesis de correspondencia, se puede afirmar que si m_1 , m_2 y m_3 son puntos correspondientes, entonces debe existir una solución no trivial para \mathbf{v} . Cabe destacar que \mathbf{G} es una matriz de 9×7 , es decir su determinante no está definido. Sin embargo si se escogen 7 cualesquiera de las 9 ecuaciones del sistema (4.34) se obtiene un nuevo sistema de ecuaciones cuya representación matricial es $\mathbf{G}_7 \mathbf{v} = \mathbf{0}$. Si $\mathbf{v} \neq \mathbf{0}$ entonces el determinante de \mathbf{G}_7 debe ser cero. Esto quiere decir que para que exista una solución no trivial para \mathbf{v} todas las submatrices de \mathbf{G} formadas a partir de 7 de sus filas, debe ser cero.

El desarrollo de los *subdeterminantes* de \mathbf{G} a partir de la fórmula de Laplace, como se hizo en la Sección 4.1.3, lleva a expresiones matemáticas que dependen de las coordenadas de los puntos m_1 , m_2 y m_3 y valores constantes para las tres imágenes que dependen sólo de las tres matrices de proyección. Estos valores constantes son los denominados *tensores trifocales* [20, 25, 26].

En este caso existen 36 posibles submatrices de \mathbf{G} obtenidas a partir de la eliminación de dos de sus filas. Estas submatrices se pueden dividir en dos tipos, aquellas que tienen sólo una fila de una matriz de proyección (9 casos) y aquellas que tienen las tres filas de una matriz de proyección (27 casos).

A manera de ejemplo, en el primer tipo de submatrices, se obtiene para el subdeterminante de la matriz \mathbf{G} en la que se han eliminado las filas 2 y 3:

$$\begin{vmatrix} \tilde{\mathbf{a}}_1 & x_1 & 0 & 0 \\ \tilde{\mathbf{b}}_1 & 0 & x_2 & 0 \\ \tilde{\mathbf{b}}_2 & 0 & y_2 & 0 \\ \tilde{\mathbf{b}}_3 & 0 & 1 & 0 \\ \tilde{\mathbf{c}}_1 & 0 & 0 & x_3 \\ \tilde{\mathbf{c}}_2 & 0 & 0 & y_3 \\ \tilde{\mathbf{c}}_3 & 0 & 0 & 1 \end{vmatrix} = x_1 \begin{vmatrix} \tilde{\mathbf{b}}_1 & x_2 & 0 \\ \tilde{\mathbf{b}}_2 & y_2 & 0 \\ \tilde{\mathbf{b}}_3 & 1 & 0 \\ \tilde{\mathbf{c}}_1 & 0 & x_3 \\ \tilde{\mathbf{c}}_2 & 0 & y_3 \\ \tilde{\mathbf{c}}_3 & 0 & 1 \end{vmatrix} = 0.$$

Esto lleva a que el segundo determinante es cero, ya que la solución $x_1 = 0$ es trivial. Se observa que este determinante no contiene información de la primera cámara, y corresponde al análisis bifocal de las imágenes 2 y 3. La expansión de este determinante es $\mathbf{m}_3^T \mathbf{F}_{23} \mathbf{m}_2 = 0$, donde \mathbf{F}_{23} es la matriz Fundamental de las imágenes 2 y 3. Como esta expresión no es trifocal, no interesa para el análisis de tres vistas.

El segundo caso de submatrices, en las que están presentes las tres filas de una matriz de proyección resulta más interesante para el análisis trifocal. A manera de ejemplo, un subdeterminante en el que están presentes todas las filas de la primera matriz de proyección sería:

$$\begin{vmatrix} \tilde{\mathbf{a}}_1 & x_1 & 0 & 0 \\ \tilde{\mathbf{a}}_2 & y_1 & 0 & 0 \\ \tilde{\mathbf{a}}_3 & 1 & 0 & 0 \\ \tilde{\mathbf{b}}_1 & 0 & x_2 & 0 \\ \tilde{\mathbf{b}}_3 & 0 & 1 & 0 \\ \tilde{\mathbf{c}}_1 & 0 & 0 & x_3 \\ \tilde{\mathbf{c}}_3 & 0 & 0 & 1 \end{vmatrix} = 0.$$

En este ejemplo se han eliminado las filas de \mathbf{G} en las que están las segundas filas de $\tilde{\mathbf{B}}$ y $\tilde{\mathbf{C}}$.

Para el caso de submatrices en los que se mantienen las tres filas de una de las matrices de proyección se obtienen tres subcasos distintos, uno para cada matriz de proyección seleccionada. Para cada subcaso existen 9 subdeterminantes posibles de los cuales sólo cuatro son linealmente independiente

del resto. Esto quiere decir que si estos cuatro subdeterminantes son cero, el resto de subdeterminantes también será cero [23]. Para el caso de escoger la primera matriz de proyección los cuatro subdeterminantes son:

$$\begin{aligned}
 D_1 = \begin{vmatrix} \tilde{\mathbf{a}}_1 & x_1 & 0 & 0 \\ \tilde{\mathbf{a}}_2 & x_2 & 0 & 0 \\ \tilde{\mathbf{a}}_3 & 1 & 0 & 0 \\ \tilde{\mathbf{b}}_1 & 0 & x_2 & 0 \\ \tilde{\mathbf{b}}_3 & 0 & 1 & 0 \\ \tilde{\mathbf{c}}_1 & 0 & 0 & x_3 \\ \tilde{\mathbf{c}}_3 & 0 & 0 & 1 \end{vmatrix} = 0 \quad D_2 = \begin{vmatrix} \tilde{\mathbf{a}}_1 & x_1 & 0 & 0 \\ \tilde{\mathbf{a}}_2 & y_1 & 0 & 0 \\ \tilde{\mathbf{a}}_3 & 1 & 0 & 0 \\ \tilde{\mathbf{b}}_1 & 0 & x_2 & 0 \\ \tilde{\mathbf{b}}_3 & 0 & 1 & 0 \\ \tilde{\mathbf{c}}_2 & 0 & 0 & y_3 \\ \tilde{\mathbf{c}}_3 & 0 & 0 & 1 \end{vmatrix} = 0 \\
 D_3 = \begin{vmatrix} \tilde{\mathbf{a}}_1 & x_1 & 0 & 0 \\ \tilde{\mathbf{a}}_2 & y_1 & 0 & 0 \\ \tilde{\mathbf{a}}_3 & 1 & 0 & 0 \\ \tilde{\mathbf{b}}_2 & 0 & y_2 & 0 \\ \tilde{\mathbf{b}}_3 & 0 & 1 & 0 \\ \tilde{\mathbf{c}}_1 & 0 & 0 & x_3 \\ \tilde{\mathbf{c}}_3 & 0 & 0 & 1 \end{vmatrix} = 0 \quad D_4 = \begin{vmatrix} \tilde{\mathbf{a}}_1 & x_1 & 0 & 0 \\ \tilde{\mathbf{a}}_2 & y_1 & 0 & 0 \\ \tilde{\mathbf{a}}_3 & 1 & 0 & 0 \\ \tilde{\mathbf{b}}_2 & 0 & y_2 & 0 \\ \tilde{\mathbf{b}}_3 & 0 & 1 & 0 \\ \tilde{\mathbf{c}}_2 & 0 & 0 & y_3 \\ \tilde{\mathbf{c}}_3 & 0 & 0 & 1 \end{vmatrix} = 0
 \end{aligned}$$

Expandiendo estos determinantes por medio de la fórmula de Laplace que debe ser utilizada tres veces (una vez para la columna de m_1 , otra vez para la columna de m_2 y otra para la columna de m_3) se obtiene la expresión:

$$\begin{cases} D_1 = \mathbf{m}_1^\top (x_3 \mathbf{T}^{13} - x_3 x_2 \mathbf{T}^{33} + x_2 \mathbf{T}^{31} - \mathbf{T}^{11}) = 0 \\ D_2 = \mathbf{m}_1^\top (y_3 \mathbf{T}^{13} - y_3 x_2 \mathbf{T}^{33} + x_2 \mathbf{T}^{32} - \mathbf{T}^{12}) = 0 \\ D_3 = \mathbf{m}_1^\top (x_3 \mathbf{T}^{23} - x_3 y_2 \mathbf{T}^{33} + y_2 \mathbf{T}^{31} - \mathbf{T}^{21}) = 0 \\ D_4 = \mathbf{m}_1^\top (y_3 \mathbf{T}^{23} - y_3 y_2 \mathbf{T}^{33} + y_2 \mathbf{T}^{32} - \mathbf{T}^{22}) = 0 \end{cases}, \quad (4.35)$$

donde

$$\mathbf{T}^{jk} = [T_1^{jk} \ T_2^{jk} \ T_3^{jk}]^\top,$$

y

$$T_i^{jk} = (-1)^{i+1} \begin{vmatrix} \sim \tilde{\mathbf{a}}_i \\ \tilde{\mathbf{b}}_j \\ \tilde{\mathbf{c}}_k \end{vmatrix} = \tilde{b}_{ji} \tilde{c}_{k4} - \tilde{b}_{j4} \tilde{c}_{ki}, \quad \text{para } i, j, k = 1, 2, 3 \quad (4.36)$$

Los elementos T_i^{jk} son los denominados *tensores trifocales* para las imágenes 1, 2 y 3.

La ecuación (4.35), conocida como las *Trilinearidades* de Shashua [46], es de suma importancia en el análisis trifocal ya que establece una relación lineal entre las coordenadas de los puntos m_1 , m_2 y m_3 para establecer su correspondencia. Si se cumplen las cuatro trilinearidades se dice entonces que m_1 , m_2 y m_3 son correspondientes (condición necesaria y suficiente). Como se observa, los tensores trifocales son dependientes exclusivamente de las matrices de proyección, no dependen de los puntos m_1 , m_2 , m_3 y M .

4.2.2. Deducción alternativa de los tensores trifocales

A continuación se presentará otra forma de deducir las trilinearidades. Se puede obtener directamente del sistema de ecuaciones (4.33) una relación trifocal de una manera más simple [19, 21]: A partir de la primera ecuación $\lambda_1 \mathbf{m}_1 = [\mathbf{I} \mid \mathbf{0}] \tilde{\mathbf{M}}$:

$$\lambda_1 \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \end{bmatrix} \quad (4.37)$$

se obtiene una expresión para $\tilde{\mathbf{M}}$:

$$\tilde{\mathbf{M}} = \begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ 1 \end{bmatrix} = \lambda_1 \begin{bmatrix} x_1 \\ y_1 \\ 1 \\ 1/\lambda_1 \end{bmatrix}. \quad (4.38)$$

El cálculo de la proyección de este punto en la segunda imagen sería entonces a partir de (4.33)

$$\lambda_2 \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \lambda_1 \begin{bmatrix} \tilde{b}_{11} & \tilde{b}_{12} & \tilde{b}_{13} & \tilde{b}_{14} \\ \tilde{b}_{21} & \tilde{b}_{22} & \tilde{b}_{23} & \tilde{b}_{24} \\ \tilde{b}_{31} & \tilde{b}_{32} & \tilde{b}_{33} & \tilde{b}_{34} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \\ 1/\lambda_1 \end{bmatrix}. \quad (4.39)$$

Sustituyendo la tercera ecuación de (4.39) en las primeras dos quedaría:

$$\begin{cases} x_2(\tilde{b}_{31}x_1 + \tilde{b}_{32}y_1 + \tilde{b}_{33} + \tilde{b}_{34}/\lambda_1) &= (\tilde{b}_{11}x_1 + \tilde{b}_{12}y_1 + \tilde{b}_{13} + \tilde{b}_{14}/\lambda_1) \\ y_2(\tilde{b}_{31}x_1 + \tilde{b}_{32}y_1 + \tilde{b}_{33} + \tilde{b}_{34}/\lambda_1) &= (\tilde{b}_{21}x_1 + \tilde{b}_{22}y_1 + \tilde{b}_{23} + \tilde{b}_{24}/\lambda_1) \end{cases} \quad (4.40)$$

Las dos posibles soluciones para λ_1 son entonces:

$$\begin{cases} \lambda_1 &= (\tilde{b}_{14} - \tilde{b}_{34}x_2)/[x_2(\tilde{b}_{31}x_1 + \tilde{b}_{32}y_1 + \tilde{b}_{33}) - (\tilde{b}_{11}x_1 + \tilde{b}_{12}y_1 + \tilde{b}_{13})] \\ \lambda_1 &= (\tilde{b}_{24} - \tilde{b}_{34}y_2)/[y_2(\tilde{b}_{31}x_1 + \tilde{b}_{32}y_1 + \tilde{b}_{33}) - (\tilde{b}_{21}x_1 + \tilde{b}_{22}y_1 + \tilde{b}_{23})] \end{cases} \quad (4.41)$$

Otras dos posibles soluciones para λ_1 pueden ser calculadas de la misma manera a partir de la tercera proyección de $\tilde{\mathbf{M}}$ en (4.33):

$$\begin{cases} \lambda_1 &= (\tilde{c}_{14} - \tilde{c}_{34}x_3)/[x_3(\tilde{c}_{31}x_1 + \tilde{c}_{32}y_1 + \tilde{c}_{33}) - (\tilde{c}_{11}x_1 + \tilde{c}_{12}y_1 + \tilde{c}_{13})] \\ \lambda_1 &= (\tilde{c}_{24} - \tilde{c}_{34}y_3)/[y_3(\tilde{c}_{31}x_1 + \tilde{c}_{32}y_1 + \tilde{c}_{33}) - (\tilde{c}_{21}x_1 + \tilde{c}_{22}y_1 + \tilde{c}_{23})] \end{cases} \quad (4.42)$$

De esta manera se obtienen cuatro *relaciones trifocales* al igualar una ecuación de (4.41) a una ecuación de (4.42):

$$\begin{cases} (\tilde{c}_{14} - \tilde{c}_{34}x_3)[x_2(\tilde{b}_{31}x_1 + \tilde{b}_{32}y_1 + \tilde{b}_{33}) - (\tilde{b}_{11}x_1 + \tilde{b}_{12}y_1 + \tilde{b}_{13})] - \\ (\tilde{b}_{14} - \tilde{b}_{34}x_2)[x_3(\tilde{c}_{31}x_1 + \tilde{c}_{32}y_1 + \tilde{c}_{33}) - (\tilde{c}_{11}x_1 + \tilde{c}_{12}y_1 + \tilde{c}_{13})] &= 0 \\ (\tilde{c}_{24} - \tilde{c}_{34}y_3)[x_2(\tilde{b}_{31}x_1 + \tilde{b}_{32}y_1 + \tilde{b}_{33}) - (\tilde{b}_{11}x_1 + \tilde{b}_{12}y_1 + \tilde{b}_{13})] - \\ (\tilde{b}_{14} - \tilde{b}_{34}x_2)[y_3(\tilde{c}_{31}x_1 + \tilde{c}_{32}y_1 + \tilde{c}_{33}) - (\tilde{c}_{21}x_1 + \tilde{c}_{22}y_1 + \tilde{c}_{23})] &= 0 \\ (\tilde{c}_{14} - \tilde{c}_{34}x_3)[y_2(\tilde{b}_{31}x_1 + \tilde{b}_{32}y_1 + \tilde{b}_{33}) - (\tilde{b}_{21}x_1 + \tilde{b}_{22}y_1 + \tilde{b}_{23})] - \\ (\tilde{b}_{24} - \tilde{b}_{34}y_2)[x_3(\tilde{c}_{31}x_1 + \tilde{c}_{32}y_1 + \tilde{c}_{33}) - (\tilde{c}_{11}x_1 + \tilde{c}_{12}y_1 + \tilde{c}_{13})] &= 0 \\ (\tilde{c}_{24} - \tilde{c}_{34}y_3)[y_2(\tilde{b}_{31}x_1 + \tilde{b}_{32}y_1 + \tilde{b}_{33}) - (\tilde{b}_{21}x_1 + \tilde{b}_{22}y_1 + \tilde{b}_{23})] - \\ (\tilde{b}_{24} - \tilde{b}_{34}y_2)[y_3(\tilde{c}_{31}x_1 + \tilde{c}_{32}y_1 + \tilde{c}_{33}) - (\tilde{c}_{21}x_1 + \tilde{c}_{22}y_1 + \tilde{c}_{23})] &= 0 \end{cases} \quad (4.43)$$

Estas cuatro ecuaciones corresponden a las *Trilinearidades* de Shashua [46] expresadas en (4.35), que a continuación se repiten:

$$\begin{cases} \mathbf{m}_1^T(x_3\mathbf{T}^{13} - x_3x_2\mathbf{T}^{33} + x_2\mathbf{T}^{31} - \mathbf{T}^{11}) &= 0 \\ \mathbf{m}_1^T(y_3\mathbf{T}^{13} - y_3x_2\mathbf{T}^{33} + x_2\mathbf{T}^{32} - \mathbf{T}^{12}) &= 0 \\ \mathbf{m}_1^T(x_3\mathbf{T}^{23} - x_3y_2\mathbf{T}^{33} + y_2\mathbf{T}^{31} - \mathbf{T}^{21}) &= 0 \\ \mathbf{m}_1^T(y_3\mathbf{T}^{23} - y_3y_2\mathbf{T}^{33} + y_2\mathbf{T}^{32} - \mathbf{T}^{22}) &= 0 \end{cases} \quad (4.44)$$

donde

$$\mathbf{T}^{jk} = [T_1^{jk} \ T_2^{jk} \ T_3^{jk}]^\top,$$

y

$$T_i^{jk} = (-1)^{i+1} \begin{vmatrix} \sim \tilde{\mathbf{a}}_i \\ \tilde{\mathbf{b}}_j \\ \tilde{\mathbf{c}}_k \end{vmatrix} = \tilde{b}_{ji}\tilde{c}_{k4} - \tilde{b}_{j4}\tilde{c}_{ki}, \quad \text{para } i, j, k = 1, 2, 3$$

Como ya se mencionó, los elementos T_i^{jk} son los *tensores trifocales* para las imágenes 1, 2 y 3. En forma tensorial se pueden escribir las Trilinearidades de Shahua de la siguiente manera:

$$m_1^i s_j^\mu r_k^\rho T_i^{jk} = 0 \quad \text{para } \mu, \rho = 1, 2 \quad (4.45)$$

mit

$$s_j^\mu = \begin{bmatrix} -1 & 0 & x_2 \\ 0 & -1 & y_2 \end{bmatrix} \quad r_k^\rho = \begin{bmatrix} -1 & 0 & x_3 \\ 0 & -1 & y_3 \end{bmatrix}.$$

4.2.3. Interpretación geométrica de las trilinearidades

A continuación se le dará una interpretación geométrica a las trilinearidades basándose en la Figura 4.5. La ecuación (4.38) corresponde al proceso de proyección de M en m_1 , esto quiere decir que esta ecuación define la recta $l_1 : \langle C_1, m_1 \rangle$ que contiene los puntos C_1 , M y m_1 . Dónde exactamente esta M sobre esta línea recta no se puede determinar a partir de m_1 y C_1 , sólo se sabe que está en algún punto perteneciente a l_1 . El parámetro λ_1 proporciona la información de la ubicación de M . En (4.39) se proyecta M en la segunda imagen. Las coordenadas (x_2, y_2) de este segundo punto se pueden determinar a partir de (4.40), o bien a partir de (4.41). Estas coordenadas tienen la forma

$$\begin{cases} x_2 &= f_{x_2}(x_1, y_1, \lambda_1) \\ y_2 &= f_{y_2}(x_1, y_1, \lambda_1) \end{cases}, \quad (4.46)$$

esto quiere decir que los puntos posibles en la segunda imagen que satisfacen (4.46a) se encuentran sobre una línea vertical en la que está el punto m_2 . Esta recta ha sido denominada como l_{x_2} en la Figura 4.5. Este mismo

planteamiento es válido para la recta horizontal l_{y_2} que se puede definir a partir de (4.46b).

Los puntos m_1 y m_2 son correspondientes si

$$l_1 \cap \Pi_{x_2} = M \quad \text{y} \quad l_1 \cap \Pi_{y_2} = M, \quad (4.47)$$

donde Π_{x_2} y Π_{y_2} denotan los planos definidos a partir del centro óptico C_2 y la recta l_{x_2} y la recta l_{y_2} respectivamente.

De esta manera se obtiene para la tercera proyección de M en la imagen 3 las rectas l_{x_3} y l_{y_3} , las que son definidas a partir de (4.42):

$$\begin{cases} x_3 = f_{x_3}(x_1, y_1, \lambda_1) \\ y_3 = f_{y_3}(x_1, y_1, \lambda_1) \end{cases}, \quad (4.48)$$

Al igual que la condición que se estableció para la correspondencia de m_1 y m_2 , se puede afirmar que m_1 y m_3 son correspondientes si

$$l_1 \cap \Pi_{x_3} = M \quad \text{y} \quad l_1 \cap \Pi_{y_3} = M, \quad (4.49)$$

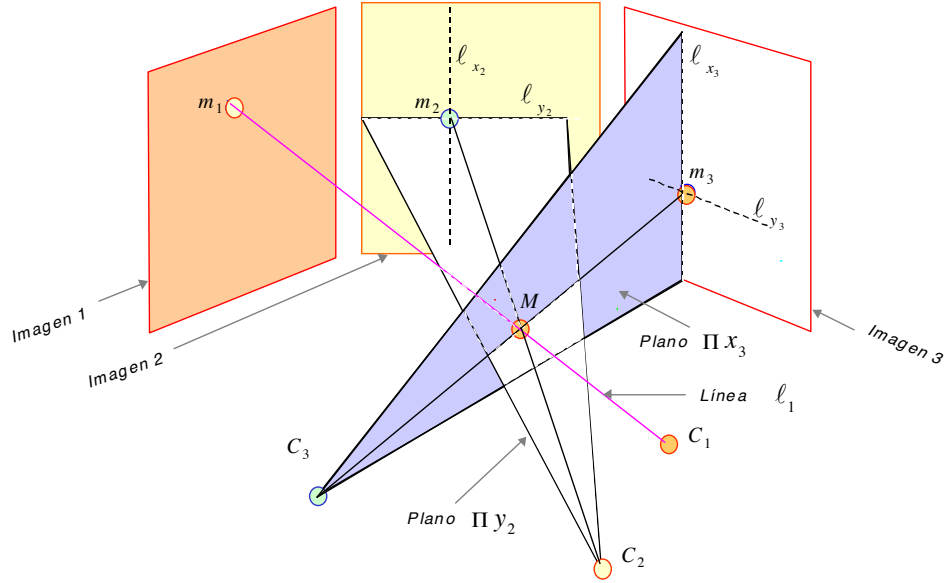


Figura 4.5: Representación geométrica de las Trilinearidades (adaptada de [1]).

donde Π_{x_3} y Π_{y_3} son los planos definidos a partir del punto C_3 con las rectas l_{x_3} y l_{y_3} respectivamente.

Si los puntos de intersección de (4.47) y los puntos de intersección de (4.49) son iguales, entonces se ha encontrado una correspondencia entre m_1 , m_2 y m_3 . De esta manera las cuatro trilinearidades en esta interpretación geométrica son:

$$\begin{cases} l_1 \cap \Pi_{x_2} = l_1 \cap \Pi_{x_3} \\ l_1 \cap \Pi_{x_2} = l_1 \cap \Pi_{y_3} \\ l_1 \cap \Pi_{y_2} = l_1 \cap \Pi_{x_3} \\ l_1 \cap \Pi_{y_2} = l_1 \cap \Pi_{y_3} \end{cases} \quad (4.50)$$

Cada trilinearidad representa una correspondencia entre el punto m_1 (en la imagen 1), una recta horizontal o vertical (en la imagen 2) que contiene a m_2 y una recta horizontal o vertical (en la imagen 3) que contiene a m_3 . El punto M es el punto de intersección de la recta l_1 con los planos que se forman a partir de las rectas mencionadas con sus centros ópticos respectivos. Como ejemplo la Figura 4.5 muestra la tercera trilinearidad.

4.2.4. Propiedades de las trilinearidades

Las principales propiedades de las cuatro trilinearidades (ver ecuaciones (4.35) y (4.44)) pueden resumirse de la siguiente manera [20, 21, 44, 45]:

- i)* Las trilinearidades representan relaciones lineales y trifocales. Estas relaciones han sido determinadas sin emplear la Geometría Epipolar (ver Sección 4.2.1).
- ii)* Los tensores trifocales son independientes de los puntos proyectados en las tres imágenes m_1 , m_2 , m_3 y también del punto 3D M . Los tensores trifocales son una función de las matrices de proyección $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ y $\tilde{\mathbf{C}}$ (ver ecuación (4.36)).
- iii)* La *reproyección* de m_3 , es decir la predicción de las coordenadas de m_3 a partir de las coordenadas de m_1 y m_2 , puede calcularse directamente de las trilinearidades. La reproyección no tiene ninguna singularidad.
- iv)* Las cuatro trilinearidades son linealmente independientes.

- v) Los puntos m_1 , m_2 y m_3 (en tres imágenes distintas) son correspondientes, si las cuatro trilinearidades son válidas (condición necesaria y suficiente) (ver Sección 4.2.3).

4.2.5. Relación entre la geometría bifocal y trifocal

Una relación entre la geometría epipolar y las trilinearidades se presenta a continuación:

Teorema 4.1 *Si m_1 , m_2 y m_3 son tres puntos en tres imágenes distintas, entonces son correspondientes si:*

- a) m_1 y m_2 satisfacen la condición epipolar, y
 b) m_1 , m_2 y m_3 satisfacen las dos primeras o las dos últimas trilinearidades.

Prueba: Las cuatro trilinearidades provienen de las ecuaciones (4.41) y (4.42), en las que una ecuación de (4.41) se ha igualado a una ecuación de (4.42). Si se igualan sin embargo las dos ecuaciones de (4.41) se obtiene entonces la condición epipolar entre m_1 y m_2 (comparar con (4.26)). Esto quiere decir que las trilinearidades (4.44a) y (4.44c) y las trilinearidades (4.44b) y (4.44d) en este caso son iguales. •

En la práctica, se puede utilizar el Teorema 4.1 como criterio de correspondencia. Es decir, m_1 , m_2 y m_3 son puntos correspondientes si m_1 y m_2 satisfacen la condición epipolar práctica (4.32) y si la distancia Euclídea entre m_3 y \hat{m}_3 (la reproyección de m_3) calculado a partir de m_1 y m_2 es suficientemente pequeña:

$$d = \|\mathbf{m}_3 - \hat{\mathbf{m}}_3\| < d_1, \quad (4.51)$$

donde la reproyección $\hat{\mathbf{m}}_3$ se calcula de las dos primeras trilinearidades (4.44a) y (4.44b):

$$\hat{\mathbf{m}}_3 = \begin{bmatrix} \hat{x}_3 \\ \hat{y}_3 \\ 1 \end{bmatrix} = \frac{1}{\mathbf{m}_1^T (\mathbf{T}^{13} - x_2 \mathbf{T}^{33})} \begin{bmatrix} \mathbf{m}_1^T (\mathbf{T}^{11} - x_2 \mathbf{T}^{31}) \\ \mathbf{m}_1^T (\mathbf{T}^{12} - x_2 \mathbf{T}^{32}) \\ \mathbf{m}_1^T (\mathbf{T}^{13} - x_2 \mathbf{T}^{33}) \end{bmatrix} \quad (4.52)$$

4.3. Correspondencia en cuatro vistas

Desde un punto vista geométrico es posible establecer una condición en la que cuatro puntos m_1 , m_2 , m_3 y m_4 (en las imágenes respectivas 1, 2, 3 y 4) son correspondientes entre sí a partir de las líneas epipolares. Se calculan las líneas epipolares de los primeros tres puntos en la cuarta imagen y se busca el punto de intersección. Si la intersección coincide con el cuarto punto se dice entonces que los cuatro puntos son correspondientes. Como ya se señaló en la Sección 4.2 para el caso de las tres vistas, esta solución presenta una singularidad cuando dos (o tres) de las tres líneas epipolares son iguales. Adicionalmente, la estimación del punto de intersección no es directo y esta solución no es rápida.

Otra forma de abordar este problema es tratarlo algebraicamente. Las cuatro proyecciones de un punto 3D M en las imágenes p , q , r y s se pueden describir como se hizo para el caso de dos o tres vistas (ver ecuaciones (4.22) y (4.33):

$$\begin{cases} \lambda_1 \mathbf{m}_1 &= [\mathbf{I} \mid \mathbf{0}] \tilde{\mathbf{M}} = \tilde{\mathbf{A}} \tilde{\mathbf{M}} \\ \lambda_2 \mathbf{m}_2 &= \tilde{\mathbf{B}} \tilde{\mathbf{M}} \\ \lambda_3 \mathbf{m}_3 &= \tilde{\mathbf{C}} \tilde{\mathbf{M}} \\ \lambda_4 \mathbf{m}_4 &= \tilde{\mathbf{D}} \tilde{\mathbf{M}} \end{cases}, \quad (4.53)$$

donde $\tilde{\mathbf{D}} = \mathbf{P}_s \mathbf{H}^{-1}$ y $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \tilde{\mathbf{M}}$ como en (4.23) y (4.33). Este sistema de ecuaciones puede ser expresado de la siguiente manera:

[illegible]

donde $\tilde{\mathbf{a}}_i$, $\tilde{\mathbf{b}}_i$, $\tilde{\mathbf{c}}_i$ y $\tilde{\mathbf{d}}_i$ son la fila i de la matriz $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, $\tilde{\mathbf{C}}$ y $\tilde{\mathbf{D}}$ respectivamente. Si m_1 , m_2 , m_3 y m_4 son puntos correspondientes, entonces debe existir una solución no trivial para M (y por ende para \mathbf{v}). La existencia de esta solución lleva a una condición necesaria y suficiente que señala que todos los determinantes de las submatrices de 8×8 de \mathbf{G} (obtenidas a partir de la eliminación de cuatro de sus filas) deben ser iguales a cero. Se pueden obtener relaciones *cuadrifocales* si una fila de cada matriz de proyección se elimina. De esta manera se llega a 16 *cuadrilinealidades* (lineales e independientes) como función de los conocidos 81 *tensores en cuadrifocales* [20, 22, 25, 26]. Una aplicación de estos tensores en ensayos radioscópicos se puede encontrar en [35, 37].

Sin embargo, si se conoce los tensores trifocales, y si además es conocida la correspondencia entre algunos de sus puntos es posible establecer la correspondencia de los cuatro puntos sin tener que calcular los tensores cuadrifocales y sin tener que evaluar las cuadrilinealidades, como se verá a continuación.

Teorema 4.2 *Si m_1 , m_2 , m_3 y m_4 son cuatro puntos en cuatro imágenes distintas, entonces son condiciones suficientes para establecer su correspondencia que*

- a) m_1 , m_2 y m_3 son correspondientes entre sí, y
- b) m_2 , m_3 y m_4 son correspondientes entre sí.

Prueba: Si m_2 y m_3 son correspondientes entre sí, entonces existe una solución única para el punto 3D M que produce los dos puntos de proyección mencionados. Esta afirmación esta contenida en las condiciones a) y b). Esto implica que si se cumplen estas dos condiciones el punto 3D M de los puntos de proyección sería el mismo. En este caso m_1 y m_4 son proyecciones de M en las imágenes 1 y 5. Esto quiere decir que m_1 , m_2 , m_3 y m_4 son puntos correspondientes. ●

Teorema 4.3 *Si m_1 , m_2 , m_3 y m_4 son cuatro puntos correspondientes en cuatro imágenes distintas, entonces son condiciones suficientes para establecer su correspondencia que*

- a) m_1 y m_2 cumplan la condición epipolar (para las imágenes 1 y 2)
- b) m_1 , m_2 y m_3 cumplan las dos primeras (o últimas) trilinealidades (para las imágenes 1, 2 y 3), y
- c) m_2 , m_3 y m_4 cumplan las dos primeras (o últimas) trilinealidades (para las imágenes 2, 3 y 4)

Prueba: De los teoremas 4.2 y 4.1 se sabe que las condiciones suficientes para establecer la correspondencia de m_1 , m_2 , m_3 y m_4 son:

- i) m_1 y m_2 deben cumplir la condición epipolar (para las imágenes 1 y 2), y
- ii) m_1 , m_2 y m_3 deben satisfacer las dos primeras (o últimas) trilinealidades (para las imágenes 1, 2 y 3), y
- iii) m_2 y m_3 deben cumplir la condición epipolar (para las imágenes 2 y 3), y
- iv) m_2 , m_3 y m_4 deben satisfacer las dos primeras (o últimas) trilinealidades (para las imágenes 2, 3 y 4).

Las condiciones i), ii) y iv) son exactamente iguales a las condiciones a), b) y c). La condición iii) no es necesario evaluarla ya que esta implícitamente incluida en i) y ii), ya que en este caso m_2 y m_3 son puntos correspondiente y por lo tanto cumplen la condición epipolar iii). •

En la práctica se puede establecer un criterio de correspondencia que señala que m_1 , m_2 , m_3 y m_4 son correspondientes entre sí, si m_1 , m_2 y m_3 satisfacen el criterio práctico de la condición trifocal (4.32) y (4.51) y que la distancia Euclídea entre m_4 y su reproyección \hat{m}_4 (calculada a partir de m_2 , m_3 y los tensores trifocales entre las imágenes 2, 3 y 4) es menor que ε_3 :

$$d_4 = \|\mathbf{m}_4 - \hat{\mathbf{m}}_4\| < \varepsilon_3, \quad (4.55)$$

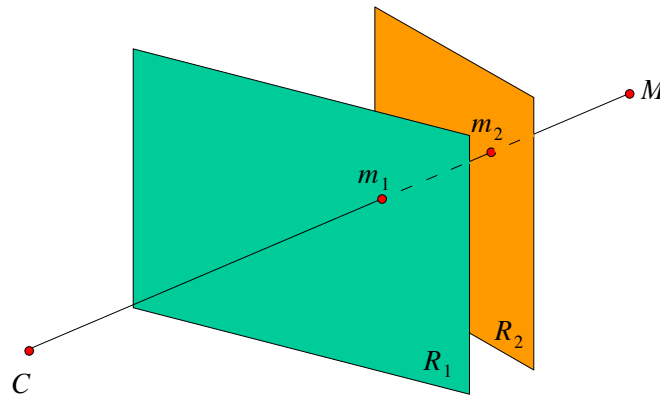


Figura 4.6: Correspondencia en dos vistas que comparten el centro óptico (ver Ejercicio 4.1).

4.4. Ejercicios

Ejercicio 4.1 En el esquema de la Figura 4.6 se muestra la formación de dos imágenes en los planos R_1 y R_2 a partir de un centro óptico C . Si se conoce la ubicación de m_1 y las matrices 3×4 de proyección \mathbf{A} y \mathbf{B} tal que $\lambda_1 \mathbf{m}_1 = \mathbf{A}\mathbf{M}$ y $\lambda_2 \mathbf{m}_2 = \mathbf{B}\mathbf{M}$, pero no se conoce la ubicación de M , a) ¿qué restricción existe para la ubicación de m_2 ? b) ¿Qué relación existe entre \mathbf{A} y \mathbf{B} ?

Ejercicio 4.2 En el plano (x, y) se tiene una recta l y un punto m_2 , cuyas representaciones homogéneas son $\mathbf{m}_2 = [x_2 \ y_2 \ 1]^T$ y $\mathbf{l} = [l_1 \ l_2 \ l_3]^T$. a) Calcule la distancia mínima d (ver Figura 4.7) que existe entre m_2 y l . b) En la práctica, dos puntos correspondientes en dos imágenes distintas no cumplen la restricción epipolar porque la medición de las coordenadas de los puntos, así como el cálculo de la Matriz Fundamental \mathbf{F} están sujetos a errores. Por lo tanto se usa como restricción práctica que dos puntos pueden ser correspondientes si el segundo punto está a menos de una distancia d_{min} de la línea epipolar del primer punto en la segunda imagen. Utilice el resultado de a) para expresar matemáticamente esta restricción epipolar práctica suponiendo que se conocen las representaciones homogéneas de los dos puntos \mathbf{m}_1 y \mathbf{m}_2 y se conoce la Matriz Fundamental \mathbf{F} (sugerencia:

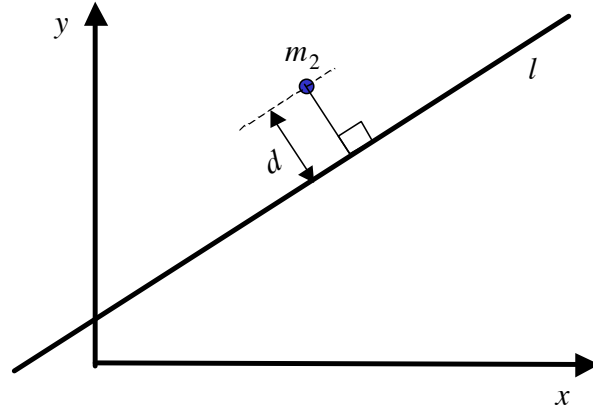


Figura 4.7: Distancia mínima de un punto m_2 a una recta l (ver Problema 4.2).

calcule la línea epipolar l del primer punto en la segunda imagen; calcule la distancia d del segundo punto a la línea epipolar usando el resultado de a); y compare esta distancia con d_{min}).

Ejercicio 4.3 Un sistema bifocal tiene la siguiente matriz fundamental entre la imagen 1 y 2:

$$\mathbf{F} = \begin{bmatrix} 1 & 2 & 2 \\ 4 & 3 & 1 \\ 3 & 1 & -1 \end{bmatrix} \quad (4.56)$$

Encuentre el epipolo en la segunda imagen.

Solución: Se puede asumir $\mathbf{e}_2 = [x_2 \ y_2 \ 1]^T$. Reemplazando este valor en (4.20b) se obtiene un sistema de tres ecuaciones con dos incógnitas. Como $\det(\mathbf{F})=0$ basta con considerar sólo dos ecuaciones linealmente independientes, ya que la tercera es linealmente dependiente de las otras dos. Tomando las dos primeras ecuaciones de $\mathbf{F}^T [x_2 \ y_2 \ 1]^T = \mathbf{0}$ se obtiene:

$$\begin{cases} x_2 + 4y_2 = -3 \\ 2x_2 + 3y_2 = -1 \end{cases}$$

Las coordenadas de \mathbf{e}_2 son entonces $(x_2 = 1; y_2 = -1)$.

Capítulo 5

Reconstrucción 3D

En este Capítulo se estudiarán algunos algoritmos para *reconstruir* las coordenadas de un punto 3D M a partir de sus vistas. Es necesario aclarar que el término *reconstrucción* empleado en tomografía computarizada es distinto al usado en este Capítulo, ya que aquí se utilizará reconstrucción en el sentido de localización de un punto 3D y no en el sentido reconstrucción de las características de la materia de este punto, como es el caso de la tomografía computarizada.

El punto 3D M se estimará entonces a partir de sus distintas proyecciones obtenidas de las imágenes. Para este problema, las coordenadas de estas proyecciones m_1, \dots, m_n (en n imágenes) son conocidas y se asumirá que estos puntos son correspondientes, es decir que son proyecciones de un mismo punto 3D. Asimismo, son conocidas las matrices de proyección que generaron estos puntos a partir del punto 3D M . Las matrices de proyección se obtienen a partir de algún proceso de calibración.

Debido a que con una sola vista la información de profundidad del espacio 3D se pierde en la proyección, para la reconstrucción 3D son necesarias por lo menos dos vistas. Detrás de los métodos de reconstrucción 3D está el concepto de *triangulación* que se usa en la estimación de un punto 3D a partir de sus vistas. La triangulación consiste en inferir la información 3D a partir de los rayos que van desde los centros ópticos de las imágenes respectivas hasta los puntos proyectados. Como es sabido que el punto 3D que produjo estas proyecciones pertenece a estos rayos se busca entonces la intersección de ellos en el espacio 3D. Un ejemplo para tres vistas se muestra en la Figura 5.1. En

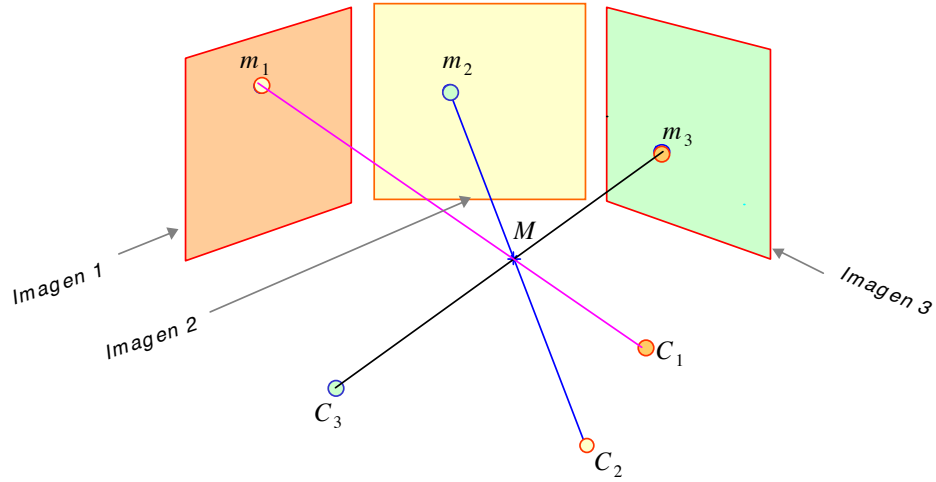


Figura 5.1: Triangulación en tres vistas: M es la intersección de los rayos $\langle C_i, m_i \rangle$.

este caso el punto M es la intersección de los tres rayos $\langle C_1, m_1 \rangle$, $\langle C_2, m_2 \rangle$ y $\langle C_3, m_3 \rangle$ ¹. Como en la gran mayoría de casos prácticos estos rayos no se intersectan (incluso para el caso de sólo dos vistas), es necesario encontrar entonces el *mejor* punto 3D que produciría las proyecciones dadas.

A continuación se presentarán dos métodos de reconstrucción 3D. El primero de ellos es una reconstrucción 3D directa y lineal a partir de dos vistas. El segundo método utiliza el criterio de los mínimos cuadrados para hacer una reconstrucción 3D a partir de dos o más vistas.

5.1. Método de reconstrucción lineal para dos vistas

En esta sección se realizará una estimación del punto 3D M a partir de dos puntos correspondientes m_1 y m_2 ubicados en la imagen 1 y la imagen 2 respectivamente (ver ilustración en Figura 4.1a). Este método fue desarrollado por Hartley en [19].

¹Otro ejemplo se puede apreciar en la Figura 1.10.

Como ya se vio en la Sección 4.1.3 las ecuaciones de proyección en dos imágenes están dadas por:

$$\begin{cases} \lambda_1 \mathbf{m}_1 = \mathbf{A}\mathbf{M} \\ \lambda_2 \mathbf{m}_2 = \mathbf{B}\mathbf{M} \end{cases} \quad (5.1)$$

donde $\mathbf{M} = [X \ Y \ Z \ 1]^\top$ y $\mathbf{m}_i = [x_i \ y_i \ 1]^\top$ son las representaciones homogéneas de M y m_i , $i = 1, 2$, y \mathbf{A} y \mathbf{B} las respectivas matrices de proyección de las imágenes 1 y 2. Haciendo una transformación de coordenadas para \mathbf{M} , la ecuación (5.1) se puede escribir como:

$$\begin{cases} \lambda_1 \mathbf{m}_1 = [\mathbf{I} \mid \mathbf{0}] \tilde{\mathbf{M}} = \tilde{\mathbf{A}} \tilde{\mathbf{M}} \\ \lambda_2 \mathbf{m}_2 = \tilde{\mathbf{B}} \tilde{\mathbf{M}} \end{cases} \quad (5.2)$$

donde

$$\begin{aligned} \mathbf{M} &= \mathbf{H}^{-1} \tilde{\mathbf{M}} \\ \tilde{\mathbf{A}} &= \mathbf{A} \mathbf{H}^{-1} = [\mathbf{I} \mid \mathbf{0}] \\ \tilde{\mathbf{B}} &= \mathbf{B} \mathbf{H}^{-1} \end{aligned} \quad (5.3)$$

La matriz \mathbf{H} , de 4×4 elementos, es una matriz regular cuyas tres primeras filas corresponden a la matriz \mathbf{A}^2 .

La primera ecuación de (5.2)

$$\lambda_1 \mathbf{m}_1 = [\mathbf{I} \mid \mathbf{0}] \tilde{\mathbf{M}} \quad (5.4)$$

se puede reescribir como

$$\lambda_1 \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \end{bmatrix}. \quad (5.5)$$

Con esta ecuación es fácil obtener una expresión para $\tilde{\mathbf{M}}$, es decir las coordenadas de M en un sistema de coordenadas transformadas, en el que la matriz de proyección tiene una forma canónica del tipo $[\mathbf{I} \mid \mathbf{0}]$:

$$\tilde{\mathbf{M}} = \begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ 1 \end{bmatrix} = \lambda_1 \begin{bmatrix} x_1 \\ y_1 \\ 1 \\ 1/\lambda_1 \end{bmatrix}. \quad (5.6)$$

²La demostración de $\mathbf{A} \mathbf{H}^{-1} = [\mathbf{I} \mid \mathbf{0}]$ se puede encontrar en la Sección 4.1.3.

Se observa que λ_1 establece en qué lugar del rayo que produce m_1 se encuentra M . Utilizando (5.2), la proyección de este punto 3D en la segunda imagen es:

$$\lambda_2 \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \lambda_1 \begin{bmatrix} \tilde{b}_{11} & \tilde{b}_{12} & \tilde{b}_{13} & \tilde{b}_{14} \\ \tilde{b}_{21} & \tilde{b}_{22} & \tilde{b}_{23} & \tilde{b}_{24} \\ \tilde{b}_{31} & \tilde{b}_{32} & \tilde{b}_{33} & \tilde{b}_{34} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \\ 1/\lambda_1 \end{bmatrix}. \quad (5.7)$$

Suponiendo que se conocen las coordenadas de m_1 y m_2 , y también se conoce la matriz de proyección $\tilde{\mathbf{B}}$, entonces la ecuación (5.7) representa un sistema de tres ecuaciones con dos incógnitas (λ_1 y λ_2). Si m_1 y m_2 son puntos correspondientes se puede establecer una solución a partir de dos de las tres ecuaciones. Las primeras dos ecuaciones son:

$$\begin{cases} \lambda_2 x_2 = \lambda_1 \tilde{b}_{11} x_1 + \lambda_1 \tilde{b}_{12} y_1 + \lambda_1 \tilde{b}_{13} + \tilde{b}_{14} \\ \lambda_2 y_2 = \lambda_1 \tilde{b}_{21} x_1 + \lambda_1 \tilde{b}_{22} y_1 + \lambda_1 \tilde{b}_{23} + \tilde{b}_{24} \end{cases}. \quad (5.8)$$

Estas ecuaciones se pueden dividir para cancelar λ_2 . De esta manera se puede obtener la siguiente ecuación:

$$\begin{aligned} x_2(\lambda_1 \tilde{b}_{21} x_1 + \lambda_1 \tilde{b}_{22} y_1 + \lambda_1 \tilde{b}_{23} + \tilde{b}_{24}) = \\ y_2(\lambda_1 \tilde{b}_{11} x_1 + \lambda_1 \tilde{b}_{12} y_1 + \lambda_1 \tilde{b}_{13} + \tilde{b}_{14}) \end{aligned} \quad (5.9)$$

A partir de esta ecuación se puede despejar λ_1 :

$$\lambda_1 = \frac{y_2 \tilde{b}_{14} - x_2 \tilde{b}_{24}}{x_1 x_2 \tilde{b}_{21} + y_1 x_2 \tilde{b}_{22} + x_2 \tilde{b}_{23} - x_1 y_2 \tilde{b}_{11} - y_1 y_2 \tilde{b}_{12} - y_2 \tilde{b}_{13}} \quad (5.10)$$

que, utilizando $\mathbf{m}_1 = [x_1 \ y_1 \ 1]^T$, puede ser escrito como:

$$\lambda_1 = \frac{(y_2 \tilde{b}_{14} - x_2 \tilde{b}_{24})}{(x_2 [\tilde{b}_{21} \ \tilde{b}_{22} \ \tilde{b}_{23}] - y_2 [\tilde{b}_{11} \ \tilde{b}_{12} \ \tilde{b}_{13}]) \mathbf{m}_1} \quad (5.11)$$

Substituyendo en (5.6) el valor obtenido para λ_1 en (5.11), y utilizando la definición de \mathbf{M} a partir de $\tilde{\mathbf{M}}$ en (5.3a) se obtiene el punto 3D reconstruido.

$$\hat{\mathbf{M}} = \mathbf{H}^{-1} \hat{\tilde{\mathbf{M}}} = \mathbf{H}^{-1} \begin{bmatrix} \frac{(y_2 \tilde{b}_{14} - x_2 \tilde{b}_{24}) \mathbf{m}_1}{(x_2 [\tilde{b}_{21} \ \tilde{b}_{22} \ \tilde{b}_{23}] - y_2 [\tilde{b}_{11} \ \tilde{b}_{12} \ \tilde{b}_{13}]) \mathbf{m}_1} \\ 1 \end{bmatrix} \quad (5.12)$$

De esta manera se obtiene una estimación de M a partir de dos puntos m_1 y m_2 cuyas coordenadas son (x_1, y_1) y (x_2, y_2) .

5.2. Reconstrucción 3D para dos o más vistas

En n vistas distintas ($n \geq 2$) se tienen los puntos correspondientes m_i , $i = 1, \dots, n$. El punto 3D M que produjo estas proyecciones se puede obtener por medio del método de los mínimos cuadrados [8].

La representación homogénea de los puntos correspondientes está dada por $\mathbf{m}_i = [x_i \ y_i \ 1]^\top$, y para M sería $\mathbf{M} = [X \ Y \ Z \ 1]^\top$. Cada proyección proporciona un sistema de ecuaciones $\lambda_i \mathbf{m}_i = \mathbf{P}_i \mathbf{M}$, donde \mathbf{P}_i es la matriz de proyección de la imagen i . Este sistema de ecuaciones es de tres ecuaciones con cuatro incógnitas X , Y , Z y λ_i . Para $i = 1, \dots, n$ se obtiene entonces el sistema de $3n$ ecuaciones con $3 + n$ incógnitas (X , Y , Z , $\lambda_1, \dots, \lambda_n$):

$$\begin{bmatrix} \lambda_1 x_1 \\ \lambda_1 y_1 \\ \lambda_1 \\ \vdots \\ \lambda_n x_n \\ \lambda_n y_n \\ \lambda_n \end{bmatrix} = \begin{bmatrix} p_{11}^1 & p_{12}^1 & p_{13}^1 & p_{14}^1 \\ p_{21}^1 & p_{22}^1 & p_{23}^1 & p_{24}^1 \\ p_{31}^1 & p_{32}^1 & p_{33}^1 & p_{34}^1 \\ \vdots & \vdots & \vdots & \vdots \\ p_{11}^n & p_{12}^n & p_{13}^n & p_{14}^n \\ p_{21}^n & p_{22}^n & p_{23}^n & p_{24}^n \\ p_{31}^n & p_{32}^n & p_{33}^n & p_{34}^n \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (5.13)$$

donde p_{jk}^i corresponde al elemento (j, k) de la matriz \mathbf{P}_i . De la tercera ecuación de cada proyección se obtiene:

$$\lambda_i = p_{31}^i X + p_{32}^i Y + p_{33}^i Z + p_{34}^i \quad (5.14)$$

Este valor encontrado para λ_i puede reemplazarse en (5.13) dando:

$$\begin{bmatrix} (p_{31}^1 X + p_{32}^1 Y + p_{33}^1 Z + p_{34}^1) x_1 \\ (p_{31}^1 X + p_{32}^1 Y + p_{33}^1 Z + p_{34}^1) y_1 \\ \vdots \\ (p_{31}^n X + p_{32}^n Y + p_{33}^n Z + p_{34}^n) x_n \\ (p_{31}^n X + p_{32}^n Y + p_{33}^n Z + p_{34}^n) y_n \end{bmatrix} = \begin{bmatrix} p_{11}^1 & p_{12}^1 & p_{13}^1 & p_{14}^1 \\ p_{21}^1 & p_{22}^1 & p_{23}^1 & p_{24}^1 \\ \vdots & \vdots & \vdots & \vdots \\ p_{11}^n & p_{12}^n & p_{13}^n & p_{14}^n \\ p_{21}^n & p_{22}^n & p_{23}^n & p_{24}^n \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (5.15)$$

Se obtiene así un nuevo sistema de ecuaciones con $2n$ ecuaciones con sólo 3 incógnitas (X , Y y Z), ya que en (5.14) se obtuvo una expresión para λ_i en función de X , Y y Z . La ecuación anterior puede ser reformulada de la

siguiente manera:

$$\underbrace{\begin{bmatrix} p_{31}^1 x_1 - p_{11}^1 & p_{32}^1 x_1 - p_{12}^1 & p_{33}^1 x_1 - p_{13}^1 \\ p_{31}^1 y_1 - p_{21}^1 & p_{32}^1 y_1 - p_{22}^1 & p_{33}^1 y_1 - p_{23}^1 \\ \vdots & \vdots & \vdots \\ p_{31}^n x_n - p_{11}^n & p_{32}^n x_n - p_{12}^n & p_{33}^n x_n - p_{13}^n \\ p_{31}^n y_n - p_{21}^n & p_{32}^n y_n - p_{22}^n & p_{33}^n y_n - p_{23}^n \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \underbrace{\begin{bmatrix} p_{14}^1 - p_{34}^1 x_1 \\ p_{24}^1 - p_{34}^1 y_1 \\ \vdots \\ p_{14}^n - p_{34}^n x_n \\ p_{24}^n - p_{34}^n y_n \end{bmatrix}}_{\mathbf{r}} \quad (5.16)$$

En caso de que el rango de \mathbf{Q} sea tres, existe una solución para $\hat{\mathbf{M}} = [\hat{X} \ \hat{Y} \ \hat{Z} \ 1]^T$ utilizando el método de los mínimos cuadrados:

$$[\hat{X} \ \hat{Y} \ \hat{Z}]^T = [\mathbf{Q}^T \mathbf{Q}]^{-1} \mathbf{Q}^T \mathbf{r} \quad (5.17)$$

Este método puede usarse también para $n = 2$ proyecciones, sin embargo el número de operaciones necesarias para obtener la reconstrucción de M es mayor que las que se necesitan utilizando el método de la Sección 5.1 donde no hay que calcular las matrices \mathbf{Q} y $[\mathbf{Q}^T \mathbf{Q}]^{-1}$.

En el método de los mínimos cuadrados se ha minimizado $\| \mathbf{QM} - \mathbf{r} \|^2$ con respecto a \mathbf{M} . Esto tiene la gran ventaja de poder obtener una solución no iterativa para $\hat{\mathbf{M}}$, sin embargo, la gran desventaja es que la expresión $\| \mathbf{QM} - \mathbf{r} \|$ no tiene una buena interpretación física. Una alternativa interesante es presentada en [8], donde se minimiza

$$J(\mathbf{M}) = \sum_{i=1}^n (\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2 \quad (5.18)$$

donde (x_i, y_i) son las coordenadas de m_i y (\hat{x}_i, \hat{y}_i) son las estimaciones de m_i obtenidas del modelo de proyección a partir de $\lambda_i \hat{\mathbf{m}}_i = \mathbf{P}_i \mathbf{M}$. En este problema se debe encontrar el mejor \mathbf{M} , es decir $\hat{\mathbf{M}}$, tal que se minimice la función de costo (5.18). La solución de este problema de optimización se puede obtener mediante métodos iterativos como los métodos de gradiente [47]. En este caso el valor inicial de la iteración puede ser el encontrado en (5.17).

Capítulo 6

Matching en dos Vistas

6.1. Introducción

En este capítulo se analizarán distintas técnicas para establecer la correspondencia entre elementos que están presentes en varias imágenes. En la Sección 6.2 se estudiará en detalle el *matching* en dos vistas. Se presenta la correspondencia punto-punto, línea-línea y región-región. En la Capítulo 7 se estudiará el seguimiento, o *tracking*, de un objeto a lo largo de una secuencia de imágenes.

6.2. Matching

El término *matching* se utiliza en visión artificial para establecer la correspondencia entre imágenes distintas de un mismo objeto. Comúnmente, se emplea como ‘objeto’ un punto, una línea o una región. Como ya se estudió con detalle en el Capítulo 4, se sabe que una condición necesaria pero no suficiente para establecer la correspondencia en dos puntos (en dos imágenes distintas) es la condición epipolar. Se plantea entonces la pregunta ¿qué otros criterios se pueden emplear para realizar el matching de dos puntos? En esta sección se estudiarán algunas posibilidades.

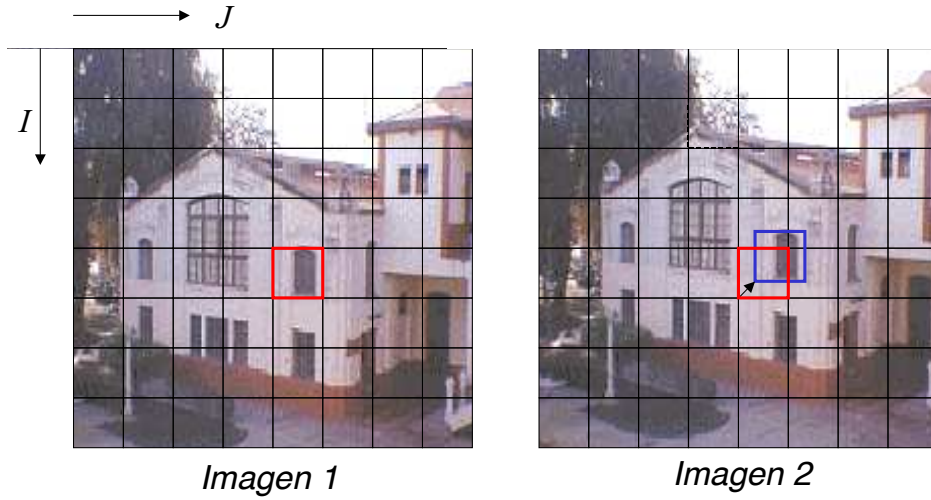


Figura 6.1: Flujo óptico: en la imagen 2 se busca la ventana roja más parecida a la ventana $I = 5, J = 5$ de la imagen 1.

6.2.1. Correspondencia entre puntos

Una posibilidad la da el *flujo óptico*, que corresponde a un conjunto de vectores que indica el movimiento de una imagen hacia otra. En el flujo óptico primero se divide la primera imagen en ventanas pequeñas. Luego, para cada una de estas ventanas se busca en la segunda imagen la ventana (de las mismas dimensiones) que sea lo más parecida posible a la ventana de la primera imagen. Finalmente, teniendo la ubicación de la ventana en la primera y en la segunda imagen, se obtiene un vector de desplazamiento. Este vector se puede interpretar como el movimiento que sufrió la ventana de la primera imagen para transformarse en la ventana de la segunda imagen. Como se puede apreciar, este método sólo se puede usar si la diferencia entre las dos imágenes es pequeña, ya que para un movimiento mayor, por ejemplo una rotación de 90° , no será posible encontrar las ventanas similares.

Matemáticamente se puede establecer que si las imágenes son I_1 e I_2 , cuyos tamaños son $N \times M$, se puede dividir la imagen en ventanas de $n \times m$. Siendo los índices $I = 1, \dots, n$ y $J = 1, \dots, m$ los que indican cuál de las ventanas de I_1 se va a analizar, se puede encontrar el vector de desplazamiento $\mathbf{v}_{IJ} = (k_I, k_J)$

minimizando la función objetivo que calcula el valor absoluto de la diferencia entre la ventana (I, J) de la imagen I_1 con una ventana en la imagen I_2 desplazada k_I píxels y k_J píxels en la dirección i y j respectivamente, es decir:

$$k_I, k_J = \arg \min \sum_{i=1}^n \sum_{j=1}^m |I_1(nI + i, mJ + j) - I_2(nI + i + k_I, mJ + j + k_J)| \quad (6.1)$$

En la Figura 6.1 se muestra a manera de ejemplo el vector de desplazamiento \mathbf{v}_{55} . El vector de desplazamiento también se puede encontrar maximizando una función objetivo que calcule la correlación entre las dos ventanas:

$$k_I, k_J = \arg \max \sum_{i=1}^n \sum_{j=1}^m I_1(nI + i, mJ + j) I_2(nI + i + k_I, mJ + j + k_J) \quad (6.2)$$

o bien utilizando la correlación normalizada:

$$k_I, k_J = \arg \max \frac{\sum_{i=1}^n \sum_{j=1}^m I_1(nI + i, mJ + j) I_2(nI + i + k_I, mJ + j + k_J)}{\sqrt{\sum_{i=1}^n \sum_{j=1}^m I_1^2(nI + i, mJ + j) \sum_{i=1}^n \sum_{j=1}^m I_2^2(nI + i + k_I, mJ + j + k_J)}} \quad (6.3)$$

En la Figura 6.2 se muestran dos imágenes en una esquina en las que se observa que ha habido un movimiento de los automóviles. Al calcular el flujo óptico, sólo en las porciones en que ha ocurrido un cambio se detectará un vector de desplazamiento mayor que cero.

Se puede pensar entonces, que esta herramienta puede ser de gran utilidad para encontrar la correspondencia entre dos puntos de dos imágenes distintas, ya que para un punto en la imagen I_1 se puede buscar su correspondiente en la imagen I_2 buscando la intersección del vector de desplazamiento del flujo óptico con la línea epipolar del primer punto en la imagen I_2 . Existen también técnicas más sofisticadas en que la ventana en la imagen I_2 no necesariamente tiene la misma orientación y tamaño de la ventana de la imagen 1 [42].

Es necesario tomar en cuenta que el flujo óptico presenta dos problemas: oclusión y ambigüedad. En el primero no es posible realizar el matching si el objeto que aparece en la ventana de búsqueda en I_1 ya no aparece en la imagen I_2 . En el problema de ambigüedad, el matching puede equivocarse si lo que se desea buscar en la imagen I_2 está repetido. Esta situación se puede

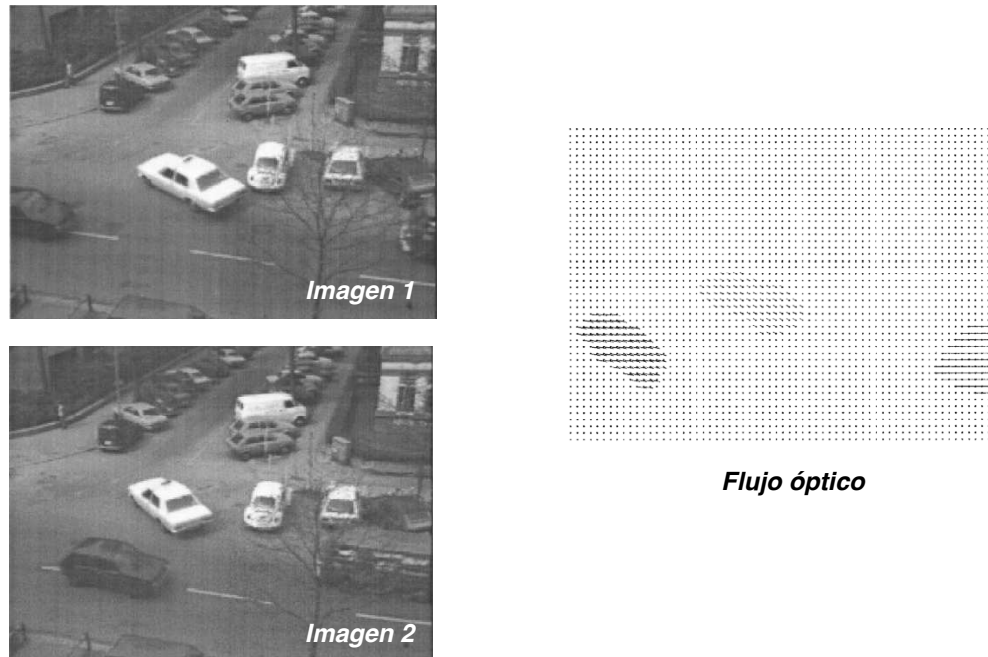


Figura 6.2: Ejemplo de flujo óptico [2].

producir por ejemplo en la Figura 6.1 si la ventana de la imagen I_1 es más pequeña a la indicada y contiene la ventana izquierda de la torre que aparece en la parte superior derecha de la imagen. Al buscar el matching en la imagen I_2 no se podrá diferenciar entre las dos ventanas que tiene la torre, es posible entonces que el algoritmo asocie la ventana izquierda con la ventana derecha.

6.2.2. Correspondencia entre líneas

Establecer la correspondencia entre dos líneas (en dos imágenes distintas) no tiene mucho sentido, ya que en la gran mayoría de los casos existe una línea 3D que podría haber producido las líneas en las imágenes. Tal como se muestra en la Figura 6.3, la intersección de los planos $\langle C_1, l_1 \rangle$ y $\langle C_2, l_2 \rangle$ proporciona una línea en el espacio 3D, que proyectada en ambas imágenes corresponde con las líneas originales l_1 y l_2 , sin embargo es posible que l_1 haya sido producida por otra línea 3D que pertenezca al plano $\langle C_1, l_1 \rangle$, en este caso la intersección de los planos no proporciona una condición suficiente

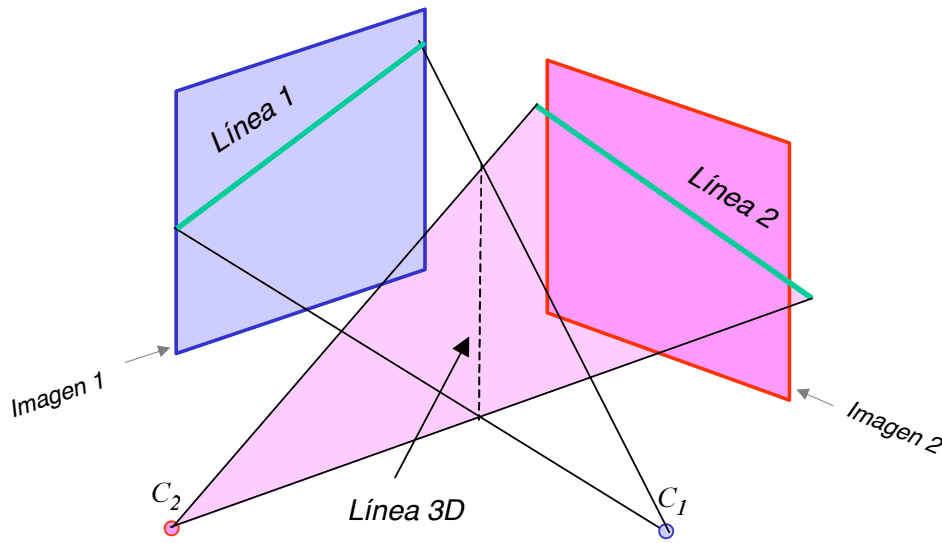


Figura 6.3: Análisis de líneas correspondientes.

para establecer la correspondencia. Un caso particular se produce cuando los planos son paralelos y no tienen una línea 3D de intersección, en este caso se dice que las líneas l_1 y l_2 no son correspondientes. Otro caso particular es cuando ambos planos son iguales, en este caso existen infinitas líneas 3D que podrían proyectarse como l_1 y l_2 , lo que quiere decir que las líneas podrían ser correspondientes.

Para establecer si dos líneas son correspondiente es conveniente introducir criterios adicionales al descrito anteriormente. Estos criterios deben analizar por ejemplo las características de color de cada una de las líneas o bien la longitud. En el segundo caso sin embargo, es difícil establecer si las líneas en las imágenes son proyecciones de la misma porción de la línea 3D.

6.2.3. Correspondencia entre regiones

Una técnica muy utilizada para establecer la correspondencia de objetos correspondientes en dos imágenes es la de búsqueda de regiones similares. En esta técnica se segmentan en cada una de las dos imágenes las regiones que sean de interés. Luego, se extraen las características de estas regiones y se comparan las características de las regiones de la imagen 1 con las

características de las regiones de la imagen 2. Aquellas regiones que tengan características similares y que estén ubicadas obedeciendo la condición epipolar serán entonces regiones correspondientes.

En esta sección se describen algunas características que pueden ser empleadas para describir cuantitativamente regiones que hayan sido segmentadas en una imagen. Se entenderá por *región* aquel conjunto de píxels que pertenezcan a una misma zona de la imagen y que esté limitado por bordes. Se asumirá que los bordes no pertenecen a la región.

Para explicar las características que se detallarán a continuación se usará el ejemplo de la Figura 6.4. En este ejemplo se presenta una región circular que ha sido segmentada. La región entonces se conforma por los píxels que pertenecen al círculo (pero no a su perímetro), es decir los píxels que han sido marcados con color gris en la Figura 6.4b. Los bordes de la región definen el límite de la región.

Las características que se pueden extraer de una región se dividen en dos categorías: *características geométricas* y *características de color*.

6.2.4. Características geométricas

A continuación se enumeran algunas características geométricas que se usan comúnmente en el reconocimiento de patrones.

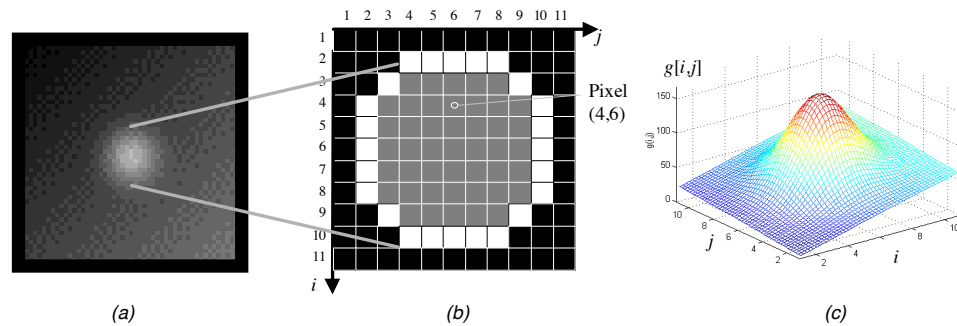


Figura 6.4: Ejemplo de una región: a) Imagen. b) Región segmentada. c) Representación 3D de los valores de gris de la región y su entorno.

• **Altura y ancho** (h y w): La altura y el ancho de una región se definen como:

$$h = i_{\max} - i_{\min} + 1 \quad \text{y} \quad w = j_{\max} - j_{\min} + 1 \quad (6.4)$$

donde i_{\max} e i_{\min} representan el valor máximo y mínimo que toma la coordenada i en la región (ver Figura 6.4), y lo mismo es válido para j_{\max} y j_{\min} . En el ejemplo mostrado $h = w = 7$ píxels.

• **Área** (A): El área de una región se define como el número de los píxels de la región. En el ejemplo $A = 45$ píxels.

• **Perímetro** (L): El perímetro de una región puede ser definido de varias maneras. Una definición práctica, mas no exacta, es tomar el perímetro como el número de píxels que pertenecen al borde de la región¹. En el ejemplo de la Figura 6.4b, L es el número de píxels marcados en color blanco, es decir $L = 24$.

• **Redondez** (R): Esta característica que indica la calidad de redondo de una región es una medida de su forma. La redondez se define como [24]:

$$R = \frac{4 \cdot A \cdot \pi}{L^2} \quad (6.5)$$

La redondez R de una región estará entre los valores 0 y 1. Teóricamente $R = 1$ para un círculo (perfecto); y $R = 0$ para una región que tenga altura y/o ancho igual a cero. En la práctica sin embargo, debido al muestreo en el espacio de la región estos valores presentan desviaciones como se puede ver en la región circular de nuestro ejemplo. En este caso $R = 4 \cdot 45 \cdot \pi / 24^2 = 0,98$.

• **Momentos**: Los momentos estadísticos se definen como

$$m_{rs} = \sum_{i,j \in \mathfrak{R}} i^r j^s \quad \text{para } r, s \in \mathcal{N} \quad (6.6)$$

donde \mathfrak{R} es el conjunto de píxels de la región. En el ejemplo de la Figura 6.4b el pixel cuyas coordenadas son $(i = 4, j = 6)$ pertenece a este conjunto. El parámetro $r + s$ denota el orden del momento. El momento de orden cero

¹Otras definiciones más exactas consideran el factor $\sqrt{2}$ para píxels del borde de la región que estén en diagonal, como por ejemplo en el caso de un borde que contenga los píxels (i, j) y $(i + 1, j + 1)$ [4]. En este caso existe un compromiso entre la precisión y el costo computacional que requiere su cálculo.

m_{00} corresponde al área de la región A . El centro de gravedad de una región queda definido por:

$$\bar{i} = \frac{m_{10}}{m_{00}} \quad \bar{j} = \frac{m_{01}}{m_{00}} \quad (6.7)$$

Con ayuda de las coordenadas del centro de gravedad se definen los momentos centrales que son invariantes al desplazamiento de la región en la imagen.

$$\mu_{rs} = \sum_{i,j \in \mathcal{R}} (i - \bar{i})^r (j - \bar{j})^s \quad \text{para } r, s \in \mathcal{N} \quad (6.8)$$

Muy conocidos en la teoría de reconocimiento de patrones son las características derivadas de los momentos centrales, denominados momentos de Hu [27, 49]:

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ &\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ \phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + \\ &\quad 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - \\ &\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \quad (6.9)$$

con

$$\eta_{rs} = \frac{\mu_{rs}}{\mu_{00}^t} \quad t = \frac{r+s}{2} + 1.$$

Los momentos de Hu son invariantes a la traslación, rotación y escalamiento. Esto quiere decir que dos regiones que tengan la misma forma pero que sean de distinto tamaño y que estén ubicados en posiciones y orientaciones distintas en la imagen tendrán momentos de Hu iguales.

A veces sin embargo, es necesario contar con características que además sean invariantes a las transformadas afines. Un conjunto alternativo de características que son invariantes a la traslación, rotación, escalamiento y también a transformaciones afines se puede derivar de los momentos de segundo y tercer orden [48]:

$$I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4}$$

$$I_2 = \frac{\mu_{30}^2 \mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 + 4\mu_{21}^3\mu_{03} - 3\mu_{21}^2\mu_{12}^2}{\mu_{00}^{10}}$$

$$I_3 = \frac{\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2)}{\mu_{00}^7} \quad (6.10)$$

$$I_4 = \frac{(\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} + 9\mu_{20}^2\mu_{02}\mu_{12}^2 + 12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} + 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} - 8\mu_{11}^3\mu_{30}\mu_{03} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + 9\mu_{20}\mu_{02}^2\mu_{21} + 12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} + \mu_{02}^3\mu_{30}^2)/\mu_{00}^{11}}$$

• **Descriptores de Fourier:** Una buena caracterización de la forma de una región se logra utilizando los *descriptores de Fourier* [5, 43, 51]. Las coordenadas (i_k, j_k) de los píxeles del borde, para $k = 0, \dots, L - 1$ de una región se agrupan en un sentido de giro conformando un número complejo $(i_k + j \cdot j_k)$ con $j = \sqrt{-1}$, donde L es el perímetro de la región definido como el número de píxeles del borde de la región. La línea continua formada por estas coordenadas corresponden a una señal periódica que puede ser transformada al dominio de Fourier por medio de la Transformada Discreta de Fourier (DFT) [4]:

$$F_n = \sum_{k=0}^{L-1} (i_k + j \cdot j_k) e^{-j \frac{2\pi kn}{L}} \quad \text{para } n = 0, \dots, L - 1. \quad (6.11)$$

Los descriptores de Fourier corresponden al módulo de los coeficientes complejos de Fourier. Como se puede apreciar los descriptores de Fourier son invariantes a la rotación de la región. El primer descriptor de Fourier $|F_0|$ da información de la ubicación de la región en la imagen. Los descriptores que son invariantes a la posición de la región son los siguientes descriptores. La fase de los coeficientes de Fourier proporciona información acerca de la orientación y de la simetría de las regiones.

En la Figura 6.5 se muestran los descriptores de Fourier para el ejemplo de la Figura 6.4. En este ejemplo el píxel de partida es $(i_0, j_0) = (6, 10)$. En el caso de un círculo ideal los descriptores serían $|F_n| = 0$ para $1 < n < L$, ya que la representación de las coordenadas (i_k, j_k) corresponderían a una senoide perfecta.

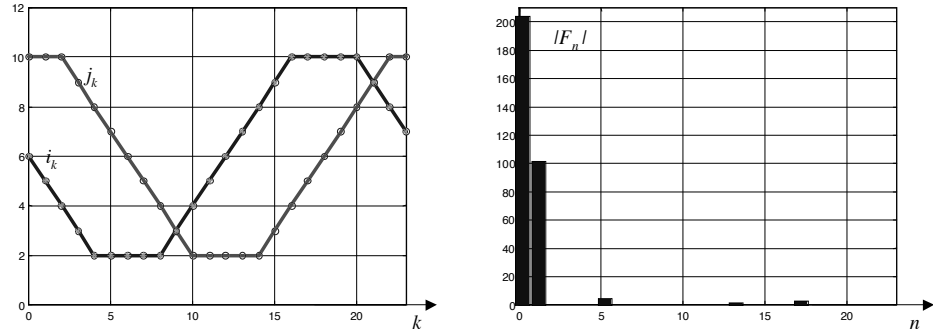


Figura 6.5: Coordenadas del borde de la región de la Figura 6.4 y sus descriptores de Fourier.

6.2.5. Características de color

Antes de entrar a definir las características del color es necesario saber si la imagen que se pretende analizar es a color o en blanco y negro. En el primer caso el color se descompone en tres componentes (rojo, verde y azul) para cada píxel de la imagen, en el segundo caso se cuenta sólo con el tono de gris en cada píxel. Las características que se mencionan a continuación son para una sola variable de color. Esta variable puede ser cada una de las componentes del color, una combinación lineal de las tres componentes o bien simplemente el tono de gris. La información entonces necesaria para calcular estas características es el valor de esta variable de color en cada píxel que es representada como $x[i, j]$ para el píxel (i, j) de la imagen. Es posible que en algunas aplicaciones sea necesario analizar de manera independiente dos variables de color, por ejemplo la componente en rojo y la componente en azul. En este tipo de aplicaciones será necesario extraer las características de color para cada una de las variables de color requeridas.

• **Color promedio (G):** Esta característica es el promedio de la variable de color que se define como:

$$G = \frac{1}{A} \sum_{i,j \in \mathfrak{R}} x[i, j] \quad (6.12)$$

donde \mathfrak{R} denota el conjunto de píxeles de la región y $x[i, j]$ el valor de la variable de color en el píxel (i, j) .

El número de píxels de la región \mathfrak{R} es A , el área de la región. Una representación 3D de la variable de color de una región y su entorno se muestra en la Figura 6.4c. En este caso se trabaja con el valor de gris, ya que la imagen es blanco y negro. Para este ejemplo el promedio es $G = 121,90$ ($G=0$ significa 100 % negro y $G=255$ corresponde a 100 % blanco).

- **Gradiente promedio en el borde (C):** Esta característica toma el valor promedio del gradiente de la variable de color en el borde de la región. Con esta característica se puede medir qué tan abrupto es el cambio en la coloración en la región con respecto a su entorno. El gradiente promedio en el borde se calcula como:

$$C = \frac{1}{L} \sum_{i,j \in \ell} x'[i, j] \quad (6.13)$$

donde $x'[i, j]$ es el módulo del gradiente de la variable de color del píxel (i, j) . Los píxels a evaluar pertenecen exclusivamente al borde. Estos píxels conforman el conjunto ℓ . El número de píxels del conjunto ℓ es L , el perímetro de la región. El gradiente puede ser calculado utilizando el operador de gradiente de Gauss [4], en este caso para el ejemplo de la Figura 6.4 $C = 35,47$.

- **Promedio de la segunda derivada (D):** Esta característica se calcula como el promedio de la segunda derivada de la variable de color en la región:

$$D = \frac{1}{A} \sum_{i,j \in \mathfrak{R}} x''[i, j] \quad (6.14)$$

donde $x''[i, j]$ denota el módulo de la segunda derivada de la variable de color en el píxel (i, j) , y \mathfrak{R} el conjunto de píxels que pertenecen a la región. Para calcular la segunda derivada se puede utilizar el operado LoG (*Laplacian-of-Gauss*) [32, 4]. Es necesario observar que $D < 0$ significa que la región es más clara que su entorno (es decir que su variable de color es mayor en la región que fuera de ella). Así mismo $D > 0$ indica una región más oscura que su entorno.

- **Contraste:** El contraste de una región es concebido como una medida para la diferencia de color entre la región y su entorno. ‘Región’ y ‘entorno’ no tienen píxels en común, y conforman una ‘zona’, que puede ser definida como un rectángulo: La zona entonces queda definida como la ventana

$$g[i, j] = x[i + i_r, j + j_r] \quad (6.15)$$

para $i = 1, \dots, 2h+1$ y $j = 1, \dots, 2w+1$, donde h y w representan la altura y el ancho de la región respectivamente (ver ecuación (6.4)). Los puntos centrales de estas zonas se definen como $i_r = \bar{i} - h - 1$ y $j_r = \bar{j} - b - 1$, donde (\bar{i}, \bar{j}) corresponde al centro de gravedad de la región (ver ecuación (6.7)).

Entre más pequeña sea la diferencia de la variable de color en la región con respecto a su entorno, más pequeño será el contraste. Para visualizar el contraste se pueden representar la variable de color de una zona como una función 3D, donde el eje x y el eje y representan el eje i y el eje j de la imagen, y el eje z el valor de la variable de color que toma el pixel correspondiente, es decir $g[i, j]$. La Figura 6.4c muestra esta representación para el ejemplo de la Figura 6.4a. Se reconoce en este ejemplo una región de alto contraste.

El contraste se define matemáticamente de diversas formas. Una definición comúnmente usada es utilizando características de textura [4], que serán explicadas posteriormente. Otras definiciones de contraste [30, 48] se dan a continuación:

$$K_1 = \frac{G - G_e}{G_e}, \quad K_2 = \frac{G - G_e}{G + G_e} \quad \text{y} \quad K_3 = \ln(G/G_e), \quad (6.16)$$

donde G y G_e representan el promedio de la variable de color en la región y en el entorno respectivamente.

Una nueva forma de calcular el contraste se muestra en [33]. El cálculo de esta característica se obtiene en tres pasos:

i) Extracción del color en los ejes principales de la zona: se calculan dos funciones de color P_1 y P_2 . La primera función P_1 toma los valores de la variable de color en la dirección i y la segunda función P_2 en la dirección j . Ambas funciones se centran en los centros de gravedad. En el ejemplo de la Figura 6.4b el centro de gravedad está en el píxel (6,6), esto quiere decir que P_1 y P_2 son los valores del tono de gris de la zona de la columna 6 y de la fila 6 respectivamente, tal como se muestra en la Figura 6.6a-b.

ii) Aislamiento de la región: Para aislar la región de su entorno se trata de eliminar el fondo de la región, que se modela como una función lineal de primer orden, es decir una rampa. Se asume entonces que los valores extremos de P_1 y P_2 pertenecen las rampas R_1 y R_2 , tal como se ilustra en la Figura 6.6a-b. Las rampas serán substraídas de las funciones originales para conformar $Q_1 = P_1 - R_1$ y $Q_2 = P_2 - R_2$ que se fusionan en la nueva función Q como se muestra en la Figura 6.6c.

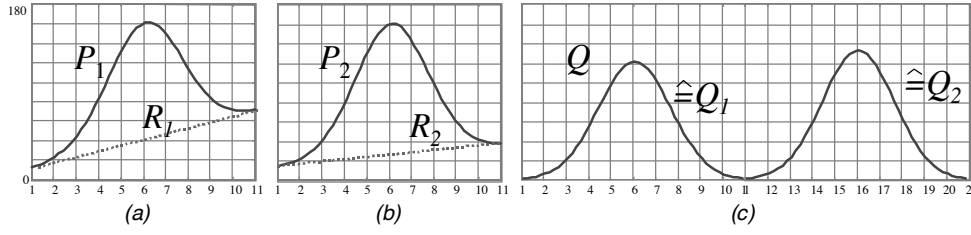


Figura 6.6: Cálculo del contraste para la Figura 6.4: a) valor de gris de la zona y rampa en dirección i ; b) valor de gris de la zona y rampa en dirección j ; c) fusión de las funciones de a) y b) sin rampas.

iii) Cálculo del contraste: a partir de la nueva función Q se definen dos nuevos contrastes:

$$K_\sigma = \sigma_Q \quad \text{y} \quad K = \ln(Q_{\max} - Q_{\min}) \quad (6.17)$$

donde σ_Q , Q_{\max} y Q_{\min} representan la desviación estándar, el máximo y el mínimo de Q respectivamente.

• **Momentos:** Los momentos definidos en la Sección 6.2.4 pueden ser utilizados en el análisis de color de la región si en la ecuación del cálculo de los momentos (6.6) se incorpora la información de la variable de color:

$$m'_{rs} = \sum_{i,j \in \mathcal{R}} i^r j^s x[i,j] \quad \text{para } r, s \in N. \quad (6.18)$$

La sumatoria se calcula sobre los pixels (i, j) de la región. De esta manera se pueden calcular las características indicadas en (6.9) y (6.10) que incorporen la información del color según (6.18), el resultado será el conjunto de características $\phi'_1 \dots \phi'_7$ e $I'_1 \dots I'_4$.

• **Características de textura:** Las características de textura proporcionan información sobre la distribución espacial del color en la imagen. Para el análisis de regiones se pueden aplicar las características de textura no a la imagen entera sino sólo a las zonas (región y entorno) como se definió en (6.15).

Una característica simple de textura es la varianza local [28] definida como:

$$\sigma_g^2 = \frac{1}{4hb + 2h + 2b} \sum_{i=1}^{2h+1} \sum_{j=1}^{2b+1} (g[i,j] - \bar{g})^2 \quad (6.19)$$

donde \bar{g} denota el valor promedio de la variable de color en la zona.

Otras características de textura se obtienen por medio de la *matriz de coocurrencia*². La matriz de coocurrencia se denotará como \mathbf{P}_{kl} , donde el elemento $P_{kl}[i, j]$ otorga el valor de frecuencia (divido por N_T) de ocurrencia de los valores de color i y j en dos píxels ubicados en una posición relativa dada por el vector (k, l) . La variable N_T significa el número de píxels que fueron necesarios para calcular \mathbf{P}_{kl} , con esto se normaliza la matriz de coocurrencia ya que la suma de todos sus elementos es uno.

Si la variable de color tiene una resolución de 256, por ejemplo de 0 a 255, el tamaño de la matriz de coocurrencia \mathbf{P}_{kl} será 256×256 . Ya que esto implica un costo computacional muy alto, es común que se utilicen matrices más pequeñas empleando sólo los bits más significativos de la variable de color [4]. A manera de ejemplo, se puede tener una matriz de coocurrencia de 8×8 agrupando el valor de la variable de color x en $[0, \dots, 31]$, $[32, \dots, 63]$, ... $[224, \dots, 255]$.

Algunas características de textura para imágenes (o zonas) cuyas matrices de coocurrencia sean de $N_x \times N_x$ elementos se presentan a continuación [4, 48, 17, 11]:

Entropía:

$$H_{kl} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_x} P_{kl}[i, j] \log(P_{kl}[i, j]) \quad (6.20)$$

Inercia o contraste:

$$I_{kl} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_x} (i - j)^2 P_{kl}[i, j] \quad (6.21)$$

Homogenidad o energía:

$$E_{kl} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_x} [P_{kl}[i, j]]^2 \quad (6.22)$$

Momento de diferencia inverso:

$$Z_{kl} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_x} \frac{P_{kl}[i, j]}{1 + (i - j)^2} \quad (6.23)$$

²del inglés co-occurrence matrix.

6.2.6. Criterios de correspondencia entre regiones

Para establecer si dos regiones r_1 y r_2 , en dos imágenes distintas, son correspondientes se utilizan los siguientes criterios:

- **Condición epipolar:**

Los centros de gravedad de las regiones deben satisfacer la condición epipolar .

Como la condición epipolar se aplica a puntos (y no a regiones) es necesario simplificar la región en un punto. Esto se hace tomando el centro de gravedad de la región. Esta simplificación está sujeta a un error, que muchas veces puede ser considerablemente grande, ya que los centros de gravedad no necesariamente corresponden a la proyección del mismo punto 3D en el espacio. A manera de ejemplo esta simplificación es cierta si se tratara de regiones esféricas, donde la proyección del centro de gravedad de la esfera coincide con los centros de gravedad de las regiones en las imágenes. Sin embargo, las regiones reales no son proyecciones de esferas, por esta razón el uso de este criterio debe usarse sabiendo que está sujeto a error. Este criterio sólo debe usarse si se trabaja con regiones circulares o si la rotación relativa del objeto (o de las cámaras) en ambas imágenes es pequeña.

Si m_1 y m_2 son los centros de gravedad de r_1 y r_2 respectivamente, es necesario comprobar entonces si satisfacen la condición epipolar. Para esto se utiliza comúnmente la restricción epipolar práctica, que señala que m_1 y m_2 satisfacen la condición epipolar si m_2 se encuentra a una distancia Euclídea de la línea epipolar de m_1 en la segunda imagen a una distancia pequeña (ver ecuación (4.32)).

$$d_2(\mathbf{m}_1, \mathbf{F}, \mathbf{m}_2) = \frac{|\mathbf{m}_2^T \mathbf{F} \mathbf{m}_1|}{\sqrt{l_1^2 + l_2^2}} < \varepsilon_2 \quad (6.24)$$

donde \mathbf{F} la matriz fundamental existente entre la imagen 1 y 2, y $[l_1 \ l_2 \ l_3]^T = \mathbf{F} \mathbf{m}_1$.

- **Coloración correcta de las regiones con su entorno:**

Ambas regiones deben ser o bien más claras, o bien más oscuras que su entorno.

Si una región en la imagen 1 es más clara que su entorno se puede

pensar que en la imagen 2 su región correspondiente también debería ser más clara que su entorno (suponiendo la misma iluminación). El mismo criterio se puede aplicar para regiones que sean más oscuras que su entorno. Por esta razón, para que las regiones sean correspondientes, la coloración relativa de las regiones con respecto a su entorno debe ser la misma. Para expresar matemáticamente este criterio se puede decir que el signo del promedio de la segunda derivada de la variable de color (ver ecuación (6.14)) de las regiones debe ser igual:

$$\text{sgn}(D_1) = \text{sgn}(D_2) \quad \text{con} \quad \text{sgn}(x) \begin{cases} 1 & \text{para } x > 0 \\ 0 & \text{para } x = 0 \\ -1 & \text{para } x < 0 \end{cases} \quad (6.25)$$

donde D_1 y D_2 son los promedios de la segunda derivada de las regiones.

■ Criterio de similitud:

Las regiones deben ser ‘similares’.

Antes de investigar qué tan similares son las regiones, es necesario hacer un estudio de cuales características son las que proporcionan información relevante, esto se logra haciendo un análisis estadístico [29, 14].

Suponiendo que se hayan extraído n características³ de cada una de las regiones r_1 y r_2 se tiene entonces los vectores de características $\mathbf{z}_1 = [z_{11} \ z_{12} \ \dots \ z_{1n}]^T$ y $\mathbf{z}_2 = [z_{21} \ z_{22} \ \dots \ z_{2n}]^T$ para r_1 y r_2 respectivamente. El criterio de similitud evalúa la distancia Euclídea ente los vectores de características \mathbf{z}_1 y \mathbf{z}_2 . De esta manera se considera que las regiones son ‘similares’ si cumplen:

$$S_d(\mathbf{z}_1, \mathbf{z}_2) = \|\mathbf{z}_1 - \mathbf{z}_2\| = \sqrt{\sum_{i=1}^n (z_{1i} - z_{2i})^2} < \varepsilon_s \quad (6.26)$$

donde ε_s es un valor pequeño.

Características de fácil computación y que sirven en muchos casos para establecer si dos regiones son similares, son el área (A), el perímetro (L), el color promedio (G) y el contraste (K) [33].

³Por lo general se utilizan características normalizadas. Un criterio de normalización es desplazando los valores que toman las características a un rango entre 0 y 1. Otro criterio es obtener nuevas características que tengan media igual a cero y varianza igual a uno.

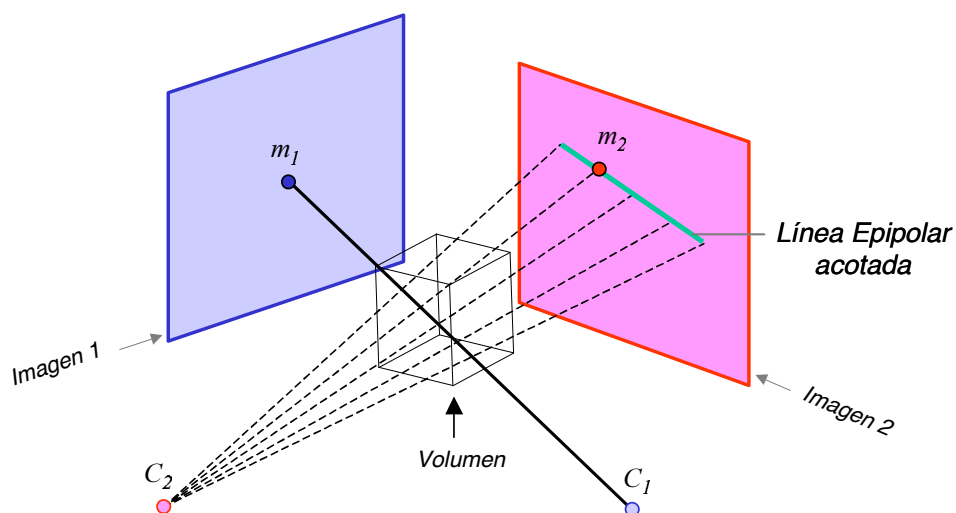


Figura 6.7: Línea epipolar acotada: la reconstrucción de M a partir de m_1 a m_2 debe pertenecer a un subespacio 3D.

■ **Localización correcta en el espacio 3D:**

El punto 3D reconstruido, obtenido por triangulación a partir de los centros de gravedad m_1 y m_2 de las regiones, debe encontrarse dentro del espacio 3D ocupado por el objeto de análisis.

Muchas veces se sabe a-priori dónde está ubicado en el espacio el objeto de análisis. Este espacio puede corresponder a un cubo, cilindro o a un volumen más complejo almacenado como un modelo CAD. Si se cuenta con esta información, es posible entonces realizar una reconstrucción del punto 3D M utilizando el método lineal explicado en la Sección 5.1 y verificar si las coordenadas de este punto pertenecen al volumen que ocupa el objeto de análisis.

Este criterio corresponde a evaluar la condición epipolar sólo en una porción de la imagen. Esta línea epipolar acotada se obtiene a partir de la proyección en la segunda imagen del segmento de recta $\langle C_1, m_1 \rangle$ que está en el subespacio 3D donde se encuentra el objeto de análisis (ver Figura 6.7).

Evaluando estos cuatro criterios se puede establecer con un grado mayor de certeza si las regiones r_1 y r_2 son correspondientes entre sí.

Capítulo 7

Tracking en una Secuencia

A continuación se presentará un algoritmo para realizar un seguimiento de regiones en una secuencia de imágenes. El problema aquí mencionado consiste en hacer el seguimiento de una serie de regiones que han sido detectadas en una secuencia de imágenes. No se sabe a priori cuáles son las regiones correspondientes, sólo se conoce las características extraídas de estas regiones y la geometría de la proyecciones de las vistas. Se asumirá entonces que se trata de una secuencia calibrada.

El seguimiento explicado en esta Sección será ilustrado con una secuencia de imágenes radioscópicas de una pieza de aluminio [38]. Las imágenes radioscópicas son utilizadas para llevar a cabo un control de calidad de la pieza, un ejemplo se puede apreciar en la Figura 7.1. En cada imagen se detectan muchas fallas hipotéticas y es necesario establecer cuáles de ellas son fallas reales y cuáles son falsas alarmas. El seguimiento de las fallas hipotéticas es un buen criterio para la eliminación de las falsas alarmas, de esta manera se establece que sólo las fallas hipotéticas que pueden ser seguidas en la secuencia son consideradas como fallas reales. Las Figuras 7.2 y 7.3 muestran una secuencia radioscópica y al respectiva segmentación de fallas hipotéticas. En esta secuencia sólo existen tres fallas reales. En esta secuencia la imagen número 5 coincide con la Figura 7.1.

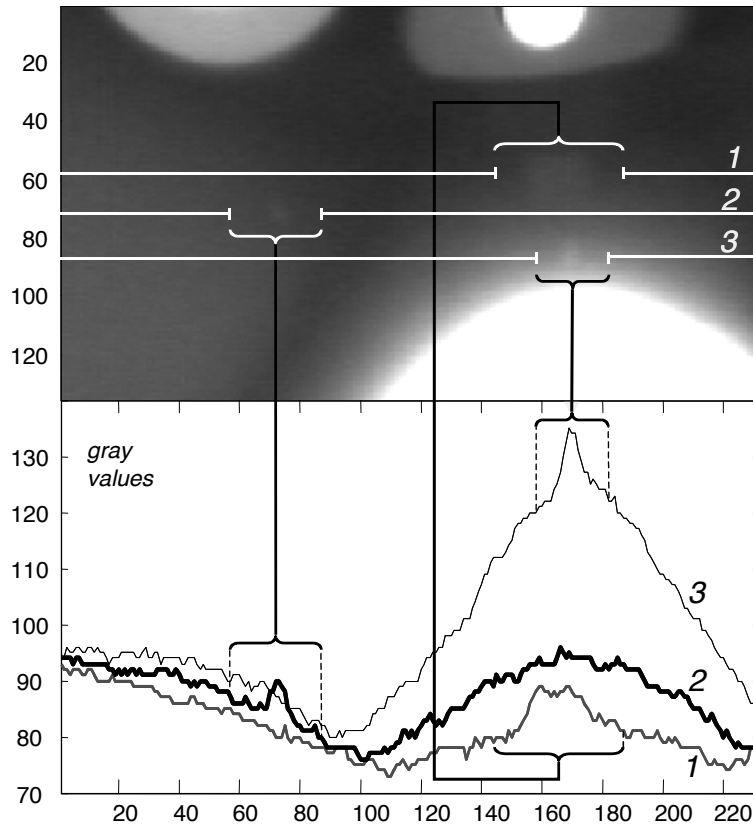


Figura 7.1: Ejemplo de una imagen radioscópica con tres fallas.

7.1. Definiciones

En esta sección se definirán las variables que serán utilizadas en el algoritmo de seguimiento de regiones.

N : Número de imágenes en las cuales se hará el seguimiento (número de imágenes de la secuencia).

N_1 : Número total de regiones detectadas en las N imágenes de la secuencia.

t_i : Número de la imagen en la que se encuentra la región i , para $i = 1, \dots, N_1$. El vector \mathbf{t} , de N_1 elementos, se define como $[t_1 \dots t_{N_1}]^T$.

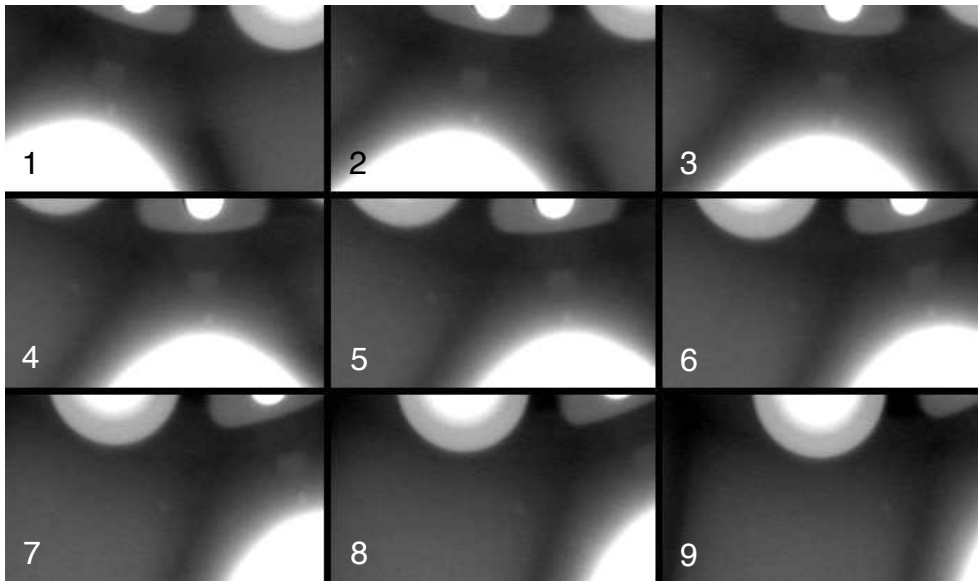


Figura 7.2: Ejemplo de una secuencia de imágenes.

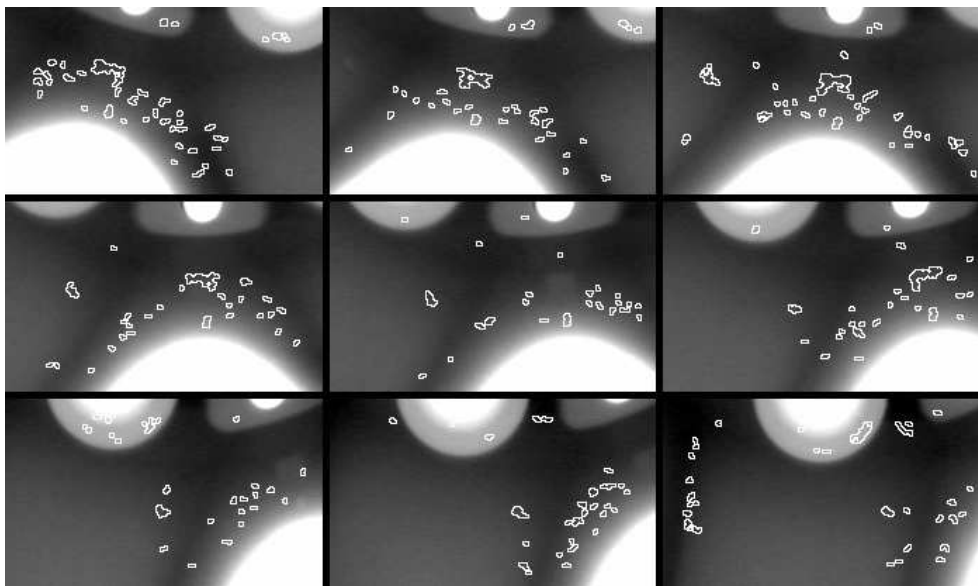


Figura 7.3: Detección de fallas hipotéticas en una secuencia.

- e_j, f_j : Número de la primera y última región detectada en la imagen j de la secuencia, para $j = 1, \dots, N$. Los vectores \mathbf{e} y \mathbf{f} de N elementos se definen como $[e_1 \dots e_N]^T$ y $[f_1 \dots f_N]^T$, donde $e_1 = 1$, $e_k = f_{k-1} + 1$ para $k = 2, \dots, N$ y $f_N = N_1$.
- (x_i, y_i) : Coordenadas del centro de gravedad de la región i en un sistema de coordenadas relativas a la imagen. El vector homogéneo correspondiente es $\mathbf{m}_i = [x_i \ y_i \ 1]^T$.
- \mathbf{z}'_i : Vector que contiene n' características extraídas de la región i : $\mathbf{z}'_i = [z'_{i1} \dots z'_{in'}]^T$.
- N_2, N_3, N_4 : Número de enlaces encontrados al realizar un seguimiento en 2, 3 y 4 imágenes respectivamente.
- B, C, D: Matrices en las que se encuentran los números de las regiones que han sido enlazadas en 2, 3 y 4 imágenes respectivamente.

7.2. Matching en dos imágenes

Una región detectada puede ser considerada como la proyección en la imagen de un (porción de un) objeto 3D. Como el objeto puede ser proyectado en varias imágenes de la secuencia, es posible que una región en una imagen tenga una región correspondiente en otra imagen. Para establecer la correspondencia entre regiones en dos imágenes distintas se pueden estudiar los cuatro criterios explicados en la Sección 6.2.6: *condición epipolar*, *coloración correcta de las regiones con su entorno*, *criterio de similitud* y *localización correcta en el espacio 3D*.

Para llevar a cabo el tracking de las regiones en la secuencia, se comienza con establecer cuáles regiones son correspondientes en dos imágenes. Esta correspondencia se determina evaluando los cuatro criterios mencionados en todas las regiones a y b de dos imágenes distintas $p = t_a$ y $q = t_b$ ($p \neq q$) en m imágenes consecutivas de la secuencia, para $p = 1, \dots, N - m$; $q = p + 1, \dots, p + m$; $a = e_p, \dots, f_p$ y $b = e_q, \dots, f_q$, como se muestra en la Fig. 7.4. Si una región no encuentra un matching con ninguna otra región en la secuencia será considerada como una región que no tiene correspondencia y por lo tanto no podrá ser seguida en la secuencia. Estas regiones serán eliminadas y no serán tomadas en cuenta en los pasos posteriores.

1. $k = 0$
2. para $p = 1, \dots, N - 1$
3. para $a = e_p, \dots, f_p$
4. para $q = p + 1, \dots, \min(p + m, N)$
5. para $b = e_q, \dots, f_q$
6. si se satisfacen los cuatro criterios de correspondencia para las regiones a y b entonces:
 $k = k + 1$ y $B_k = [a \ b]$.
7. fin lazos b, q, a y p .

Figura 7.4: Algoritmo para el matching en dos imágenes.

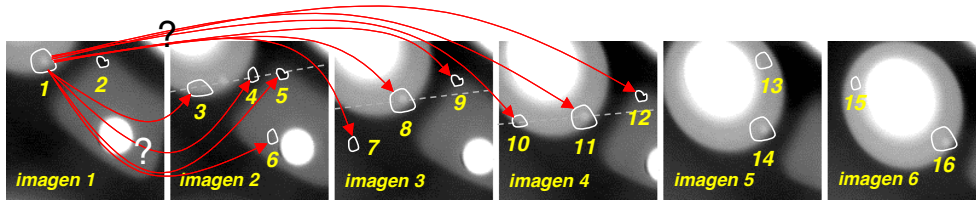


Figura 7.5: Ejemplo de matching en dos regiones. Las líneas epipolares del centro de gravedad de la primera región se representan en las imágenes 2, 3 y 4 como líneas punteadas.

Como se ilustra en la Fig. 7.4, se define una matriz B de $N_2 \times 2$ elementos, cuya k -ésima fila $[b_{k1} \ b_{k2}]$ significa un matching entre la regiones b_{k1} y la region b_{k2} (para $k = 1, \dots, N_2$), donde N_2 significa el número de pares de regiones que encontraron matching en la secuencia.

En el ejemplo de la Fig. 7.5 se muestra el matching de una región en la imagen 1 con las regiones de las siguientes imágenes. En este ejemplo, $N = 6$, $m = 3$, $N_1 = 16$, $\mathbf{t} = [1 \ 1 \ 2 \ 2 \ 2 \ 2 \ 3 \ 3 \ 3 \ 4 \ 4 \ 4 \ 5 \ 5 \ 6 \ 6]^T$, $\mathbf{e} = [1 \ 3 \ 7 \ 10 \ 13 \ 15]^T$ y $\mathbf{f} = [2 \ 6 \ 9 \ 12 \ 14 \ 16]^T$.

A partir de la región 1 detectada en la imagen 1 se evalúa entonces cuáles de las regiones de las siguientes tres imágenes pueden ser sus correspondientes. Como la región 1 es más clara que su entorno, entonces las regiones 5 y 9 no cumplen el criterio del contraste parecido ya que figuran como más oscuras. Adicionalmente las regiones 6, 7 y 12 no satisfacen la condición epipolar ya que no están sobre la recta epipolar correspondiente. Por otra parte la región 4 no cumple el criterio de la similitud, ya que la región en comparación con la región 1 es mucho más oscura. Por último, el punto 3D reconstruido a partir de las posibles regiones restantes no reside dentro del volumen de la pieza analizada para la región 10. Esto quiere decir que sólo las regiones 3, 8 y 11 satisfacen los cuatro criterios de matching. Este procedimiento se repite para las siguientes regiones y se obtiene finalmente la Matriz \mathbf{B} con los siguientes $N_2=17$ enlaces:

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 5 & 5 & 8 & 8 & 8 & 9 & 11 & 11 & 14 \\ 3 & 8 & 11 & 5 & 9 & 12 & 8 & 11 & 9 & 12 & 11 & 14 & 16 & 12 & 14 & 16 & 16 \end{bmatrix}^T$$

La consideración de m imágenes consecutivas en este análisis puede superar el problema de no segmentación y oclusión de las regiones. Una región que es detectada en la imagen p y no es detectada en la imagen $p + 1$ puede igualmente ser seguida en la secuencia si aparece en la imagen $p + 2, \dots, p + m$. En el ejemplo de la Fig. 7.5 no se encontró la región correspondiente a la región 1 en la segunda imagen, sin embargo si se encontró un enlace entre la región 1 y regiones de la tercera y cuarta imagen.

Después de este procedimiento muchas falsas detecciones fueron eliminadas, sin discriminar las fallas reales existentes en la secuencia. Sin embargo, existe todavía un número considerable de enlaces producidos de falsas alarmas, como el caso del matching entre las regiones 1 y 3. Un resultado en una secuencia verdadera se muestra en 7.6.

7.3. Seguimiento en más imágenes

En el seguimiento se determinan *trayectorias* de regiones correspondientes en la secuencia de imágenes. A partir de los matchings encontrados en dos imágenes se puede ahora buscar si existen trayectorias con más de dos regiones, cuyos centros de gravedad son proyecciones de un mismo punto 3D.

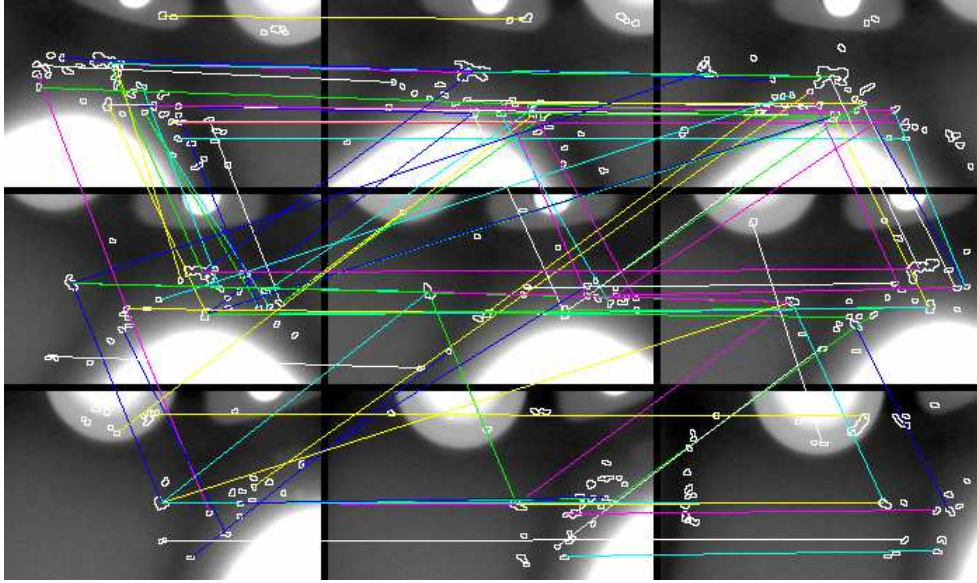


Figura 7.6: Matching de regiones en una secuencia real.

Como se hizo en el procedimiento anterior, aquellas regiones que no pueden ser seguidas en la secuencia serán clasificadas como falsas detecciones.

7.3.1. Seguimiento en tres imágenes

A continuación se buscan todos los posibles enlaces de tres regiones en la matriz B (definida en la Sección 7.2) que cumplan la condición de correspondencia en tres imágenes. Se busca entonces las filas i y j de la matriz B (para $i, j = 1, \dots, N_2$ y $i \neq j$), que

$$b_{i2} = b_{j1}. \quad (7.1)$$

En el caso de que las filas i y j cumplen la condición (7.1), es decir si $B_i = [a \ b]$ y $B_j = [b \ c]$, se establecerá entonces la correspondencia de las tres regiones (a, b, c) con coordenadas \mathbf{m}_a , \mathbf{m}_b y \mathbf{m}_c en las imágenes $p = t_a$, $q = t_b$ y $r = t_c$, si la condición trifocal (4.51) es verdadera:

$$\|\hat{\mathbf{m}}_c - \mathbf{m}_c\| < \varepsilon_3, \quad (7.2)$$

```

1.  $k = 0$ 

2. para  $p = 1, \dots, N - 1$ 

3. para  $i = 1, \dots, N_2 - 1$ 

4. para  $j = i + 1, \dots, N_2$ 

5. si  $b_{i2} = b_{j1}$ 

    5.1  $a = b_{j1}$ 
         $b = b_{j2}$ 
         $c = b_{i2}$ 

    5.2 si  $\| \hat{\mathbf{m}}_c - \mathbf{m}_c \| < \varepsilon_3$ 

        5.2.1  $k = k + 1$ 
        5.2.2  $C_k = [a \ b \ c]$ 

6. fin lazos  $j, i$  y  $p$ .

```

Figura 7.7: Algoritmo para el seguimiento en tres imágenes.

donde $\hat{\mathbf{m}}_c$ representa la reproyección de las coordenadas de la tercera región en la imagen r . La estimación de $\hat{\mathbf{m}}_c$ se realiza a partir del Teorema 4.1, que señala que sólo es necesario evaluar las dos primeras (o últimas) trilinearidades de Shashua (4.44), ya que \mathbf{m}_a y \mathbf{m}_b ya cumplen la condición epipolar (ver procedimiento Matching). A partir de \mathbf{m}_a y \mathbf{m}_b se calcula $\hat{\mathbf{m}}_c$ utilizando la ecuación (4.52).

Los *tríos* encontrados se agrupan en una nueva matriz de $N_3 \times 3$ elementos: $C = [c_{k1} \ c_{k2} \ c_{k3}]$, para $k = 1, \dots, N_3$. El algoritmo para el seguimiento en tres imágenes es mostrado en la Figura 7.7. El lazo para j puede empezar a partir de $i + 1$ ya que la matriz B ya está ordenada de menor a mayor. El seguimiento en tres imágenes para el ejemplo de la Figura 7.5 se encuentra representado en la Figura 7.8. En este ejemplo, la matriz C es:

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 5 & 8 & 8 & 8 & 11 \\ 3 & 3 & 8 & 8 & 11 & 11 & 5 & 5 & 9 & 9 & 11 & 11 & 14 & 14 \\ 8 & 11 & 11 & 14 & 14 & 16 & 9 & 12 & 12 & 12 & 14 & 16 & 16 & 16 \end{bmatrix}^T$$

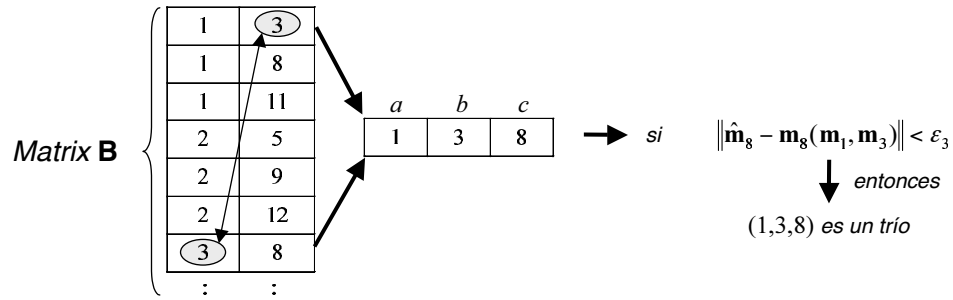


Figura 7.8: Ejemplo de seguimiento en tres imágenes.

En este caso $N_3 = 14$. Se aprecia que en este ejemplo los tríos $(1, 8, 16)$, $(3, 8, 11)$, $(3, 8, 14)$, $(3, 8, 16)$, $(3, 11, 14)$ y $(3, 11, 16)$ no cumplieron la condición trifocal y por ende no fueron agrupados en la matriz **C**.

En la Figura 7.9 se muestran los enlaces en tres imágenes encontrados en la secuencia real de la Figura 7.6. Se puede observar que un gran número de matchings de falsas alarmas fueron eliminados.

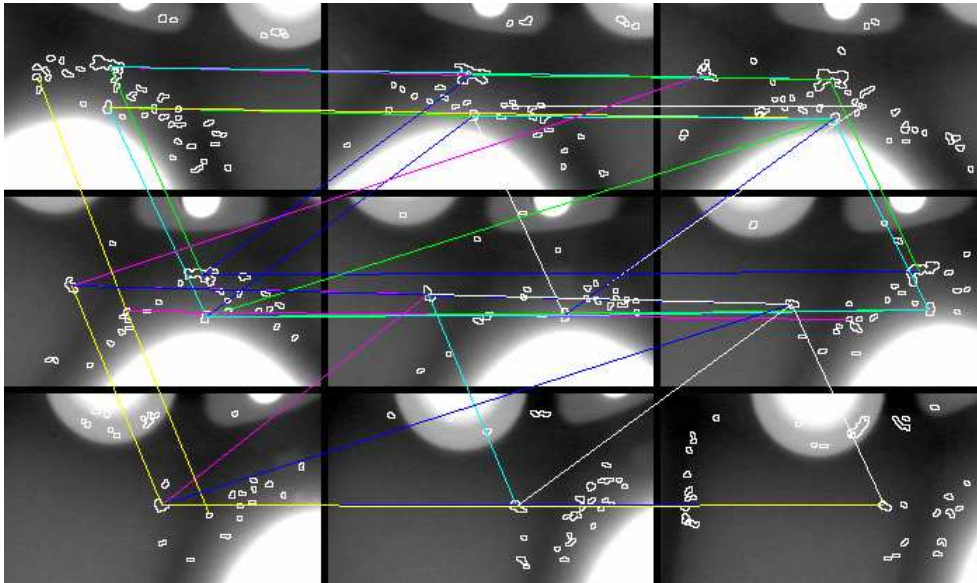


Figura 7.9: Seguimiento de fallas hipotéticas en tres imágenes.

1. $k = 0$
2. para $i = 1, \dots, N_3 - 1$
3. para $j = i + 1, \dots, N_3$
4. si $(c_{i2} = c_{j1})$ y $(c_{i3} = c_{j2})$ entonces
 - 4.1 $k = k + 1$
 - 4.2 $D_k = [a \ b \ c \ d]$
5. fin lazos j e i .

Figura 7.10: Algoritmo para el seguimiento en cuatro imágenes.

7.3.2. Seguimiento en cuatro imágenes

Para encontrar las trayectorias de cuatro regiones, se buscan *cuartetos*, es decir cuatro regiones correspondientes en cuatro imágenes distintas. Para encontrar los cuartetos se puede proceder de la siguiente manera: se buscan las filas i de la matriz **B** y las filas j de la matriz **C** (para $i = 1, \dots, N_2$ y $j = 1, \dots, N_3$) que cumplan la condición:

$$b_{i1} = c_{j3}. \quad (7.3)$$

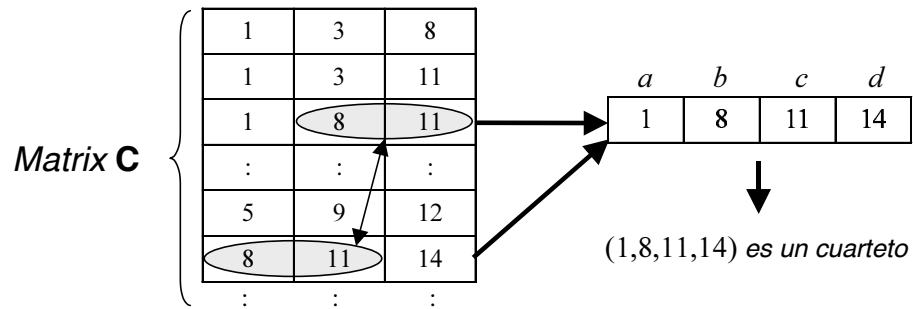


Figura 7.11: Ejemplo de seguimiento en cuatro imágenes.

En caso de que las filas i y j cumplan la condición (7.3) ($B_i = [c \ d]$ y $C_j = [a \ b \ c]$), se puede afirmar que se han encontrado cuatro regiones correspondientes con coordenadas \mathbf{m}_a , \mathbf{m}_b , \mathbf{m}_c y \mathbf{m}_d en las imágenes $p = t_a$, $q = t_b$, $r = t_c$ y $s = t_d$, si las regiones satisfacen la condición de correspondencia en cuatro proyecciones [34]. Sin embargo, los tres primeros puntos \mathbf{m}_a , \mathbf{m}_b und \mathbf{m}_c se sabe que son correspondientes entre sí porque ellos en el seguimiento en tres imágenes ya cumplieron la condición trifocal. Se sabe entonces (a partir del Teorema 4.3) que las cuatro regiones serán correspondientes si \mathbf{m}_b , \mathbf{m}_c y \mathbf{m}_d satisfacen las dos primeras (o últimas) trilinearidades de Shashua (4.44). Como la matriz C contiene en cada fila los tríos encontrados en la secuencia, es decir los grupos de tres regiones que satisfacen la condición de correspondencia, se puede encontrar en ella los grupos de cuatro regiones buscando las filas i y j de C (para $i, j = 1, \dots, N_3$ y $i \neq j$) que

$$c_{i2} = c_{j1} \quad \text{und} \quad c_{i3} = c_{j2}, \quad (7.4)$$

Un diagram de flujo se muestra en la Figura 7.10. Esto quiere decir que $a = c_{i1}$, $b = c_{i2} = c_{j1}$, $c = c_{i3} = c_{j2}$ y $d = c_{j3}$.

Los cuartetos encontrados se agrupan en una nueva matriz de $N_4 \times 4$ elementos: $D = [d_{k1} \ d_{k2} \ d_{k3} \ d_{k4}]$, $k = 1, \dots, N_4$. El seguimiento en cuatro imágenes para el ejemplo de la Figura 7.5 se muestra en la Figura 7.11. En este caso la matriz D es:

$$D = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 8 \\ 8 & 8 & 8 & 11 & 5 & 11 \\ 11 & 11 & 14 & 14 & 9 & 14 \\ 14 & 16 & 16 & 16 & 12 & 16 \end{bmatrix}^T$$

Se observa que $N_4 = 6$. En este ejemplo la falsa alarma de la región 3 no encontró correspondencia alguna en cuatro proyecciones, por lo tanto es eliminada.

El resultado en la secuencia real de la Figura 7.9 se muestra en la Figura 7.12. Una repetición de este procedimiento para cinco imágenes puede llevar a la eliminación de fallas reales ya que no siempre se puede garantizar la segmentación de un falla real en más de cuatro imágenes.

7.3.3. Resumen de trayectorias

Una falla que aparece en más de cuatro imágenes, puede pertenecer a más de un cuarteto. Éste es el caso del ejemplo mostrado en la Figura 7.5, ya que aquí fueron encontrados los cuartetos (1, 8, 11, 14), (1, 8, 11, 16), (1, 8, 14, 16), (1, 11, 14, 16) y (8, 11, 14, 16) que son representaciones de la misma falla en distintas imágenes. Un resumen de estas trayectorias es en este caso: (1, 8, 11, 14, 16), como se muestra en las Figuras 7.13 y 7.14.

Estas trayectorias pueden entonces agruparse en una trayectoria de más de cuatro regiones. El resultado en la secuencia real de la Figura 7.12 se muestra en la Figura 7.15. Se puede apreciar que existen aún dos falsas alarmas.

7.4. Verificación

Una trayectoria representa los enlaces de proyecciones de fallas hipotéticas correspondientes a lo largo de una secuencia de imágenes. Definiendo el término *subseuencia* de una falla como las imágenes de una secuencia en

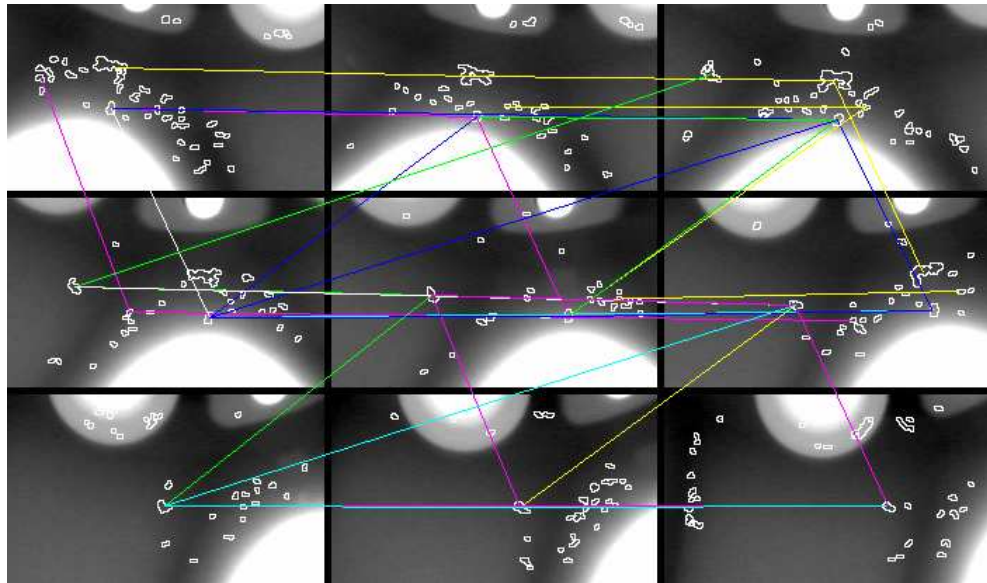


Figura 7.12: Seguimiento de fallas hipotéticas en cuatro imágenes.

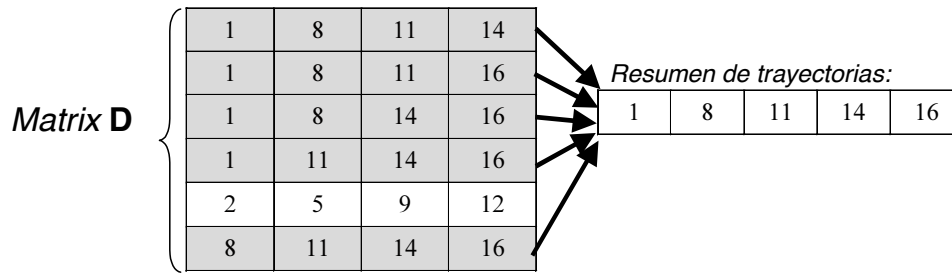


Figura 7.13: Ejemplo de resumen de trayectorias.

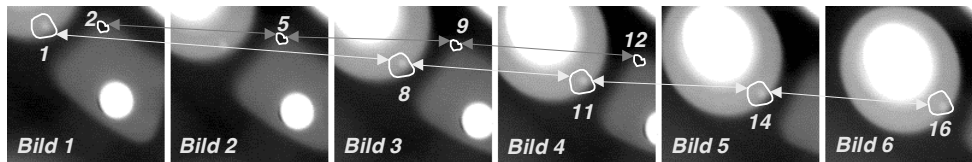


Figura 7.14: Ejemplo de seguimiento en más imágenes.

las que la falla está presente, se puede observar que una trayectoria algunas veces está interrumpida en su subsecuencia. Esto debido a que una falla, a pesar de estar presente en la imagen, no siempre puede ser segmentada. Este es el caso de la trayectoria (1, 8, 11, 14, 16) de la Figura 7.14, donde la falla no fue detectada en la segunda imagen.

A partir de cada trayectoria se estima ahora el punto 3D que corresponde al punto que habría producido cada uno de los centros de gravedad de las regiones de la trayectoria. Esta estimación se realiza con el método de reconstrucción 3D de la Sección 5.2, en el que se estima el punto 3D \hat{M} a partir de las coordenadas de las regiones de la trayectoria. Este punto 3D puede ser proyectado en la imágenes de subsecuencia en las que la segmentación de las fallas no tuvo éxito para obtener así la posición de la falla en cada imagen de la subsecuencia. El tamaño de la falla puede ser estimado como el promedio de los tamaños de las fallas que sí fueron detectadas.

En todas las imágenes de la subsecuencia se puede definir ahora una *zona* en donde los centros de gravedad de las regiones (detectadas y estimadas) son el centro de estas zonas. Una zona, como fue descrita en la Sección 6.2.5

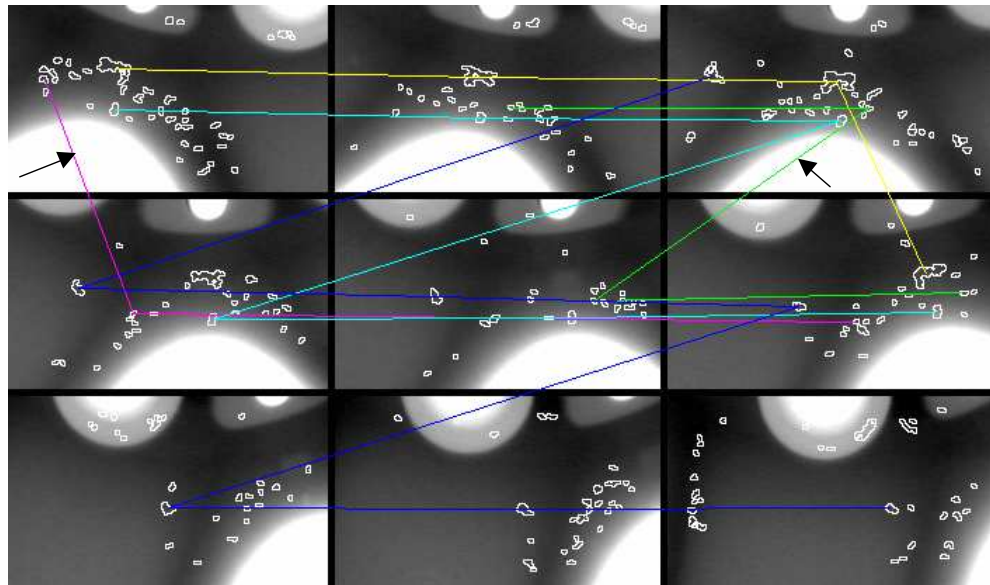


Figura 7.15: Resumen de trayectorias.

(ver ecuación(6.15)), es un rectángulo pequeño que contiene la región y su entorno. La zonas para la primera trayectoria de la Figura 7.14 son mostradas en la Figura 7.16. La zona de la falla no detectada en la segunda imagen fue estimada a partir de la reproyección de $\hat{\mathbf{M}}$ en esta imagen.

Se calcula ahora una zona *suavizada* a partir del promedio de todas las zonas que pertenecen a una trayectoria en la respectiva subsecuencia (ver ejemplo en la Figura 7.16). Esta operación es muy útil para reducir el ruido que pueda estar presente en las imágenes. La verificación consiste en evaluar el contraste de la zona suavizada, si es suficientemente alto se dice entonces que

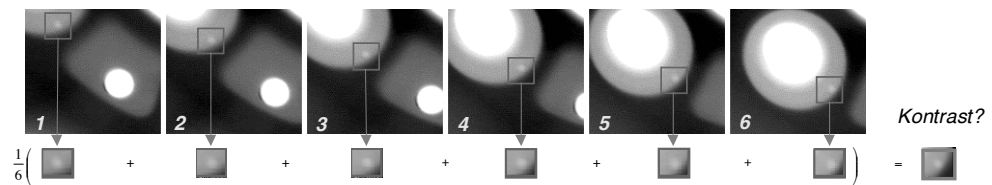


Figura 7.16: Verificación de una trayectoria.

la trayectoria corresponde a una falla real.

La Figura 7.17 muestra la falla reconstruida en la secuencia de imágenes radioscópicas. Se aprecia que la existencia de dos falsas alarmas (comparar con Figura 7.1). Estas dos falsas alarmas se eliminaron después de la etapa de verificación como se aprecia en la Figura 7.18. En este ejemplo se consiguió el objetivo: se eliminaron las falsas alarmas sin discriminar las fallas reales.

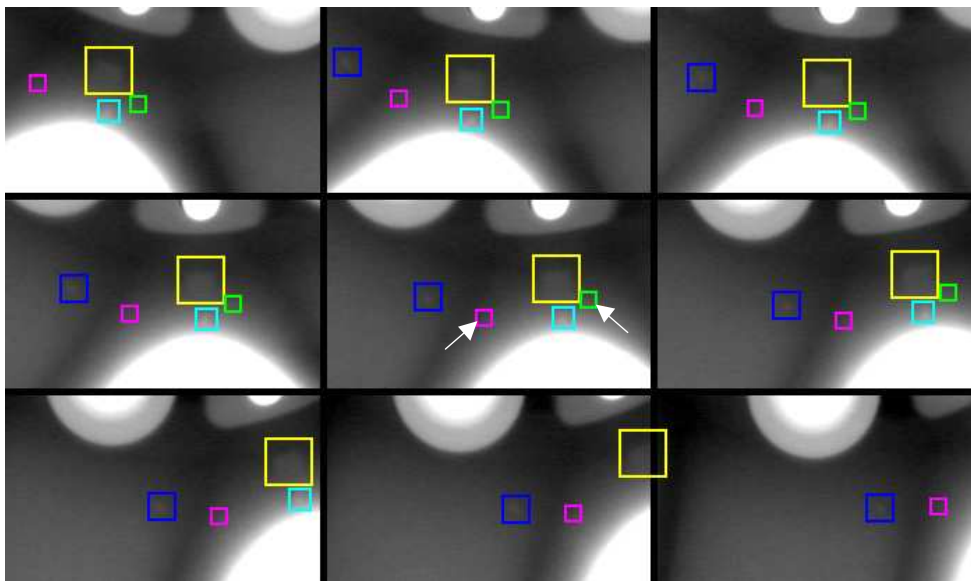


Figura 7.17: Rekonstruierte Fehler. Die Pfeile zeigen zwei Fehldetektionen.

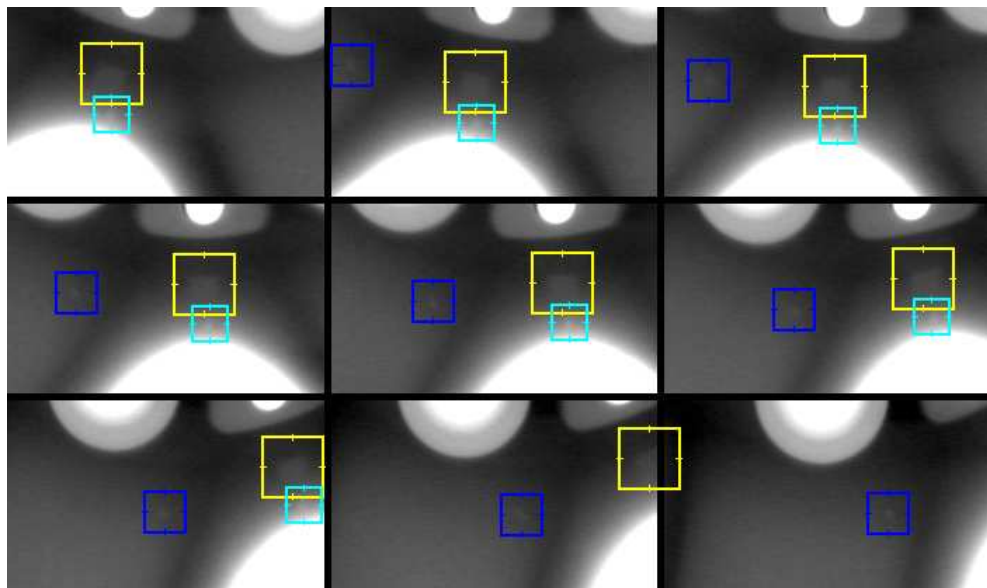


Figura 7.18: Detektierte Fehler.

Índice de figuras

1.1. Procesamiento de imágenes (ejemplo 1)	1
1.2. Procesamiento de imágenes (ejemplo 2)	2
1.3. Análisis de imágenes	2
1.4. Análisis de imágenes	3
1.5. Reconocimiento de patrones	5
1.6. Computación Gráfica	6
1.7. Tópicos de Visión por Computador	7
1.8. Ejemplo de rectificación de perspectiva	8
1.9. Ejemplo de rectificación de distorsión de lente	9
1.10. Triangulación	9
1.11. Correspondencia en tres puntos	11
1.12. Pintura pre-renacentista y renacentista	12
1.13. Máquina de perspectiva por Albrecht Dürer	13
1.14. Cámara oscura	14
2.1. Proyección $(x_1, x_2, x_3) \rightarrow (x, y)$	20
2.2. Proyección en dos planos paralelos	22
2.3. Proyección en dos planos no paralelos	23
2.4. Rectificación de distorsión proyectiva	24
2.5. Transformación 2D isométrica (Euclídea)	26
2.6. Transformaciones proyectivas 2D	29

2.7. Transformaciones proyectivas 3D	31
2.8. Transformación 3D Euclídea	31
2.9. Rotación de los ejes Z , Y , y X	32
2.10. Figura del Ejercicio 2.8: Transformación $(X_4, Y_4, Z_4) \rightarrow (X_1, Y_1, Z_1)$	38
2.11. Distorsión de perspectiva (ver Ejercicio 2.9)	38
2.12. Transformación Euclídea 2D (Ejercicio 2.10)	39
3.1. Sistema de visión por computador	42
3.2. Modelo geométrico de cámara <i>pinhole</i>	44
3.3. Modelo geométrico de proyección rayos X y cámara oscura	45
3.4. Modelo geométrico de proyección con rotación de ejes	46
3.5. Construcción de un arreglo CCD	48
3.6. Ejemplo de Distorsión	51
3.7. Modelación de distorsión de lente	53
3.8. Péndulo invertido	55
3.9. Rotación del eje X' en el péndulo invertido	56
3.10. Distorsión no lineal	58
4.1. Geometría epipolar	62
4.2. Líneas epipolares y epipolos	64
4.3. Planos epipolares	65
4.4. Geometría Epipolar para tres vistas	74
4.5. Representación geométrica de las Trilinearidades	82
4.6. Correspondencia en dos vistas que comparten el centro óptico	88
4.7. Distancia mínima de un punto a una recta	89
5.1. Triangulación	92
6.1. Flujo óptico	98
6.2. Ejemplo de flujo óptico	100

6.3. Análisis de líneas correspondientes	101
6.4. Ejemplo de una región	102
6.5. Descriptores de Fourier	106
6.6. Cálculo del contraste	109
6.7. Línea epipolar acotada	113
7.1. Ejemplo de una imagen radioscópica con tres fallas	116
7.2. Ejemplo de una secuencia de imágenes	117
7.3. Detección de fallas hipotéticas de la secuencia mostrada en la Fig. 7.2	117
7.4. Algoritmo para el matching en dos imágenes	119
7.5. Ejemplo de matching en dos regiones	119
7.6. Matching de regiones	121
7.7. Algoritmo para el seguimiento en tres imágenes	122
7.8. Ejemplo de seguimiento en tres imágenes	123
7.9. Seguimiento de fallas hipotéticas en tres imágenes	123
7.10. Algoritmo para el seguimiento en cuatro imágenes	124
7.11. Ejemplo de seguimiento en cuatro imágenes	124
7.12. Seguimiento de fallas hipotéticas en cuatro imágenes	126
7.13. Ejemplo de resumen de trayectorias	127
7.14. Ejemplo de seguimiento en más imágenes	127
7.15. Resumen de trayectorias	128
7.16. Verificación de una trayectoria	128
7.17. Rekonstruierte Fehler	129
7.18. Detektierte Fehler	130

Bibliografía

- [1] A. Avidan and A. Shashua. Novel view synthesis in tensor space. In *Conference on Computer Vision and Pattern Recognition (CVPR-97)*, pages 1034–1040, Puerto Rico, 1997.
- [2] A.G. Bors and I. Pitas. Prediction and tracking of moving objects in image sequences. *IEEE Trans. Image Processing*, 9(8):1441–1445, 2000.
- [3] I.N. Bronstein and K.A. Semendjajew. *Taschenbuch der Mathematik*. Harri Deutsch, Thun-Frankfurt, Main, 24 edition, 1989.
- [4] K.R. Castleman. *Digital Image Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 1996.
- [5] R. Chellappa and R. Bagdazian. Fourier coding of image boundaries. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-6(1):102–105, 1984.
- [6] A. de la Escalera. *Visión por Computador: Fundamentos y Métodos*. Prentice Hall, Madrid, 2001.
- [7] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., New York, 2 edition, 2001.
- [8] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, Cambridge MA, London, 1993.
- [9] O. Faugeras, Q.-T. Luong, and T. Papadopoulos. *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. The MIT Press, Cambridge MA, London, 2001.
- [10] O. Faugeras and T. Papadopoulos. A nonlinear method for estimating the projective geometry of 3 views. In *6th International Conference on Computer Vision (ICCV-98)*, pages 477–484, Bombay, India, 1998.
- [11] O. Faugeras and W. Pratt. Decorrelation methods of texture feature extraction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-2(4):323–332, 1980.
- [12] O. Faugeras and L. Robert. What can two images tell us about a third one. *International Journal of Computer Vision*, 18(1):5–20, Apr. 1996.

- [13] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes. *Computer graphics, principles and practice*. Addison-Wesley, London, 2 edition, 1996.
- [14] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, Inc., San Diego, 2 edition, 1990.
- [15] R. Gonzalez and O. Wintz. *Digital Image Processing*. Addison-Wesley Publishing Co., Massachusetts, 3 edition, 1991.
- [16] J. Gribbin. *Historia de la Ciencia (1543-2001)*. Editorial Crítica, Barcelona, 2003.
- [17] R.M. Haralick. Statistical and structural approaches to texture. *Proc. IEEE*, 67(5):786–804, 1979.
- [18] R.M. Haralick and L.G. Shapiro. *Computer and robot vision*. Addison-Wesley Publishing Co., New York, 1992.
- [19] R. Hartley. A linear method for reconstruction from lines and points. In *5th International Conference on Computer Vision (ICCV-95)*, pages 882–887, Cambridge, MA, 1995.
- [20] R. Hartley. Multilinear relationships between coordinates of corresponding image points and lines. In *Proceedings of the International Workshop on Computer Vision and Applied Geometry*, International Sophus Lie Center, Nordfjordeid, Norway, Aug. 1995.
- [21] R. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–150, 1997.
- [22] R. Hartley. Computation of the quadrifocal tensor. In *5th European Conference on Computer Vision (ECCV-98)*, pages 20–35, Jun. 1998.
- [23] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [24] I. Hartmann. *Mustererkennung*. Skriptreihe Regelungstechnik und Bildverarbeitung, Technische Universität Berlin, 1996.
- [25] A. Heyden. A common framework for multiple view tensors. In *5th European Conference on Computer Vision (ECCV-98)*, pages 3–19, Jun. 1998.
- [26] A. Heyden. Multiple view geometry using multifocal tensors. In *DSAGM*, Köpenhamn, 1999.
- [27] M.-K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. Info. Theory*, IT(8):179–187, 1962.
- [28] B. Jähne. *Digitale Bildverarbeitung*. Springer, Berlin, Heidelberg, 2 edition, 1995.
- [29] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.
- [30] K.-F Kamm. Grundlagen der Röntgenabbildung. In K. Ewen, editor, *Moderne Bildgebung: Physik, Gerätetechnik, Bildbearbeitung und -kommunikation, Strahlenschutz, Qualitätskontrolle*, pages 45–62, Stuttgart, New York, 1998. Georg Thieme Verlag.

- [31] Q.-T. Luong and O. Faugeras. The fundamental matrix: theory, algorithms and stability analysis. *International Journal of Computer Vision*, 17(1):43–76, 1996.
- [32] D. Marr and E. Hildreth. Theory of edge detection. *Proc. Roy. Soc. London*, B(207):187–217, 1980.
- [33] D. Mery. *Automated Flaw Detection in Castings from Digital Radioscopic Image Sequences*. Verlag Dr. Köster, Berlin, 2001. (Ph.D. Thesis in German).
- [34] D. Mery. Exploiting multiple view geometry in X-ray testing: Part I, theory. *Materials Evaluation*, 61(11):1226–1233, November 2003.
- [35] D. Mery and D. Filbert. Flaw tracking in the X-ray automated inspection of aluminum castings. Technical Report 189, Institute for Measurement Technology and Control Engineering, Technical University of Berlin, 1999.
- [36] D. Mery and D. Filbert. A fast non-iterative algorithm for the removal of blur caused by uniform linear motion in X-ray images. In *Proceedings of the 15th World Conference on Non-Destructive Testing (15th–WCNDT)*, Rome, Oct. 15-21 2000.
- [37] D. Mery and D. Filbert. Mehr-Bild-Analyse in der Röntgendurchleuchtungsprüfung. In *Berichtsband der Jahrestagung der Deutschen, Österreichischen und Schweizerischen Gesellschaft für Zerstörungsfreie Prüfung (DACH)*, Innsbruck, Austria, 29.-31. Mai 2000.
- [38] D. Mery and D. Filbert. Automated flaw detection in aluminum castings based on the tracking of potential defects in a radioscopic image sequence. *IEEE Trans. Robotics and Automation*, 18(6):890–901, December 2002.
- [39] D. Mery, D. Hahn, and N. Hitschfeld. Interactive environment for the simulation of defects in radioscopic images of aluminum castings. *Computer Vision and Image Understanding*, March 2004. (submitted).
- [40] D. Mery and F. Pedreschi. Segmentation of colour food images using a robust algorithm. *Journal of Food Engineering*, 2004. (accepted April 2004).
- [41] T.S. Newman and A.K. Jain. A survey of automated visual inspection. *Computer Vision and Image Understanding*, 61(2):231–262, 1995.
- [42] J.-R. Ohm. *Digitale Bildcodierung*. Springer, Berlin Heidelberg, 1995.
- [43] E. Persoon and K.S. Fu. Shape discrimination using Fourier descriptors. *IEEE Trans. Systems, Man, and Cybernetics*, SMC-7(3):170–179, 1977.
- [44] A. Shashua. Algebraic functions for recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI*, 17(8):779–789, 1995.
- [45] A. Shashua. Trilinear tensor: The fundamental construct of multiple-view geometry and its applications. In *International Workshop on Algebraic Frames For The Perception Action Cycle (AFPAC)*, Kiel, Sep. 8-9 1997.
- [46] A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *5th International Conference on Computer Vision (ICCV-95)*, Boston MA, Jun. 1995.

- [47] T. Söderström and P. Stoica. *System Identification*. Prentice Hall, New York, 1989.
- [48] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. PWS Publishing, Pacific Grove, CA, 2 edition, 1998.
- [49] C.H. Teh and R.T. Chin. On digital approximation of moment invariants. *Computer Vision, Graphics and Image Processing*, 33(3):318–326, 1986.
- [50] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 4(10):965–980, 1992.
- [51] C.T. Zahn and R.Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Trans. Computers*, C-21(3):269–281, 1971.