

SÍNTESIS DE IMÁGENES A PARTIR DE FOTOGRAFÍAS

GUSTAVO OLAGUE

Departamento de Ciencias de la Computación, División de Física Aplicada
Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California- México

RAMIRO RAMÍREZ

Departamento de Ciencias de la Computación, División de Física Aplicada
Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California- México

Recibido para revisión 3 de Julio de 2001, aceptado 10 de Septiembre de 2001, versión final recibida 19 de Octubre de 2001

RESUMEN: La visualización de información tridimensional obtenida a partir de cámaras digitales es una de las mayores inquietudes de los fotogrametras, de los científicos dedicados a la visión por computadora y recientemente, de los investigadores en el campo de la graficación por computadora. Varios métodos referentes a ambientes de realidad virtual pueden ser encontrados en la literatura. Básicamente se consideran dos diferentes métodos para el renderizado de un objeto 3D usando imágenes digitales, estos son: a) el método basado en el modelo y b) el método basado en la imagen. Sin embargo, uno de los grandes problemas en ambos métodos es la correspondencia de imágenes. En este trabajo se describe un sistema, el cual permite la generación de imágenes para puntos de vista desde los cuales no ha sido tomada ninguna fotografía usando el primer método. Ejemplos y resultados experimentales ilustran la calidad y complejidad de la solución.

PALABRAS CLAVE: geometría epipolar, apareamiento, calibración de cámaras.

ABSTRACT: Visualization of three-dimensional data obtained from digital cameras is a major concern of photogrammetrists, computer vision scientists and now recently computer graphic researchers. Several methods toward a virtual reality environment can be found in the literature. Basically two different approaches for rendering a 3D-object using digital images are considered: a) the model based approach and b) the image based approach. Within both methods a problem of main concern is the image correspondence. This article describes a system which allows the generation of image for viewers position from which no photograph was taken, using the first approach. Examples and experimental results illustrate the quality and complexity of the solution.

KEY WORDS: epipolar geometry, matching, camera calibration.

1 INTRODUCCIÓN

La sintetización de imágenes, que es parte de la visión por computadora, tiene como propósito el cálculo de nuevas vistas a partir de un modelo geométrico tridimensional conocido, apoyándose en cierta información fotométrica: color, textura, materiales, y su interacción con la luz. Para lograr hacer la reconstrucción de dichas vistas es necesario conocer el modelo (escena), lo que significa contar con las coordenadas tridimensionales de los puntos o elementos que lo conforman.

Dentro de las técnicas de sintetización de imágenes siempre están presentes dos problemas:

- 1.Cálculo de la representación tridimensional de la escena.
- 2.Generación de nuevas imágenes.

La solución del primer problema ataña directamente la solución del segundo, ya que nuevas vistas no pueden ser generadas si no se cuenta con un modelo de la escena en tres dimensiones (Blanc, 1998). Además, un modelo tridimensional permite tratamientos geométricos más generales como: las deformaciones, el escalamiento o la

adicción de nuevos objetos a una escena.

Es imposible recuperar, geométricamente, la información tridimensional a partir de una sola imagen debido a la perdida de la información referente a la profundidad durante la formación de la imagen (Xu and Zhang, 1996). Sin embargo, muchas aplicaciones basadas en visión, como lo son la modelización de objetos, la navegación de vehículos y la inspección geométrica, requieren información tridimensional (tanto métrica como no métrica). Una de las soluciones es el uso de múltiples vistas. Sin embargo, nuevos problemas aparecen; algunos de los cuales son: cómo identificar las mismas características entre diferentes vistas, cómo fusionar la alta redundancia de información presente en diferentes vistas, cómo representar eficientemente la escena percibida. La Figura 1 muestra el acercamiento a una sección de una imagen. En ella podemos percibir el grado de dificultad que representa identificar elementos que integran la imagen, lo cual conlleva a uno de los principales problemas de nuestro trabajo que es el apareamiento entre 2 imágenes, el cual consiste en determinar que punto de la primera imagen corresponde al mismo punto físico de la segunda imagen.

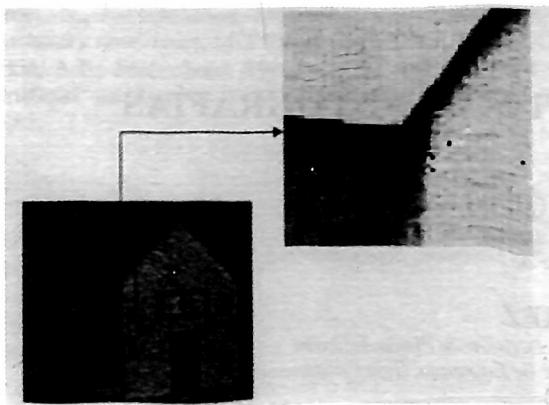


Figura 1. Acercamiento a una sección de la imagen

Actualmente, un número considerable de investigadores han estudiado la visión por computadora para explorar como se puede reconocer un mundo tridimensional al analizar sus imágenes. Este estudio ha dado lugar a la geometría epipolar, la cual no es más que una simple extensión de las reglas de la geometría perspectiva que provee reglas generales entre las formas de los objetos en tres dimensiones y sus imágenes (Xu and Zhang, 1996).

El objetivo de este trabajo consiste en que a partir de dos imágenes en dos dimensiones obtener una representación en tres dimensiones de la escena, de la cual se puedan generar nuevas vistas en dos dimensiones (imágenes o fotografías con un ángulo de visión diferente al de las imágenes iniciales), apoyándonos en la geometría epipolar de las imágenes.

2 REPRESENTACIÓN DIGITAL DE UNA IMAGEN

El término imagen monocromática o simplemente imagen se refiere a una función bidimensional de intensidad $f(x, y)$, donde x y y representan las coordenadas espaciales, y el valor de $f(x, y)$ en un punto cualquiera (x, y) , es proporcional al brillo (o nivel de gris) de la imagen en ese punto.

Una imagen digital es una imagen $f(x, y)$ que se ha discretizado tanto en las coordenadas espaciales como en el brillo. Una imagen digital puede considerarse como una matriz cuyos índices de filas y columnas identifican un punto de la imagen, y el valor del elemento correspondiente de la matriz indica el nivel de gris en ese punto. Sin embargo, errores en el proceso de discretización están presentes, principalmente en su parte óptica y electrónica.

2.1 Operadores Diferenciales

La detección de bordes es una de las operaciones más usadas en el análisis de imágenes, esto, debido a que los bordes forman el contorno de un objeto. Un borde es el límite entre un objeto y el fondo, además de que indica el límite del traslape entre objetos.

Debido a que un borde está definido por un cambio en el

nivel de gris, un operador que es sensible a este cambio operará como un detector de bordes; un operador diferencial hace esto. Una derivada puede ser interpretada como la razón de cambio de una función. La razón de cambio en los niveles de grises en una imagen es mayor cerca de los bordes y más pequeña en áreas con niveles de grises constantes.

Ya que una imagen es bidimensional, es importante considerar el cambio de niveles en muchas direcciones. Por esta razón, las derivadas parciales de una imagen se realizan con respecto a las principales direcciones, en este caso x y y . Un estimador de la dirección de un borde puede ser obtenido al calcular las derivadas en x y en y . De esta manera, un cambio en el nivel de gris o vector suma estará representado por sus componentes dadas en las direcciones a lo largo de los ejes coordenadas x y y . El operador involucrado en este caso es el gradiente.

2.2 Detección de Esquinas

La detección de esquinas es una parte esencial del procesamiento de imágenes de bajo nivel y de la visión por computadora (Lu and Zhou, 1999). Las esquinas son características importantes en aparcamiento de imágenes, descripción de formas y detección de movimiento, debido a que éstas indican posiciones en la imagen con un alto contenido de información (González and Woods, 1996, Kitchen and Rosenfeld, 1982).

Aunque no existe una definición exacta en una imagen de un punto que representa una esquina, podemos reconocer que este punto se encuentra sobre un borde donde la magnitud del gradiente es alta (Lu and Zhou, 1999). La información que proporcionan la diferenciación de primer orden es ampliamente usada en los detectores de esquinas.

Uno de los detectores de esquinas utilizados con más frecuencia es el llamado detector K&R, el cual trabaja directamente sobre la escala de niveles de gris de las imágenes. Kitchen y Rosenfeld (1982) proponen una manera de medir las esquinas, basada en el producto de la magnitud del gradiente y en el cambio de su dirección a lo largo de un borde, la cual es la siguiente:

$$C_{K\&R} = \frac{(I_{xx}I_x^2 + I_{yy}I_y^2 - 2I_{xy}I_xI_y)}{(I_x^2 + I_y^2)} \quad (1)$$

donde I es la intensidad o nivel de gris de la imagen, I_x denota la derivada parcial de I con respecto a la dirección de x , y así sucesivamente.

El detector K&R usa directamente derivadas de primer y segundo orden para obtener la medición de la esquina. Despues de que se aplica la Ecuación (1) en cada uno de los puntos de la imagen se realiza una umbralización para aislar las esquinas, la cual consiste en efectuar un mapeo de 0 a 255, que son los valores de los niveles de gris, entre los valores resultantes de las mediciones de las esquinas. Finalmente se obtiene una nueva imagen en la cual las esquinas están representadas por los pixeles más luminosos.

Para poder realizar la sintetización de imágenes es necesario que conozcamos las coordenadas de algunos puntos sobre la escena, para ello nos apoyamos en una mira de calibración. Esta mira la constituyen dos grupos de 12 cuadros cada uno de ellos, de dimensiones 4.5cm por lado, y una separación de 1.5 cm entre cada cuadro. Los dos grupos de cuadros se encuentran montados sobre una estructura de madera, la cual está compuesta de dos láminas que forman un ángulo de 90°, ver Figura 2. El propósito de utilizar esta mira es que podamos conocer coordenadas tridimensionales en la escena, las cuales las obtenemos a partir de las esquinas de los cuadros.

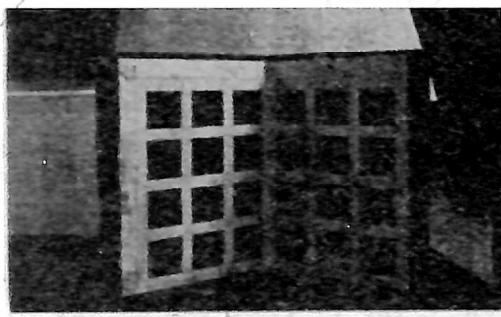


Figura 2: Mira de calibración

Al realizar un acercamiento dentro de la imagen para identificar una esquina, es difícil determinar que pixel corresponde a dicha esquina, ver Figura 3a. Para poder determinar dicho pixel, aplicamos a la imagen el detector de esquinas K&R, Ecuación (1), con el cual se logra resaltar las esquinas de una imagen en la escala de niveles de gris y por consiguiente el pixel que representa dicha esquina. Las esquinas son representadas por los pixeles más luminosos de la imagen, ver Figura 3b.

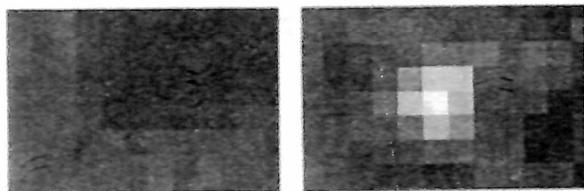


Figura 3. Acercamiento a una esquina, a) imagen original b) imagen después de aplicar el detector de esquinas

3 GEOMETRÍA E IMÁGENES

3.1 Formación de la imagen

Limitándonos al aspecto geométrico, una imagen en niveles de gris obtenida con una cámara es el resultado de una transformación geométrica. Esta transformación permite pasar de una representación tridimensional de la escena a una representación bidimensional de la misma. Para pasar de coordenadas definidas en el sistema coordenado de la escena a coordenadas de la imagen expresadas en pixeles, son necesarios tres pasos:

1. Transformación tridimensional: los puntos tridimensionales expresados en el sistema coordenado de la escena experimentan un cambio al pasar al sistema coordenado de la cámara. Este cambio de sistema coordenado comprende 6 parámetros: 3 para las rotaciones y 3 para las translaciones. Estos parámetros son la posición y la orientación de la cámara, y son llamados parámetros extrínsecos.

$$A = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

2. Proyección 3D-2D: después del cambio de sistema coordenado, los puntos tridimensionales expresados en el sistema coordenado de la cámara son proyectados sobre el plano de la imagen. La nuevas coordenadas reciben el nombre de coordenadas normalizadas.

3. Cambio de coordenadas: para pasar a coordenadas expresadas en pixeles, las coordenadas normalizadas se someten a una transformación afín del plano. Esta transformación, para el modelo que seguiremos en el presente trabajo, está compuesta de 4 parámetros llamados parámetros intrínsecos. Estos parámetros son expresados en una matriz de 3 X 3:

$$C = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix},$$

donde α_u y α_v proporcionan la distancia focal en unidades de pixeles a lo largo de los ejes x y y respectivamente, y (u_0, v_0) es el punto principal de la imagen (Faugeras, 1996).

3.2 Relación geométrica Entre Dos Imágenes

Consideremos el caso de la estereovisión, el cual se representa a través de un sistema de dos cámaras que observan una misma escena. Llamemos a cada una de las imágenes, imagen derecha e imagen izquierda, ver Figura 4.

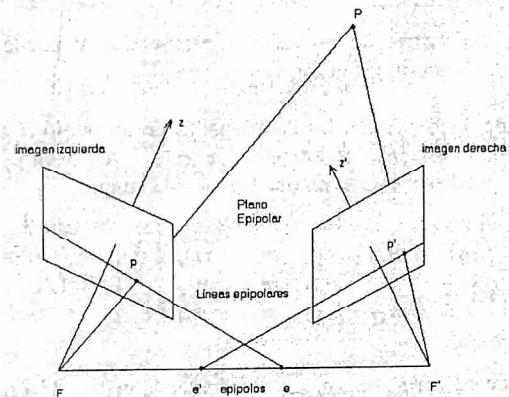


Figura 4. Relación geométrica entre un par de imágenes: geometría epipolar

Es claro que el punto p de la imagen izquierda representa la proyección de todos los puntos del espacio que pertenecen a la recta FP . Las posibles correspondencias de p en la imagen derecha están limitadas a encontrarse sobre una línea, llamada línea epipolar. Esta línea no es otra cosa que la proyección de la recta FP en la imagen derecha.

Todas las líneas epipoles de la imagen derecha pasan por e' , que es la intersección de la recta FF' con el plano de la imagen derecha, e' es llamado epipolo de la imagen derecha con relación a la imagen izquierda.

De manera simétrica, el epipolo e , de la imagen izquierda con relación a la imagen derecha, se define como la intersección de FF' con el plano de la imagen izquierda. En otros términos, un punto P del espacio define con los dos centros de proyección, F y F' , un plano llamado plano epipolar. Este plano corta los dos planos de las imágenes en dos rectas, las líneas epipoles. Todos los planos epipoles pasan por la recta FF' y por los epipolos e y e' . Esta relación entre dos imágenes es llamada geometría epipolar.

3.3 Transformación Mira → Imagen

La transformación mira → imagen puede ser expresada en forma de una matriz de 3×4 , llamada matriz de proyección perspectiva, la cual puede ser descompuesta como sigue:

$$M = I_c A$$

$$M = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3)$$

La matriz A puede escribirse en una forma más compacta al utilizar la notación $r_1 = (r_{11} \ r_{12} \ r_{13})$:

$$A = \begin{pmatrix} r_1 & t_x \\ r_2 & t_y \\ r_3 & t_z \\ 0 & 1 \end{pmatrix}$$

M es la matriz de proyección perspectiva y puede, en general, ser escrita de la forma siguiente:

Donde X, Y y Z son las coordenadas de un punto B de la mira en el sistema coordenado de ésta. Asimismo, esta matriz puede ser escrita en una forma más simple:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (4)$$

Donde $m_i = (m_{i1} \ m_{i2} \ m_{i3})$. Al calcular la matriz M a partir de I, A , teniendo en cuenta algunas propiedades de ortonormalidad de la rotación y resaltando que se debe obtener un valor negativo para α_u , obtenemos un conjunto de ecuaciones que permiten calcular los parámetros intrínsecos

y extrínsecos en función de los coeficientes de M (Horaud and Monga, 1995). Así tenemos:

$$\left\{ \begin{array}{l} r_3 = m_3 \\ u_0 = m_1 \cdot m_3 \\ v_0 = m_2 \cdot m_3 \\ \alpha_u = -m_1 \wedge m_3 \\ \alpha_u = m_2 \wedge m_3 \\ r_1 = I \alpha_u (m_1 - u_0 m_3) \\ r_2 = I \alpha_v (m_2 - v_0 m_3) \\ t_x = I \alpha_u (m_{14} - u_0 m_{34}) \\ t_y = I \alpha_v (m_{24} - v_0 m_{34}) \\ t_z = m_{34} \end{array} \right. \quad (5)$$

4 CALIBRACIÓN DE UNA CÁMARA

Utilizando la Ecuación 4 podemos escribir las coordenadas de la imagen de un punto de la escena de la siguiente forma:

$$u = m_{11}X + m_{12}Y + m_{13}Z + m_{14} \quad (6)$$

$$v = \begin{cases} m_{31}X + m_{32}Y + m_{33}Z + m_{34} \\ m_{21}X + m_{22}Y + m_{23}Z + m_{24} \\ m_{31}X + m_{32}Y + m_{33}Z + m_{34} \end{cases} \quad (7)$$

Estas dos ecuaciones describen la línea que pasa por el centro de proyección y el punto de la imagen (u, v) , en el sistema coordenado de la mira, sin tomar en cuenta la distorsión proyectiva. Esta línea es la recta que pasa por F y b en la Figura 5. Además, es también llamada línea de visión asociada a un punto de la imagen con coordenadas (u, v) .

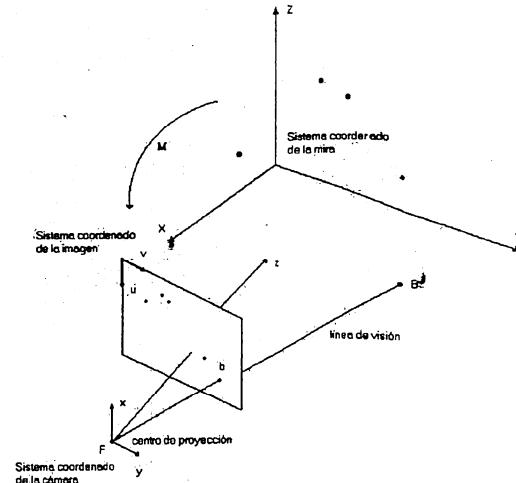


Figura 5. El principio de calibración de una cámara

Con el fin de calcular los coeficientes de la matriz M , es necesario escribir un sistema de ecuaciones a partir de los puntos de la mira y su proyección en la imagen. Cada punto (X_i, Y_i, Z_i) se proyecta en (u_i, v_i) generándose dos ecuaciones. Estas ecuaciones son lineales con respecto a los coeficientes de la matriz. De esta forma necesitamos al menos 6 puntos para determinar los 12 coeficientes de la matriz M que nos interesa.

Las ecuaciones 6 y 7 pueden ser reescritas como una combinación lineal de los m_{ij} parámetros:

$$X_i m_{11} + Y_i m_{12} + Z_i m_{13} + m_{14} - u_i X_i m_{31} - u_i Y_i m_{32} - u_i Z_i m_{33} = u_i m_{34} \quad (8)$$

$$X_i m_{21} + Y_i m_{22} + Z_i m_{23} + m_{24} - v_i X_i m_{31} - v_i Y_i m_{32} - v_i Z_i m_{33} = v_i m_{34} \quad (9)$$

De esta forma obtenemos $2n$ ecuaciones para n puntos. Podemos escribir estas ecuaciones en su forma matricial (los subíndices indican el tamaño de la matriz):

$$K_{2n \times 11} x_{11} = u_{2n} \quad (10)$$

4.1 Utilización de la Restricción $\| M3 \| = 1$

Método Faugeras -Toscani

Al calcular los coeficientes de M en función de los coeficientes que componen las matrices I_C y A , encontramos que, ver la Ecuación (5)

$$m_{31} = r_{31}, \quad m_{32} = r_{32}, \quad m_{33} = r_{33}.$$

Por otro lado, aplicando identidades trigonométricas es fácil verificar que: $r_{31}^2 + r_{32}^2 + r_{33}^2 = 1$. Obteniendo así:

$$m_3^2 = m_{31}^2 + m_{32}^2 + m_{33}^2$$

A continuación mostramos como calcular la matriz M tomando en cuenta la restricción anterior.

La Ecuación (10) se puede reescribir de la siguiente manera:

$$B_{2n \times 9} x_9 + C_{2n \times 3} x_3 = 0, \quad (11)$$

donde:

$$B_{2n \times 9} = \begin{pmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i \end{pmatrix} \quad \vdots$$

$$y \quad C_{2n \times 3} = \begin{pmatrix} -u_i X_i & -u_i Y_i & -u_i Z_i \\ -v_i X_i & -v_i Y_i & -v_i Z_i \end{pmatrix} \quad \vdots$$

Como se puede observar en la Ecuación (11), es necesario descomponer x_{11} , en dos incógnitas, x_9 y x_3 , así tenemos:

$$\begin{aligned} x_9 &= (m_1 \ m_{14} \ m_2 \ m_{24} \ m_{34}) \\ y \\ x_3 &= (m_3) \end{aligned}$$

El segundo vector que es necesario encontrar, X_3 , representa a las variables involucradas al aplicar la restricción de que se vale este método.

El criterio a minimizar para poder solucionar de una manera más óptima el sistema de ecuaciones es el siguiente:

$$Q = B_{2n \times 9} x_9 + C_{2n \times 3} x_3^2, \quad (12)$$

con la restricción:

$$x_3^2 = 1$$

El criterio puede ser escrito de la siguiente manera:

$$Q = B x_9 + C x_3^2 + \lambda (1 - x_3^2),$$

Desarrollando esta expresión obtenemos:

$$\begin{aligned} Q = & x_9' B' B x_9 + x_3' C' C x_3 + x_9' B' C x_3 + \\ & x_3' C' B x_9 + \lambda (1 - x_3' x_3) \end{aligned} \quad (13)$$

Igualando a cero las derivadas parciales de x_9 y x_3 , resultan las siguientes dos ecuaciones:

$$\begin{aligned} B' B x_9 + B' C x_3 &= 0 \\ C' C x_3 + C' B x_9 &= 0, \end{aligned}$$

de las cuales obtenemos:

$$\begin{aligned} x_9 &= -(B' B)^{-1} B' C x_3 \\ D x_3 &= \lambda x_3 \\ D &= C' C - C' B (B' B)^{-1} B' C \end{aligned}$$

Finalmente, sustituyendo estas ecuaciones en la ecuación 13 tenemos:

$$Q = x_3' D x_3 = \lambda x_3' x_3 = \lambda$$

Cabe hacer notar que D es una matriz simétrica y positiva de 3×3 . Ésta tiene sus valores propios reales y positivos. X_3 es un vector propio de D asociado al valor propio de la variable A .

Para minimizar el criterio es necesario seguir los siguientes pasos:

1. Calcular los valores propios de la matriz D .
2. Escoger el valor propio más pequeño (este valor es el que minimiza el criterio).
3. Calcular el vector propio (X_3) asociado al valor propio seleccionado.
4. Normalizar el vector propio (X_3).
5. Finalmente, calcular el vector x_9 .

Los coeficientes de M están dados por los vectores X_3 y X_9 . El signo del vector propio, X_3 , no está definido, teniendo así dos soluciones: M y $-M$. Podemos escoger una de estas

soluciones tomando en cuenta que la mira a calibrar se encuentra en frente de la cámara y no atrás de ésta. Así pues $m_{34} = t_z > 0$.

5 VISIÓN ESTEREOOSCÓPICA Y GEOMETRÍA EPIPOLAR

Uno de los objetivos de la visión por computadora es la de reconstruir la estructura tridimensional (3D) del espacio a partir de un grupo de imágenes. La visión estereoscópica utiliza dos imágenes tomadas con dos cámaras. Conociendo el modelo de proyección de cada cámara y la relación espacial entre ellas, sólo resta calcular las coordenadas 3D de un punto a partir de su proyección en las dos imágenes.

5.1 Calibración Estereoscópica

La calibración estereoscópica consiste en determinar la matriz de transformación entre el sistema coordenado de la cámara izquierda y el sistema coordenado de la cámara derecha. Sea A , dicha matriz de transformación, la cual está compuesta por una matriz de rotación y una matriz de translación, del mismo tipo que la matriz A de la Ecuación (2). La matriz A , está representada en la Figura 6. Esta figura además sugiere los pasos a seguir para calibrar el sistema estereoscópico, los cuales son los siguientes:

1. Primero, calibrar cada cámara a partir de un sistema coordenada único. Esto proporciona los coeficientes de las matrices, M y M' .
2. Extraer enseguida los parámetros intrínsecos y extrínsecos de cada una de las cámaras.
3. Con la ayuda de los parámetros extrínsecos construimos dos matrices, A y A' ; la primera es la transformación del sistema coordenado de la mira al sistema coordenado de la cámara izquierda, y la segunda es la transformación del sistema coordenado de la mira al sistema coordenado de la cámara derecha.
4. Enseguida se calcula A_s :

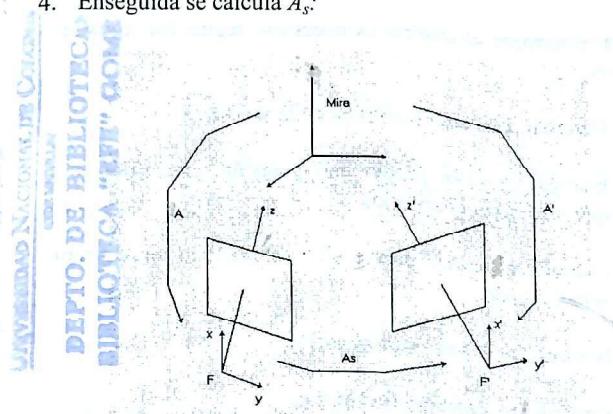


Figura 6. Calibración de dos cámaras formando un sistema estereoscópico a partir de una sola escena

$$A_s = A'A^{-1}$$

En forma matricial tenemos:

$$A_s = \begin{pmatrix} r_{11} & r_{12} & r_{13} & b_x \\ r_{21} & r_{22} & r_{23} & b_y \\ r_{31} & r_{32} & r_{33} & b_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} r_1 & b_x \\ r_2 & b_y \\ r_3 & b_z \\ 0 & 1 \end{pmatrix}$$

El vector $b = (b_x, b_y, b_z)^t$ es el vector que va de F a F' . Éste representa las coordenadas de F en el sistema coordenado de la cámara derecha, ver Figura 4.

5.2 Relación Izquierda→Derecha

Ahora vamos a establecer una relación simple entre un punto de la imagen de la izquierda y un punto de la imagen de la derecha. A partir de este momento vamos a trabajar en el sistema coordenado de la cámara y no en el sistema coordenado de la imagen. Retomando que a partir de un pixel de la imagen con coordenadas u y v , podemos fácilmente deducir sus coordenadas en la cámara izquierda:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = C'^{-1} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix}$$

De igual forma, tenemos una expresión similar para la cámara derecha:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = C'^{-1} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix}$$

Ahora podemos expresar un punto P de la escena al mismo tiempo en los dos sistemas coordenados. Siendo (X, Y, Z) sus coordenadas en el sistema de la cámara izquierda y (X', Y', Z') sus coordenadas en el sistema de la cámara derecha. La relación entre los dos sistemas de coordenadas se expresa de la siguiente forma:

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = A_s \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Las coordenadas de p (proyección de P en la imagen izquierda) son (x, y, l) donde $x = X/Z$ y $y = Y/Z$. De igual manera tenemos que para p' (proyección de P en la imagen derecha) sus coordenadas son (x', y', l) . x' y y' pueden ser escritas como sigue:

$$\begin{aligned} x' &= \frac{X'}{Z'} = \frac{r_{11}X + r_{12}Y + r_{13}Z + b_x}{r_{31}X + r_{32}Y + r_{33}Z + b_z} \\ y' &= \frac{Y'}{Z'} = \frac{r_{21}X + r_{22}Y + r_{23}Z + b_y}{r_{31}X + r_{32}Y + r_{33}Z + b_z} \end{aligned}$$

Tomando en cuenta que $X = xZ$ y que $Y = yZ$, y con la notación $p = (x \ y \ l)^t$ podemos simplificar estas fórmulas:

$$x' = \frac{Zr_1 \cdot p + b_x}{Zr_3 \cdot p + b_z} \quad (15)$$

$$x' = \frac{Zr_2 \cdot p + b_y}{Zr_3 \cdot p + b_z} \quad (16)$$

Podemos expresar entonces la posición de un punto en la imagen de la derecha en función de su punto correspondiente en la imagen de la izquierda, a partir de los parámetros del sistema y de la profundidad de Z del punto P .

5.3 La Restricción Epipolar

Al eliminar Z de las Ecuaciones (15) y (16), obtenemos una relación lineal entre x' y y' :

$$(b_z r_2 \cdot p - b_z r_3 \cdot p)x' + (b_z r_3 \cdot p - b_z r_1 \cdot p)y' = b_x r_2 \cdot p - b_y r_1 \cdot p \quad (17)$$

Esta ecuación describe el lugar de los puntos de la imagen derecha, que pueden corresponder a un punto p de la imagen izquierda: esta línea es llamada la línea epipolar derecha. Para cada punto de la imagen de la izquierda existe una línea epipolar derecha y , recíprocamente, para cada punto de la imagen derecha existe una línea epipolar izquierda. Además, ésta es la ecuación de una recta de la forma:

$$a' x' + b' y' + c' = 0 \quad (18)$$

Los $(')$ indican que es una recta expresada en el sistema coordenada de la cámara derecha. Esta recta no es otra, que la restricción epipolar presentada en la sección anterior. Además, podemos notar que los parámetros de esta recta, a' , b' y c' , son iguales a las expresiones siguientes:

$$\begin{aligned} a' &= (b_y r_{31} - b_z r_{21})x + (b_y r_{32} - b_z r_{22})y + (b_y r_{33} - b_z r_{23}) \\ b' &= (b_z r_{11} - b_x r_{31})x + (b_z r_{12} - b_x r_{32})y + (b_z r_{13} - b_x r_{33}) \\ c' &= (b_x r_{21} - b_y r_{11})x + (b_x r_{22} - b_y r_{12})y + (b_x r_{23} - b_y r_{13}) \end{aligned}$$

5.4 La Matriz Esencial

Las expresiones que se han establecido para los parámetros a' , b' y c' pueden ser representadas en forma matricial como sigue:

$$\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} 0 & -b_z & b_y \\ b_z & 0 & -b_x \\ -b_y & b_x & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (19)$$

En esta ecuación, el producto de las dos matrices, una matriz antisimétrica (de rango 2) y una matriz ortonormal (de rango 3), generan como resultado la matriz E , llamada matriz esencial. Podemos notar que esta matriz puede ser calculada a partir de los parámetros b_x , b_y , b_z y \mathbf{r}_1 , \mathbf{r}_2 , \mathbf{r}_3 . Estos parámetros son obtenidos durante la calibración del sistema estereoscópico. De acuerdo a lo anterior, la Ecuación (19) puede también ser expresada de la siguiente manera:

$$\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = E \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (20)$$

Esta es la transformación epipolar en la cual a un punto de la imagen izquierda $(x \ y \ 1)^t$, corresponde una recta sobre la imagen derecha descrita por los parámetros $(a' \ b' \ c')$. La ecuación de la línea epipolar, Ecuación (18), puede ser expresada de la siguiente manera:

$$(x' \ y' \ 1) \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = 0,$$

o también, sustituyendo en la Ecuación (20) tenemos

$$(x' \ y' \ 1) E \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0,$$

siendo

$$p^t E p = 0 \quad (21)$$

La matriz E describe la transformación epipolar izquierda-derecha, la cual permite calcular la ecuación de una línea epipolar que pasa por la imagen derecha asociada a un punto de la imagen izquierda. Observemos que la transformación epipolar derecha-izquierda está dada por la matriz transpuesta:

$$p^t E^t p = 0. \quad (22)$$

5.5 La Matriz Fundamental

La matriz fundamental describe una relación geométrica entre dos imágenes tomadas desde diferentes ángulos de visión: geometría epipolar (Boufama, 1994).

Recordemos que la relación entre las coordenadas de la cámara y las coordenadas de la imagen, en el caso izquierdo, están dadas por:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

la cual, utilizando notación matricial se puede expresar como:

$$m = Cp$$

Tenemos la misma expresión para la cámara derecha:

$$m' = C'p'$$

Los elementos de las matrices C y C' son los parámetros intrínsecos de las dos cámaras.

Al sustituir en la ecuación 21 tenemos:

$$m^t (C'^t)^t E C^t m = 0$$

De la ecuación anterior obtenemos la matriz fundamental, y se define como sigue:

$$F = (C^I)^T E C^I. \quad (23)$$

F es la matriz fundamental que describe la geometría epipolar. La ecuación:

$$m^T F m = 0, \quad (24)$$

no es otra que la ecuación de una línea epipolar en el sistema coordenada de la imagen, y no en el sistema coordenada de la cámara como en la Ecuación (21).

De la misma forma que la matriz esencial, la matriz fundamental puede ser calculada por un sistema estereoscópico previamente calibrado.

6 CORRELACIÓN

La correlación de dos funciones continuas complejas $f(x)$ y $g(x)$, representada por $f(x)$ o $g(x)$, se define por la relación (González and Woods, 1996):

$$f(x) \circ g(x) = \int_{-\infty}^{+\infty} f^*(\alpha) g(x + \alpha) d\alpha, \quad (25)$$

donde $*$ es el complejo conjugado. Para realizar la correlación, simplemente se desplaza $g(x)$ sobre $f(x)$ y se integra el producto desde $-\infty$ hasta $+\infty$ para cada valor de desplazamiento en x .

El equivalente discreto de la Ecuación (25) se define como:

$$f_e(x) \circ g_e(x) = \sum_{m=0}^{M-1} f_e^*(M) g_e(x + m), \quad (26)$$

donde $x = 0, 1, 2, \dots, M-1$, y M es el tamaño del intervalo en el que es evaluada la función.

Para el caso bidimensional siguen siendo válidas expresiones similares. Así, si $f(x, y)$ y $g(x, y)$ son funciones de variables continuas, su correlación se define como:

$$f(x, y) \circ g(x, y) = \int \int_{-\infty}^{+\infty} f^*(\alpha, \beta) g(x + \alpha, y + \beta) d\alpha d\beta, \quad (27)$$

y para el caso discreto:

$$f_e(x, y) \circ g_e(x, y) = \sum_{n=0}^{M-1} \sum_{m=0}^{M-1} f_e^*(m, n) g_e(x + m, y + n), \quad (28)$$

6.1 Apareamiento

El apareamiento de puntos entre dos imágenes se basa en el principio siguiente: un punto de la primera imagen representa el mismo punto físico que otro punto de la segunda imagen (decimos que los dos puntos se corresponden) si los dos puntos se asemejan (Boufarna, 1994). Esta semejanza debe tomar en cuenta los puntos vecinos debido al ruido en las imágenes, los ocultamientos, el cambio de punto de visión de las imágenes, etc.

En ausencia de todo conocimiento a priori la semejanza entre dos porciones de dos imágenes de una misma escena se puede cuantificar gracias a la medición de la correlación, la cual es una medida de semejanza. La correlación se calcula entre dos ventanas, generalmente cuadradas y centradas en el punto que se está poniendo en correspondencia o apareando.

Después de experimentos desarrollados por Boufarna (1994), para determinar cual es la medida de correlación que ofrece los mejores resultados, se encontró que la correlación llamada ZNCC (Zero Mean Normalized CrossCorrelation) es la que muestra mayor robustez.

La correlación ZNCC entre el punto $p(x, y)$ de una imagen y el punto $p'(x', y')$ de otra imagen está dada por la fórmula:

$$C(p, p') = \frac{\sum_{i=-n}^n \sum_{j=-n}^n A \cdot B}{\sum \sum C \cdot \sum \sum D} \quad (29)$$

siendo

$$\begin{aligned} A &= (I(x+i, y+j) - \bar{I}(x, y))^2, \\ B &= (I(x'+i, y'+j) - \bar{I}(x', y'))^2, \\ C &= (I(x+i, y+j) - \bar{I}(x, y))^2, \\ D &= (I(x'+i, y'+j) - \bar{I}(x', y'))^2, \end{aligned}$$

donde $I(x, y)$ es la intensidad o nivel de gris en el punto (x, y) , y $\bar{I}(x, y)$ es la media de las intensidades de los puntos que se encuentran en la ventana de tamaño $(2n + 1) \times (2n + 1)$ con centro en (x, y) . El apareamiento busca para un punto dado en la primera imagen, el punto en la segunda que tenga la mayor respuesta de correlación.

7 RECONSTRUCCIÓN

La reconstrucción tridimensional a partir de un par de imágenes supone que existe correspondencia entre ambas. El tipo de reconstrucción que podemos efectuar dependerá del tipo de calibración que se tenga. En nuestro trabajo, el sistema estereoscópico está calibrado y tenemos los parámetros internos de cada cámara así como la transformación rígida entre las dos cámaras. En este caso obtenemos una reconstrucción Euclídea en el sistema coordenada que se está calibrando. Tomando este caso y si el punto p de la imagen izquierda tiene en correspondencia un punto p' en la imagen derecha, tenemos dos pares de ecuaciones, basándonos en 6 y 7:

$$\begin{aligned} u &= m_{11}X + m_{12}Y + m_{13}Z + m_{14} \\ &\quad m_{31}X + m_{32}Y + m_{33}Z + m_{34} \\ v &= m_{21}X + m_{22}Y + m_{23}Z + m_{24} \\ &\quad m_{31}X + m_{32}Y + m_{33}Z + m_{34} \\ u' &= m'_{11}X + m'_{12}Y + m'_{13}Z + m'_{14} \\ &\quad m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34} \\ v' &= m'_{21}X + m'_{22}Y + m'_{23}Z + m'_{24} \\ &\quad m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34} \end{aligned}$$

Las coordenadas X , Y y Z del punto P , que será reconstruido en el sistema coordenado que se está calibrando, se calculan al resolver el siguiente sistema de 4 ecuaciones lineales que se forma a partir de las ecuaciones anteriores:

$$(um_{31} - m_{11})X + (um_{32} - m_{12})Y + \\ (um_{33} - m_{13})Z = m_{14} - um_{34}$$

$$(vm_{31} - m_{21})X + (vm_{32} - m_{22})Y + \\ (vm_{33} - m_{23})Z = m_{24} - vm_{34}$$

$$(u'm'_{31} - m'_{11})X + (u'm'_{32} - m'_{12})Y + \\ (u'm'_{33} - m'_{13})Z = m'_{14} - u'm'_{34}$$

$$(v'm'_{31} - m'_{21})X + (v'm'_{32} - m'_{22})Y + \\ (v'm'_{33} - m'_{23})Z = m'_{24} - v'm'_{34}.$$

Como podemos observar, es un sistema sobre determinado de 4 ecuaciones y 3 incógnitas de la forma,

$$Ax = b,$$

el cual podemos resolver al multiplicar a cada uno de los términos de la ecuación por A' :

$$A'Ax = A'b,$$

despejando el vector solución, tenemos que:

$$x = (A'A)^{-1}A'b.$$

Al contar con las coordenadas tridimensionales del punto P , estamos en posibilidades de proyectarlo en una nueva imagen, lo cual nos conduce a lograr el propósito de este trabajo, generar una vista a partir de vistas existentes.

8 GENERACIÓN DE NUEVAS VISTAS

Después de haber realizado los cálculos de la calibración de la cámaras para dos fotografías y encontrar la matriz fundamental, F , la cual define la geometría epipolar de éstas, se está en posibilidad de poder generar una nueva imagen con un ángulo de visión diferente al de las imágenes iniciales.

Para generar la nueva vista es necesario conocer las coordenadas tridimensionales de cada uno de los puntos que serán proyectados en la nueva imagen. El cálculo de las coordenadas tridimensionales del punto P a partir de dos imágenes se realiza de la siguiente manera:

1. Conociendo las coordenadas u y v del punto p de la imagen izquierda, correspondiente al punto P de la escena, procedemos a calcular la ecuación de la línea epipolar sobre la imagen derecha apoyándonos en la siguiente ecuación:

$$m^T F m = 0 \quad (30)$$

donde $m = p$ y $m' = p'$. La ecuación anterior también puede ser expresada de la siguiente manera:

$$\begin{pmatrix} u' & v' & 1 \end{pmatrix} \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0. \quad (31)$$

Al efectuar la multiplicación de la matriz fundamental, F , por el vector correspondiente al punto p , cuyos valores son conocidos, y luego por el vector del punto p' , obtenemos una recta de la forma:

$$Au' + Bv' + C = 0. \quad (32)$$

Esta recta sobre la imagen derecha, define los posibles puntos JY con coordenadas u' y v' , que corresponden al punto p de la imagen izquierda.

2. El siguiente paso es identificar el punto p' que corresponde a p , y por consiguiente al punto P . Para poder localizar el punto p' es necesario utilizar la Ecuación (29), la cual indica la correlación existente entre dos puntos, esta ecuación es aplicada en cada uno de los puntos que se encuentran en la recta definida por la Ecuación (32). El punto con mayor correlación (más cercano a l), es considerado el punto p' .
3. Luego de conocer las coordenadas u y v del punto p , y obtener las coordenadas u' y v' del punto p' , se está en condiciones de calcular las coordenadas X , Y y Z del punto P , únicamente es necesario resolver un sistema de 4 ecuaciones y 3 incógnitas generado por las coordenadas de los puntos p , p' y P , y por los elementos de las matrices de proyección M y M' , ver Sección 7.
4. Despues de obtener las coordenadas de P , este punto puede ser proyectado; sólo es necesario definir la matriz de proyección de la nueva imagen, lo que significa, determinar sus parámetros intrínsecos y extrínsecos, para luego multiplicar las respectivas matrices que incluyen a estos parámetros:

$$M = I_c A,$$

Para definir estos nuevos parámetros es necesario tomar como base los parámetros de las dos imágenes a partir de las cuales se realizaron todos los cálculos hasta llegar a definir la matriz fundamental. Por ejemplo, si se desea generar la imagen intermedia de estas dos imágenes, el cálculo de los elementos de la matriz I , (parámetros intrínsecos) es sencillo, únicamente es necesario efectuar la suma de las matrices I , de las dos imágenes iniciales, y luego cada elemento de la nueva matriz dividirlo entre 2. En cambio, para definir la matriz A (parámetros extrínsecos) de la nueva imagen, la cual está compuesta de 3 rotaciones y 3 translaciones, se deben de conocer los ángulos de rotación de los ejes y las

translaciones en cada una de las 2 fotos. Los ángulos son calculados en base a los elementos de la matriz A correspondientes a las de rotación, donde:

$$\operatorname{sen}\beta = r_{31}$$

$$\tan\alpha = -r_{32}/r_{33}$$

$$\tan\theta = -r_{21}/r_{11}.$$

Al conocer los ángulos de las 2 imágenes, los correspondientes de ambas se suman y se dividen entre 2. A partir de los ángulos obtenidos se genera la sección de la matriz A de la nueva imagen, correspondiente a las rotaciones, aplicando la siguiente ecuación:

$$\mathbf{R}_{\alpha\beta\theta} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

donde

$$r_{11} = \cos\beta\cos\theta$$

$$r_{12} = \operatorname{sen}\alpha\operatorname{sen}\beta\cos\theta + \cos\alpha\operatorname{sen}\theta$$

$$r_{13} = -\cos\alpha\operatorname{sen}\beta\cos\theta + \operatorname{sen}\alpha\operatorname{sen}\theta$$

$$r_{21} = -\cos\beta\operatorname{sen}\theta$$

$$r_{22} = \operatorname{sen}\alpha\operatorname{sen}\beta\operatorname{sen}\theta + \cos\alpha\cos\theta$$

$$r_{23} = \cos\alpha\operatorname{sen}\beta\operatorname{sen}\theta + \operatorname{sen}\alpha\cos\theta$$

$$r_{31} = \operatorname{sen}\beta$$

$$r_{32} = -\operatorname{sen}\alpha\cos\beta$$

$$r_{33} = \cos\alpha\cos\beta$$

El cálculo de la sección de la matriz A de la nueva imagen, correspondiente a las translaciones, también es sencillo, únicamente se suman cada una de las translaciones de ambas imágenes, y también se dividen entre dos.

Cabe hacer notar que todos los datos (parámetros intrínsecos y extrínsecos) de las dos imágenes originales son conocidos, ya que fueron calculados al momento de calibrar la cámara en base a cada una de ellas, para luego calcular la matriz fundamental.

5. Despues del paso anterior sólo resta calcular las coordenadas u y v del punto P en la nueva imagen, las cuales se definen a partir de la siguiente ecuación:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

despejando u y v tenemos:

$$u = \frac{su}{s}, \quad v = \frac{sv}{s}.$$

Los pasos anteriores deben de repetirse tantas veces como puntos p tenga la imagen izquierda, excepto el paso 4, ya que

la matriz M de la nueva imagen, es la misma para todos los puntos de ésta.

Para lograr obtener un mayor número de puntos P , es necesario aplicar el caso inverso, es decir, trabajar con la relación derecha-izquierda. Para ello es necesario tomar cada punto p' de la imagen derecha, y tratar de buscar el correspondiente punto p en la imagen izquierda, para luego calcular P . Los pasos a seguir son los mismos que los anteriores, con la excepción de que para encontrar la línea epipolar sobre la imagen izquierda es necesario evaluar la siguiente ecuación:

$$m'F'm' = 0.$$

Además, tampoco es necesario calcular la matriz de proyección de la nueva imagen (paso 4), ya que se estará trabajando sobre la misma nueva imagen.

Para agilizar la búsqueda de las coordenadas del punto p' a lo largo de la línea epipolar, es necesario conocer el máximo desplazamiento entre los puntos de la imagen derecha con respecto a los de la imagen izquierda, esto con el fin de realizar la búsqueda de dicho punto sólo en una sección de la línea epipolar y no a lo largo de ésta.

Para determinar esta sección se pueden seguir los siguientes pasos:

1. A partir de la visualización de las dos imágenes, determinar la zona en la que existe un mayor desplazamiento entre ambas imágenes.
2. Extraer las coordenadas (u, v) de un punto en dicha zona sobre la imagen izquierda.
3. De igual manera, extraer las coordenadas (u', v') de un punto en esa misma zona, pero ahora sobre la imagen derecha.
4. Al conocer ambas coordenadas, realizar las siguientes diferencias: $u - u'$ y $v - v'$
5. Llamamos D al mayor resultado entre las diferencias anteriores.
6. La sección de búsqueda para todos los puntos p' sobre la imagen derecha, a partir de las coordenadas (u, v) del punto p de la imagen izquierda, es $(u \pm D, v \pm D)$.

9 EXPERIMENTO I

El objetivo de este experimento es estimar la cantidad de puntos que se proyectan sobre la nueva imagen utilizando ventanas de diferentes tamaños para el cálculo de la correlación. En este experimento se trabajo con dos imágenes en las cuales la mayor diferencia entre los puntos p de la imagen izquierda, y los punto p' de la imagen derecha es de 3 píxeles; esto es, si la posición del punto p en la imagen izquierda es (x, y) , en la imagen derecha p' se

encuentra como máximo en $(x \pm 3, y \pm 3)$. La razón de que se trabaje con una diferencia mínima entre los puntos correspondientes de las imágenes, es ver el comportamiento del apareamiento entre los puntos que las constituyen, ya que al aplicar una medida de correlación para determinar el grado de similitud entre los puntos de ambas imágenes, el obtener el mayor valor entre dos de ellos, no garantiza que sean los que se corresponden. Para este experimento, debido que existe poca diferencia entre los puntos p y los puntos p' , se consideró como correcta la relación entre p y p' , si al calcular el punto sobre la nueva imagen la diferencia de sus coordenadas x y y con respecto a la de las 2 imágenes iniciales no excedían los 3 píxeles. Se determinó como límite 3 píxeles, ya que el ángulo de visión de la nueva imagen se encuentra en medio del que tienen las imágenes iniciales; por consiguiente, la diferencia máxima debe ser menor a 3 píxeles. El total de puntos que trataron de ser proyectados sobre la nueva imagen fue de 354,850. En las Tablas 1 y 2 se muestran los resultados obtenidos.

Tabla 1. Resultados Experimento I

Tamaño Ventana	Medida de Correlación	Puntos Proyectados Apaream. Izq-der	% de puntos proyectados Apaream. Izq-Der
5x5	0.9	305,711	86.15%
7x7	0.95	310,140	87.40%
9x9	0.995	313,045	88.22%
11x11	0.995	315,166	88.81%
13x13	0.999	316,901	89.31%
15x15	0.999	317,070	89.64%
17x17	0.999	318,960	89.89%
19x19	0.999	319,314	89.99%

Tabla 2: Resultados Experimento I

Tamaño Ventana	Medida de Correlación	Puntos Proyectados Apaream. Der-Izq	% de puntos proyectados Apaream. Der-Izq
5x5	0.9	313,012	88.21%
7x7	0.95	316,315	89.14%
9x9	0.995	319,810	90%
11x11	0.995	322,663	90.92%
13x13	0.999	325,501	91.73%
15x15	0.999	327,852	92.39%
17x17	0.999	329,609	92.89%
19x19	0.999	330,599	93.16%

En la primer columna observamos que se utilizaron diferentes tamaños de ventanas para realizar el apareamiento, además de que la medida de correlación utilizada (segunda columna), fue variable dependiendo del tamaño de la ventana, esto debido a que para lograr obtener los correspondientes 354,850 puntos de la imagen izquierda sobre la imagen derecha, o viceversa, en las ventanas de menor tamaño se necesita ser más flexible con la medida de correlación empleada, ya que de no hacerlo, no es posible encontrar el total de puntos buscados. La columna 3 presenta el total de puntos que al ser proyectados sobre la

nueva imagen, cumplen con el requisito de no estar desplazados más de 3 píxeles con respecto a su correspondiente píxel en las dos imágenes iniciales. La columna 4 muestra el porcentaje de píxeles desplegados en la nueva imagen (columna 3), con respecto al total de puntos esperados, en este caso 354,850.

Las Figuras 7 y 8 muestran las imágenes a partir de las cuales se efectuó este experimento. Se observa que son casi iguales, esto debido a que la cámara con la que fueron adquiridas se desplazó muy poco, de tal forma que al analizar los puntos entre ambas imágenes, la mayor diferencia entre puntos correspondientes es de 3 píxeles.

La Figura 9 muestra las líneas epipolares sobre la imagen derecha en la cual se destacan las esquinas. Además, en esta imagen se puede apreciar perfectamente el epipolo derecho, el cual se encuentra en el cruce de las líneas epipolares.

Al hacer una comparación entre el total de puntos desplegados en cada una de las imágenes podemos observar que estos se incrementan al ir aumentando el tamaño de la ventana, ver Tabla 1, pero si observamos las imágenes resultantes, ver Figuras 10 y 11, nos damos cuenta que al emplear las ventanas de menor tamaño, los puntos faltantes están dispersos a lo largo de la imagen, en cambio cuando se emplean ventanas de mayor tamaño, los puntos faltantes se concentran en zonas específicas de la imagen.

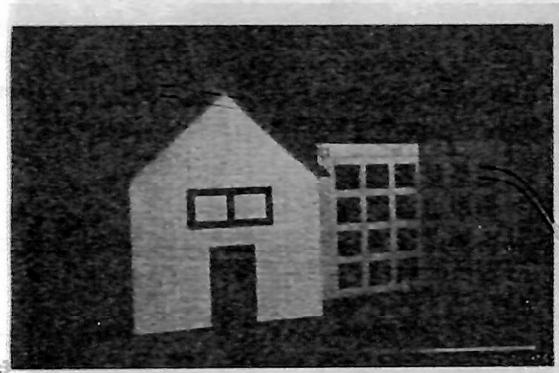


Figura 7. Imagen izquierda del experimento I

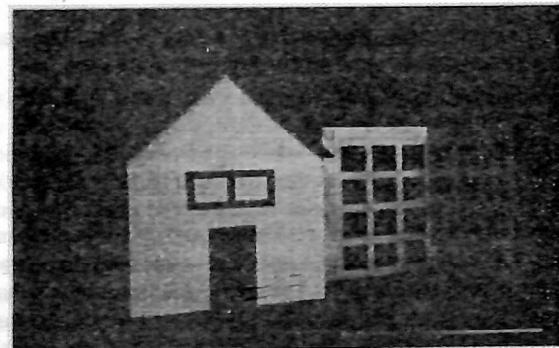


Figura 8. Imagen derecha del experimento I

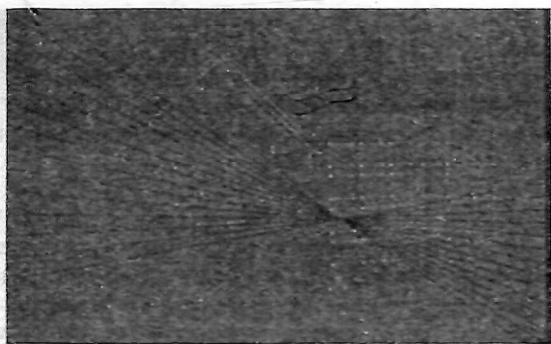


Figura 9. Líneas epipolares sobre la imagen derecha del experimento I

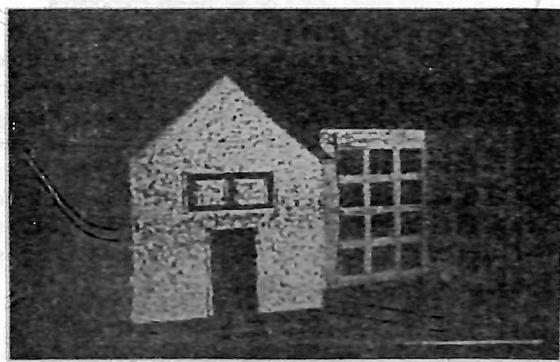


Figura 10. Imagen resultante del experimento I al utilizar una ventana de 5×5



Figura 11. Imagen resultante del experimento I al utilizar una ventana de 19×19

10 MEJORA DE LA IMAGEN

Al observar las imágenes resultantes del experimento anterior, ver Figuras 10 a la 11, es notorio que los puntos que no fueron proyectados se encuentran dispersos para algunas imágenes, mientras que para otras se encuentran concentrados en ciertas zonas. Cualquiera que sea el caso para una imagen, en todas ellas se observa que los puntos que sí fueron proyectados describen en forma detallada los bordes, es decir, zonas donde existen cambios de

contraste, lo que significa que estas zonas son las que tienen un mayor número de puntos proyectados, sin embargo, en las que existe muy poca variación se presenta un mayor número de puntos que no fueron proyectados. Con base a lo anterior, una forma de mejorar la imagen sería asignarle a los píxeles de los puntos que no fueron proyectados, un valor igual al promedio de los píxeles vecinos de los puntos que sí fueron proyectados. En las Figuras 12 y 13 se muestran los resultados obtenidos sobre los casos extremos del experimento: para el caso en que los puntos no proyectados se encuentran dispersos o separados dentro de la imagen, ver Figura 10, y para el caso en que los puntos no proyectados se encuentran juntos o concentrados en ciertas zonas de la imagen, ver Figura 11.

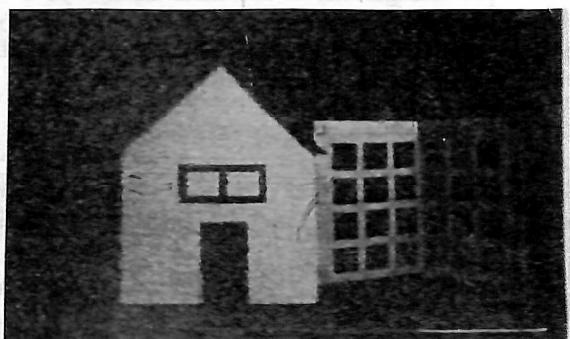


Figura 12. Imagen mejorada tomando como base la imagen de la figura 10

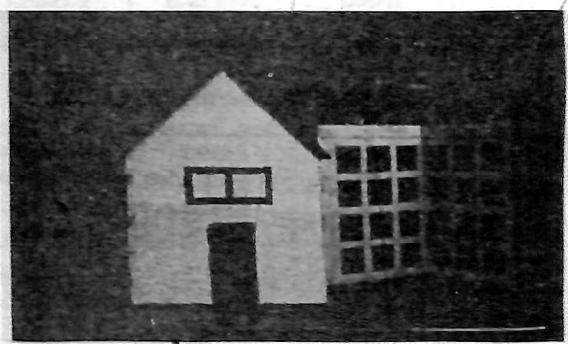


Figura 13. Imagen mejorada tomando como base la imagen de la Figura 11

11 EXPERIMENTO II

En el experimento anterior observamos que en las zonas de la imagen donde existe un mayor contraste, la proyección de los puntos que forman la imagen es mejor comparada con las zonas donde el contraste varía muy poco, y también que el uso de una ventana de mayor tamaño provoca que los puntos no proyectados se encuentren concentrados en ciertas zonas de la imagen, mientras que cuando se utiliza una ventana de menor tamaño, los puntos no proyectados se encuentran dispersos en la imagen.

Tomando como base las observaciones anteriores, para este experimento fue capturada por la cámara una escena real donde existen una gran cantidad de cambios de contraste. Las Figuras 14 y 15, son las imágenes que se tomaron como base para generar la vista intermedia.

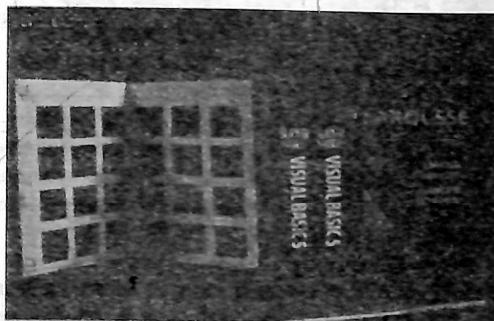


Figura 14. Imagen izquierda, Experimento II

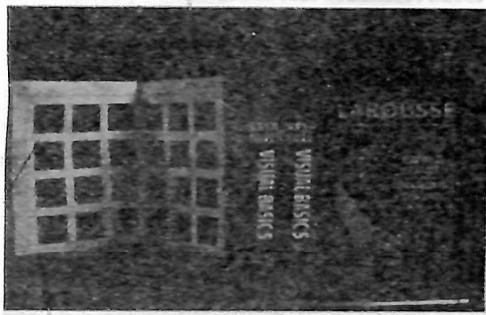


Figura 15. Imagen derecha, Experimento II

Para este experimento el tamaño de la ventana utilizada fue de 11×11 , con una correlación mínima de 0.90. Los parámetros extrínsecos de la cámara, correspondientes a las imágenes que se utilizaron como base para este experimento, son mostrados en la Tabla 3.

Tabla 3: Parámetros extrínsecos, experimento II

Imagen	R_x	R_y	R_z
1	-84.616°	-52.317°	84.671°
2	-86.306°	-36.042°	86.653°
	t_x	t_y	t_z
1	1109.666	244.127	226.360
2	975.014	240.030	258.410

Cuando observamos la imagen resultante y la mejorada, ver Figuras 17 y 18, es evidente que los mejores resultados se logran en las zonas donde existe un mayor cambio de contraste, mientras que en las zonas donde los niveles de gris son homogéneos, la imagen es difusa. Además, los bordes de los objetos que integran la imagen siempre están bien definidos.

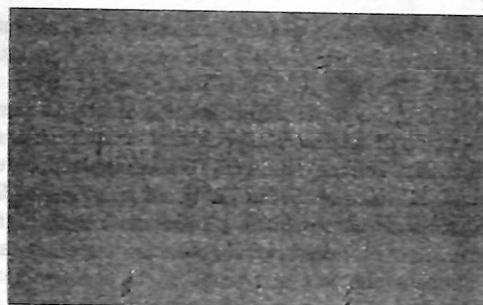


Figura 16. Imagen derecha con algunas líneas epipolares, Experimento II

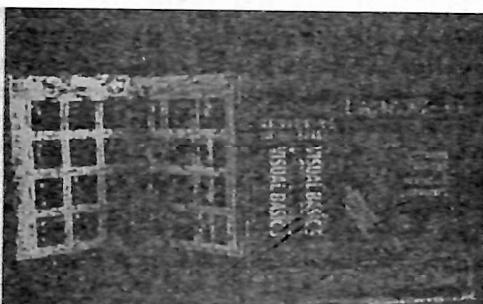


Figura 17. Imagen resultante, Experimento II

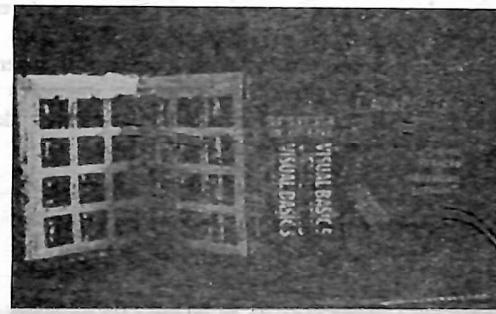


Figura 18. Imagen mejorada, Experimento II

12 CONCLUSIONES

En este trabajo hemos tratado principalmente el problema de la reconstrucción tridimensional apoyándonos en la calibración de cámaras a partir de imágenes. La calibración de una cámara es un proceso delicado y difícil de realizar debido a que factores como: iluminación, temperatura, vibraciones, calidad de las lentes, variación del voltaje, entre otros, afectan el proceso de adquisición de una imagen.

Dentro de la visión por computadora, la geometría es de suma importancia para la formación de imágenes al ser adquiridas por una cámara, así como también en la relación existente entre las imágenes de una escena. A la relación geométrica entre imágenes se le conoce como geometría epipolar, la cual es una simple extensión de las reglas de la

geometría perspectiva de las cuales se derivan reglas generales entre las formas de los objetos y sus imágenes.

El aplicar la geometría epipolar para lograr generar la reconstrucción de una escena, permite conocer a partir de un punto p en una imagen, una recta (línea epipolar) sobre otra imagen, en la cual se encuentra el punto p' correspondiente. Pero el conocer esta recta no es suficiente para lograr la reconstrucción, es necesario conocer las coordenadas del punto p' , para ello nos apoyamos en el apareamiento de puntos sobre las imágenes. A partir del punto p de la primera imagen y para cada uno de los puntos de la recta, calcularlos la correlación existente, asignando a p' las coordenadas del punto sobre la recta con el cual se logró obtener la mayor medida de correlación.

El apareamiento es un aspecto de suma importancia en la reconstrucción tridimensional, así como también para la síntesis de nuevas vistas de una escena, ya que es necesario conocer los puntos p y p' correspondientes sobre las imágenes, con el fin de calcular las coordenadas tridimensionales del punto P , para luego poder proyectarlo en una nueva imagen. Cabe mencionar que el apareamiento es un proceso en el cual existe el inconveniente de que los puntos con mayor correlación no siempre son los puntos que se corresponden, lo que implica tener inconsistencias en la reconstrucción de la escena, y por consiguiente en la generación de una nueva imagen. Al realizar el aparcamiento, los mejores resultados dentro de la imagen se obtienen en las zonas en que existe un gran contraste entre los objetos (bordes), y al referirnos a la imagen por completo, los mejores resultados se obtienen al tener un menor ángulo de desplazamiento entre las imágenes que sirven como base para realizar la reconstrucción de la escena.

TRABAJOS FUTUROS

Un trabajo por realizar consiste en integrar las problemáticas planteadas en este artículo al trabajo realizado por Olague (1998), en relación a la planificación de redes fotogramétricas, con el propósito de obtener mejores reconstrucciones.

Para la calibración de una cámara de acuerdo al método utilizado en este trabajo, es necesario conocer las coordenadas tridimensionales de ciertos puntos en la escena que se desea reconstruir, así como también las coordenadas bidimensionales de la proyección de estos puntos sobre la imagen. Además, las coordenadas de los puntos sobre la imagen se trabajan a nivel de píxeles. Existen métodos que permiten efectuar el cálculo de dichos puntos a nivel subpixel, lo cual ayuda en gran medida a lograr una mejor calibración de una cámara, Olague y Hernández (2001).

La restricción epipolar, o más general la geometría epipolar, juega un papel fundamental en la visión estereoscópica. Sería interesante analizar en qué medida es posible estimar la matriz fundamental sin calibrar la cámara para cada una de las imágenes, apoyándose únicamente en la

correspondencia izquierda-derecha de los puntos de las imágenes sin ningún conocimiento de las coordenadas tridimensionales. El método más directo es un método lineal (Hartley, 1997). En este método para cada correspondencia $(p \leftrightarrow p')$ existe una ecuación lineal de la forma:

$$f_{11}uu' + f_{12}vu' + f_{13}u' + f_{21}uv' + f_{22}vv' + f_{23}v' + f_{31}u + f_{32}v + f_{33}uu' = 0,$$

a partir de la cual, conociendo ciertos puntos con coordenadas (u, v) de una imagen y sus correspondientes (u', v') de otra imagen, se podría formar un sistema de ecuaciones lineales que serviría para calcular la matriz fundamental.

El apareamiento, uno de los aspectos de mayor importancia en la reconstrucción, presenta ciertos inconvenientes debido a que los puntos con una mayor medida de correlación algunas veces no son los puntos que se corresponden en las imágenes, por lo que se podría tratar de utilizar y comparar nuevos métodos con el propósito de encontrar aquellos que nos ofrezcan resultados más robustos.

AGRADECIMIENTOS:

Esta investigación fue patrocinada a través del contrato 35267-A de CONACYT, México.

REFERENCIAS

- Blanc, J., *Synthèse de nouvelle vues d'une scène à partir d'images existentes*, PhD thesis, Institut National Polytechnique de Grenoble, Rance, 1998.
- Boufama, B., *Reconstruction @dimensionnelle en Vision par Ordinateur*, PhD thesis, Institut National Polytechnique de Grenoble, France, 1994.
- Faugeras, O., *Three-Dimensional Computer Vision a Geometric Point of View*, MIT Press, Cambridge, Massachusetts, 1996.
- González, R. and Woods, R., *Tratamiento Digital de Imágenes*, Addison-Wesley/Díaz de Santos, Wilniington, Delawere, 1996.
- Horaud, R. and Monga, O., *Vision par Ordinateur*, Hermès, Paris, FY-ance, 1995.
- Kitchen, L. and Rosenfeld A., *Gray level corner detection*, *Pattern Recognition Letters*, No. 1, pp 95-102, 1982.
- Lu, X. and Zhou, J., *Robust corner detection under varying illumination*, Department of Automatitación, Tsinghua University, Beijing 100084, P.R. China, 1999.
- Olague, G., *Planification du placement de cameras pour des mesures 3D de précision*, PhD thesis, Institut National Polytechnique de Grenoble, France, 1998.
- Olague, G. and Ramírez, R., *Artificial Image Generation from Digital Images*, In American Society for Photogrammetry and Remote Sensing, Annilal Conference Proceedings, St. Louis, MO, 12 pages, April 23-27, 2001.

Olague, G. and Hernández, B., *Autonomous Model Based Corner Detection using Evolutionary Algorithms*. In American Society for Photogrammetry and Remote Sensing, Annual Conference Proceedings, St. Louis, MO, 12 pages, April 23-27, 2001.

Olague, G., *Automated Photogrammetric Network Design using Genetic Algorithms*, Photogrammetric Engineering & Remote Sensing, (in press) 2002.

Olague, G. and Mohr, R., *Optimal Camera Placement for Accurate Reconstruction*. Pattern Recognition, Vol. 35, No. 4, pp. 927-944, April 2002.

Hartley, R., *In Defense of the Eight-Point Algorithm*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 6, June 1997.

Xu, G. and Zhang, Z., *Epipolar Geometry in Stereo, motion and Object Recognition*, Kluwer Academic Publisher, Dordrecht, Netherlands, 1996.

