

Conceptos y Técnicas de Estereovisión por Computador

José M. López-Valles¹, Antonio Fernández-Caballero², Miguel A. Fernández²

¹Universidad de Castilla-La Mancha, Escuela Universitaria Politécnica de Cuenca
Departamento de Ingeniería de Telecomunicación, Cuenca, 13071
josemaria.lopez@uclm.es

²Universidad de Castilla-La Mancha, Escuela Politécnica Superior de Albacete
Departamento de Informática, Albacete, 02071
{caballer, miki}@info-ab.uclm.es

Resumen

La extracción de la estructura tridimensional de una escena a partir de imágenes estereoscópicas es un problema que ha sido estudiado en las últimas décadas por la comunidad de visión por computador. Los primeros trabajos estaban enfocados hacia los fundamentos de la correspondencia entre las imágenes y la geometría de la estereoscopia. La investigación ha madurado significativamente a través del tiempo y son muchos los avances que siguen produciéndose en estereovisión computacional, lo que ha permitido que la estereoscopia pueda ser aplicada a nuevos problemas. Este artículo presenta una panorámica general de los principales conceptos y técnicas de estereovisión por computador. Se incide fundamentalmente en el análisis geométrico y la calibración en estereovisión, en los principales retos que supone el problema de la correspondencia, así como en las principales técnicas usadas en visión estéreo por computador.

Palabras clave: Estereovisión, Correspondencia, Geometría, Calibración.

1. Introducción

Vivimos en un mundo tridimensional y dinámico en el que los hechos acaecen a nuestro alrededor en las dimensiones horizontal, vertical y en profundidad. Sin embargo, la información de profundidad se pierde al captar una escena a través de una cámara. Esto se debe a la integración que se produce en cada píxel de la información del ángulo sólido subtendido por el centro óptico del objetivo de la cámara y los límites físicos de dicho píxel en el dispositivo fotosensible. A raíz de esto, una imagen bidimensional aislada no da una información suficiente en cuanto a la profundidad de los objetos de la escena. Los humanos y los animales son capaces de recuperar la tridimensionalidad perdida debido a que disponen de dos ojos a modo de cámaras situados en dos puntos diferentes y captando perspectivas de la escena también distintas. Si no fuera así, la percepción visual se

reduciría a planos transversales a la dirección en la que estamos viendo. Es cierto que los humanos, cerrando uno de los ojos, somos capaces de estimar la profundidad a la que están los objetos, aunque con una escasa precisión. Esto lo hacemos de una manera creativa y relativamente sencilla, que consiste en reconocer los objetos que aparecen en la imagen singular y compararlos en tamaño con el recuerdo que tenemos sobre ellos cuando están más o menos lejos. Lo mismo ocurre cuando vemos una fotografía. La estimación de la profundidad la realizamos también de este modo creativo.

Este proceso de estimación, que para los humanos parece casi trivial por lo habitual que nos resulta y por la rapidez y eficacia con la que lo realizamos, para un ordenador supone un trabajo importantísimo si tenemos en cuenta el tiempo que se tardaría en realizarlo y la cantidad de recursos necesarios [Brown et al. 03].

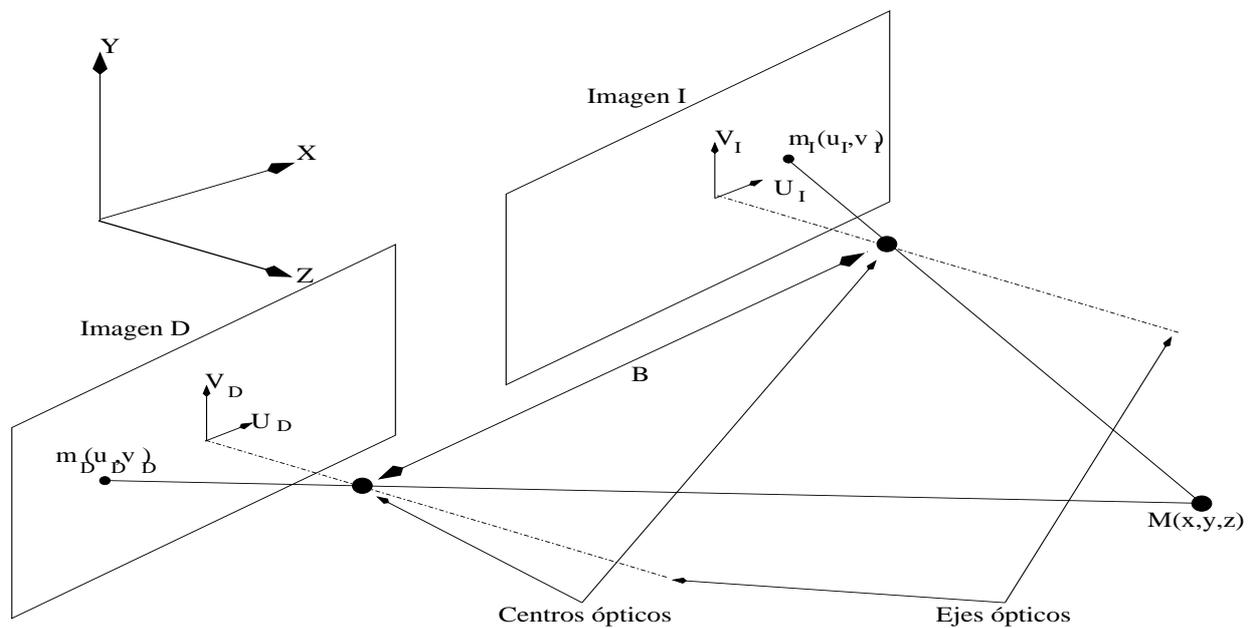


Figura 1. Geometría estereoa: Marco de referencia y sistemas de referencia de las cámaras.

Para los ordenadores y otras máquinas, esta tridimensionalidad es más sencilla y precisa recuperarla utilizando, al menos, una segunda cámara orientada hacia la misma escena, y que esté mirando desde otra perspectiva cercana pero diferente de la primera. Analizando estas dos imágenes, y la geometría del sistema formado por las dos cámaras y la escena, es como se puede recuperar la profundidad de cada uno de los objetos visibles. Este es el objetivo de la estereovisión.

Tradicionalmente, el análisis de la estereovisión se ha dividido en dos problemas diferentes. Partiendo de dos imágenes bidimensionales en el espacio de coordenadas (u, v) , la imagen (*I*) izquierda y (*D*) derecha de la figura 1, el problema de la correspondencia trata de buscar qué dos puntos $m_I(u_I, v_I)$ de la imagen izquierda y $m_D(u_D, v_D)$ de la imagen derecha corresponden a un mismo punto *M* del espacio tridimensional (X, Y, Z) . Una vez encontrados estos dos puntos, el problema de la reconstrucción trata de encontrar las coordenadas del punto *M*. En la figura 1 se observa el eje óptico que es la línea imaginaria ortogonal al plano imagen y que contiene el centro óptico del objetivo.

La tarea más difícil es sin duda responder al problema de la correspondencia. Como, en general, hay varias posibilidades para escoger el elemento correspondiente en la imagen *D* de un elemento de la imagen *I*, el problema de la correspondencia

estéreo se dice que es ambiguo. Debido a esta ambigüedad, se hace necesario averiguar qué elementos, qué características, qué restricciones y qué consideraciones se pueden aplicar para reducirla al máximo. A ello se dedicará una parte importante de este apartado; pero en primer lugar se analiza el problema de la reconstrucción, que es un problema cerrado y determinista.

La extracción de la estructura tridimensional de una escena a partir de imágenes estereoscópicas es un problema que ha sido estudiado ampliamente en las últimas décadas por la comunidad de visión por computador. Los primeros trabajos estaban enfocados hacia los fundamentos de la correspondencia entre las imágenes y la geometría de la estereoscopia. La investigación ha madurado significativamente a través del tiempo y son muchos los avances que siguen produciéndose en estereovisión computacional, lo que ha permitido que la estereoscopia pueda ser aplicada a nuevos problemas.

Este artículo presenta una panorámica general de los principales conceptos y técnicas de estereovisión por computador. Se incide fundamentalmente en el análisis geométrico y la calibración en estereovisión, en los principales retos que supone el problema de la correspondencia, así como en las principales técnicas usadas en estereo por computador.

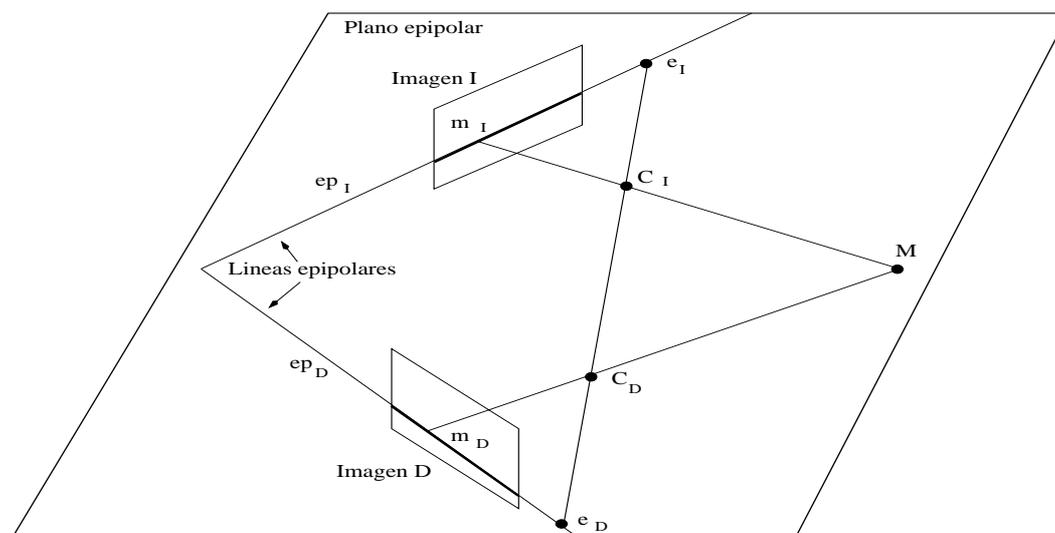


Figura 2. Restricción epipolar. Se representa el plano epipolar correspondiente al punto P de la escena y las líneas epipolares correspondientes.

2. Análisis Geométrico y Calibración

En este apartado se plantea la resolución del problema de la reconstrucción. Esta resolución parte de la geometría del sistema y puede plantearse como un problema determinista con una solución única basada en la triangulación.

En la figura 1 se ha representado un sistema de referencia (U,V) para cada cámara, y un sistema de referencia (X,Y,Z) general de la escena; éste último es el llamado *marco de referencia* del sistema estéreo. Para este estudio consideraremos que la imagen se forma a la distancia focal por detrás de la lente objetivo, y que los rayos que pasan por el centro de dicha lente conservan su dirección.

2.1. Plano y Líneas Epipolares

Antes de entrar a analizar las dos geometrías básicas en estereovisión, es decir, la geometría de cámaras paralelas y la geometría de cámaras convergentes, es necesario describir una serie de elementos fundamentales y comunes a ambas geometrías. Estos elementos son el plano epipolar, las líneas epipolares y los epipolos. En la figura 2, consideramos *plano epipolar* aquel que forman los dos centros ópticos C_I y C_D de los objetivos de las cámaras con cualquier punto M del espacio objeto. Por otra parte, el plano epipolar ($C_I M C_D$) corta a las dos superficies imagen I e imagen D en sendas

líneas rectas ep_I y ep_D , que se denominan *líneas epipolares*. Finalmente, la proyección del centro óptico de cada cámara sobre la otra cámara define el llamado epipolo e . Los epipolos de cada una de las cámaras (e_I y e_D) van a ser los puntos por los que van a pasar todas las líneas epipolares.

Utilizando cualquier plano epipolar como base, todos los puntos de la escena que pertenecen a este plano tendrán su imagen en cada una de las líneas epipolares de las dos imágenes. Esto implica que cualquiera de los píxeles de una línea epipolar, sea derecha o izquierda, tendrá su correspondiente dentro de la línea epipolar correspondiente en la otra imagen.

En general, las líneas epipolares son oblicuas respecto del plano retinal o superficie fotosensible de la cámara. Pero existe una excepción muy interesante que simplifica de gran manera el análisis. Si consideramos que los planos retinales son coincidentes entre sí, y paralelos a la línea base (C_I, C_D) que une los dos centros ópticos de las cámaras, observaremos que los epipolos de ambas cámaras se situarán en el infinito, y por lo tanto, las líneas epipolares serán todas paralelas entre sí, y paralelas a su vez a la línea base. Esta configuración especial se denomina configuración de cámaras paralelas. Con esta configuración, y un adecuado posicionamiento de los planos retinales se puede conseguir que las líneas epipolares coincidan con las filas de las imágenes digitales tomadas [Sáez et al. 03].

2.2. Geometría de Cámaras Paralelas

La configuración de cámaras paralelas es la más sencilla de tratar en cuanto a geometría; por ello, es la primera que se analiza en este trabajo. Para realizar este sencillo análisis se parte de la figura 1 anterior, en la que se considera que el marco de referencia, o el sistema de referencia de la escena tiene su origen en el centro óptico de la cámara izquierda, su eje X coincide con la línea base, y su eje Z coincide con el eje óptico de dicha cámara. En la figura 3 se representa esta configuración considerando exclusivamente un plano epipolar.

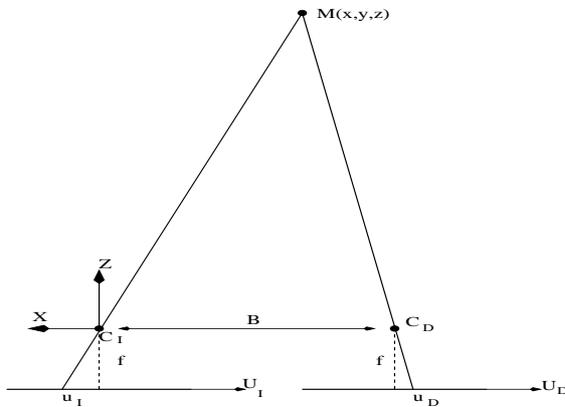


Figura 3. Geometría en el plano epipolar para cámaras paralelas.

El objetivo es hallar las coordenadas del punto $M(x, y, z)$ partiendo de las coordenadas de las proyecciones de dicho punto sobre los planos de imagen (u_1, v_1) y (u_2, v_2) . Para ello se necesita la distancia focal f de las cámaras y la distancia entre sus dos centros ópticos o línea base B .

La reconstrucción tridimensional puede resolverse, bien mediante geometría proyectiva, o bien mediante geometría euclídea. En el caso de la geometría euclídea, que es la forma más sencilla y ofrece unos resultados suficientemente aproximados tenemos, para la imagen I , por semejanza de triángulos:

$$\frac{u_I}{f} = \frac{x}{z} \quad ; \quad x = \frac{u_I}{f} \cdot z$$

Y de la misma manera para la dirección Y :

$$\frac{v_I}{f} = \frac{y}{z} \quad ; \quad y = \frac{v_I}{f} \cdot z$$

Para la imagen D se tiene:

$$x = \frac{u_D}{f} \cdot z - B \quad ; \quad y = \frac{v_D}{f} \cdot z$$

Desarrollando estas ecuaciones podemos llegar a las expresiones siguientes:

$$x = \frac{u_I \cdot B}{d} \quad ; \quad y = \frac{v_I \cdot B}{d} \quad ; \quad z = \frac{f \cdot B}{d}$$

En estas expresiones el valor d es la denominada *disparidad* que hace referencia a la diferencia entre las coordenadas u_I y u_D respecto del centro de sus imágenes. Para la obtención de dichas coordenadas u_I y u_D se ha tomado como origen de coordenadas de cada imagen (U, V) el punto de corte del eje óptico y el plano retinal.

$$d = u_D - u_I$$

Además, al conjunto de todas las disparidades entre dos imágenes de un par estéreo se denomina *mapa de disparidad*. Claramente, las disparidades sólo se pueden calcular a partir de aquellas características que son visibles en las dos imágenes; las características que sólo se ven en una imagen y no en la otra se denominan *oclusiones*. Con estas expresiones podemos concluir que, una vez conocidas la distancia focal de las cámaras, la línea base y la disparidad entre los píxeles correspondientes, es sencillo calcular las coordenadas (x, y, z) del punto M del espacio para la configuración de cámaras paralelas.

La digitalización de las imágenes estéreo obliga a que la imagen tridimensional de la escena también sea discretizada. Si queremos calcular el error que se comete en el cálculo de las coordenadas por el hecho de tener un dispositivo fotosensible con un tamaño de píxel discreto, derivando las expresiones anteriores llegamos a las expresiones siguientes:

$$\delta x = -\frac{u_I B}{d^2} \cdot \delta d \quad ; \quad \delta y = -\frac{v_I B}{d^2} \cdot \delta d \quad ;$$

$$\delta z = -\frac{f \cdot B}{d^2} \cdot \delta d$$

Esto implica que cuanto menor sea la disparidad, el punto del espacio estará más lejos y, a su vez, la precisión con la que podemos calcular las coordenadas (x, y, z) tendrá un valor finito que es inversamente proporcional al cuadrado de la disparidad. Para orientarnos en la magnitud de las medidas que se han realizado, quizá lo mejor es ver lo que ocurre con datos numéricos concretos. Supongamos que tenemos un sistema cuyas cámaras tienen una distancia focal de 35 mm, un tamaño de píxel de 25 μm , y están separadas entre sí una distancia de 70 mm (línea base). Con estos parámetros, para distintas disparidades tendremos diferentes medidas de profundidad y diferentes

errores de estimación. Por ejemplo, para una disparidad de 1 píxel ($d=25 \mu\text{m}$) el punto físico M tendrá una coordenada z según las expresiones anteriores de 98 m, con un error debido a la discretización de 98 m (± 49 m). Si la disparidad es algo mayor, significa que el objeto estará más cerca. Una disparidad de 5 píxeles, da unos resultados de $z=19,6$ m y un error $\delta z=3,9\text{m}$. En el caso de disparidades de 10 píxeles, los resultados son $z=9,8$ m y un error $\delta z=0,98\text{m}$.

Todo esto significa que, por el hecho de discretizar las imágenes tomadas por las cámaras con un tamaño de píxel determinado, también se está discretizando la imagen tridimensional en celdas de tamaño finito en las tres direcciones del espacio.

2.3. Geometría de Cámaras Convergentes y Rectificado

La configuración de cámaras paralelas se utiliza con mucha frecuencia debido a su simplicidad; pero siendo físicamente posible, en la práctica resulta dificultoso alinear dos sistemas ópticos de forma tan precisa y suficientemente estable. En el caso general en el que, como se puede ver en la figura 2, los ejes ópticos de las cámaras convergen hacia un punto finito del espacio objeto, el procedimiento a seguir sería el siguiente: en primer lugar calcular la posición de los epipolos e_I y e_D de cada una de las cámaras. Con ello, cualquier punto de la imagen izquierda m_I estará contenido en una línea epipolar ep_I que lo une con el epipolo izquierdo. Dicha línea epipolar cortará al plano imagen de la cámara derecha en un punto, que junto con el epipolo derecho e_D formará la línea epipolar derecha ep_D , sobre la que habrá que buscar su correspondiente. Además, la relación entre cada sistema de referencia (U,V) de cada una de las imágenes con el marco de referencia (X,Y,Z) no es tan trivial como en el caso anterior. Es necesario transformar las componentes de los vectores (u,v) en coordenadas (x,y,z) de la escena con la complejidad que esto supone.

Habida cuenta de esta complejidad, lo que se suele hacer es aplicar el proceso de rectificado para convertir una geometría general de cámaras convergentes en una geometría más simple de cámaras paralelas. Esto es posible, como puede verse en la figura 4, volviendo a proyectar las imágenes sobre un mismo plano imagen' virtual utilizando los mismos centros ópticos de las cámaras. En este caso, el plano imagen' es el mismo para las dos nuevas retinas imagen I' e imagen D'.

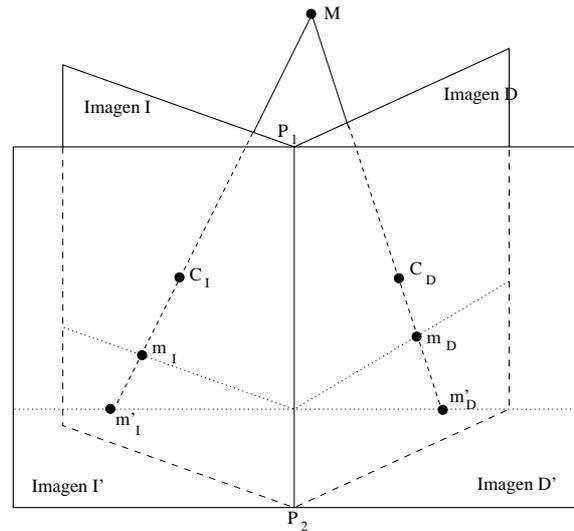


Figura 4. Rectificación de planos.

Para que los dos epipolos estén en el infinito, el plano R debe ser paralelo a la línea que une los dos centros ópticos (C_I, C_D). Y para minimizar la distorsión de las imágenes re proyectadas sobre el nuevo plano R , interesa que éste sea también paralelo a la línea $\langle P_1, P_2 \rangle$ de intersección de las dos retinas originales.

Por otra parte, la distancia de este plano imagen' a la línea que une los centros ópticos $\langle C_I, C_D \rangle$ corresponde a un cambio de escala; aunque quizá para minimizar este cambio de escala, lo propio será utilizar como medida de esta distancia la distancia focal real de las cámaras. Una vez proyectadas las imágenes sobre el plano imagen' se tendrán las líneas epipolares paralelas a las filas de la imagen. Pueden consultarse las referencias [Trucco et al. 98] y [Zitnick et al. 00] para una discusión más profunda sobre la rectificación.

2.4. Calibración

El problema de la reconstrucción puede resolverse mediante la intersección de las rectas $\langle C_I, m_I \rangle$ y $\langle C_D, m_D \rangle$ de la figura 2. El resultado dependerá de con qué precisión se conocen las posiciones de C_I y C_D y los planos derecho e izquierdo en el sistema de coordenadas. Esto nos lleva al problema de la calibración, puesto que, si las posiciones m_I y m_D no se conocen con precisión, las rectas $\langle C_I, m_I \rangle$ y $\langle C_D, m_D \rangle$ podrían no cortarse.

Mediante la calibración de un sistema de estereovisión se estiman los parámetros intrínsecos

(distancia focal, centro óptico y distorsiones de las lentes) y extrínsecos (posiciones relativas y orientaciones) de las cámaras que lo componen. Existen dos métodos comúnmente usados para la calibración: la autocalibración y la calibración fotogramétrica. En la autocalibración se toman varias imágenes de una misma escena y mediante la correspondencia entre puntos de distintas imágenes se pueden encontrar los mejores parámetros del modelo que puedan otorgar dicha correspondencia. La reconstrucción de la escena tridimensional realizada con el modelo encontrado no es precisa, ya que ésta está afectada por un factor de escala. Con este método no se puede saber cuál es el tamaño real de los objetos captados por las cámaras. Esto es así puesto que un objeto pequeño cerca del centro óptico podrá tener la misma imagen que el mismo objeto más grande alejado de dicho centro óptico.

Si lo que se busca es una reconstrucción precisa, como es el caso en muchas de las aplicaciones de la robótica, es recomendable utilizar la calibración fotogramétrica. Esta calibración utiliza un objeto tridimensional de referencia cuya geometría es conocida a la perfección.

En la figura 5 aparecen dos ejemplos. Se escogen N puntos de interés del objeto de referencia, obteniéndose así las coordenadas $M'_i = (x'_i, y'_i, z'_i)$, para $i = 1, \dots, N$. El objeto es a continuación captado por el sistema de cámaras, y sus puntos de interés son vistos como puntos bidimensionales con coordenadas $m_i = (u_i, v_i)$. A partir de las medidas reales y las estimadas por el sistema estereoscópico, es posible obtener una estimación teórica de los puntos en tres dimensiones.

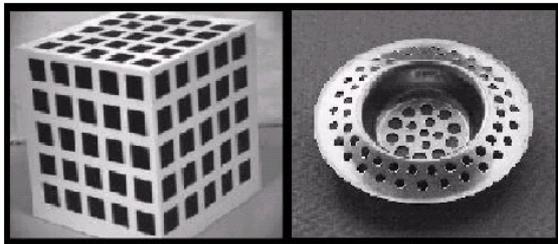


Figura 5. Ejemplos de objetos para calibración.

Así pues, se calculan los puntos en función de los parámetros desconocidos, como la línea base B y la distancia focal de las cámaras f . Una vez llegado a este punto, el problema de la calibración se transforma en un problema estadístico mediante el cual se calculan los parámetros B y f que minimizan el error cuadrático medio de las medidas tomadas

respecto de los resultados obtenidos, partiendo de las imágenes.

Para una mayor profundidad en la discusión acerca de la calibración en trabajos recientes puede verse [Faugeras et al. 01] y [Hartley et al. 00].

3. Primitivas de Correspondencia

Al contrario del problema geométrico, que es un problema con solución cerrada, el problema de la correspondencia es un problema más complejo de resolver por la ambigüedad que existe al buscar posibles correspondencias entre los píxeles de las dos imágenes. Con el fin de comenzar a minimizar al máximo esta ambigüedad, en primer lugar se buscan características de las imágenes (primitivas) que sirvan para establecer con mayor robustez dichas correspondencias. Si nos quedamos simplemente comparando los píxeles uno a uno, encontraremos varios píxeles iguales entre los que escoger, y sus emparejamientos nos llevarían a varios puntos posibles de la escena 3D. En cambio, si utilizamos primitivas más complejas, formadas por conjuntos de píxeles, como contornos o regiones, será menos probable encontrar falsas correspondencias.

El píxel individual es la primitiva más simple de todas. El píxel se utiliza en las técnicas de correlación, pero para ello hemos de caracterizarlo, de forma que si los píxeles m_l y m_p se corresponden, sus características sean iguales o aproximadamente iguales. La característica más simple es una medición de la radiación. Existe un problema a la hora de utilizar píxeles individuales como primitivas, y es que la luz reflejada por una superficie de un objeto es función de la posición de las fuentes de luz, de la orientación de la superficie y de la dirección de la visión. Así pues, las imágenes tomadas de un mismo punto físico, pueden variar en luminosidad dependiendo de estos factores. Este comportamiento se resume en la función de reflectancia. Para superficies Lambertianas (las que son totalmente mates) la luz reflectante es igual en todas las direcciones. Así que como primera aproximación se puede considerar que todas las superficies son de este tipo, para después observar los errores que se producen con esta simplificación. Como resultado, las intensidades de los dos puntos correspondientes de una superficie Lambertiana son las mismas.

Si se quiere tener una mayor información sobre la correspondencia entre dos píxeles, una posibilidad utilizada habitualmente es analizar éstos junto con sus entornos. De esta manera, resulta más sencillo descartar correspondencias falsas producidas por la igualdad entre píxeles individuales, ya que el entorno proporcionará una mayor información.

La siguiente primitiva a considerar es el píxel de contorno. Los píxeles de contorno pueden ser caracterizados mediante medidas basadas en su intensidad y en su geometría, como puede ser su longitud, su orientación, su curvatura y el contraste medio a lo largo de él. También se puede pensar en agrupar los píxeles de contorno para formar curvas o trozos de curvas. La curva más simple y que además es invariante mediante proyección perspectiva es la línea recta.

No todos los contornos son buenos para la visión estéreo. Dos casos particulares son los que podemos llamar los contornos ocultos y las discontinuidades suaves, donde el objeto termina de forma redondeada, en cuyo caso las dos cámaras no captan el contorno en la misma parte del objeto. Excepto en estos casos, los contornos son fuentes muy válidas y muy robustas de información para guiar el proceso de correspondencia estéreo. El inconveniente más importante que tiene el trabajo con contornos es lo sensibles que son al ruido. Dependiendo de la calidad de las imágenes, los contornos pueden quedar perfectamente definidos y uniformes, o, por contra, discontinuos y llenos de irregularidades. En este segundo caso es necesario acudir a información o procesamientos posteriores para conseguir píxeles de contorno fiables.

Finalmente, podemos utilizar regiones de las imágenes como primitivas para analizar su correspondencia en el procesamiento estéreo. Existen muchas características que se pueden extraer de dichas regiones para establecer las correspondencias: su nivel medio de brillo, su tamaño, su perímetro, su forma, etc. El proceso de dividir una imagen en regiones se denomina segmentación y pueden encontrarse muchos métodos en la bibliografía. Generalmente, los algoritmos de segmentación de imágenes se basan en dos propiedades fundamentales de la intensidad de los píxeles. Por un lado, los cambios abruptos, como son las líneas y los contornos de los objetos, y por otro lado, las regiones cuyos píxeles son similares de acuerdo con una serie de criterios predefinidos.

La segmentación basada en los contornos de las figuras ha sido el método principal utilizado por los algoritmos de segmentación durante muchos años. Los dos operadores matemáticos más habituales para la detección de discontinuidades son el operador gradiente y el Laplaciano de un Gaussiano. Tomando como base estos operadores, la forma más común de buscar discontinuidades de niveles de gris en imágenes digitales es mediante la convolución con máscaras de tamaños 2x2, y más habitualmente 3x3. Varios autores han desarrollado diversas máscaras tomando como base el operador gradiente. Algunos ejemplos clásicos son: Roberts [Roberts 65], Prewitt [Prewitt 70], Sobel [Sobel 78] y Canny [Canny 86]. Por otra parte, las propiedades del cruce por cero del operador Laplaciano en contornos las podemos encontrar en un artículo de Marr y Hildreth [Marr et al. 80] y en el clásico libro de Marr [Marr 82].

La umbralización es una técnica que cuenta con una gran popularidad, ya que es muy sencilla de implementar. Existe una cantidad considerable de trabajo en este ámbito, que puede verse en los artículos de revisión de Sahoo et al. [Sahoo et al. 88], de Lee et al. [Lee et al. 90] y de Russ [Russ 95]. En ellos, se proponen métodos automáticos para el ajuste de los umbrales, utilizando bien el histograma o bien la propia imagen como guía.

Finalmente hemos de considerar la segmentación basada en regiones como tales. Dos de los métodos más referenciados son el de “unión y división” y el “crecimiento de regiones”, que son métodos opuestos por el vértice.

El método de “unión y división” es un método “top-down” o descendente, que comienza considerando como región la imagen completa. Eligiendo una característica concreta como criterio de uniformidad, se evalúa si toda la región es uniforme. A menudo este criterio se basa en las características del histograma de brillo. Si se considera que la región no es uniforme, se divide en varias subregiones, examinándose cada una de la misma manera.

En cuanto al crecimiento de regiones, éste comienza desde abajo, o a nivel de píxel. A partir de un determinado píxel, se examinan los píxeles vecinos y, si son suficientemente similares, se añaden a la región creciente. Esta similitud puede ser evaluada considerando toda la región, o simplemente los píxeles adyacentes, permitiendo regiones con

cambios de brillo graduales. El proceso termina cuando no se puede añadir ningún píxel más.

4. Restricciones a las Correspondencias

La búsqueda del píxel correspondiente a m_I , en teoría podría abarcar toda la imagen derecha, trabajo que es, sin duda, muy laborioso. Afortunadamente, existen algunas características de la geometría del sistema y de la geometría de los objetos que nos van a restringir considerablemente esta búsqueda, reduciendo el número de posibles correspondencias potenciales de m_I . Estas restricciones son de tres tipos básicos.

- Restricciones geométricas impuestas por el sistema de captación de las imágenes. Probablemente la restricción más importante sea la restricción epipolar, gracias a la cual podemos transformar una búsqueda en 2 dimensiones en otra de 1 dimensión.
- Restricciones geométricas impuestas por los objetos a los que estamos mirando. Podemos asumir que, dada la continuidad de los objetos de las escenas, las distancias al origen de los puntos cercanos de la escena varían lentamente en todas las direcciones excepto en las discontinuidades producidas en los contornos. Este es el origen de la restricción de disparidad y del gradiente de disparidad. Otra restricción de este tipo aparece si asumimos que los objetos a los que estamos mirando se aproximan a poliedros.
- Restricciones físicas como aquellas que proceden de modelos por los que los objetos interactúan con la iluminación: fuentes y reflexiones en los objetos. De todos ellos, el modelo más simple y el más usado es el modelo Lambertiano ya comentado.

4.1. Restricción Epipolar

Esta restricción se basa en la condición vista en el apartado 2.1. Cada punto M del espacio forma con los centros ópticos de las cámaras, C_I y C_D , el plano epipolar $\langle C_I M C_D \rangle$. Este plano epipolar corta a las superficies fotosensibles imagen I e imagen D de las cámaras en sendas líneas rectas ep_I y ep_D , que llamamos líneas epipolares. La existencia de estas líneas epipolares va a implicar que el correspondiente m_D en la imagen derecha de un píxel m_I de la imagen izquierda caerá

inevitablemente en la línea epipolar ep_D correspondiente de ep_I , y viceversa.

Esto restringe la búsqueda de correspondencias que anteriormente se suponía en todo el plano imagen contrario a exclusivamente la línea epipolar, reduciendo considerablemente los cálculos a realizar. Además, si se utiliza la configuración de cámaras paralelas, las líneas epipolares pueden hacerse coincidir con las filas de las imágenes.

4.2. Restricción de Unicidad

Esta es una restricción basada en la geometría del sistema de visión y en la naturaleza de los objetos de la escena. Viene a decir que, en el caso general en el que consideramos opacos todos los objetos de la escena, a cualquier píxel de la imagen derecha sólo le ha de corresponder uno de la imagen izquierda. Esto significa que en el caso de tener varios píxeles potencialmente correspondientes, habrá que escoger aquel que nos dé una mayor confianza. El análisis de confianza se puede realizar considerando cualquiera de las otras restricciones aquí expuestas.

4.3. Restricción de Ordenamiento

La restricción de ordenamiento indica que, si dos píxeles m_I y n_I de la línea epipolar de la imagen izquierda están situados m_I a la derecha de n_I , sus correspondientes m_D y n_D estarán situados en la línea epipolar derecha de forma que m_D estará a la derecha también de n_D . Esta restricción se cumple en un muy alto porcentaje de los casos, pero ahora veremos que existen algunas excepciones.

Considerando la figura 6, en ella aparece un punto M de la escena tridimensional y sus proyecciones m_I y m_D en las retinas izquierda y derecha respectivamente. Si se escoge otro punto N en el triángulo definido por $\langle MC_I C_D \rangle$, véase la zona sombreada de la figura, N tiene las imágenes n_I y n_D en distinto orden respecto de M en las retinas izquierda y derecha. Es fácil ver que sucede lo contrario con cualquier punto de la zona no rayada. Existen dos argumentos para llamar a la zona rayada como zona prohibida asociada a M :

- Lo primero es que la distancia de M a las cámaras suele ser mucho mayor que la longitud B de la línea base. De esta manera, el ángulo $C_I M C_D$ es muy pequeño y la probabilidad de que un punto N caiga dentro de él es muy baja.

- b) El segundo argumento es que, si asumimos que M y N están situados en objetos opacos de espesor no despreciable, entonces M esconderá detrás de sí a N , no siendo posible verlos simultáneamente en las retinas izquierda y derecha. Así que, podemos decir que N está en la zona prohibida definida por M , considerando el orden de sus imágenes a lo largo de las líneas epipolares.

En la práctica es difícil eliminar toda la zona rayada, ya que existirán configuraciones como la de la figura 6, en la que los puntos M y N pertenecen a objetos distintos y ambos pueden verse. En este caso la restricción de orden no se aplicaría. Así parece razonable forzar sólo a los vecinos de M , con una cercanía pequeña, a pertenecer a la zona prohibida.

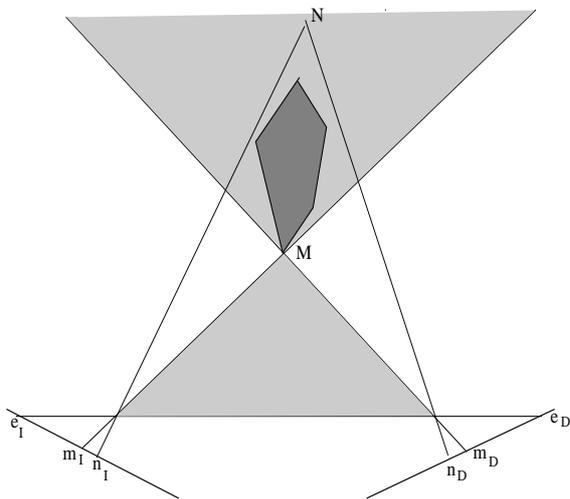


Figura 6. Restricción de ordenamiento: zona prohibida.

4.4. Restricción de Disparidad

Esta restricción está basada en la geometría de los objetos e impone un rango de profundidades posibles para los objetos de la escena (ver figura 7). Esta restricción considera que no es probable que existan objetos muy cerca de las cámaras, ya que la escena suele tomarse a media distancia. Si se descarta la posibilidad de que existan objetos muy cercanos, estamos eliminando disparidades excesivamente grandes. Nuevamente se está limitando la búsqueda en toda la línea epipolar a una búsqueda restringida a un segmento de dicha línea ([Sumi et al. 02] [Hong et al. 04]). Esta restricción está basada en la geometría de la escena.

La figura 7 muestra, como se ha mencionado antes, que sólo objetos muy cercanos producen disparidades grandes. Como la probabilidad de que existan objetos tan cercanos es baja, suele aplicarse la restricción de disparidad.

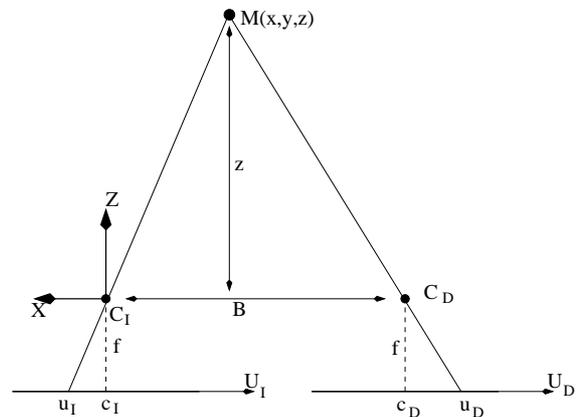


Figura 7. Disparidad en geometría de cámaras paralelas.

Los mejores trabajos iniciales relacionados con la restricción de disparidad son el de Koenderink y van Doorn [Koenderink et al. 76], quienes expresaron la teoría necesaria, así como el de Wildes [Wildes 91] quien implementó algunas de sus ideas [Wilson et al. 95]. Más recientemente, podemos ver cómo la disparidad en estereoscopia sigue suscitando gran interés (véase, por ejemplo, [Alvarez et al. 02], [Mühlmann et al. 02], [Mayer 03] y [Gutiérrez et al. 04]).

4.5. Restricción del Gradiente de Disparidad

La idea de la zona prohibida que se apuntaba en la restricción de ordenamiento, se puede ver como una forma particular de la idea del gradiente de disparidad. Si consideramos la figura 8 donde el eje x es paralelo a los ejes u_1 y u_2 con su origen en O , el punto medio de la línea $C_1 C_2$, podemos imaginar una retina virtual paralela a las dos reales con el centro óptico en O .

Esta retina se llama a menudo retina ciclópea [Julesz 71]. Si el punto M de la escena tridimensional tiene las imágenes m_1 y m_2 en las dos retinas reales con coordenadas u_1 y u_2 , entonces su imagen m_c en la retina virtual tendrá las coordenadas:

$$u_c = \frac{u_1 + u_2}{2}$$

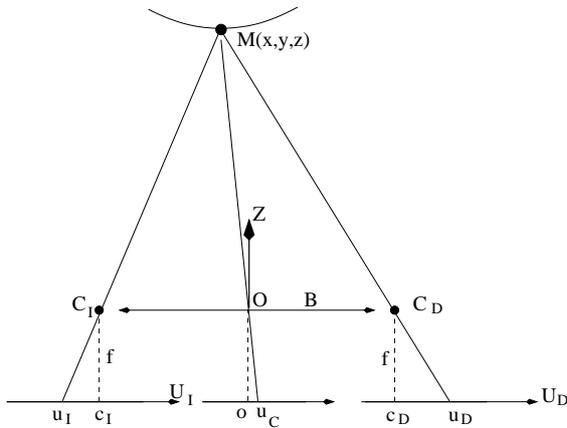


Figura 8. Definición del gradiente de disparidad.

Asumiendo que el punto M puede variar en un objeto suavizado definido por la ecuación $z=h(x)$, las ecuaciones vistas para la configuración de cámaras paralelas de las coordenadas (x,y,z) definen la disparidad d como una función suavizada de v_C .

$$d = u_D - u_I = \frac{B \cdot f}{z} \quad ; \quad x = \frac{B}{2d} (u_I + u_D)$$

Ahora consideramos dos puntos M y N en el mismo objeto con coordenadas ciclopeas u_{MC} y u_{NC} y disparidades d_M y d_N . La magnitud de la derivada de la disparidad con respecto a la coordenada ciclopea, el gradiente de disparidad puede aproximarse a lo siguiente:

$$DG = \left| \frac{d_M - d_N}{u_{MC} - u_{NC}} \right|$$

Utilizando las ecuaciones anteriores de la disparidad d y la coordenada x , obtenemos:

$$DG = \frac{B \cdot |z_M - z_N|}{|z_N x_M - z_M x_N|}$$

Experimentos psicofísicos [Pollard et al. 85] permiten conjeturar que la percepción humana impone la restricción que el gradiente de disparidad está limitado por arriba por una constante ($DG < K$). Esto significa que, si un punto M de un objeto se percibe, los puntos N del mismo objeto que están cerca de M de forma que $DG > K$ no se perciben correctamente. Supongamos que $M = [x_M, z_M]^T$ es fijo, y veamos qué tipo de condiciones obtenemos en $N = [x_N, z_N]^T$ imponiendo que $DG < K$.

La condición $DG < K$ puede escribirse como

$$B \cdot |z_M - z_N| < K \cdot |z_N x_M - z_M x_N|$$

Así obtenemos el cono definido por las líneas de ecuaciones

$$z_M - z_N = \pm \frac{K}{B} (z_N x_M - z_M x_N)$$

El vértice del cono es el punto M . Su complemento es el cono prohibido para el que el gradiente de disparidad es menor que K . El caso especial $K=2$ es interesante porque las dos líneas son $\langle C_I, M \rangle$ y $\langle C_D, M \rangle$ y la correspondiente zona prohibida es el cono prohibido de la restricción de ordenamiento. Si $K < 2$ obtenemos una zona prohibida menor.

Nótese que la restricción del gradiente de disparidad impone una condición a la tangente a la superficie del objeto: ésta debe quedar fuera del cono prohibido. Así, la restricción del gradiente de disparidad es una restricción geométrica del tipo de objetos posibles. Es decir, los tipos de objetos que pueden ser reconstruidos mediante el proceso estéreo. Nótese también, en general, que la restricción del gradiente de disparidad implica la restricción de unicidad vista anteriormente.

4.6. Otras Restricciones Geométricas

En las restricciones a la correspondencia podemos ir un poco más lejos que simplemente imponer un gradiente de disparidad a la escena. Además podemos imponer que las superficies de los objetos tengan localmente alguna forma simple. Esta imposición puede llegar a restringir que los objetos sean localmente planos, es decir que ellos se aproximen por sus planos tangentes en todas las direcciones excepto en las discontinuidades. Nótese que esto no es un caso especial del gradiente de disparidad ya que, dado un umbral al gradiente de disparidad, algunos planos quedarían excluidos.

4.6.1. Continuidad de superficie y continuidad figural

Estas restricciones van en la misma dirección que la del gradiente de disparidad. La restricción de continuidad de superficie se basa en que los puntos proyectados sobre la imagen pertenecen a las superficies de los objetos de la escena, superficies que se asumen continuas presentando únicamente discontinuidades en los contornos de los objetos. Esta continuidad de las superficies se traduce en una continuidad en el mapa de profundidades y, por lo tanto, en el mapa de disparidad, pues la disparidad está estrechamente relacionada con la profundidad.

La restricción figural fue propuesta por Mayhew y Frisby, y establece la continuidad de las superficies formulada como la continuidad de la disparidad a lo largo de los contornos de las figuras, y no a través de ellas. Esta formulación evita los problemas derivados de la discontinuidad en los límites de las superficies. Esta restricción impone que si dos puntos M y N pertenecen al contorno de un mismo objeto, las disparidades de sus puntos imagen han de ser iguales o al menos semejantes, de forma que la diferencia de las mismas sea inferior a cierto valor umbral.

4.6.2. Posición general

Ciertos eventos ocurren bastante infrecuentemente, en un sentido estadístico, de forma que permiten desestimar correspondencias asociadas a disposiciones geométricas improbables de los objetos. Se trata de una restricción específica de cada aplicación, pues la escena a captar y los objetos que la componen pueden poseer ciertas características tanto en su geometría como en su posición, de forma que permitan predecir posibles correspondencias correctas y desestimar o eliminar posibles correspondencias erróneas. Por ejemplo, si se está utilizando la estereovisión para realizar un mapa topográfico de una zona no urbana, como puede ser una cordillera, la escena que se tiene, en su mayoría, no estará compuesta por grandes discontinuidades, sino que será un terreno continuo, en ocasiones abrupto, pero continuo, permitiendo así descartar correspondencias que indicaran en el mapa de disparidad obtenido alguna gran discontinuidad.

Otra posibilidad dentro de las aplicaciones de la estereovisión, es la del control de calidad. Un sistema de inspección de piezas tridimensionales conocidas permitiría guiar el proceso de búsqueda de correspondencias.

4.6.3. Restricciones fotométricas

Las restricciones fotométricas son restricciones que están basadas en los modelos de interacción de los objetos con la iluminación. Entre ellas podemos destacar:

- (1) La restricción de reflectancia superficial, que asume un modelo de reflectancia Lambertiano para las superficies, de modo que la intensidad de la proyección de cualquier punto tridimensional no dependa del punto de vista. En realidad, esto no es del todo cierto, aunque es una buena aproximación.

- (2) La restricción de compatibilidad fotométrica, que impone que la distribución de intensidades entre puntos correspondientes debe ser similar.
- (3) Y la restricción de compatibilidad fotométrica diferencial, que utiliza la continuidad de superficie para asumir que, dados dos puntos cercanos de una imagen, la diferencia de intensidades entre ambos puntos ha de ser similar a la diferencia de intensidades de sus correspondientes.

Estas restricciones deben aplicarse sobre entornos de vecindad o regiones, pues los valores puntuales de intensidad en un píxel están sujetos al ruido.

4.6.4. Restricciones locales de primitiva

Estas restricciones son indicadores de la compatibilidad entre las características geométricas de pares de posibles correspondencias en función del tipo de primitivas utilizadas. Por ejemplo, en los puntos de borde correspondientes, la orientación del gradiente (signo) debe ser similar. En el caso de los segmentos de borde, la orientación de éste, su longitud y su gradiente también deben ser compatibles. Y, finalmente, en las regiones, sus dimensiones y su forma también deben estar relacionadas.

5. Técnicas de Correspondencia Utilizadas

El análisis de la visión en estereo tiene una historia relativamente corta. Los primeros artículos específicos que se encuentran en la literatura datan de los años 70. Desde entonces han surgido multitud de ideas para resolver el problema de la correspondencia que es el más esquivo, y quizá por ello el más importante. Actualmente se siguen buscando soluciones a este problema, ya que no se ha encontrado una solución que funcione bien con imágenes sintéticas y con imágenes reales, en ausencia y en presencia de ruido, etc.

Debido a esa multitud de intentos de resolución que han aparecido hasta la fecha y a que todavía se siguen buscando nuevos métodos, el intentar abarcar absolutamente todos los intentos sería un trabajo inacabable. Por ello, en este apartado, se van a repasar algunos de los intentos de solución más representativos. Se van a intentar clasificar los distintos algoritmos según su afinidad con el fin de

centrar más la atención en los conceptos que se presentan en cada modalidad.

5.1. Técnicas Basadas en la Correlación

Las técnicas de área basadas en intensidad han sido investigadas extensamente para aplicaciones comerciales en estereofotogrametría ([Kelly et al. 77] y [Forstner et al. 86]), pero éstos son tan sólo algunos métodos antiguos utilizados en visión por ordenador. La principal de este tipo de técnicas es la técnica de correlación de área. Ésta se basa en considerar los valores de intensidad de los píxeles de las imágenes como una señal bidimensional, que en una de las dos imágenes ha sufrido una traslación, lo que nos lleva al concepto de disparidad. Se trata de obtener, para cada punto de la imagen, dicha traslación, minimizando una función de coste, que comúnmente tiene que ver con la correlación.

Para cada píxel de una imagen se calcula la correlación entre la distribución de intensidades de una ventana centrada en dicho píxel y una ventana del mismo tamaño centrada en el píxel a corresponder de la otra imagen. Esta técnica aplica, además de la restricción epipolar, las restricciones Lambertiana, de continuidad y otra restricción conocida como frontoparalela, que asume que la disparidad es constante localmente, por lo que las superficies deben ser paralelas a los planos de imagen de las cámaras, o al menos tener una pendiente pequeña. Las ventajas de utilizar este método de correlación de área, es que se obtienen unos buenos resultados en imágenes con texturas importantes y son fáciles de paralelizar. Además, permite crear mapas densos de disparidad, es decir, se obtendrá una disparidad para todos los puntos de la escena, y no solo para los contornos, esquinas u otras primitivas de mayor nivel.

También es cierto que presenta problemas con imágenes que contienen elevadas discontinuidades de superficie y es una técnica muy sensible a variaciones fométricas debidas a sombras o reflejos. Tiene además problemas con las oclusiones y requiere de un proceso posterior de eliminación de falsas correspondencias. También es posible utilizar esta técnica como complemento de otras e incluso realizar algún tipo de postprocesado sobre el mapa de disparidad hallado, que permita reducir los inconvenientes de la correlación de área como técnica de correspondencia.

5.1.1. El principio general de las técnicas de correlación

Para encontrar las coordenadas de un píxel en la imagen D que corresponde con el píxel de coordenadas (u_I, v_I) en la imagen I , consideramos una ventana rectangular de tamaño $(2V+1) \cdot (2H+1)$ centrada en (u_I, v_I) y calculamos su similitud con una parte de la segunda imagen, la zona de búsqueda. Esta zona de búsqueda, atendiendo a la restricción epipolar, estará dentro de la línea $v_D = v_I$, y, atendiendo a la restricción de disparidad, podremos centrarnos en un segmento de dicha línea. El método estadístico más utilizado para determinar dicha similitud es la correlación cruzada normalizada $C_{ID}(\tau)$.

$$C_{ID}(\tau) = \frac{1}{k} \sum_{i=-H}^H \sum_{j=-V}^V \left(I_I(u_I + i, v_I + j) - \overline{I_I(u_I, v_I)} \right) \cdot \left(I_D(u_I + i + \tau, v_I + j) - \overline{I_D(u_I + \tau, v_I)} \right)$$

donde

$$k = (2H + 1)(2V + 2) \cdot \sigma_I(u_I, v_I) \cdot \sigma_D(u_I + \tau, v_I)$$

En estas fórmulas, $\overline{I_I(u_I, v_I)}$ y $\sigma_I(u_I, v_I)$ son la intensidad media y la desviación estándar en la imagen I del entorno centrado en (u_I, v_I) .

$$\overline{I_I(u_I, v_I)} = \frac{1}{(2H + 1)(2V + 1)} \sum_{i=-H}^H \sum_{j=-V}^V I_I(u_I + i, v_I + j)$$

$$\sigma_I^2(u_I, v_I) = \frac{1}{(2H + 1)(2V + 1)} \cdot \sum_{i=-H}^H \sum_{j=-V}^V \left(I_I(u_I + i, v_I + j) - \overline{I_I(u_I, v_I)} \right)^2$$

Las formulas son similares para $\overline{I_D}$ y σ_D^2 . Debido a la normalización por la media y la varianza, esta función es relativamente insensible a cambios en la iluminación. La curva $C_{ID}(\tau)$ habitualmente tiene un máximo que se alcanza para el valor τ_0 de τ . En este caso, la disparidad del píxel (u_I, v_I) debe ser τ_0 . Esto puede verse en la figura 9.

Con esta técnica es necesario hacer algunas consideraciones: (1) Las líneas epipolares deben ser filas de la imagen, ya que el estudio se hace por filas. Esto implica que las imágenes deben estar rectificadas. (2) Si el máximo de la función de correlación $C_{ID}(\tau)$ no está muy definido, o existen varios máximos, la disparidad τ_0 puede no ser muy precisa o completamente errónea. (3) La disparidad se asume como constante en la ventana de análisis.

Es posible tener problemas cuando el gradiente de disparidad sea diferente de 0. Este tercer problema es el más importante, ya que debemos asumir que las superficies de la escena deben tener una disparidad constante. Esto equivale a decir que estas superficies deben representarse localmente como planos frontales.

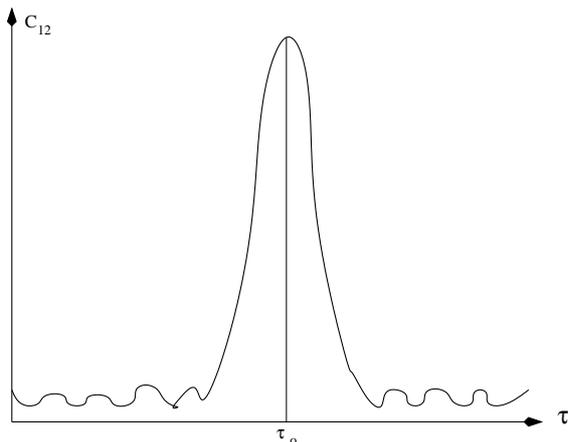


Figura 9. Forma general de la función de correlación.

Por otra parte, podemos considerar que la fórmula de $C_{ID}(\tau)$ no es simétrica respecto de I_I e I_D . Por ello podemos también calcular la función de correlación $C_{DI}(\tau)$ intercambiando los roles de las imágenes I y D . En este caso, los píxeles en los que $C_{ID}(\tau)$ y $C_{DI}(\tau)$ son distintos, probablemente sean píxeles de contorno o cerca de discontinuidades, en los que puede haber oclusiones y el gradiente de disparidad ser grande. En el algoritmo desarrollado por Fua [Fua 91], los puntos donde C_{ID} y C_{DI} difieren, no se asocia correspondencia entre ellos, sino que se les considera oclusiones, y el resultado final se deja a una función de suavizado.

5.1.2. Algoritmo de Nishihara

Nishihara [Nishihara et al. 84] propuso una técnica de correlación basada en el signo de la convolución de las imágenes con el operador Laplaciano de un Gaussiano. El uso del Laplaciano del Gaussiano fue motivado por el trabajo de Marr y Hildreth [Marr et al. 80] en detección de contornos, donde los contornos se detectan en los cruces por cero de la salida de dicha convolución. Otra extensión notable del método de correlación básico es el propuesto por Kass [Kass 83], [Kass 88], quién procesó las dos imágenes con filtros lineales independientes y después correlacionó los vectores de las salidas.

Las posiciones de los cruces por cero de las imágenes en el algoritmo de Nishihara son muy sensibles al ruido. Estos cruces por cero se desplazan en cantidades que son proporcionales a la amplitud del ruido e inversamente proporcionales al gradiente de la convolución en el cruce por cero [Nishihara et al. 84], [Nishihara et al. 82]. Por otro lado, si el espacio entre cruces por cero es relativamente grande comparado con la cantidad de este desplazamiento, la región de signo constante entre ceros será estable en un rango grande de relaciones señal-ruido. La idea es realizar la convolución de las imágenes derecha e izquierda con $\nabla^2 G_{\sigma}$, donde G_{σ} es un ruido Gaussiano de varianza σ^2 , y después correlacionar las imágenes de salida. Éste tiene el efecto de hacer más abrupto el pico de la función de autocorrelación. En [Faugeras 93] puede extenderse el estudio de la sensibilidad al ruido y la forma de las funciones de correlación, utilizando algunos modelos de imágenes de estadística simple.

5.1.3. Función de similitud basada en la suma de diferencias

La función de correlación es una función muy costosa computacionalmente. Por ello, los algoritmos que la utilizan suelen tardar mucho tiempo en ejecutarse. Debido a ello, se han planteado dos alternativas a esta función, alternativas que en principio son menos costosas computacionalmente y que a la postre han resultado ser tanto o más fiables que la propia función de correlación. Estas funciones son la suma de diferencias absolutas (SAD) y la suma de diferencias al cuadrado (SSD) y se aplican de la misma manera a entornos bidimensionales de los píxeles a corresponder.

Mediante la suma de diferencias absolutas se comparan los entornos bidimensionales de los píxeles a analizar, calculando sus diferencias píxel a píxel y sumándolas todas entre sí. Aquellos entornos que tengan un mínimo en esta suma serán los que tengan una correspondencia de más confianza. Por su parte, la suma de diferencias al cuadrado se diferencia del anterior, en que previamente a la suma, las diferencias de los píxeles de los entornos bidimensionales se elevan al cuadrado. En [Aschwanden et al. 93] puede verse una comparación extensa de estos métodos.

5.1.4. Funciones basadas en Rango y Censo

Zabih y Woodfill [Zabih et al. 94] propusieron un método alternativo para analizar la correspondencia, aplicando una transformación previa a las imágenes con el fin de eliminar dependencias de la iluminación. Ambas funciones se aplican de forma local a las regiones a corresponder de las dos imágenes. La transformada del rango de una región centrada en un píxel se define como el número de píxeles de esta región cuya intensidad es menor que la del píxel central. El resultado está más basado en la cantidad de píxeles que en los píxeles en sí. Después de la transformación del rango, se procede a una comparación de las regiones mediante técnicas de correlación, SAD o SSD.

La transformada del rango es interesante porque reduce la sensibilidad a la iluminación, pero reduce el poder de discriminación en las correspondencias, ya que se pierde información sobre la posición del orden relativo de los píxeles. Los propios Zabih y Woodfill propusieron una variación de la transformación del rango llamada transformación del censo. Ésta conserva la distribución espacial de los píxeles, codificándolos mediante una cadena de bits. En este caso, el análisis de correspondencia se realiza mediante la distancia de Hamming, número de bits diferentes entre cadenas de bits. Esta transformación incrementa la cantidad de información necesaria por un factor que depende del tamaño local de la región, haciéndola algo más costosa computacionalmente.

Banks y Croke [Banks et al. 01] compararon las prestaciones de las correspondencias mediante censos y rangos con las analizadas mediante correlaciones y medidas de la diferencia. Sus resultados indican que los métodos del rango y del censo consiguen métricas comparables con las anteriores y son más robustos frente a la iluminación y a las oclusiones. Para la mayoría de las escenas analizadas, las diferencias entre la correlación cruzada y la correspondencia mediante el censo están entre el 5 y el 9 por ciento de errores del número total de píxeles.

5.2. Técnicas de Relajación

La técnica de correlación de área por sí sola presenta numerosos errores en las correspondencias, que bien pueden ser eliminados mediante un postprocesado o parcialmente evitados mediante un proceso que se conoce con el nombre de relajación o

algoritmo cooperativo. La idea básica de las técnicas de relajación es permitir a los píxeles que se van a poner en correspondencia, realizar “estimaciones controladas” de cómo debe ser su correspondencia y, después, permite a las correspondencias reorganizarse propagando algunas de las restricciones descritas en los apartados anteriores.

Para este tipo de proceso, no solamente importa el valor de la correlación obtenida para los píxeles de la línea que se analiza, sino que también otorga importancia a los valores de correlación obtenidos para una cierta vecindad que se conocerá con el nombre de región excitatoria, y un grupo de píxeles que se conocerán como región inhibitoria.

Este algoritmo se implementa a partir del denominado cubo de correlación, que será una matriz de tres dimensiones (filas x columnas x disparidad). Cada uno de los elementos de la matriz almacenará el valor de la correlación obtenido para la ventana de la imagen 1 centrada en el píxel marcado por las coordenadas (fila, columna) con la ventana de la imagen 2 centrada en el píxel de coordenadas (fila, columna + disparidad). Este cubo de correlación se muestra en la figura 10.

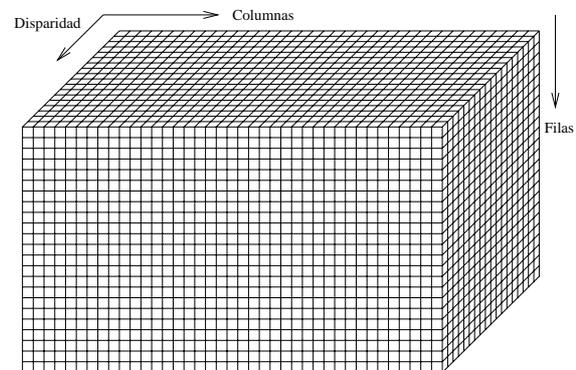


Figura 10. Cubo de correlaciones.

Una vez creado el cubo de correlación para cada uno de los píxeles de la imagen marcado por su fila y su columna, se tiene un vector unidimensional del tamaño el límite de disparidad elegido en el análisis, y que almacena los valores de la correlación.

El proceso de relajación se realizará ahora línea por línea, o fila por fila, de modo que, para cada una de las filas, se tiene una matriz bidimensional de la magnitud horizontal y la disparidad. Es en esta matriz donde se definirán las que antes se han llamado región inhibitoria y excitatoria.

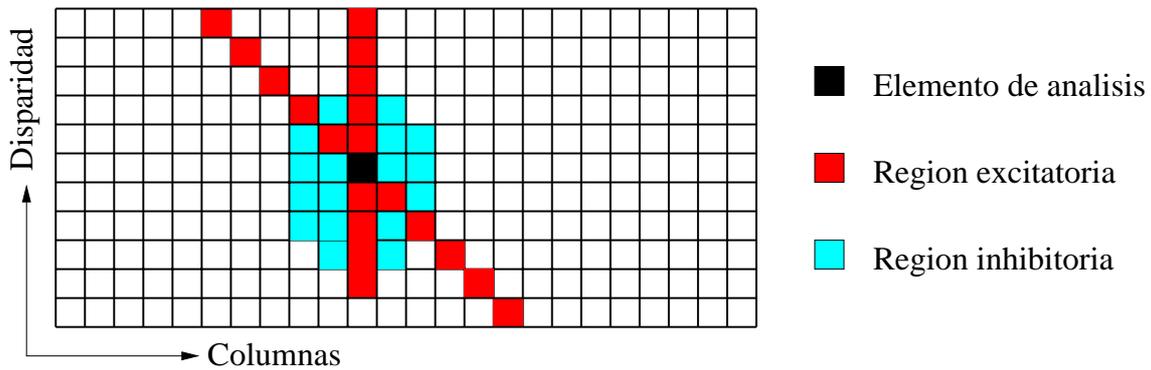


Figura 11. Regiones excitatoria e inhibitoria.

Las regiones excitatoria e inhibitoria se definen utilizando las restricciones de continuidad, unicidad y, de forma indirecta, también de ordenamiento, además de la restricción epipolar que se ha utilizado para crear el cubo de correlación. Para cada píxel de coordenadas (u_i, v_i) , se tratará de buscar un valor de disparidad d para el que se encuentre el píxel correspondiente en la imagen D.

Si ese punto es realmente una correspondencia tendrá un alto valor de correlación y, a su vez, atendiendo a la restricción de continuidad, los puntos cercanos a él también tendrán valores altos, de modo que esos puntos cercanos serán la región excitatoria. Por el contrario, si dicho punto es correcto, los demás elementos de la matriz cuya coordenada X sea la misma pero difieran en la disparidad, serán correspondencias falsas, de modo que tendrán valores pequeños de disparidad. Si a su vez se aplica la restricción de ordenamiento y unicidad, los puntos que tengan una coordenada X diferente pero tengan una disparidad tal que lleven al mismo píxel, no serán correspondencias válidas, por lo que sus valores de correlación también serán pequeños. Estos dos últimos grupos de puntos se corresponderán con la región inhibitoria.

En la figura 11 se muestra gráficamente todo esto. Conociendo las características de las regiones comentadas, será posible utilizarlas para mejorar el proceso de la búsqueda de las correspondencias.

5.2.1. El algoritmo de Marr-Poggio.

El algoritmo de Marr y Poggio [Marr et al. 76] [Marr et al. 79] fuerza las restricciones de unicidad y de disparidad. Para cada par de píxeles m_I, m_D de las imágenes I y D , se calcula un conjunto inicial de medidas de confianza $C(m_I, m_D)$ que estiman cuál de

los m_D puede ser puesto en correspondencia con dicho m_I de la imagen I . Existen muchas formas de calcular los valores de la confianza C . El más simple es comparar los valores de intensidad de m_I y m_D :

- $C(m_I, m_D) = 1$ si la intensidad de m_I y m_D están lo suficientemente cerca
- $C(m_I, m_D) = 0$ de otra forma.

Los puntos m_D son elegidos de la línea epipolar correspondiente de m_I , y las medidas de confianza se van actualizando secuencialmente para comprobar la restricción de continuidad así:

- $C(m_I, m_D) (n+1) = 1$, si el número de píxeles m'_I en el vecindario V_I de m_I son tales que
 - $C(m'_I, m'_D) (n) = 1$ para m'_D en el vecindario V_D de m_D está por encima de un umbral.
- $C(m_I, m_D) (n+1) = 0$ en otro caso

Los vecindarios V_I y V_D son usualmente aproximaciones a discos con un radio de unos pocos píxeles centrados en m_I y m_D .

El algoritmo ha sido probado sobre todo en estereogramas de puntos donde se han obtenido resultados muy satisfactorios. En unas pocas iteraciones éste converge; es decir, la confianza medida permanece estable. Más aún, la confianza medida es habitualmente no ambigua en el sentido de que, para la mayoría de los píxeles m_I de la primera imagen sólo hay un píxel m_D de la línea epipolar correspondiente de forma que $C(m_I, m_D) = 1$. El punto m_I es entonces puesto en correspondencia con m_D . Si existen varios m_D candidatos, se debe aplicar la restricción de unicidad.

El algoritmo no funciona tan bien con imágenes reales, básicamente porque las características que utiliza, la intensidad en el punto, y las restricciones que impone (geometría plana y disparidad) no son suficientes para utilizarlas en la mayoría de las imágenes reales. Su implementación y muchas aportaciones interesantes son debidas a Eric Grimson [Grimson 81], [Grimson 85].

5.2.2. El algoritmo de Pollard, Mayhew, Frisby

La mayor diferencia del algoritmo PMF [Pollard et al. 85] con el de Marr y Poggio es que éste fuerza la restricción del gradiente de disparidad. El algoritmo PMF primero extrae de las dos imágenes varios elementos junto a sus características. Por ejemplo, los puntos de contorno son caracterizados por su contraste y su orientación. Una correspondencia entre el detalle t_I de la imagen I y el detalle t_D de la imagen D se caracteriza por una medida C_{ID} de su confianza tras comparar los valores de las características de t_I y de t_D .

Para cada elemento localizado en el píxel m_I y cada posible correspondiente en la imagen D (aplicando la restricción epipolar) se calcula una “fuerza de la correspondencia” $SM(m_I, m_D)$. Se consideran todos los elementos t_i localizados tanto en los píxeles m_I como en algún vecino de m_I . Para cada detalle, las correspondencias t_j localizadas en los píxeles m_D que satisfacen la restricción $DG(m_I, m_D, n_I, n_D) < \sigma$ hacen que se consideren como candidatos a ser correspondientes al par (m_I, m_D) . En el caso donde hay más de un detalle con esta restricción $DG < \sigma$, sólo se considera aquel con mayor valor de C_{ij} .

Los detalles son correspondientes mediante un proceso en que el que gana se queda con todo para asegurar la unicidad. El método funciona como se describe a continuación. Para cada iteración se consideran correctos aquellos cuya fuerza de correspondencia es máxima para los dos detalles de las imágenes. Después, debido a la restricción de unicidad, todas las correspondencias asociadas a los dos puntos se eliminan para la siguiente iteración. Esto permite que las correspondencias se elijan tan correctas como sea posible a partir de la fuerza de la correspondencia para los dos detalles (uno de cada imagen). Normalmente, con cuatro o cinco iteraciones es suficiente para conseguir todas las correspondencias. El algoritmo se ha probado en gran número de imágenes sintéticas y reales, dando resultados precisos y fiables.

5.2.3. Técnicas de grueso a fino

Dentro de las técnicas basadas en área, existe otra bastante común que se conoce como técnica grueso a fino y que puede emplearse con diferentes técnicas de correspondencia, siendo lo más habitual su uso junto con la técnica de correlación de Marr-Poggio. Esta técnica se basa en crear una estructura piramidal a partir de las imágenes estéreo reduciendo su tamaño y por lo tanto, su resolución. El proceso se realizará para K niveles de reducción, comenzando la búsqueda de correspondencias en el nivel más bajo, es decir, el de menor tamaño o resolución. Los resultados obtenidos en este nivel serán utilizados para guiar y restringir la búsqueda de las correspondencias en el nivel superior, y así sucesivamente hasta alcanzar el nivel $K-1$, que serán las imágenes originales.

Esta técnica tiene el inconveniente de que un pequeño error en un nivel bajo, se convertirá en un gran error en el mapa de disparidad final. Su ventaja es que al realizar gran parte de la búsqueda de las correspondencias en imágenes de bajas resoluciones, el tiempo de ejecución se reduce, pues al llegar a las imágenes de los niveles altos, la búsqueda queda muy restringida a unos pocos píxeles.

5.3. Métodos Basados en el Gradiente

Los métodos basados en el gradiente o en el flujo óptico persiguen determinar disparidades locales entre dos imágenes en base a formular una ecuación diferencial que relaciona el movimiento y el brillo en las imágenes. Para hacer esto, se asume que el brillo en la imagen de un punto de la escena es constante entre las dos vistas. Así la traslación horizontal que se produce debida a la diferente perspectiva se analiza a partir de la ecuación

$$(\nabla_x E)v + E_t = 0,$$

donde $\nabla_x E$ es la componente horizontal del gradiente en la imagen, E_t es la diferencia de intensidades entre las imágenes derecha e izquierda, y v es el desplazamiento entre las dos imágenes.

Puede notarse que con esta expresión solamente es posible estimar el desplazamiento en la dirección del gradiente. Para solucionar esto, es conveniente estudiar el gradiente en todos los píxeles del entorno del punto a estudiar, para de esta manera tener una información más fiable para realizar las correspondencias [Choi et al. 03] [Wang et al. 03].

5.4. Correspondencias entre Características

Las técnicas relacionadas con el área como la correlación y las basadas en el gradiente son muy sensibles a errores provocados por las discontinuidades en la profundidad, así como a las regiones uniformes. Para solventar estos problemas, las técnicas basadas en características limitan las regiones a analizar a características fiables de las imágenes, como pueden ser los contornos, las curvas, etc. Esta técnica reduce la densidad de puntos sobre los que se estima la profundidad. Por eso, debido a la necesidad de conseguir mapas de disparidad densos para distintas aplicaciones y también al mejorar la eficiencia y robustez en los métodos basados en áreas, en la última década el estudio de los métodos basados en características ha decaído. Los dos métodos que han recibido más atención en los últimos tiempos han sido los métodos jerárquicos y la segmentación basada en regiones.

Venkateswar y Chellappa [Venkateswar et al. 95] propusieron un algoritmo de correspondencias basadas en características explotando cuatro tipos de ellas: líneas, vértices, contornos y superficies. La correspondencia comienza en el nivel más alto de la jerarquía, las superficies, y continúa hasta el más bajo, o sea las líneas. Una vez que los componentes de la jerarquía más alta han sido puestos en correspondencia, las características que lo componen no son vueltas a analizar en los niveles más bajos.

Otra aproximación basada en características es primero segmentar en regiones las imágenes y después poner en correspondencia dichas regiones [Wei et al. 04]. Como todos los métodos basados en características, los mapas densos se definen en base a desplazamientos de las regiones. Por lo tanto, estos métodos son sensibles sobre todo a la calidad de la segmentación.

5.5. Técnicas de Programación Dinámica

El problema de la correspondencia de primitivas entre imágenes puede ser abordado como minimización de una función de coste. La programación dinámica es una forma eficiente de minimizar (o maximizar) funciones de gran número de variables discretas. Intentos satisfactorios utilizando programación dinámica para resolver el problema de la correspondencia estéreo son los de Baker y Binford [Baker et al. 81] y Ohta y Kanade [Ohta et al. 85]. En ambos casos se utilizan los contornos como primitivas básicas.

Asumamos que las líneas epipolares son paralelas a las filas de las imágenes y consideremos dos líneas correspondientes en las imágenes derecha e izquierda. En cada fila se identifican varios píxeles de contorno, y se incluyen los dos finales de las líneas por conveniencia. La correspondencia de estos píxeles de contorno puede considerarse como el problema de corresponder los intervalos entre ellos de la siguiente manera: Ordenamos los píxeles de contorno de izquierda a derecha en cada línea y los numeramos entre 0 y $N-1$ en la imagen izquierda y 0 a $M-1$ en la derecha.

En la figura 12 se representan los pares (i,j) de puntos de contorno de las líneas derecha e izquierda como puntos que forman una rejilla. Corresponder el intervalo $[i,i']$ de la izquierda con el $[j,j']$ de la derecha es equivalente a dibujar un segmento entre los puntos $m=(i,j)$ y $m'=(i',j')$ en la rejilla. El objetivo es encontrar una secuencia de segmentos (un camino) desde el punto $m_0=(0,0)$ hasta el punto $m_e=(N-1, M-1)$. En esta búsqueda podemos aplicar restricciones del estilo de las vistas en apartados anteriores. La restricción de orden es interesante ya que es equivalente a decir que los caminos admisibles son caminos monótonos decrecientes.

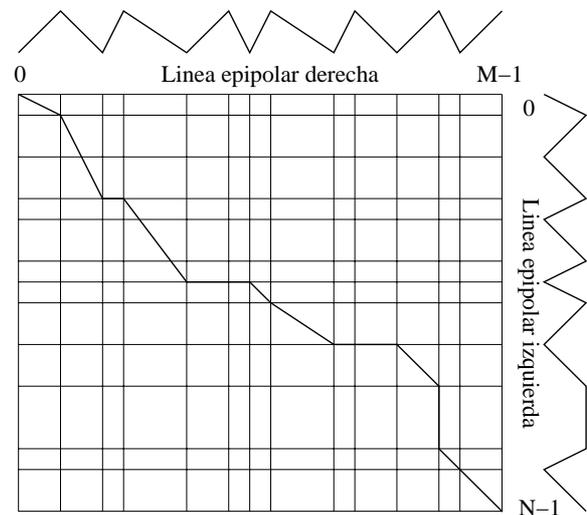


Figura 12. Técnicas de programación dinámica.

A pesar de esta restricción todavía existen muchos caminos posibles. Definiremos el mejor camino como el que minimice una función de coste. Primero definiremos el coste $c(m,m')$ de un segmento entre los puntos $m=(i,j)$ y $m'=(i',j')$. Por supuesto, existen varias formas de definir esta función de coste. En general esta función de coste debe medir dos cosas: (1) la similitud de las características de los píxeles de

la imagen derecha y la izquierda, características que pueden ser, por ejemplo, la orientación de los contornos o contraste a través de ellos, y (2) la similitud de intensidades a lo largo de los intervalos entre contornos.

Si existe un contorno entre dos filas, las correspondencias de una fila deben depender bastante de las vecinas. Reforzar la consistencia es equivalente a aplicar la continuidad de las figuras, y hay varias formas de hacerlo: (1) Baker [Baker et al. 81] utiliza un proceso cooperativo para detectar y corregir los resultados de la correspondencia, mientras que (2) Ohta y Kanade [Ohta et al. 85] lo incluyen en la función de coste y resuelven una programación dinámica en un espacio 3D en vez de un espacio previo 2D.

Además de la minimización de la función de coste para las líneas epipolares, también se pueden aplicar relaciones de correspondencia entre líneas epipolares vecinas (superiores e inferiores) con el fin de reducir la ambigüedad. Algunos ejemplos que avanzan en esta dirección son los llamados "graph cuts", que tratan de minimizar una función de coste que implica tanto a la dirección horizontal como a la vertical. La principal desventaja de la programación dinámica es la probabilidad de que errores locales se puedan propagar a lo largo de la línea epipolar descartándose correspondencias potencialmente correctas.

5.6. Técnica de las Curvas Íntimas

Tomasi y Manduchi [Tomasi et al. 98] propusieron una alternativa a la búsqueda convencional de correspondencias utilizando una representación diferente de las líneas epipolares de las imágenes. Se trata de la técnica de las curvas íntimas. Una curva íntima es una representación de descriptores definidos por la aplicación de operadores de borde o de esquinas. Las curvas íntimas se definen representando gráficamente las intensidades de cada línea epipolar frente a sus respectivas derivadas. Esta representación es invariante con la disparidad, así que, en el caso ideal, los píxeles correspondientes se sitúan superpuestos en las dos representaciones de las líneas epipolares.

En el caso general, debido al ruido y a las diferencias de perspectivas, los puntos correspondientes no siempre se superponen exactamente en el mismo sitio. De esta manera se recurre a identificar a los puntos correspondientes como los puntos más cercanos de la gráfica. Una vez que se ha encontrado

la correspondencia mediante las curvas íntimas, debe deshacerse la representación para asociar a cada correspondencia la disparidad apropiada. La principal ventaja de las curvas íntimas es su invarianza respecto de la disparidad. Algunos métodos utilizan precisamente la restricción de disparidad para reducir su coste computacional, pero en esta aproximación, eso no es necesario. Por supuesto, esta representación también está afectada por las oclusiones y por texturas homogéneas en las imágenes. Para compensar, sobre todo el segundo problema, se añade un mecanismo de programación dinámica que se aplica en aquellos puntos de la curva íntima que por su nivel de brillo constante y su derivada también constante no quedan del todo situados.

Las oclusiones en las curvas íntimas quedan patentes como discontinuidades en la representación. Y, aunque es visiblemente un indicador interesante, todavía no se ha publicado ningún algoritmo para su tratamiento. Tampoco se han publicado análisis comparativos con otros métodos.

5.7. Métodos de Predicción y Verificación

Este es un ejemplo de algoritmo estéreo donde los elementos puestos en correspondencia son de un nivel simbólico mayor que los píxeles. Esta aproximación ha sido seguida particularmente por Medioni y Nevatia [Medioni et al. 85]. Describiremos un algoritmo estéreo desarrollado por Ayache y Faverjon [Ayache et al. 87] que realiza la correspondencia de segmentos entre dos imágenes por una técnica de predicción y verificación de hipótesis.

El tomar segmentos de línea como primitivas tiene dos ventajas: (1) El número de primitivas a poner en correspondencia es más pequeño que el número usual de píxeles de contorno, y (2) en la correspondencia de primitivas simbólicas podemos utilizar características de las mismas, como por ejemplo características geométricas que son robustas y fiables.

El utilizar segmentos también tiene desventajas: (1) Se asume que los contornos se pueden aproximar bien por segmentos de línea, y que además los objetos de la escena son poliedros o parecidos. Si no lo son, el número de segmentos puede aumentar para una tolerancia dada, reduciendo la primera ventaja vista anteriormente. (2) Si los objetos no son poliedros, no hay garantía de que los vértices de los

poliedros aproximados en las imágenes del par estéreo se correspondan. Es decir, que estén en las mismas líneas epipolares. (3) La densidad de correspondencias es menor que las obtenidas en el caso de puntos. A pesar de esto, las correspondencias son habitualmente más precisas y fiables que las tomadas por técnicas basadas en correlación y esta ventaja compensa la desventaja.

La primera desventaja puede minorarse considerando aproximaciones por curvas en vez de segmentos de línea. La segunda desventaja también puede aliviarse mediante técnicas de interpolación.

Este método utiliza tres restricciones para reducir el tamaño del espacio de búsqueda: la restricción epipolar modificada para el caso de segmentos de línea que se verá a continuación, la restricción de continuidad y la restricción de unicidad.

5.7.1. Construcción de descripciones simbólicas

Las dos imágenes del par estéreo se representan como conjuntos de segmentos. Estos segmentos se extraen de una primera búsqueda de puntos de contorno y después se aproximan dichos puntos por líneas rectas. Cada segmento se representa por varias características geométricas (coordenadas del punto medio, longitud, orientación) y basadas en intensidad (contraste, gradiente de intensidad media de la imagen a lo largo del segmento). También se introduce una estructura de vecindario en los conjuntos de segmentos [Asano et al. 85]. Esto significa que cada imagen se divide en ventanas cuadradas no solapadas. A cada ventana se le asigna una lista de segmentos que intersectan. Ello da un acceso rápido a los segmentos que están cerca de un segmento dado y se utilizan para implementar el proceso de propagar las hipótesis que se describirán a continuación.

5.7.2. Definición de correspondencias

Estrictamente hablando, una correspondencia es un par (D, I) de los segmentos de la derecha y la izquierda que satisfacen la restricción epipolar. En la práctica, la probabilidad de que dos aproximaciones poligonales produzcan segmentos cuyos finales estén en el mismo plano epipolar es muy pequeña. Es mejor observar situaciones en las que las líneas epipolares de los finales de segmentos intersectan con otros segmentos s_1 y s_2 en algún punto distinto de los finales.

Una correspondencia aceptable puede definirse de la manera más rigurosa considerando los puntos finales de I y sus correspondientes líneas epipolares que intersectan en D a lo largo del segmento D' (más largo). Debe existir la intersección D'' de D' y D , si la correspondencia se acepta. Utilizando un método similar, los puntos finales de D'' corresponderán con I' que estará incluida en I . Los segmentos I' y D'' existirán y tendrán sus características en correspondencia.

Como esta forma de proceder es muy costosa, definimos una restricción epipolar modificada como se detalla a continuación. La restricción epipolar para segmentos de línea implica que los segmentos homólogos tienen al menos un punto análogo. Para hacerlo más sencillo, este punto se escoge como el punto medio del segmento. Imponemos que la línea epipolar del punto medio del segmento I intersecte al segmento D para que la correspondencia (I, D) sea válida. Esta definición no es simétrica con respecto a I y D , y tiene la desventaja de permitir potencialmente una correspondencia global entre contornos aproximados por distintos segmentos. Finalmente, para evitar falsas correspondencias espurias posibles, las características de los segmentos se comparan.

Como resumen, una correspondencia es un par (D, I) de segmentos de la derecha y la izquierda que satisfacen la restricción epipolar modificada. Además, las características de los segmentos derecho e izquierdo deben estar suficientemente semejantes.

5.8. Métodos Basados en Diferencias de Fase

Otro de los enfoques quizás menos extendido sea el fundamentado en el cálculo de la disparidad binocular en base a la diferencia de fase a la salida de filtros paso banda locales aplicados a las imágenes estéreo. Este enfoque es analizado con mayor profundidad en [Jepson et al. 89], [Sanger 88]. Además existen evidencias biológicas incluidas en [Wilson et al. 95] que soportan diferentes aspectos de este método. Estos métodos han surgido debido a que se piensa que la forma en que el cortex visual primario procesa las imágenes visuales puede ser modelada adecuadamente mediante filtros paso banda espacio-temporales [Coren et al. 93].

5.9. Oclusiones

La mayor parte de la investigación en estereovisión en la última década se ha orientado a la detección y medida de regiones ocultas y en recuperar la profundidad precisa para estas regiones. Este apartado define el problema de la oclusión en estereovisión y contempla tres clases de algoritmos para el manejo de oclusiones: los métodos que detectan oclusiones, los métodos que reducen la sensibilidad a las oclusiones y los métodos que modelan la geometría de las oclusiones.

El problema de las oclusiones en estereovisión se refiere al hecho de que algunos puntos de la escena son visibles por una cámara y, sin embargo, no por la otra, debido a la propia escena y a la geometría del sistema. En estos casos, la estimación de la profundidad no es posible si no se añaden más vistas en las que el punto no esté oculto, o se asuman ciertas características de la escena.

5.9.1. Métodos de detección de oclusiones

Los acercamientos más simples al manejo de oclusiones comienzan por su detección previa o posteriormente a las correspondencias. Estas regiones en algunos casos resultan interpoladas cuando se pretende conseguir un mapa denso de disparidades, o simplemente no se toman en consideración cuando se busca un mapa menos denso. La aproximación más común es detectar discontinuidades en el mapa de profundidad después del análisis de correspondencias.

Habitualmente se utilizan filtros de mediana para eliminar dichas discontinuidades que son producidas generalmente por oclusiones. Por ejemplo, Hoff y Ahuja [Hoff et al. 89] detectan la profundidad y la orientación de las discontinuidades comparando los resultados con superficies planas. Cuando dos de estas superficies difieren en profundidad o en orientación más de un determinado umbral, se considera que existe una oclusión.

Chang et al. [Chang et al. 91] calculan dos mapas de disparidad, uno basado en las correspondencias de la imagen izquierda con las de la derecha, y otro basado en las correspondencias de la derecha con las de la izquierda. Las disparidades inconsistentes se consideran producidas por oclusiones en la escena. Existen otras muchas causas posibles de inconsistencias, incluyendo diferencias de perspectiva, iluminación no uniforme o ruido en los sensores. La inconsistencia de izquierda a derecha

trata todos estos fenómenos por igual, pero es un método que al utilizar las funciones SAD o SSD vistas anteriormente tiene un coste computacional razonable. Por ello, estas funciones son comúnmente utilizadas en sistemas en tiempo real.

La restricción de ordenamiento también se puede utilizar para detectar oclusiones. El ordenamiento relativo de los puntos a lo largo de las líneas epipolares es monótono, asumiendo que no existen objetos excesivamente estrechos en la escena. Silva y Santos-Victor [Silva et al. 00] han propuesto un método que busca correspondencias fuera de orden que pueden indicar la presencia de oclusiones.

Otra aproximación a la detección de oclusiones se basa en la observación de discontinuidades en la profundidad y orientación que aparecen en torno a los bordes de los objetos. Los mapas de disparidad se suavizan, manteniendo exclusivamente sin suavizar las disparidades asociadas a los bordes. Entonces, aquellos puntos con grandes diferencias de disparidad entre la versión original y la suavizada se consideran como regiones ocultas.

La programación dinámica de Ohta y Kanade [Ohta et al. 85] que hace corresponder a las regiones a través de la interpolación de las profundidades de los contornos no sólo detecta sino que también evita el problema de las oclusiones.

5.9.2. Métodos para reducir la sensibilidad a las oclusiones

El uso de métodos robustos es un camino para conseguir reducir la sensibilidad a las oclusiones en las correspondencias y otras diferencias en las imágenes. La presencia de oclusiones en pares de imágenes estereó produce discontinuidades en la disparidad que por otro lado son coherentes. Es decir, existen regiones que por un lado tienen una discontinuidad grande en la disparidad, pero en otra dirección su función disparidad es suave. Esta suavidad introduce un nuevo umbral en la detección de oclusiones. Stewart [Stewart 97] llama a estas discontinuidades “pseudoutliers” y ofrece sugerencias para una selección cuidadosa de estimadores robustos para manejarlas.

Zabih y Woodfill [Zabih et al. 94] propusieron las transformaciones del rango y el censo que se aplican a las áreas locales antes de la correlación. Como estos métodos se basan en la intensidad relativa en vez de en las intensidades absolutas, son bastante robustos a las discontinuidades de disparidad.

Autores	Método	Restricciones
Nishihara	Búsqueda de contornos y correlación.	Epipolar Unicidad
Zabih-Woodfill	Transformada del rango y correlación. Transformada censo y distancia de Hamming.	Epipolar Unicidad
Marr-Poggio	Comparación de intensidades y análisis de vecindario.	Epipolar Unicidad Disparidad
Pollard-Mayhew-Frisby	Extracción de características de píxeles y comparación.	Epipolar Unicidad Disparidad Gradiente de disparidad
Venkatesvar-Chellappa	Correspondencias basadas en características: regiones, contornos, vértices, líneas.	Epipolar Unicidad Ordenamiento Disparidad
Baker-Binford, Otha-Kanade	Programación dinámica: Detección de contornos y estimación de superficie en profundidad mediante minimización de función de coste.	Epipolar Unicidad Ordenamiento
Tomasi-Manduchi	Representación de nivel de brillo frente a derivada (curvas íntimas).	Epipolar Unicidad
Medioni-Nevatia, Ayache-Faverjon	Comparación de segmentos de línea aplicando restricción epipolar modificada	Epipolar Unicidad Disparidad
Sanger, Jepson-Jemkin	Filtrado paso banda local y comparación de diferencias de fase.	Epipolar Unicidad Disparidad Gradiente de disparidad

Tabla 1. Resumen de Métodos y Restricciones.

Otra aproximación para reducir la sensibilidad a las oclusiones es redimensionar la ventana de correlación para optimizar la similitud de las correspondencias cerca de las oclusiones. Kanade y Okutomi [Kanade et al. 94] proponen un método iterativo para determinar el tamaño óptimo de las ventanas. El tamaño de las ventanas es inicialmente muy pequeño y se analiza la incertidumbre de la correspondencia. Entonces, el tamaño de la ventana se expande un píxel por cada dirección independientemente, y se vuelve a calcular la incertidumbre. Si esta incertidumbre ha crecido al expandir la ventana en una dirección, esa dirección se declara prohibida para expandir. Este procedimiento se aplica de forma iterativa hasta que

todas las direcciones se declaran prohibidas. Otro método más reciente ha sido propuesto en [Luo et al. 03].

5.9.3. Métodos para modelar la geometría de las oclusiones

Aunque los métodos anteriores para la detección y reducción de la sensibilidad a las oclusiones ofrecen distintas ventajas y todos son computacionalmente abordables, éstos no aprovechan todas las posibilidades que aportan las restricciones a la estereovisión. Es deseable integrar el conocimiento de la geometría de las oclusiones dentro del proceso de búsqueda.

Belhumeur [Belhumeur 96] define las bases de una serie de estimadores Bayesianos, cada uno de los cuales maneja un modelo más complicado del mundo. Los estimadores se utilizan para definir funciones de coste para utilizar en programación dinámica. El modelo más simple supone que las superficies son suaves. El segundo asume que además de la suavidad de las superficies existen contornos abruptos en los objetos. El modelo tercero es más realista e incluye superficies con inclinaciones además de los contornos en los objetos. Variaciones de estos modelos, sobre todo del segundo, han sido utilizadas tanto en la programación dinámica como en los “graph cuts” para determinar el mapa de disparidad óptimo.

Otro método para detectar oclusiones y recuperar las profundidades de estas regiones es explotar la posibilidad de tener varias cámaras. Kanade et al. [Kanade et al. 92] proponen un algoritmo que utiliza la función SSD y tiene como base la posibilidad de desplazar las cámaras obteniendo diferentes líneas base conocidas. La ambigüedad en el par estéreo debida a las oclusiones no está presente en otros pares, en los que el objeto es visible desde las dos perspectivas.

También se puede utilizar visión activa para detectar oclusiones y recuperar la profundidad de éstas, en base al estudio del movimiento de las cámaras, para conseguir que el punto oculto pase a ser visible. Ching [Ching 94] utiliza la visión activa para discernir entre oclusiones y reflejos especulares. Una de las dos cámaras del par estéreo se rota, y se analiza el cambio de las dimensiones de las regiones supuestamente ocultas. Aquellas áreas que no crecen o decrecen de una forma predecible se asume que son debidas a reflejos especulares.

Una aproximación más general para manejar oclusiones utilizando visión activa fue propuesta por Reed y Allen [Reed et al. 00], quienes describen la planificación de un sistema sensor que minimiza el número de cámaras requeridas para capturar todos los puntos de la escena.

5.10. Cuadro resumen de Técnicas de Correspondencia

En el cuadro resumen de la tabla 1 pueden observarse las aportaciones más significativas dentro de este campo de la estereovisión, así como las restricciones básicas en las que se basa cada uno de los métodos propuestos.

6. Color y Estereovisión

La relación entre el color y la estereovisión se merece un tratamiento especial, ya que son varios los motivos por los que usar la información cromática. En primer lugar, el color desempeña un papel importante en la percepción humana. Livingstone y Hubel [Livingstone et al. 87] mostraron que los humanos no pueden percibir la profundidad en estereogramas de color cuando los colores se encuentran en equi-iluminancia. Aunque el papel del color en la visión estereoscópica es aún una cuestión por resolver, queda poco lugar a la duda de que los atributos del color influyen en la estereopsis humana (véase [Brockelbank et al. 89]). En segundo lugar, es obvio que los píxeles rojos no casan con los píxeles azules, aunque sus intensidades son iguales o similares.

Por otra parte, la mayor parte de los algoritmos de correspondencia estéreo no son capaces de calcular mapas correctos de profundidad densos en regiones de imagen homogéneas. De ahí que podamos intuir que usando una iluminación cromática, los resultados de correspondencia mejoren sensiblemente [Koschan et al. 96]. Los enfoques computacionales existentes de correspondencia estéreo en color han mostrado que los resultados de correspondencia pueden mejorarse de un modo considerable utilizando información del color.

6.1. Métodos

Como se va a poder apreciar a lo largo de este apartado, la información del color aporta una fuente importante en la correspondencia estéreo. Los resultados de todos los métodos mencionados en las secciones anteriores (basados en características o basados en el área) pueden mejorarse considerablemente añadiendo información del color en el proceso de correspondencia.

Drumheller y Poggio [Drumheller et al. 86] presentaron uno de los primeros enfoques estéreo que utilizaba el color. Usaron el signo de la convolución como una primitiva de correspondencia en cada uno de los tres canales del color. No es que encontraran grandes mejoras en los resultados de escenas naturales típicas, pero sí que animaron el interés por el uso del color en estéreo. En 1988, Jordan y Bovik [Jordan et al. 88] mejoraron sus resultados de correspondencia de bordes usando gradientes cromáticos de tres espectros diferentes

normalizados (rojo menos verde, verde menos azul y azul menos rojo). En 1989, junto a Geisler [Jordan et al. 89], los mismos investigadores presentaron la denominada restricción de correspondencia del gradiente cromático, una extensión del límite del gradiente de disparidad [Pollard et al. 85] desde imágenes en niveles de gris a imágenes en color. La calidad de los resultados de la correspondencia con el algoritmo basado en bordes PMF [Pollard et al. 85] pueden mejorarse grandemente usando imágenes color y la restricción de correspondencia del gradiente cromático. El número de correspondencias correctas se incrementa desde un 100% hasta un 500% dependiendo de las imágenes color de prueba [Jordan et al. 91].

Brockelbank y Yang [Brockelbank et al. 89] sugirieron un enfoque basado en investigaciones fisiológicas y psicofísicas en percepción humana. Usaron el modelo de colores opuestos para la representación de las imágenes. A cada una de las correspondencias potenciales entre bordes se le ofrece una probabilidad respecto de la orientación del borde, del contraste del borde y un valor de disparidad máximo definido por el área de Panum. Tras asumir una continuidad figural y un suavizado de superficie global, se usa un etiquetado relajado para incrementar las probabilidades de las posibles correspondencias que están de acuerdo en disparidad con la mayoría de las posibles correspondencias dentro de una vecindad local.

Todos los métodos que incorporan el color vistos hasta aquí están basados en características, con lo que producen mapas de disparidad poco densos. No obstante existe una necesidad real de trabajar con algoritmos que computen mapas de densidad densos. Por ello, Nguyen y Cohen [Nguyen et al. 92] sugirieron hacer corresponder las regiones con características de curvatura similares de ambas imágenes. La correspondencia se realiza separadamente para los canales R (red, rojo), G (green, verde) y B (blue, azul) y los resultados son promediados. Se calcula un mapa de disparidad denso por medio de una interpolación sucesiva de esplines. No se trata de un algoritmo estéreo en el sentido de que el problema de la correspondencia se resuelve en todos los píxeles de la imagen simultáneamente, sin embargo, se calcula un mapa de profundidad denso.

En 1992, Jordan y Bovik [Jordan et al. 92] presentaron la restricción cromática fotométrica, que requiere de colores similares para hacer corresponder píxeles de ambas imágenes. La

similitud de correspondencias potenciales se mide mediante el cómputo de las diferencias absolutas entre los valores de intensidad en las componentes R, G y B. Junto con una restricción de suavizado de disparidad, aportan la suposición de que los valores de disparidad son similares en un vecindario de cuatro píxeles conectados, y definen la suma de estas restricciones como una función de energía.

Okutomi, Yoshizaki y Tomita [Okutomi et al. 92] sugirieron una técnica basada en el área. Utilizaron la suma de diferencias al cuadrado (SSD, sum of squared differences) de los valores de intensidad en los tres canales de color R, G y B para la correspondencia. Mostraron (para señales unidimensionales) que la varianza de los valores de disparidad estimados es menor – y por tanto la precisión mayor – usando imágenes en color en vez de imágenes en gris. La precisión de las disparidades estimadas aumenta cuando las imágenes contienen colores variados.

Otro enfoque es el de extender la técnica de correspondencia de bloques a la estereoscopia en color [Koschan 93]. Se investigaron tres modelos de color diferentes (RGB, $I_1I_2I_3$, HSI) y cinco métricas de color diferentes de cara a su conveniencia en la correspondencia estéreo y se compararon con los resultados del algoritmo para niveles de gris. La precisión de los resultados de correspondencia mejoró en todos los casos entre un 20 y un 30% usando información color frente a la información en niveles de gris.

6.2. Aplicaciones

Son muchas las aplicaciones que han ido apareciendo en las que se conjuga el color y la estereoscopia.

Más concretamente puede hablarse, en primer lugar, de una serie de investigaciones en procesamiento y análisis de imágenes. Así, por ejemplo, en [Xie 97] se investiga el problema de la correspondencia estéreo en el marco de la visión estéreo color para el desarrollo de un algoritmo de correspondencia automatizado de características. El autor propone el uso de una restricción de consistencia de correspondencia (MC, matching consistence) en el espacio de color RGB y de la geometría epipolar generalizada. En otro artículo se describe un sistema de identificación de objetos basado en modelos [Mustafa et al. 99]. Dado un conjunto de objetos 3D y una escena que contiene uno o más de estos

objetos, el sistema identifica qué objetos aparecen en la escena por medio de la correspondencia de firmas de superficie. Las firmas de superficie son vectores de características que reflejan la probabilidad de ocurrencia de las características para una superficie dada. Se emplean dos tipos de firmas de superficies: firmas de curvatura y firmas espectrales (de color). Sanfeliu y otros [Sanfeliu et al. 02] discuten el uso de representaciones y técnicas basadas en grafos para la segmentación de imágenes, la agrupación perceptual de imágenes y el reconocimiento de objetos. En su trabajo se describe una fusión novedosa de segmentación basada en color y obtención de la profundidad estéreo que produce un grafo que representa cada objeto en la escena.

El seguimiento de objetos también se está beneficiando de la simbiosis entre el estéreo y el color. Así, se ha presentado un sistema que realiza el seguimiento de una cabeza humana en 3D en tiempo real [Moreno et al. 02]. La fortaleza del método propuesto reside en el uso de un filtro predictivo que fusiona la información del color y de la profundidad, refinando iterativamente el lugar de la cabeza en 3D y los parámetros del histograma de colores de la cabeza. En otro artículo [Kwolek 03] se presenta igualmente un sistema de visión que realiza el seguimiento de una cara humana en 3D. Para ello se combinan cuestiones de color y estéreo para encontrar regiones de imagen similares en las que pueda encontrarse una cara. Se ha propuesto también recientemente el uso de modelos de la mano en visión por computador con la intención de recuperar gestos usando técnicas de reconstrucción 3D. En este trabajo [Yin et al 03], se presenta un método para estimar la geometría epipolar entre dos cámaras no calibradas a partir de imágenes estéreo de las manos.

Una mención muy especial de utilización mixta de estéreo y color se halla en el área de la robótica. Chen y otros [Chen et al. 98] presentan un sistema de tele-robot de asistencia de propósito general diseñado para manipular objetos en un entorno tridimensional usando color y visión estereoscópica. Vandapel y otros cuentan su experiencia con visión artificial en la Antártida en navegación robótica [Vandapel et al. 02]. Las condiciones climáticas especiales y la unicidad del terreno son estudiadas a través de la evaluación de un algoritmo de segmentación color, de un algoritmo de visión estéreo y de un algoritmo de detección de puntos y posterior seguimiento.

Agradecimientos

Este trabajo ha sido financiado en parte por el proyecto CICYT TIN2004-07661-C02-02.

Referencias

- [Alvarez et al. 02] Alvarez, L., Deriche, R., Sánchez, J. & Weickert, J., "Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach", *JVCIR* vol. 13, no. 1/2 (2002): 3-21.
- [Asano et al. 85] Asano, T., Edahiro, H. & Imai, H., "Bucketting techniques in computer geometry", *Computer Geometry*, Toussaint, G. Ed. (1985).
- [Aschwarden et al. 93] Aschwarden, P. & Guggenbuhl, W., "Experimental results from a comparative study on correlation-type registration algorithms", *Robust Computer Vision*. Frostner and Ruwiedel, Eds. (1993): 268-289.
- [Ayache et al. 87] Ayache, N. & Faverjon, B., "Efficient registration of stereo images by matching graph descriptions of edge segments", *The International Journal of Computer Vision*, (1987): 107-131.
- [Banks et al. 01] Banks, J. & Corke, P., "Quantitative evaluation of matching methods and validity measures for stereo vision", *International Journal on Robotics Research*, vol. 20, no. 7, (2001): 427-442.
- [Baker et al. 81] Baker, H.H. & Binford, T.O., "Depth from edge- and intensity-based stereo", *Proceedings 7th Joint Conference on Artificial Intelligence*, (1981): 631-636.
- [Belhumeur 96] Belhumeur, P. N., "A bayesian approach to binocular stereopsis", *International Journal on Computer Vision*, vol. 19, no. 3, (1996): 237-260.
- [Brockelbank et al. 89] Brockelbank, D.C., & Yang, Y.H., "An experimental investigation in the use of color in computational stereopsis", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, (1989): 1365-1383.
- [Brown et al. 03] Brown, M.Z., Burschka, D., Hager, G.D., "Advances in computational stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, (2003): 993-1008.

- [Canny 86] Canny, J., "A computational approach for edge detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 8, no. 6, (1986): 679-698.
- [Chang et al. 91] Chang, C., Chatterjee, S. & Kube, P. R., "On an analysis of static occlusion in stereo vision", Proceedings on Computer Vision and Pattern Recognition", (1991): 722-723.
- [Chen et al. 97] Chen, C., Hunga, Y.-P., Chianga, C.-C. & Wub, J.L., "Range data acquisition using color structured lighting and stereo vision", Image and Vision Computing, vol. 15, no. 6, (1997): 445-456.
- [Chen et al. 98] Chen, S., Kazi, K., Foulds, R. & Chester, D., "Color and three-dimensional vision-based assistive telemanipulation", Image and Vision Computing, vol. 16, no. 4, (1998): 265-274.
- [Ching 94] Ching, W. S., "A new method of identifying occlusion and specular highlights using active vision", Proceedings of the International Symposium on Speech, Image Processing and Neural Networks, (1994): 437-440.
- [Choi et al. 03] Choi, I., Yoon, J-G., Lee, Y.-B. & Chien, S.I., "Stereo system for tracking moving object using log-polar transformation and zero disparity filtering", CAIP (2003): 182-189.
- [Coren et al. 93] Coren, S., Ward, L. M. & Enns, J. T., "Brightness and Spatial Frequency in Sensation and Perception", Chap. 4, Harcourt Brace (1993)
- [Drumheller et al. 86] Drumheller, M. & Poggio, T., "On parallel stereo", Proceedings of the IEEE Conference on Robotics and Automation, (1986): 1439-1448.
- [Faugeras 93] Faugeras, O., "Three-Dimensional Computer Vision. A Geometric Viewpoint", The MIT Press, (1993).
- [Faugeras et al. 01] Faugeras, O. & Luong, Q. T., "The Geometry of Multiple Images", The MIT Press, (2001).
- [Forstner et al. 86] Forstner, W. & Pertl, A., "Photogrammetric standard methods and digital image matching techniques for high precision surface measurements", Pattern Recognition in Practice II, Elsevier Science Publishers, (1986): 57-72.
- [Fua 91] Fua, P., "Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities", Proceedings of the 12th International Joint Conference on Artificial Intelligence, (1991): 1292-1298.
- [Grimson 81] Grimson, W.E.L., "A computer implementation of a theory of human stereo vision", Philosophical Transactions of the Royal Society of London, Series B, vol. 292, (1981): 217-253.
- [Grimson 85] Grimson, W.E.L., "Computational experiments with a feature based stereo algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7, (1985): 17-34.
- [Gutiérrez et al. 04] Gutiérrez, S. & Marroquín, J.L., "Robust approach for disparity estimation in stereo vision", Image and Vision Computing, vol. 22, no. 3, (2004): 183-195.
- [Hartley et al. 00] Hartley R. & Zisserman, A., "Multiple View Geometry in Computer Vision", Cambridge University Press, (2000).
- [Hoff et al. 89] Hoff, W. & Ahuja, N., "Surfaces from stereo: Integrating feature matching, disparity, estimation and contour detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 2, (1989): 121-136.
- [Hong et al. 04] Hong, L. & Chen, G., "Segment-based stereo matching using graph cuts", CVPR 2004, (2004).
- [Jepson et al. 89] Jepson, A. D. & Jemkin, M. R. M., "The fast computation of disparity from phase differences", Proceedings of the Conference on Computer Vision and Pattern Recognition, (1989): 398-403.
- [Jordan et al. 88] Jordan III, J.R. & Bovik, A.C., "Computational stereo vision using color", IEEE Control Systems Magazine, (1988): 31-36.
- [Jordan et al. 89] Jordan III, J.R., Bovik, A.C. & Geisler, W.S., "Chromatic stereopsis", Proceedings of the International Joint Conference on Artificial Intelligence, (1989): 1649-1654.
- [Jordan et al. 91] Jordan III, J.R. & Bovik, A.C., "Using chromatic information in edge-based stereo correspondence", CVGIP: Image Understanding, vol. 54, no. 1, (1991): 98-118.
- [Jordan et al. 92] Jordan III, J.R. & Bovik, A.C., "Using chromatic information in dense stereo correspondence", Pattern Recognition, vol. 25, no. 4, (1992): 367-383.

- [Julesz 71] Julesz, B., "Foundations of Cyclopean Perception", University of Chicago Press, (1971).
- [Kanade et al. 92] Kanade, T., Okutomi, M. & Nakahara, T., "A multiple-baseline stereo method", Proceedings ARPA Image Understanding Workshop, (1992): 409-426.
- [Kanade et al. 94] Kanade, T. & Okutomi, M., "A stereo matching algorithm with an adaptative window: Theory and experiments", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 9, (1994): 920-932.
- [Kass 83] Kass, M., "A computational framework for the visual correspondence problem", Proceedings of the Eighth International Joint Conference on Artificial intelligence, (1983): 1043-1045.
- [Kass 88] Kass, M., "Linear image features in stereopsis", The International Journal of Computer Vision, (1988): 357-368.
- [Kelly et al. 77] Kelly R.E., McConnell, P.R.H. & Mildenerger, S.J., "The gestalt photomapping system", Photogrammetric Engineering and Remote Sensing", vol. 43, no. 11, (1977): 1407-1417.
- [Koenderink et al. 76] Koenderink, J.A. & van Doorn, A.J., "Geometry of binocular vision and a model for stereopsis", Biological Cybernetics, vol. 21, (1976): 29-35.
- [Koschan 93] Koschan, A., "Chromatic block matching for dense stereo correspondence", Proceedings of the 7th International Conference on Image Analysis and Processing, (1993): 641-648.
- [Koschan et al. 96] Koschan, A., Rodehorst, V. & Spiller, K., "Color stereo vision using hierarchical block matching and active color illumination", Proceedings of the 13th International Conference on Pattern Recognition, vol. I, (1996): 835-839.
- [Kwolek 03] Kwolek, B., "Face tracking system based on color, stereovision and elliptical shape features", Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, (2003): 21-26.
- [Kwolek 04] Kwolek, B., "Real-time head tracker using color, stereovision and ellipse fitting in a particle filter", Informatica, vol. 15, no. 2, (2004): 219-230.
- [Lee et al. 90] Lee, S. U., Chung, S. Y. & Park, R.H., "A comparative performance study of several global thresholding techniques for segmentation", Computer Vision, Graphics and Image Processing, vol. 52, no.2, (1990): 171-190.
- [Livingstone et al. 87] Livingstone, M.S. & Hubel, D.H., "Psychophysical evidence for separate channels for the perception of form, color, movement and depth", Journal of Neuroscience, vol. 7, (1987): 3416-3468.
- [Luo et al. 03] Luo, Q., Zhou, J., Yu, S., Xiao, D., "Stereo matching and occlusion detection with integrity and illusion sensitivity", Pattern Recognition Letters, vol. 24, no. 9-10, (2003): 1143-1149.
- [Marr et al. 76] Marr, D. & Poggio, T., "Cooperative computation of stereo disparity", Science, vol. 194., (1976): 283-287.
- [Marr et al. 79] Marr, D. & Poggio, T., "A computational theory of human stereo vision", Proceedings of the Royal Society of London, Series B, vol. 204, (1979): 301-328.
- [Marr et al. 80] Marr, D. & Hildreth, E., "Theory of edge detection", Proceedings of the Royal Society of London, Series B, vol. 207, (1980): 187-217.
- [Marr 82] Marr, D. "Vision", Freeman, San Francisco. (1982).
- [Mustafa et al. 99] Mustafa, A.A.Y., Shapiro, L.G. & Ganter, M.A., "3D object identification with color and curvature signatures", Pattern Recognition, vol. 32, no. 3, (1999): 339-355.
- [Mayer 03] Mayer, H. "Analysis of means to improve cooperative disparity estimation", ISPRS Conference on Photogrammetric Image Analysis, (2003): 25-31.
- [Medioni et al. 85] Medioni, G. & Nevatia, R., "Segment-based stereo matching", Computer Vision, Graphics, and Image Processing, vol. 31, (1985): 2-18.
- [Moreno et al 02] Moreno, F., Tarrida, A., Andrade-Cetto, J. & Sanfeliu, A., "3D real-time head tracking fusing color histograms and stereovision", Proceedings of the 16th International Conference on Pattern Recognition, vol. 1, (2002): 368-371.
- [Mühlmann et al. 02] Mühlmann, K., Maier, D., Hesser, J., Männer, R., "Calculating dense disparity maps from color stereo images, an efficient implementation", International Journal of Computer Vision, vol. 47, no. 1-3, (2002): 79-88.

- [Nguyen et al. 92] Nguyen, H.H. & Cohen, P., "Correspondence from Color Shading", Proceedings of the 11th IAPR International Conference on Pattern Recognition, vol. 1, (1992): 113-144.
- [Nishihara et al. 82] Nishihara, H.K. & Poggio, T., "Hidden cues in random line stereograms", Nature, vol. 300, (1982): 347-349.
- [Nishihara et al. 84] Nishihara, H.K., "PRISM, a practical real-time imaging stereo matcher; Technical Report", A.I. Memo 780 MIT, MIT Press. (1984).
- [Ohta et al. 80] Ohta, Y.-I., Kanade, T. & Sakai, T., "Color information for region segmentation", Computer Graphics and Image Processing, vol.13, no. 3, (1980): 222-241.
- [Ohta et al. 85] Ohta, Y. & Kanade T., "Stereo by intra- and interscanline search", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7, no. 2, (1985): 139-154.
- [Okutomi et al. 92] Okutomi, M., Yoshizaki, O. & Tomita, G., "Color stereo matching and its application to 3-D measurement of optic nerve head", Proceedings of the 11th IAPR International Conference on Pattern Recognition, vol. I, (1992): 509-513.
- [Pollard et al. 85] Pollard, S.B., Mayhew, J.E.W. & Frisby, J.P., "A stereo correspondence algorithm using a disparity gradient constraint", Perception, vol. 14, (1985): 445-470.
- [Prewitt 70] Prewitt, J.M.S., "Object enhancement and extraction", Picture Processing and Psychopictorics, Lipkin, B.S. and Rosenfeld, A. (eds.), Academic Press, New York, (1970): 75-149.
- [Reed et al. 00] Reed, M. K. & Allen, P. K., "Constraint-based sensor planning for scene modeling", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, (2000): 1460-1467.
- [Roberts 65] Roberts, L.G., "Machine perceptions of three-dimensional solids", Optical and Electro-Optical Information Processing, Tippet, J. T. (ed), MIT Press (1965).
- [Russ 95] Russ, J. C., "Thresholding images", Computer Assisted Microscopy, vol. 7, no. 3, (1995): 41-164.
- [Sáez et al. 03] Sáez, J.M., Peñalver, A. & Escolano, F., "Compact mapping in plane-parallel environments using stereo Vision", CIARP 2003, (2003): 659-666.
- [Sahoo et al. 88] Sahoo, P. K., Soltani, S., Wong, A. K. & Chan, Y. C., "A survey of thresholding techniques.", Computer Vision, Graphics and Image Processing, vol. 4, (1988): 233-260.
- [Sanfeliu et al. 02] Sanfeliu, A., Alquézar, R., Andrade, J., Climent, J., Serratos, F. & Vergés, J., "Graph-based representations and techniques for image processing and image analysis", Pattern Recognition, vol. 35, no. 3, (2002): 639-650.
- [Sanger 88] Sanger, T.D., "Stereo disparity computation using Gabor filters", Biological Cybernetics, no. 59, (1988): 405-418.
- [Silva et al. 00] Silva, C. & Santos-Victor, J., "Intrinsic images for dense stereo matching with occlusions", Proceedings on European Computer Vision, (2000): 100-114.
- [Sobel 78] Sobel, I., "Neighbourhood coding of binary images for fast contour following and general array binary processing", Computer Graphics and Image Processing, vol. 8, (1978): 127-135.
- [Stewart 97] Stewart, C. V., "Bias in robust estimation caused by discontinuous and multiple structures", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 8, (1997): 818-833.
- [Sumi et al. 02] Sumi, Y., Kawai, Y., Yoshimi, T. & Tomita, F., "3D object recognition in cluttered environments by segment-based stereo vision", International Journal of Computer Vision, vol. 46, no. 1, (2002): 5-23.
- [Tomasi et al. 98] Tomasi, C. & Manduchi, R., "Stereo matching as a nearest neighbor problem", IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 20, (1998): 333-340.
- [Trucco et al. 98] Trucco, E. & Verri, A., "Introductory Techniques for 3-D Computer Vision", Prentice Hall, (1998).
- [Vandapel et al. 02] Vandapel, N. & Chatila, R., "Qualitative evaluation of computer vision algorithms in polar terrains", Robotics and Autonomous Systems, vol. 40, no. 2-3, (2002): 139-149.
- [Venkateswar et al. 95] Venkateswar, V. & Chellappa, R., "Hierarchical stereo and motion correspondence using feature groupings", International Journal on Computer Vision, vol. 15, (1995): 245-269.

- [Wang et al. 03] Wang, J.-G., Venkateswar, R. & Lim, E.T., "Face tracking and recognition from stereo sequence", AVBPA 2003, (2003): 145-153.
- [Wei et al. 04] Wei, Y. & Quan, L., "Region-based progressive stereo matching", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, (2004): 106-113.
- [Wilson et al. 95] Wilson, R. & Knutsson, H., "A multiresolution stereopsis algorithm based on Gabor representation", Proceedings of the IEE International Conference on Image Processing and Applications, (1989): 19-22.
- [Wildes 91] Wildes, R. P., "Direct recovery of three-dimensional scene geometry from binocular stereo disparity", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 8, (1991): 761-774.
- [Xie 97] Xie, M., "Automatic feature matching in uncalibrated stereo vision through the use of color", Robotics and Autonomous Systems, vol. 21, no. 4, (1997): 355-364.
- [Yin et al. 03] Yin, X. & Xie, M. "Estimation of the fundamental matrix from uncalibrated stereo hand images for 3D hand gesture recognition", Pattern Recognition, vol. 36, no. 3, (2003): 567-584.
- [Zabih et al. 94] Zabih, R. & Woodfill, J., "Non parametric local transforms for computing visual correspondence", Proceedings Third European Conference on Computer Vision", (1994): 150-158.
- [Zitnick et al. 00] Zitnick, C. L. & Kanade, T., "A cooperative algorithm for stereo matching and occlusion detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no.7, (2000): 675-684.