

RECUPERACIÓN DE LA PROFUNDIDAD POR PIXEL EN ESTEREOGRAMAS

Leonardo Manucci y Néstor Calvo

*Universidad Nacional del Litoral, Facultad de Ingeniería y Ciencias Hídricas, Santa Fe, Argentina,
manucci@leonardo.net.ar, <http://fich.unl.edu.ar>*

Palabras Clave: Reconstrucción tridimensional, estereografía, fotogrametría.

Resumen. En este trabajo se describe la implementación y los algoritmos desarrollados para la recuperación de la información de profundidad por píxel en un par de imágenes estereoscópicas. A diferencia de los métodos que realizan un matching o identificación de pares de píxeles correspondientes a un vértice o detalle de la escena, aquí se realiza la reconstrucción del buffer de profundidad por zonas, mediante un algoritmo recurrente que subdivide las zonas obteniendo una mayor resolución en cada iteración, hasta llegar al nivel del píxel. Las zonas se aparean mediante una métrica de similitud basada en la suma de diferencias absolutas de intensidades en los tres canales de color.

1 INTRODUCCIÓN

Se denomina reconstrucción tridimensional al proceso mediante el cual se obtiene una representación digital de la forma geométrica de las superficies de objetos de la vida real así como también su ubicación espacial relativa.

Este trabajo presenta un método de reconstrucción tridimensional de superficies a partir de un par de imágenes estereoscópicas (Figura 1). Estas imágenes son obtenidas a través de un par de cámaras posicionadas y orientadas de tal manera, que los puntos de vista obtenidos se aproximen a los captados por el sistema visual humano. La principal característica de estas imágenes es que difieren levemente una de la otra, lo cual facilita la correspondencia de puntos.

El tipo de sistema presentado en este trabajo se denomina de visión estéreo debido a que se utilizan dos puntos de vista de una misma escena para determinar las coordenadas tridimensionales. Funcionan de igual manera que el sistema visual humano, restaurando la información de profundidad perdida durante el proceso de adquisición de las imágenes.



Figura 1: Imágenes estereoscópicas tomadas desde puntos de vista similares a los captados por el sistema visual humano.

Este sistema puede ser utilizado en la reconstrucción de objetos para diversos usos, como por ejemplo: la creación de modelos para entornos de realidad virtual, el diseño y la fabricación asistida por computadora, la creación de paseos virtuales, el relevamiento arqueológico, el modelado de personajes u objetos para películas o videos juegos, etc.

La idea básica de estos sistemas es considerar a la fotografía como la proyección de un objeto tridimensional sobre un plano (Hartley y Zisserman, 2003; Trucco y Verri, 1998). Para ello adoptamos el modelo de cámara ideal (cámara sin lentes) que consiste en un punto central o punto de vista de tamaño nulo, en el cual convergen todos los rayos visuales provenientes de la escena, atravesando antes un plano virtual donde se forma la imagen. Cada punto de la fotografía corresponde a un punto de la escena, que ha sido proyectado sobre el plano de la imagen. Esta forma de abstracción (idéntica a la utilizada en la generación de gráficos por computadora) permite trabajar de manera simple con las propiedades geométricas y físicas del sistema de cámaras.

El modelo abstracto que utilizaremos para describir las propiedades físicas del sistema de cámaras está compuesto por dos clases de parámetros, los intrínsecos y los extrínsecos. Los parámetros intrínsecos describen las características físicas y ópticas de cada cámara, y los extrínsecos definen la ubicación y orientación de las cámaras en la escena. En la Figura 2 se puede ver un ejemplo en forma esquemática, de un sistema de cámaras ideales captando un objeto tridimensional desde dos puntos de vista diferentes, donde las proyecciones sobre los planos representan las imágenes capturadas.

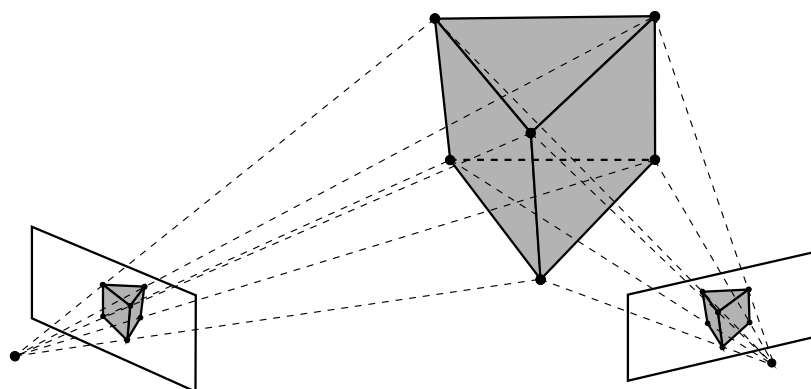


Figura 2: Sistema estéreo. Dos cámaras ideales capturan sendas imágenes de un objeto tridimensional desde dos puntos de vista diferentes.

Si contamos con dos fotografías tomadas desde diferentes puntos de vista conocidos e identificamos un mismo punto de la escena en cada una de ellas, se hace posible restablecer las coordenadas reales de dicho punto. El proceso de identificación de puntos homólogos en ambas fotografías se denomina correspondencia (*matching*) y es el proceso más importante debido a que los resultados finales dependerán directamente de él.

1.1 Correspondencia de Imágenes (*matching*)

El proceso de correspondencia consiste en determinar para un punto en una de las imágenes, cuál es el punto homólogo en la otra imagen. Dicho de otro modo, se busca en cada imagen las proyecciones generadas por un punto de la escena.

La búsqueda de correspondencias es un problema mal condicionado. Pueden existir infinitas soluciones (alineación óptica) o ninguna (oclusiones).

En este trabajo se propone un método basado en regiones, esto que significa que para obtener una correspondencia se extrae de una de las imágenes una pequeña región rectangular de píxeles denominada plantilla, la cual se utiliza para buscar en la otra imagen una región de iguales dimensiones, llamada ventana, que maximice algún criterio de similitud determinado. Debido a las características de las imágenes estereoscópicas, la búsqueda de correspondencias en la segunda imagen puede restringirse a una región o vecindario centrado en las mismas coordenadas que la plantilla. Durante la búsqueda, la ventana se desplaza dentro del vecindario en la segunda imagen, en cada posición por la que se la desplaza se calcula la similitud entre los píxeles contenidos en la plantilla y los contenidos en la ventana. Una vez recorrido todo el vecindario se determina en que posición de la ventana se ha obtenido la mayor similitud entre los píxeles. Dicha posición se toma como correspondiente y se continúa la búsqueda con la siguiente plantilla.

1.2 Mapa de disparidades

Las correspondencias son representadas mediante el conjunto de coordenadas píxel de los puntos en cada imagen. Se define a la disparidad como la diferencia vectorial en píxeles entre dos puntos correspondientes.

Para representar las disparidades se utiliza frecuentemente una imagen, con dimensiones iguales a las de las imágenes procesadas, en la cual la intensidad de los píxeles representa el valor de disparidad para cada punto. Dicha imagen se denomina “mapa de disparidades” y permite visualizar rápidamente la información tridimensional. Valores mayores de disparidad se corresponden con los tonos claros y viceversa.

2 DESCRIPCIÓN DEL TRABAJO REALIZADO

Primero se hará una descripción del método de búsqueda de correspondencias diseñado y luego se describirá el método de restablecimiento tridimensional utilizado.

2.1 Método de búsqueda de correspondencias.

Para la búsqueda de correspondencias se diseñó un método iterativo basado en la correspondencia de áreas (*area based*, *window-based* o *block matching* ver Brown, Burschka y Hager, 2003; Hirschmüller y Scharstein, 2007) y el análisis en múltiples resoluciones de las imágenes. Este método recibe como entrada el par de imágenes estereoscópicas y retorna el conjunto de puntos correspondientes.

El diseño propuesto está basado en el conocimiento previo de las características de las imágenes estereoscópicas. Se asume que las imágenes fueron tomadas con una pequeña diferencia de posición horizontal, por lo tanto las disparidades serán pequeñas y con mayor componente horizontal que vertical. Además se asume que los objetos de la escena poseen superficies suaves, por lo que la disparidad individual de los píxeles contenidos en una pequeña plantilla de búsqueda, no se desvía mucho del promedio de disparidades en la zona.

El método planteado es iterativo. En la primera iteración se trabaja con una escala reducida de las imágenes y se realiza una aproximación gruesa del mapa de disparidades. En las siguientes iteraciones se aumenta la escala de trabajo y se refinan los resultados hasta que se decida finalizar el proceso. Para determinar las correspondencias se toma la imagen origen y se la divide en sectores rectangulares (S_o), los cuales se buscan en la imagen destino, dando como resultado otro grupo de sectores (S_d).

En la primera iteración se inicializan las disparidades en cero, es decir que la búsqueda se realiza en un vecindario centrado en las mismas coordenadas que el S_o . Los desplazamientos resultantes de la búsqueda se almacenan para ser utilizados luego.

En las siguientes iteraciones, los sectores se subdividen y los valores de desplazamiento previos se utilizan para corregir la posición de los vecindarios de búsqueda de los subsectores resultantes. En cada nueva iteración los sectores se subdividen aumentando la resolución del mapa de disparidades.

Cada iteración del proceso de búsqueda de correspondencias tiene como entrada el conjunto de pares de sectores S_o y S_d , la escala a utilizar y los valores máximos de desplazamiento posible que definen el vecindario de búsqueda para los subsectores.

Para determinar la correspondencia se recorta el sector de la imagen origen (S_o) y se lo compara contra una ventana en la imagen destino, de iguales dimensiones, y que se desplaza por un vecindario centrado en las mismas coordenadas que el S_d . El criterio utilizado para determinar la correspondencia entre dos sectores consiste en minimizar la suma de los valores absolutos de las diferencias entre los valores de las componentes RGB de los píxeles individuales.

La ventana correspondiente, la de menor diferencia, se selecciona como el nuevo S_d y sus coordenadas se almacenan junto con las coordenadas del S_o . Al comenzar cada iteración se parte del conjunto de pares de sectores obtenidos del proceso de subdivisión y al finalizar la iteración se obtiene un nuevo conjunto, cuyos sectores destino han sido desplazados hacia nuevas posiciones dentro de un vecindario donde maximizan la similitud.

Para reducir los errores en los resultados se realizó un suavizado del mapa de disparidades, con este proceso se logra disminuir los picos y las variaciones bruscas de disparidad que se suponen producidas por falsas correspondencias.

El proceso iterativo de búsqueda de correspondencias puede ser descrito mediante el

siguiente pseudocódigo:

- 1) Cargar las imágenes y los parámetros de las cámaras.
- 2) Inicializar el arreglo de pares (S_o , S_d) con un par de sectores que abarquen toda la imagen origen y destino respectivamente.
- 3) Seccionar la escala, la cantidad de divisiones a realizar y el tamaño del vecindario de búsqueda.
- 4) Escalar las imágenes.
- 5) Para cada sector (S_o , S_d) en el arreglo:
 - {
 - 6) Dividir ambos sectores.
 - 7) Para cada sub-sector origen resultante:
 - {
 - 8) Buscar la mejor correspondencia del Sub- S_o en un vecindario centrado en el Sub- S_d .
 - 9) Almacenar el Sub- S_o y el nuevo Sub- S_d resultante de la búsqueda.
 - }
 - }
- 10) Con el nuevo arreglo de sectores S_o , S_d se vuelve al punto 3) hasta que el usuario decida terminar con el proceso.

2.2 Restablecimiento tridimensional

A través de este proceso se pretende recuperar la información de profundidad de cada punto en la escena (z-buffer) a partir de las coordenadas píxel de sus correspondientes proyecciones en cada una de las imágenes. Existen diferentes métodos dependiendo de la cantidad de vistas utilizadas y si se conoce o no la geometría del sistema (Elias y Laganière, 2001).

El método implementado en este trabajo recibe como entrada los parámetros intrínsecos y extrínsecos del sistema de cámaras y el conjunto de correspondencias en coordenadas píxel, y retorna las coordenadas tridimensionales de cada uno de los puntos.

El primer paso consiste en ubicar las cámaras en la escena y determinar las coordenadas tridimensionales de los puntos correspondientes. Luego, se determinan las coordenadas tridimensionales del punto que originó dichas correspondencias, a través del cálculo de la intersección de los rayos ópticos formados por los puntos de vista o puntos centrales de las cámaras y los respectivos puntos correspondientes. Por lo general ésta intersección no se produce debido a errores de discretización, deformación de las imágenes o errores de correspondencia, y el problema se convierte en la búsqueda del punto de mínima distancia a ambos rayos (Trucco y Verri, 1998) o triangulación mediante rayos visuales no intersecados (Figura 3).

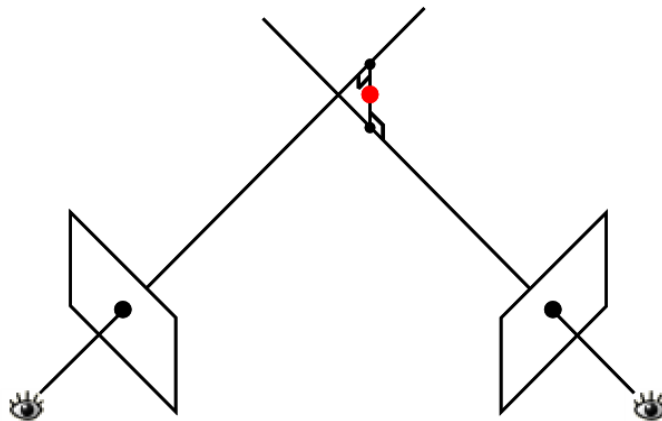


Figura 3: Triangulación mediante rayos visuales no intersecados.

2.3 Generación del modelo tridimensional.

A partir del conjunto de puntos en coordenadas píxel extraídos de las imágenes construimos una malla bidimensional de elementos triangulares, utilizando el método de triangulación de Delaunay. Con la topología obtenida renderizamos la malla pero en lugar de utilizar las coordenadas en píxeles, utilizamos las coordenadas tridimensionales reconstruidas, y de esta manera generamos el modelo tridimensional. A su vez, para cada punto reconstruido se tienen las coordenadas píxel de la correspondencia en la imagen que originó dicho punto, estas últimas coordenadas son utilizadas para aplicar la textura a las superficies del objeto reconstruido dándole un aspecto realista al modelo tridimensional generado.

3 RESULTADOS

El método presentado en este trabajo permite obtener un mapa de disparidades denso y relativamente suave gracias al refinamiento recursivo de las disparidades. La cantidad de puntos reconstruidos depende de la cantidad de divisiones que se realicen sobre las imágenes. A medida que se realiza mayor cantidad de divisiones se aumenta la precisión del mapa de disparidades. A su vez el escalado de las imágenes permite obtener aproximaciones con precisión sub-píxel.

El método de correspondencia diseñado en este trabajo es sensible a los cambios bruscos en las disparidades, los mejores resultados se logran cuando los objetos en la escena poseen superficies suaves y con texturas, además es muy importante que la escena se encuentre bien iluminada, tratando de evitar reflejos de luz sobre las superficies de interés. (ver Figura 4).

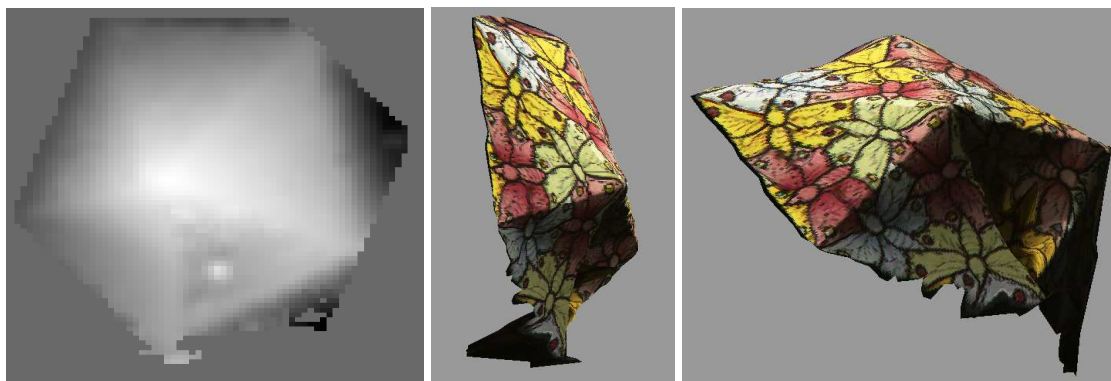


Figura 4: Mapa de disparidades suavizado y dos vistas del modelo reconstruido a partir de imágenes capturadas en condiciones de laboratorio (Geusebroek, Burghouts y Smeulders, 2005).

En la Figura 5 se puede observar el error producido por la gran diferencia de disparidad entre superficies solapadas. Este error se produce debido a que en un mismo sector origen existen zonas de píxeles correspondientes a superficies con gran diferencia de disparidad entre sí, y no es posible obtener una correspondencia correcta a menos que se separen estas zonas y se busquen individualmente. Como resultado de esta situación las zonas que “pesan más” arrastran a las otras dejándolas muy “alejadas” de su destino real, produciéndose de esta forma una falsa correspondencia.



Figura 5: Mapa de disparidades suavizado y vistas del modelo reconstruido. Se señalan los errores producidos por la discontinuidad de superficies y la gran diferencia de disparidades.

Por último cabe destacar que el método presentado nos permite determinar un mapa de disparidades, y reconstruir un modelo tridimensional sin escala cuando se parte de imágenes cuyos parámetros, no conocidos, han sido aproximados de manera empírica. En la Figura 6 se pueden apreciar los resultados logrados con un par de imágenes estereoscópicas obtenidas del Sitio Web Flickr (www.flickr.com/groups/stereo/).

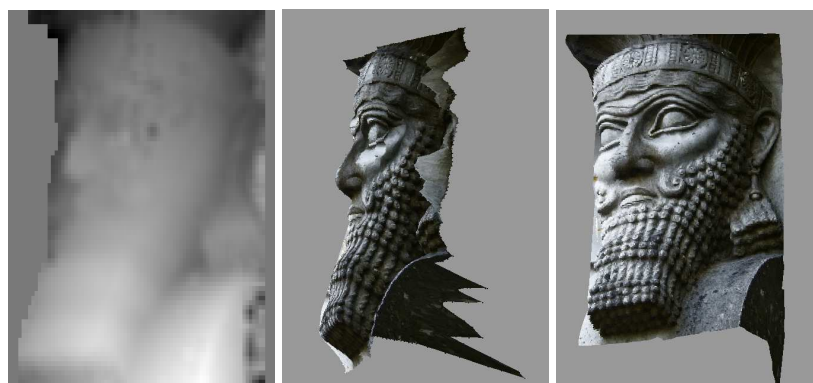


Figura 6: Mapa de disparidades suavizado y dos vistas del modelo reconstruido a partir de imágenes de parámetros desconocidos obtenidas de la Web.

REFERENCIAS

- Brown M., Burschka D. and Hager G. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.
- Elias R. and Laganière R. Projective Geometry for Three-Dimensional Computer Vision. *Proceedings of the Seventh World Multiconference on Systemics, Cybernetics and Informatics*. 5:99-104, 2003.
- Geusebroek J., Burghouts G. and Smeulders A. Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61(1):103–112, 2005.
- Hartley R. and Zisserman A. *Multiple View Geometry in Computer Vision Second Edition*. Cambridge Press, 2003.
- Hirschmüller H. and Scharstein D. Evaluation of Cost Functions for Stereo Matching, *IEEE Conference on Computer Vision and Pattern Recognition*, 18-23, 2007.
- Trucco E. and Verri A. *Introductory Techniques for 3D Computer Vision*. Prentice Hall, 1998.