Modelado 3D de escenas estructuradas mediante visión estereoscópica binocular

Angeles López, Filiberto Pla, José Ribelles
Dept. Informática, Universitat Jaume I, E-12080 Castelló, SPAIN
{lopeza, pla, ribelles}@inf.uji.es

Resumen

Las escenas estructuradas están compuestas normalmente por objetos fabricados por el hombre, que suelen tener superficies homogéneas, con poca textura, lo que hace que la correspondencia estereoscópica sea más difícil. En este trabajo presentamos un nuevo algoritmo de correspondencia que obtiene directamente una representación 3D de escenas estructuradas calculando correspondencias de regiones y oclusiones cooperativamente.

Palabras clave: visión, robótica i percepción, realidad virtual.

1 Introducción

En este trabajo se presenta un nuevo método de visión estereoscópica, cuyo objetivo es obtener una representación 3D de una escena a partir de dos vistas de la misma. El interés de esta aproximación es obtener un modelo 3D del entorno, que puede ser útil para múltiples aplicaciones: planificación del movimiento de un robot móvil o manipulador, modelado 3D de escenas y objetos para aplicaciones de realidad virtual, presentaciones de productos, etc.

Nuestro trabajo se enfoca en escenas estructuradas, principalmente interiores, que están compuestas de objetos hechos por el hombre. Estas escenas suelen contener superficies planas, que a menudo no tienen textura. Los métodos de visión estéreo son capaces de obtener un mapa de profundidades de la escena, que es más denso cuanto más textura tienen las superficies. Para obtener una representación 3D útil (por ejemplo, un modelo de malla de polígonos), debe aplicarse un método de segmentación al mapa de profundidades. Además, tanto el método de visión estéreo como el de segmentación de datos de rango tienen que ser capaces de detectar y manejar oclusiones y discontinuidades en la profundidad.

Nuestro método se basa en correspondencia de regiones, para poder reconocer y establecer la correspondencia de áreas sin textura de las imágenes. La estrategia de correspondencia consiste en un esquema cooperativo donde las correspondencias, las oclusiones y la reconstrucción 3D se obtienen en un solo proceso. En todo momento se mantiene una representación de las superficies 3D de la escena que se modifica iterativamente con el objetivo de minimizar las diferencias entre regiones correspondientes. Esta estrategia está sujeta a ciertas restricciones sobre las superficies de la escena. En este trabajo se estudia un caso sencillo para verificar la convergencia del algoritmo, analizar el coste computacional y evaluar la utilidad de los resultados.

Las principales ventajas del método que aquí se presenta son:

- se asigna una profundidad a todos los pixeles de la imagen, incluyendo aquellos que están ocultos en la otra imagen, y
- se obtiene un modelo 3D directamente, sin necesidad de segmentar los datos de rango.

2 Trabajo previo

2.1 Estructura 3D a partir de visión estereoscópica

Las técnicas de visión estereoscópica normalmente se clasifican en dos amplias familias [1, 8, 10, 15].

Los métodos basados en áreas explotan la resolución radiométrica de los pixels mediante el uso de ventanas, y obtienen buenos resultados cuando las escenas se componen de superficies bien texturadas. Estos métodos proporcionan un mapa de disparidades denso, lo cual es su principal ventaja, junto con la posibilidad de obtener precisión subpixel. Su principal desventaja es que implícitamente asumen que la superficie es continua y por tanto, tienen problemas en presencia de discontinuidades en las superficies tridimensionales. Generalmente, los datos de profundidad se transforman en descriptores estructurales a través de métodos de segmentación de datos de rango [22, 3, 7].

Los métodos basados en características explotan las primitivas de alto nivel de la imagen, que son elementos (pixeles de los bordes, segmentos lineales o curvos de los bordes, regiones, etc.) con atributos distintivos (posición, orientación, curvatura, etc.) que se usan para encontrar las correspondencias. El principal problema de estos métodos es que proporcionan un mapa de disparidades disperso. A medida que el nivel de la primitiva es mayor, crece la robustez de las correspondencias, y también la dispersión del mapa de disparidades. Es preciso un post-proceso para obtener descriptores de la superficie a partir de la información dispersa [22, 9] mediante algún método de interpolación que debe ser capaz de detectar las discontinuidades. También se han desarrollado métodos jerárquicos [14] para reducir la dispersión de los resultados. Solamente unos pocos métodos sacan provecho de las primitivas correspondientes para obtener la estructura 3D directamente [21, 20, 6].

2.2 Detección de oclusiones

Geiger *et al.* [11] mostraron que las oclusiones pueden ayudar al cálculo de la correspondencia. Es posible modelar las oclusiones y las discontinuidades en la profundidad explícitamente, de modo que sean parte del problema a resolver, y por tanto, de la solución. Así, las oclusiones y las discontinuidades en la profundidad no son problemas a evitar, sino una fuente de información a tener en cuenta. Belhumeur [2] también propuso mantener internamente un mapa detallado de la geometría de la escena (profundidad, orientación, etc.), para que todos estos elementos cooperen en la optimización de la correspondencia.

A pesar de la importancia de la detección de oclusiones, sólo unos pocos métodos la integran en el proceso de correspondencia. La mayoría de estos métodos modelan las oclusiones como elementos en el espacio de correspondencias, y usan una estrategia de programación dinámica para obtener la solución [11, 12, 13, 2, 4]. Estos métodos son técnicas basadas en áreas, que suelen fallar cuando la superficie es homogénea. Se necesita una restricción de suavizado para evitar correspondencias erróneas en las zonas homogéneas y las zonas semiocultas. La restricción de suavizado no debería aplicarse cuando hay características destacadas de la geometría de la escena, como son las discontinuidades en la profundidad y las superficies de fuerte inclinación, que deben preservarse para producir reconstrucciones precisas. Muchos de los métodos utilizados para minimizar estas complicaciones con las zonas homogéneas suavizan estas características destacadas [4]. Belhumeur [2] consiguió relacionar las características destacadas, las oclusiones y la correspondencia estéreo en un modelo muy completo que proporciona una solución global. Sin embargo, la búsqueda de la solución óptima es muy costosa computacionalmente.

Olsen [19] integró la detección de las discontinuidades en la disparidad y las áreas ocultas en un método basado en características, analizando las derivadas parciales de la superficie de disparidades reconstruida. La reconstrucción y la correspondencia se integran en un único proceso, con una estrategia de grueso a fino. Sin embargo, se basa en la correspondencia de bordes de las imágenes, que normalmente son escasas en las superficies homogéneas.

2.3 Uso de regiones como primitiva de la correspondencia

El interés del uso de regiones como primitiva de la correspondencia se puede resumir en tres puntos:

1. Tienen un nivel semántico mayor que otras primitivas. Su estabilidad y capacidad descriptiva reduce el número de ambigüedades, incrementa la tolerancia al ruido y proporciona correspondencias más fiables [17, 18].

2. Representan áreas de intensidad homogénea con discontinuidades de intensidad en sus fronteras. Debido a que las discontinuidades en la profundidad y las oclusiones están localizadas en las discontinuidades de la intensidad de la imagen, las regiones representan áreas de profundidad continua, mientras que en sus fronteras debe permitirse la aparición de discontinuidades en la profundidad.

3. Un modelo 3D compuesto de superficies planas es a menudo suficiente para representar escenas estructuradas. Los métodos existentes que reconstruyen la superficie 3D a partir de dos regiones [21, 20, 6] asumen que cada región es la proyección de una superficie plana tridimensional.

Sin embargo, las primitivas de mayor nivel son más difíciles de extraer de la imagen. Debido al ruido, a las oclusiones y a las limitaciones de las técnicas de segmentación, pueden aparecer importantes diferencias entre las regiones obtenidas a partir de ambas imágenes. Esto hace que la reconstrucción de la superficie 3D sea más difícil, si no imposible. Por ejemplo, Tarel *et al.* [21, 20] aplican un test de coherencia basado en invariantes a los pares de regiones correspondientes para rechazar los pares que no estén bien segmentados, y Chabbi *et al.* [6] necesitan un sistema trinocular para obtener tripletes de caras 3D que se validan a través de principios de geometría proyectiva.

3 Contribuciones

El método que se propone aquí está basado en características de alto nivel que son las regiones, pero la similitud de las regiones se calcula teniendo en cuenta la resolución radiométrica de los pixeles de la región (como los métodos basados en áreas) mediante el uso de técnicas de correlación. De este modo se intenta evitar las desventajas de ambos tipos de métodos.

Proponemos el uso de regiones como primitiva de la correspondencia para obtener la estructura 3D de la escena directamente. Para evitar el problema de las diferencias entre la segmentación de ambas imágenes, proponemos segmentar solamente una imagen, a la que llamamos *imagen de referencia*.

La búsqueda de la región correspondiente de cada región en la imagen de referencia se realiza minimizando el error en la similitud de ambas. Debido a la restricción de unicidad y la existencia de oclusiones, la minimización para una región no es independiente de la minimización para sus regiones adyacentes. Por tanto, éste es un problema de optimización n-dimensional, donde n es el número de regiones en la imagen de referencia.

Este problema de optimización puede expresarse en términos de cálculo variacional. Sin embargo, para obtener un algoritmo que garantice la convergencia es necesario asumir alguna restricción en la forma de las superficies 3D de la escena. Una suposición que se ha hecho habitualmente es que cada región es la proyección de una superficie plana 3D [21, 20, 6]. Proponemos una aproximación preliminar para este método, que consiste en una suposición más restrictiva: la superficie 3D es plana y paralela a la imagen de referencia. Esta aproximación permite estudiar más fácilmente la convergencia, el coste y los resultados de nuestro algoritmo. En [16] se presentó un algoritmo inicial que obtuvo resultados satisfactorios. Sin embargo, la precisión de los resultados decrecía cuando la minimización de unas regiones convergía mucho más rápidamente que la de otras.

En este trabajo, presentamos un nuevo algoritmo que controla la minimización de cada región en varios pasos, para guiar el proceso cooperativamente hacia una convergencia global. Además, mostramos que el orden del coste computacional teórico es similar a los métodos basados en áreas que utilizan programación dinámica para obtener soluciones globales. Finalmente, se presentan algunos resultados con escenas sintéticas y reales para evaluar el alcance de esta aproximación y las posibilidades de una generalización del método a otros tipos de superficies.

4 Correspondencia de regiones y detección de oclusiones

A cada región R de la imagen de referencia se le asocia una superficie 3D que se ajusta a la restricción formulada, en este caso, un plano caracterizado por su profundidad, $Z(R)$. La región correspondiente R' de una región cualquiera de la imagen de referencia, R, consiste en la proyección de su superficie asociada en la otra imagen. Por tanto, R' es una función de R y su profundidad.

$$R' = f(R, Z(R)) \tag{1}$$

La estrategia de la búsqueda de la correspondencia se basa en minimizar el error en la similitud entre regiones correspondientes, que puede calcularse mediante cualquier técnica de correlación, por ejemplo, la correlación cruzada normalizada de media cero (ZNCC).

El interés de esta estrategia es que la búsqueda de las correspondencias se puede expresar en términos de cálculo variacional, donde las incógnitas son los parámetros de las superficies asociadas, en este caso, la profundidad de cada región R, $Z(R)$.

Otra ventaja es que la detección de oclusiones es sencilla: cuando, dadas dos regiones de la imagen de referencia, R_1 y R_2, sus regiones correspondientes, R'_1 y R'_2, intersectan, el área intersectada es un área oculta de una de las dos superficies asociadas, debido a la restricción de unicidad. La región más lejana está parcialmente oculta por la región más cercana. Debido a que la existencia de una oclusión puede introducir errores en la medida de correlación de la región, las áreas ocultas deben ser detectadas y eliminadas del cálculo de la correlación. De esta forma, la detección de oclusiones durante el proceso de correspondencia facilita la búsqueda de la misma.

La función a minimizar consiste en el error en la similitud (calculado mediante el método ZNCC) de cada región de la imagen de referencia con su región correspondiente (calculada en función de una estimación de su profundidad). Es decir,

$$E(\mathcal{Z}) = \int_{R\subset I_1} -C_{ZNCC}(R, Z(R))dR \qquad (2)$$

donde $\mathcal{Z}$ es el conjunto de profundidades de todas las regiones de la imagen, $\mathcal{Z} = \{Z(R), R \subset I_1\}$, que conforman el mapa de profundidades de la imagen de referencia, I_1.

Utilizando la ecuación de Euler, se obtiene que la profundidad $Z(R)$ que hace mínima la función $E(\mathcal{Z})$ es solución de la siguiente ecuación

$$C_{ZNCC}(R, R')H(R', R') - H(R, R') = 0 \qquad (3)$$

donde R' es la región correspondiente a R y varía según la profundidad actual $Z(R)$, $C_{ZNCC}(R, R')$ es la correlación del nivel de gris entre pixeles de R y los puntos correspondientes en R', y $H(P, Q)$ es una medida de la correlación entre nivel de gris en P y la derivada parcial del nivel de gris respecto a la profundidad en Q (para más detalles, ver [16]).

Sea $F_z(R, Z(R))$ el error calculado por la ecuación anterior para una profundidad dada de una región R.

$$\begin{aligned} F_z(R, Z(R)) &= C_{ZNCC}(R, R')H(R', R') \\ &- H(R, R') \end{aligned} \qquad (4)$$

Si dibujamos F_z respecto a $Z(R)$ podemos observar un paso por cero en cada mínimo de la función $E(\mathcal{Z})$, con valores positivos a la derecha y negativos a la izquierda del paso por cero. Por tanto, se puede diseñar un algoritmo iterativo de minimización que, a partir de una estimación inicial de la profundidad, incremente/decremente iterativamente la profundidad actual hacia la solución. Es importante señalar que la profundidad inicial no debe estar lejos de la solución para evitar otros mínimos locales diferentes de la solución global. Esta es la razón por la que se necesita un esquema multiescala, donde el mapa de profundidades que resulta de cada nivel de una estructura piramidal de las imágenes se usa para inicializar el mapa de profundidades del siguiente nivel (Fig. 1).

El algoritmo de correspondencia en cada nivel es un algoritmo de optimización n-dimensional, donde n es el número de regiones de la imagen de referencia. Las n minimizaciones no son independientes entre sí debido a las oclusiones. Las regiones adyacentes a una región dada R pueden producir oclusiones en R que influyen

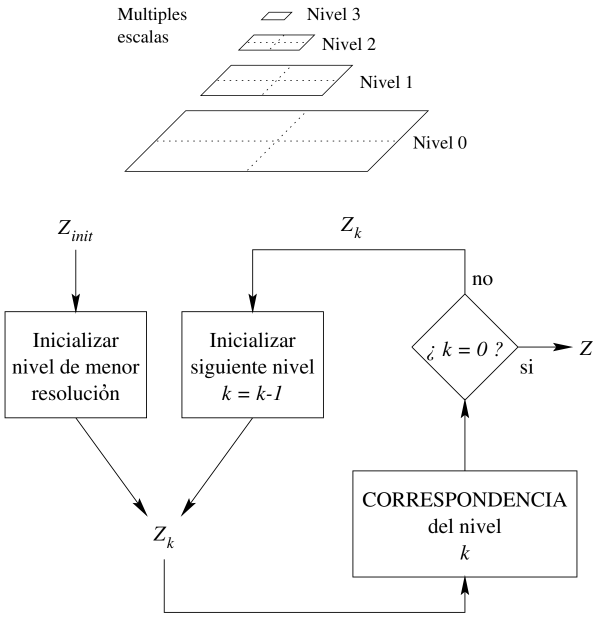

Figura 1: Esquema multinivel.

en el cálculo de F_z. Por tanto, las profundidades de las regiones se deben modificar muy poco a poco para que el algoritmo de optimización iterativo converja.

La convergencia se consigue mediante pequeños incrementos/decrementos en cada profundidad hasta que se alcanza la solución. Fijar el tamaño de los incrementos de profundidad, Δz, puede llevar a incrementos en la disparidad mayores que 1 pixel cuando los objetos están cerca, lo que hay que evitar. Proponemos fijar incrementos de disparidad discretos, Δd, que deben ser menores que 1 pixel y que deberían ser suficientemente pequeños para acercarse lentamente hacia la solución. Sin embargo, el tiempo de computación se incrementa cuando decrece Δd. Un esquema multirresolución, donde se utilizan distintos valores de Δd, decrecientes, permite acelerar los cálculos, a la vez que se refinan los resultados. Este esquema puede aplicarse de diferentes maneras: utilizando un Δd local de cada región, o un Δd global que cambia al mismo tiempo para todas las regiones. Las posibles estrategias son:

1. Cuando una región converge en cada resolución, su Δd se reduce y la convergencia empieza de nuevo.

2. Cuando una región converge en cada resolución, su Δd se reduce y la convergencia continua en la dirección inicial.

3. Cuando una región converge en cada resolución, se continua la convergencia de las demás. Cuando

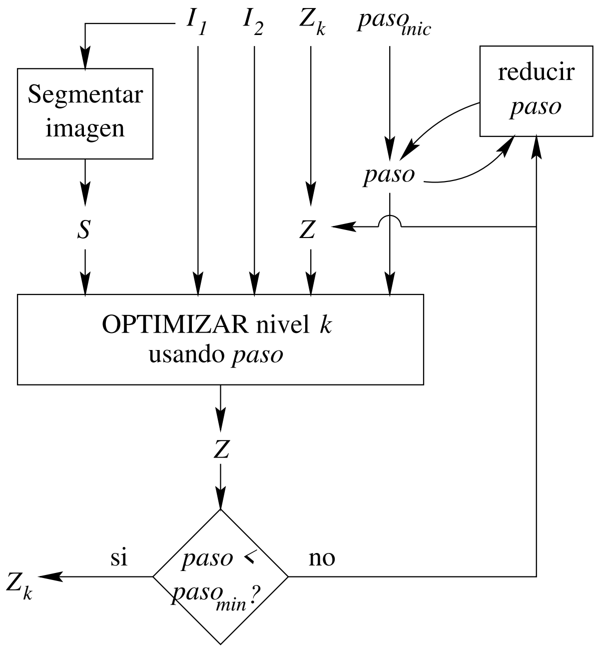

Figura 2: Correspondencia en cada nivel.

todas han convergido, el Δd global se reduce y la convergencia empieza de nuevo.

El método 1 no garantiza la convergencia del algoritmo, dado que la interdependencia de los resultados puede llevar a ciclos en su comportamiento. El método 2 garantiza la convergencia, pero la precisión de los resultados decrece cuando una región converge mucho más rápidamente que otra. Por ejemplo, cuando una región R que tiene una zona oculta en la otra imagen converge deprisa a un mínimo local, $Z(R) = z$, las oclusiones pueden detectarse mediante la convergencia correcta de sus regiones adyacentes, pero z no tiene por qué seguir siendo el valor de $Z(R)$ que hace mínima la función, ya que la profundidad de sus regiones adyacentes influye, a través de las oclusiones, en la similitud de R con su región correspondiente. Es decir, la posición del paso por cero en F_z varía en función de la profundidad de las regiones adyacentes y, aunque la profundidad de las regiones adyacentes varíe, la profundidad de R no es corregida una vez ha convergido. Finalmente, el método 3 garantiza la convergencia al mismo tiempo que la aproximación al mínimo global es controlado en varios pasos En la figura 2 se puede observar el esquema multirresolución propuesto, que consiste en un bucle que realiza la optimización con un Δd dado (llamado *paso*) y reduce el tamaño de Δd en cada iteración.

El algoritmo de correspondencia en cada paso de optimización consiste en la iteración de dos operaciones

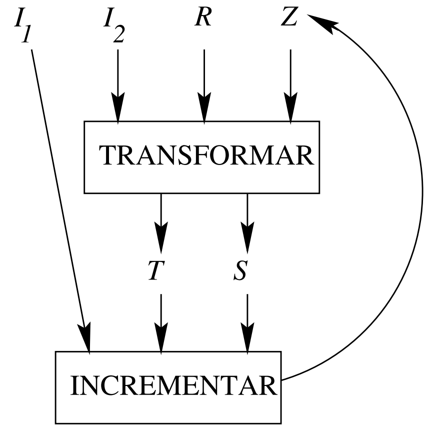

Figura 3: Matching at each level and each step.

(Fig. 3):

- *Transformar* la segunda imagen en dos mapas: T contiene los puntos correspondientes para cada pixel de la imagen de referencia, y S contiene una clasificación de todos los pixeles de I_1: *activo*, *oculto* o *fuera de límites*.

- *Incrementar* (o decrementar) la profundidad dependiendo de los pasos previos y la comparación de la imagen de referencia con T, teniendo en cuenta solamente los pixeles *activos* en S.

La precisión de la detección de oclusiones (y por tanto, de los resultados de profundidad) puede incrementarse también mediante el uso de detección subpixel de oclusiones, que consiste en usar $u \times v$ celdas para el cálculo de la clasificación de cada pixel.

4.1 Análisis del coste computacional

El coste computacional en el peor caso, suponiendo que Δd se reduce dividiéndolo por 2, es $O\left(\frac{D}{\Delta d_{min}} NMuv\right)$ donde $N \times M$ es el tamaño de las imágenes, $u \times v$ es el número de secciones subpixel para la detección de oclusiones, D es el tamaño del rango de disparidades, y Δd_{min} es el Δd mínimo requerido. Es decir, el coste del algoritmo en el peor caso es el producto del número de pixeles a la resolución subpixel seleccionada para las oclusiones ($NMuv$, que es el coste de cada iteración el bucle interno) por el número de disparidades a la resolución subpixel de las disparidades ($\frac{D}{\Delta d_{min}}$, que es el número total de iteraciones del bucle externo). Sin embargo, es importante señalar que este coste teórico se calcula asumiendo que existe una región que atraviesa

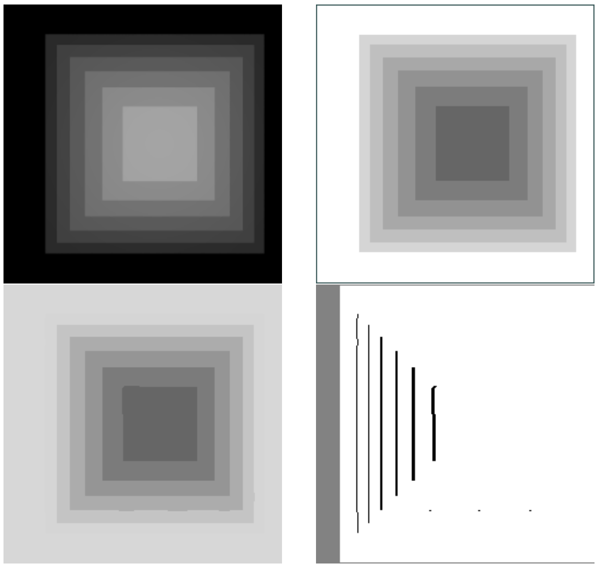

Figura 4: Un ejemplo de imágenes sintéticas. Arriba: imagen de referencia y mapa de profundidades conocido. Abajo: mapa de profundidades y mapa de oclusiones resultantes.

Figura 5: Ejemplo *parking meter* del conjunto de test JISCT [5]. Imagen de referencia, y resultados con el algoritmo anterior (izquierda) y el nuevo (derecha).

el intervalo de profundidades en cada iteración del bucle externo, incluyendo la última iteración del bucle, que es la más costosa. Esto no es verdad en la práctica, donde la primera iteración de bucle externo realiza más iteraciones del bucle interno que las siguientes iteraciones. La primera iteración del bucle externo es una etapa de acercamiento, mientras que el resto de iteraciones son etapas de refinamiento que normalmente requieren pocas iteraciones.

5 Resultados experimentales

Los experimentos con imágenes sintéticas con profundidades conocidas muestran que con el nuevo algoritmo se aumenta la precisión de los resultados respecto a los obtenidos anteriormente [16]. En la figura 4, se muestra un ejemplo de una pirámide con profundidades entre 82 y 117 cm. El error medio relativo del mapa de profundidades obtenido, sin considerar la región del fondo, es 0.62% con $u = 5$, mientras que era 2.31% con el algoritmo anterior y 3.78% sin la detección de oclusiones. En los mapas de profundidades de todos los ejemplos que se muestran, las zonas más oscuras son las más cercanas y las claras, las lejanas.

En la figura 5 se muestra otro ejemplo que permite comparar el mapa de profundidades obtenido con los resultados anteriores [16]. Hay algunas regiones que obtienen una profundidad errónea, pero en general la profundidad obtenida con el nuevo algoritmo cambia más suavemente de izquierda a derecha que en los resultados anteriores. Por ejemplo, la profundidad de ambos parquímetros y el seto del fonto se consiguen con más precisión. Algunas profundidades erróneas se deben a errores en las regiones obtenidas a partir del método de segmentación, que mezcla áreas de profundidad diferente. El resto de regiones erróneas son regiones verticales muy delgadas a las que se asignan oclusiones inexistentes debido al paralelismo de las superficies asumido en la restricción impuesta.

Los experimentos con imágenes reales de escenas estructuradas muestran que se obtiene una representación completa de la escena. En las figuras 6 y 8 se muestran imágenes de un despacho y el resultado de la correspondencia. La figura 10 muestra dos vistas de la escena, donde se puede observar que las profundidades obtenidas son satisfactorias. Tal como se esperaba, la superficie del suelo no puede ser modelada con la restricción impuesta y por tanto todo lo relacionado con él (oclusión, profundidad) es erróneo. El algoritmo detecta correctamente las oclusiones de este ejemplo (por ejemplo, la parte derecha y abajo del armario), y también algunas falsas (por ejemplo, la oclusión de la pared de encima de la cajonera, debida a una segmentación incorrecta de la zona más oscura, que es unida al objeto que oscuro que hay encima de la cajonera).

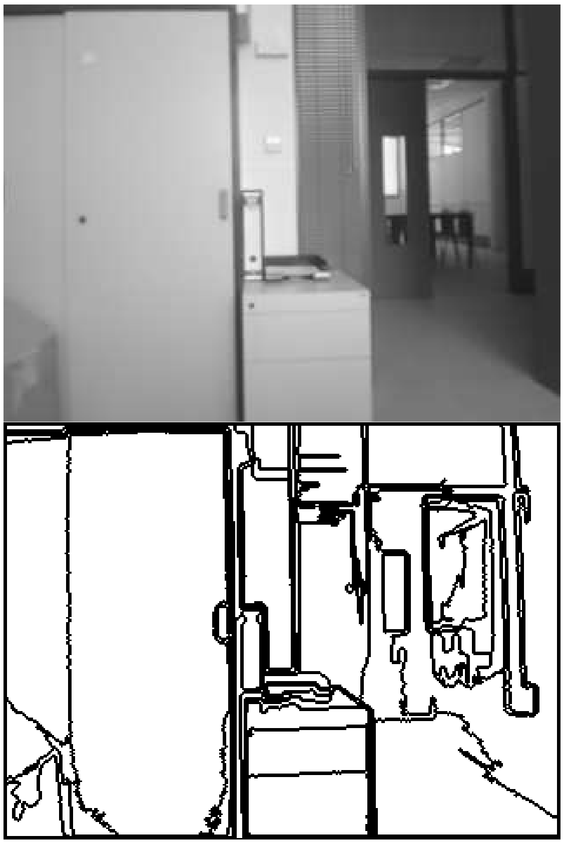

Figura 6: Par estéreo *despacho*: imagen de referencia, y segmentación por unión de regiones.

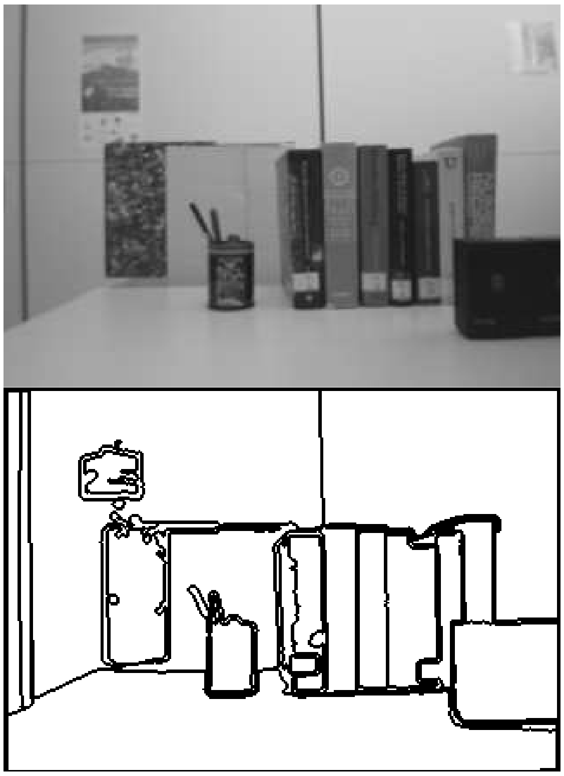

Figura 7: Par estéreo *libros*: imagen de referencia y segmentación por unión de regiones.

En la figura 7 se muestra otro ejemplo, donde las zonas ocultas son mucho mayores, y por tanto influyen todavía más en la correspondencia. En la figura 9 se muestran los mapas de profundidades y oclusiones obtenidos donde se puede observar que se han encontrado todas las oclusiones correctamente, salvo las oclusiones generadas sobre la mesa que, como el suelo del ejemplo anterior, no son correctas debido a la restricción impuesta. La figura 11 muestra dos vistas del modelo 3D obtenido, donde se puede observar la profundidad de los objetos en la escena. A pesar de la restricción de paralelismo de las superficies, los resultados son bastante satisfactorios y nos animan a generalizar el método a otros tipos de superficies.

6 Conclusiones

Hemos presentado un nuevo algoritmo de correspondencia que es capaz de obtener directamente un modelo 3D de una escena estructurada a partir de dos vistas, donde las discontinuidades en la profundidad no son alisadas, y la detección de oclusiones tiene un importante papel. La precisión de los resultados se ha incrementado mediante el control de la convergencia del algoritmo de optimización. El orden del coste computacional es similar al de las técnicas basadas en áreas, que proporcionan una solución global al mapa de profundidades denso.

La utilidad de los resultados está limitada por la restricción asumida. La suposición de una escena hecha de superficies planas de cualquier orientación podría ser suficiente para múltiples aplicaciones. El trabajo futuro se dirige a generalizar el método a este tipo de superficies y a superficies de segundo grado. Otra línea de trabajo es el uso de la consistencia izquierda-derecha para mejorar la segmentación inicial del nivel de gris de la imagen en una segmentación más consistente basada no sólo en el nivel de gris sino tambien en información de rango.

Agradecimientos

Agradecemos la financiación recibida de la Fundació Caixa-Castelló (proyecto P1B98-03) y de la CICYT, del Ministerio de Educación y Ciencia (mediante los proyectos TIC98-0677-C02-01 y TAP1999-0590-C02-01), que ha permitido la realización de este trabajo.

Referencias

[1] S. T. Barnard and M. A. Fishler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, Dec. 1982.

[2] P. Belhumeur. *A Bayesian Approach to the Stereo Correspondence Problem*. PhD thesis, Electrical Engineering, Yale University, May 1993.

[3] P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting. *IEEE Trans. on PAMI*, 10(2):167–192, Mar. 1988.

[4] S. Birchfield and C. Tomasi. Depth Discontinuities by Pixel-to-Pixel Stereo. In *Proceedings of the IEEE International Conference on Computer Vision, Bombay, India*, 1998.

[5] R. Bolles, H. Baker, and M. Hannah. The JISCT stereo evaluation. In *Proc. ARPA Image Understanding Workshop*, pages 263–274, Washington, DC, Apr. 18-21 1993. Morgan Kaufmann.

[6] H. Chabbi and M. Berger. Using projective geometry to recover planar surfaces in stereovision. *Pattern Recognition*, 29(4):533–548, 1996.

[7] L.-H. Chen and W.-C. Lin. Visual surface segmentation from stereo. *Image and Vision Computing*, 15:95–106, 1997.

[8] U. Dhond and J. Aggarwal. Structure from stereo - a review. *IEEE Trans. on Systems, Man, and Cybernetics*, 19(6):1489–1510, 1989.

[9] C. Dillon and T. Caelli. Generating complete depth maps in passive vision systems. pages 562–566, 1992.

[10] O. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat, and Z. Zhang. Quantitative and qualitative comparison of some area and feature-based stereo algorithms. In W. Förstner and S. Ruwiedel, editors, *Robust Computer Vision: Quality of Vision Algorithms*, pages 1–26. Wichmann, Karlsruhe, Germany, 1992.

[11] D. Geiger, B. Ladendorf, and A. Yuile. Occlusions and binocular stereo. *Int. J. of Computer Vision*, pages 221–226, 1995.

[12] S. Intille and A. Bobick. Disparity-space images and large occlusion stereo. In J.-O. Eklundh, editor, *Proc. 3rd European Conf. on Computer Vision*, volume B of *Lecture Notes in Computer Science*, pages 674–677, Stockholm, Sweden, May 1994. Springer Verlag. Extended version in M.I.T Media Lab Computing Group Technical Report No. 220.

[13] S. Intille and A. Bobick. Incorporating intensity edges in the recovery of occlusion regions. In *Int. Conf. on Pattern Recognition*, volume A, pages 674–677, 1994. Also M.I.T Media Lab Computing Group Technical Report No. 246.

[14] G. Jones. Constraint, optimization, and hierarchy: Reviewing stereoscopic correspondence of complex features. *IEEE Trans. on PAMI*, 65(1):57–78, 1997.

[15] R. A. Lane and N. A. Thacker. Stereo vision research: An algorithm survey. Technical Report 94/16, University of Sheffield, Electronic Systems Group, 1994.

[16] A. López and F. Pla. A minimization approach for 3D recovery in region-based stereo vision. In *Proc. 10th Image Processing and its Applications*, pages 47–51, Manchester, UK, 1999.

[17] S. Marapane and M. Trivedi. Region-based stereo analysis for robotics applications. *IEEE Trans. on Systems, Man, and Cybernetics*, 19(6):1447–1464, 1989.

[18] S. Marapane and M. Trivedi. Multi-primitive hierarchical (mph) stereo analysis. *IEEE Trans. on PAMI*, 16(3):227–240, 1994.

[19] S. Olsen. Stereo correspondence by surface reconstruction. *IEEE Trans. on PAMI*, 12(3):309–315, 1990.

[20] J.-P. Tarel. Reconstruction globale et robuste de facettes 3D. Rapport de recherche 2813, INRIA, Feb. 1996.

[21] J.-P. Tarel and J.-M. Vézien. A generic approach for planar patches stereo reconstruction. In *Proc. 11th Scandinavian Conf. on Image Analysis*, pages 1061–1070, Norway, 1995.

[22] D. Terzopoulos. Reconstruction of visual surfaces: Variational principles and finite element representations. Technical Report A. I. Memo 671, MIT, 1982.

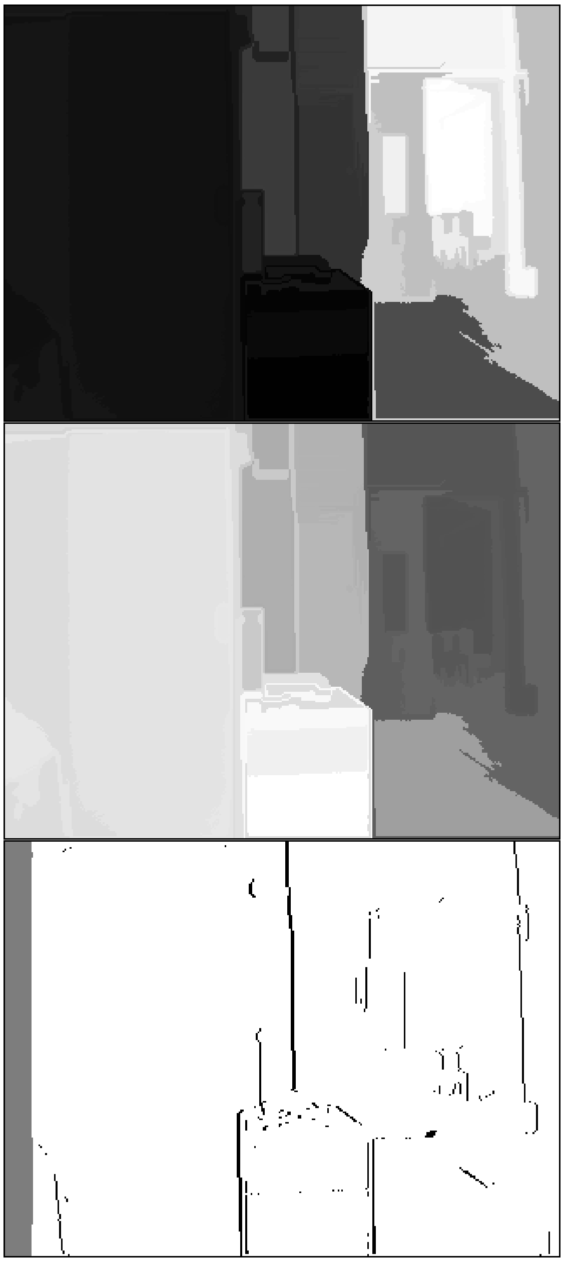

Figura 8: Par estéreo *despacho*. De arriba a abajo: mapa de profundidades, mapa de disparidades y mapa de oclusiones.

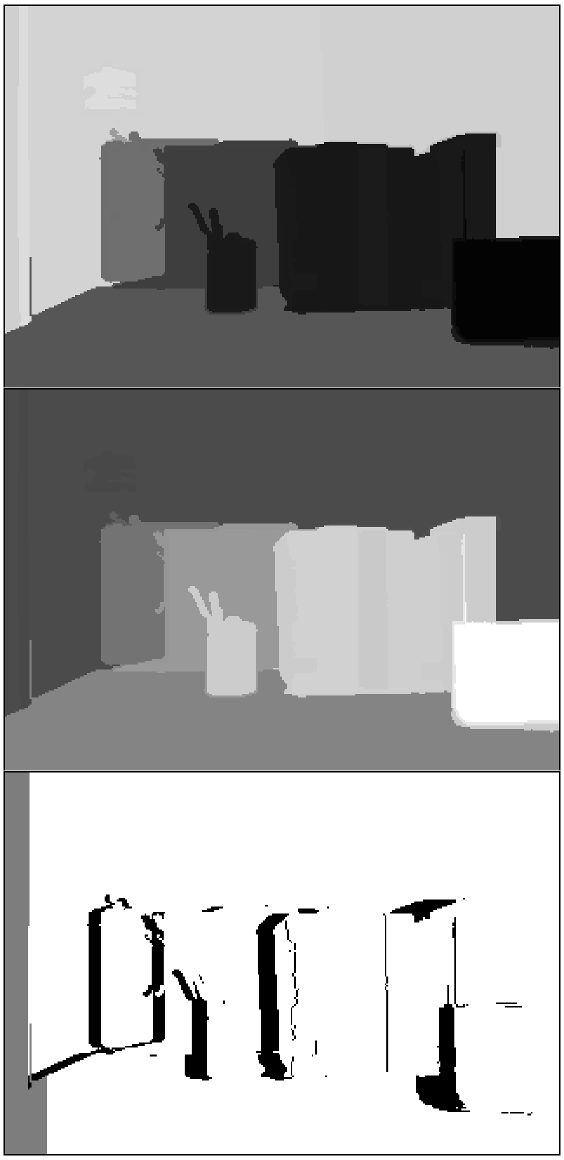

Figura 9: Par estéreo *libros*. De arriba a abajo: mapa de profundidades, mapa de disparidades y mapa de oclusiones.