Científica, Vol.14 Núm. 2, pp. 67-73, abril-junio 2010.
ISSN 1665-0654, ESIME Instituto Politécnico Nacional México.

Aplicación de un algoritmo de cálculo de disparidad para la estimación de profundidades usando cámaras estéreo

Jesús Arturo Escobedo-Cabello
Volodymyr Ponomaryov
Enrique Escamilla-Hernández

Sección de Estudios de Posgrado e Investigación (SEPI)
Escuela Superior de Ingeniería Mecánica y Eléctrica Culhuacan,
Instituto Politécnico Nacional, México.
Av. Santa Ana núm. 1000, Col. San Fco. Culhuacan,
CP 04430, México DF,
MÉXICO.

Tel. Fax (+52) 55 5729 6000
Correo electrónico arturoescobedo_iq@hotmail.com

Recibido el 4 de agosto de 2009; aceptado el 29 de enero de 2010.

1. Resumen

En este trabajo se presenta la implementación de un sistema para estimación de profundidad usando un par estéreo de cámaras. El sistema está basado en un algoritmo de Block Matching para el cálculo del mapa denso de disparidad. Para poder utilizar el algoritmo de Block Matching en imágenes reales se realiza la calibración de las cámaras y rectificación de las imágenes del par estéreo. El software fue desarrollado en el lenguaje de programación C y es capaz de realizar procesamiento en tiempo real.

Palabras clave: disparidad, profundidad, *block matching*, visión estéreo, calibración, rectificación

2. Abstract (Implementation of a Depth Estimation System Using an Stereo Pair)

This work presents the implementation of a depth estimation system using an stereo pair. The system developed is based on Block Matching method for disparity map calculation. In order to be able to use Block Matching algorithm over real images camera calibration and image rectification are done. The software was developed in C programming language and is able to do real time processing.

Key words: disparity, depth, block matching, stereo vision, camera calibration, image rectification.

3. Introducción

En muchas aplicaciones de la robótica móvil se requieren sistemas de visión estéreo con el propósito de explorar el área alrededor del robot debido a que el control de la navegación se puede mejorar mediante la resolución de tareas como la detección de obstáculos, el reconocimiento de formas y la estimación de distancias. [1, 2]. Las imágenes son sistemas que brindan una gran cantidad de información acerca del entorno que nos rodea (color, formas, etc.), no obstante la profundidad de un punto en una escena no puede ser directamente accesible a partir de una sola imagen, para ello son necesarias al menos dos. La visión estéreo o estereoscópica se define como aquella en la que se emplea más de una imagen para obtener una idea de tridimensionalidad [3]. Según el número de imágenes que se emplee, se habla de visión bifocal (dos imágenes), trifocal (tres imágenes), cuadrifocal (cuatro imágenes) o *n*-focal (*n* imágenes).

4. Desarrollo

A. Disparidad

Una forma de estimar la profundidad de cada uno de los puntos en una escena es mediante el cálculo de la disparidad (diferencia de posición) entre las imágenes de la misma [4]. Para definir la disparidad asumiremos que la escena es estática, es decir, los objetos visibles en ella no cambian su posición, ni sufren deformaciones, y las imágenes son tomadas usando una configuración de dos cámaras con características similares, formando un par estéreo (véase figura 1).

Para una cámara que cumple con el modelo *pinhole* y cuyos ejes ópticos son paralelos O_1o_1.

Ambas cámaras tienen la misma distancia focal *f*, con centros O_l y O_r separados una distancia *b*, llamada línea base *baseline*, de modo que las imágenes izquierda (I_l) y derecha (I_r), que se forman, estén en planos paralelos.

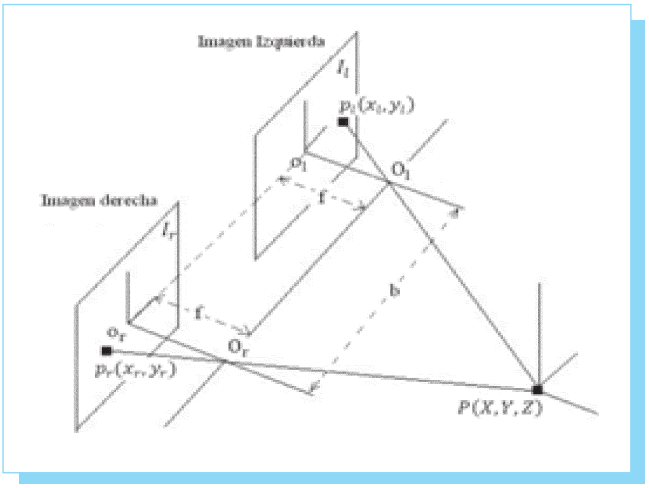

Fig. 1. Configuración de cámaras en un par estéreo.

De esta manera la línea base es paralela a la coordenada x de las imágenes. En el modelo *pinhole* considerado, un punto en el espacio tridimensional P, de coordenadas homogéneas $(X,Y,Z,1)^T$ se proyecta en cada una de las imágenes bidimensionales sobre los puntos p_l y p_r con coordenadas homogéneas $(x_l,y_l,1)^T$ y $(x_r,y_r,1)^T$, respectivamente.

La geometría proyectiva brinda las herramientas para trabajar analíticamente con las relaciones geométricas que gobiernan la proyección de un punto en 3D sobre una imagen [5]. El uso de las coordenadas homogéneas permite considerar el modelo *pinhole* de la cámara como una transformación lineal entre la escena y el plano de la imagen mediante la matriz de cámara Π_C.

$$kp = \Pi_C \mathrm{P} \tag{1}$$

$$(kx,ky,k)^T = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} (X,Y,Z,1)^T \tag{2}$$

La forma mostrada en (2) es la más sencilla para una matriz de cámara, esto es válido para todo $k\neq 0$ y permite hallar las relaciones entre las coordenadas de P y p.

Dependiendo el sistema de referencia utilizado en las imágenes, la definición de la disparidad puede cambiar, de forma que el signo sea siempre positivo.

$$Z = b\ \frac{f}{x_l - x_r} \tag{3}$$

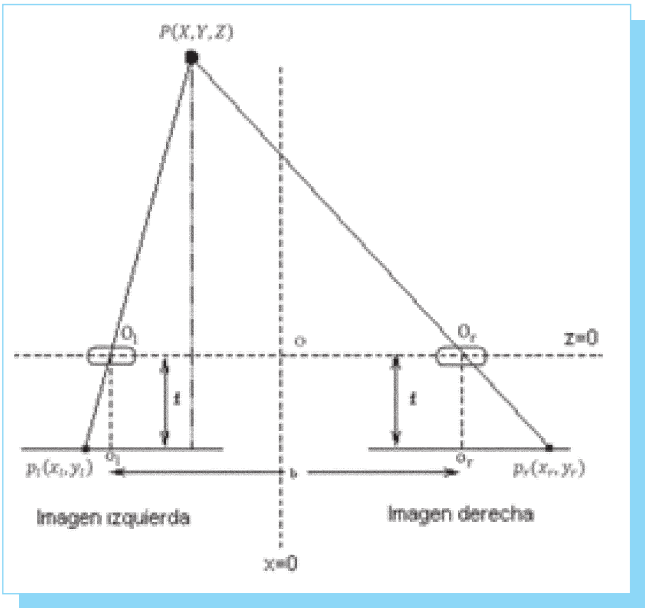

Fig. 2. Relación geométrica para obtener la profundidad Z a partir de la disparidad.

Debido a que la profundidad es inversamente proporcional a la disparidad, existe una relación no lineal entre estos dos términos. Cuando la disparidad es cercana a 0, pequeñas diferencias de disparidad producen grandes diferencias en profundidad. Cuando la disparidad es grande, pequeñas diferencias en disparidad casi no producen cambios en el valor de profundidad. La consecuencia es que los sistemas de visión estéreo tienen una alta resolución en profundidad solo para objetos relativamente cerca a la cámara.

Dado el mínimo incremento de disparidad permitido Δd, podemos determinar el valor de profundidad mínimo que se puede medir (resolución) ΔZ usando la fórmula:

$$\Delta Z = \frac{Z^2}{f_b}\,\Delta d \tag{4}$$

Las cámaras CCD no ideales pueden modelarse como sigue.

$$K = \begin{bmatrix} f_x & \dots & c_x \\ . & f_y & c_y \\ . & \dots & 1 \end{bmatrix} \tag{5}$$

Donde $f_x = f\,s_x$ y $f_y = f\,s_y$ representan la distancia focal medida en pixeles considerando los errores producidos al no tener pixeles cuadrados como dos factores de escala s_x y s_y. Los términos c_x y c_y consideran la falta de alineación entre el centro del sensor y el eje óptico de la cámara.

Estos parámetros son conocidos como los parámetros intrínsecos de la cámara, y son calculados durante el proceso de calibración de la misma.

Las cámaras reales (especialmente aquellas de bajo costo) también sufren de los efectos de la distorsión causada por las lentes, que se modelan como se muestra a continuación, la ecuación (6) muestra los coeficientes de distorsión radial.

$$x_{corregida} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$
$$y_{corregida} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (6)$$

Donde r es la distancia del pixel distorsionado con respecto al centro de la imagen y k_1, k_2 y k_3 se conocen como coeficientes de distorsión y también deben ser calculados. La distorsión tangencial está caracterizada por dos parámetros p_1 y p_2, tal que:

$$x_{corregida} = x + [2\, p_1 y + p_2(r^4 + 2x^2)]$$
$$y_{corregida} = y + [\, p_1(r^2 + 2y^2) + 2\, p_2 x] \quad (7)$$

Por lo tanto hay cinco parámetros de distorsion que deben ser calculados durante el proceso de calibración de la cámara, adicionalmente a los términos de la matriz de calibración. Para una mayor información acerca del modelado de cámaras no ideales véase [5].

B. Calibración y rectificación

La calibración de las cámaras del par estéreo es el proceso por medio del cual se determinan los parámetros intrínsecos de cada una de ellas, así como la posición en la que se encuentra una con respecto a la otra.

Para calibrar las cámaras suele utilizarse un patrón de puntos ordenados a una distancia regular, por ejemplo las esquinas de la cuadrícula de un tablero de ajedrez, este método se conoce como el método Tsai y es presentado en [6]. Una vez que se conocen los parámetros internos de las cámaras y la relación geométrica existente entre ellas, es posible llevar a cabo la rectificación de las imágenes. La rectificación de las imágenes consiste en proyectar las imágenes del par estéreo, de forma que los planos de las imágenes en cada cámara sean paralelos entre sí, y paralelos a la dirección en la cual existe el desplazamiento entre las imágenes. Luego de la rectificación las imágenes de la escena quedan en posición (fronto-paralela) como se muestra en la Fig. 3. Con esta configuración se simplifica la búsqueda de los puntos correspondientes pues se asegura que el correspondiente de un punto con coordenada vertical y_L en la imagen izquierda, se encuentra en la fila de coordenada $y_R = y_L$ de la imagen derecha, que se denomina *scanline*. La solución al problema de la rectificación de las imágenes de un par estéreo ha sido ampliamente tratado y existen soluciones en la literatura [7]. Lo importante es que siempre es posible hacer la rectificación de las imágenes llevándolas a la configuración de la Fig. 3.

C. Algoritmo de Block Matching

El algoritmo de búsqueda de correspondencias estéreo de Kurt Konolige, es un algoritmo rápido de búsqueda de bloques correspondientes (*Block Matching,* BM), que calcula sus resultados en un solo paso usando ventanas que computan la semejanza por medio de suma de diferencias absolutas (SAD) entre pixeles en la imagen izquierda y pixeles en la imagen derecha, desplazando la ventana una cierta cantidad de pixeles sobre la imagen izquierda con respecto a la imagen derecha (desde un valor de disparidad mínimo a un valor máximo. Con el fin de mejorar la calidad y confiabilidad del mapa de disparidad, el algoritmo incluye etapas de pre filtrado y posfiltrado. Este algoritmo encuentra sólo puntos altamente relacionados (con alta textura) entre las dos imágenes. Por lo tanto, en una escena con alta textura como puede ocurrir al exterior en un bosque, cada pixel podría tener una profundidad estimada. En una escena con muy baja textura como en un salón interior, muy pocos puntos podrán registrar una profundidad.

D. Etapas del algoritmo de BM

Este algoritmo cumple con las siguientes etapas.

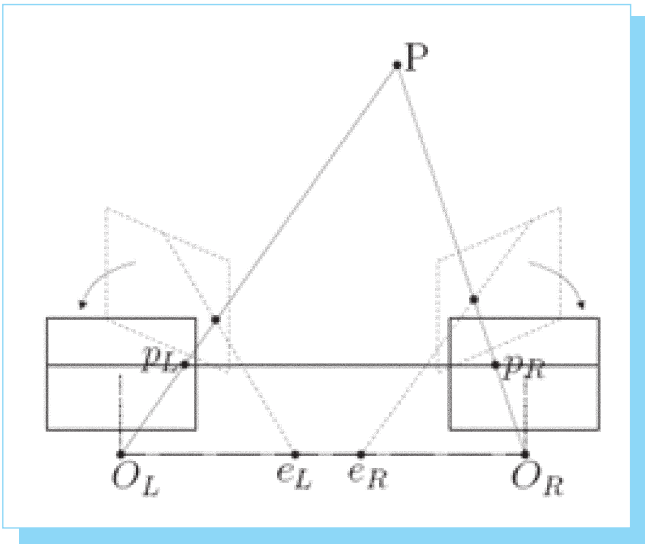

Fig. 3. Esquema del proceso de rectificación de imágenes de un par estéreo.

- *Etapa de prefiltrado*. Utiliza un operador local que transforma cada pixel en el par estéreo, normalizándolo basado en la intensidad promedio de los pixeles adyacentes.
- *Búsqueda de correspondientes utilizando suma de diferencias absolutas* (SAD).
- *Extracción del valor extremo*. Se determina el valor extremo de la correlación para cada pixel, esto genera el mapa de disparidad.
- *Etapa de posfiltrado*.

E. Etapa de prefiltrado

Esta etapa se lleva a cabo para disminuir los efectos de la diferencia en iluminación y perspectiva entre las imágenes del par estéreo. Los métodos de correlación usualmente tratan de compensarlo al realizar la correlación no sobre las imágenes originales sino sobre alguna imagen transformada. En este caso se usa el método de normalización de intensidades.

Normalización de intensidades. Cada una de las intensidades en un área correlacionada es normalizada por el promedio de las intensidades en el área. Es decir el pixel central I_c de la ventana es remplazado por:

$$\text{mín}\,(\text{máx}(I_c - \bar{I}, -I_{cap}), I_{cap}) \tag{8}$$

donde $\bar{I}$ es el valor de intensidad promedio en la ventana e I_{cap} es un límite numérico positivo.

F. Etapa de posfiltrado

En esta etapa, el algoritmo emplea un operador de interés para eliminar las zonas con poca textura y de un chequeo de consistencia izquierda/derecha para eliminar los errores causados en zonas ocluidas en alguna de las dos imágenes. Aunque el operador de interés requiere un umbral [8] muestra que es simple fijar su valor basado en el nivel de ruido presente en la entrada de video. Poniendo un área gris lisa enfrente de las cámaras produce un nivel de interés referido únicamente al ruido en el video, el umbral es definido un poco arriba de dicho nivel.

En la práctica, la combinación de las dos técnicas mencionadas han probado ser altamente efectivas eliminando malos correspondientes. El operador empleado es descrito en [2].

La variación horizontal promedio de una ventana centrada en (y,x) está dada por:

$$\sigma^2_h = \frac{1}{4W^2}\sum_{j=-W}^{W}\sum_{i=-W}^{W}[I(y+j,\, x+i) - \bar{I}(y+j,\, x)]^2 \tag{9}$$

donde y denota una *scanline* particular de la imagen y:

$$\bar{I}(y,x) = \frac{1}{2W}\sum_{j=-W}^{W} I(y,\, x+i) \tag{10}$$

denota el valor medio de la j-ésima *scanline* de la ventana centrada en (y,x).

Chequeo de consistencia. Aun después de aplicar el operador de interés para zonas con baja textura, se siguen teniendo errores en porciones de la imagen con discontinuidades en el valor de disparidad (bordes de objetos) debido a que frecuentemente una parte de la escena puede ser visible únicamente en una de las imágenes. Esto resulta en que los algoritmos encuentran correspondientes de forma aleatoria o espuria, debido a que no existe un correspondiente correcto.

La aplicación de un chequeo izquierdo/derecho puede eliminar estos errores. Este chequeo puede ser implementado eficientemente almacenando suficiente información al momento de hacer la correlación de disparidad original.

Chequeo de unicidad. Finalmente se realiza un chequeo que filtra todos aquellos pixeles que no arrojan valores de la función de costo lo suficientemente diferenciados como para ser aceptados como un resultado confiable, para ello se fija un parámetro *uniquenessRatio* que define el umbral para el valor de la función de costo $match_{val}$, definido por:

$$uniquenessRatio > (match_{val} - \text{mín}\,[match])/\,\text{mín}\,[match] \tag{11}$$

G. Refinamiento a nivel subpixel

Finalmente el algoritmo realiza un refinamiento de los valores de disparidad usando interpolación, con el propósito de mejorar la resolución del sistema al momento de estimar los valores de profundidad. Después del refinamiento se obtienen medidas de disparidad con una resolución de 1/16 pixeles [9].

H. Implementación y simulación

La plataforma en la cual se ha probado este algoritmo utiliza un procesador a 2.4 GHz, usando la cámara estéreo BUMBLE-

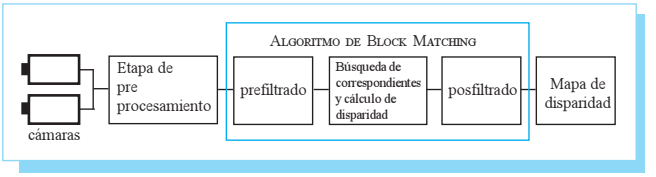

Fig. 4. Esquema de bloques del sistema creado.

BEE® de la empresa Point Gray Research® con una conexión de video IEEE 1394. El algoritmo implementado realiza la rectificación de las imágenes y el cálculo del mapa denso de disparidad (véase figura 4).

I. Etapa de preprocesamiento

Esta etapa consiste en todas las transformaciones previas, que deben hacerse antes de aplicar el algoritmo estéreo.

La cámara utilizada, envía las dos imágenes del par estéreo como una sola imagen entrelazada byte a byte con lo cual se consigue que las imágenes se reciban sincronizadas (es decir que se tomen en el mismo instante de tiempo, véase figura 5).

El primer paso del sistema de preprocesamiento consiste en separar las dos imágenes izquierda y derecha. La imagen se adquiere mediante un sensor con un filtro Bayer, lo que reduce la cantidad de información que se tiene que transmitir de la cámara a la estación de trabajo, disminuyendo los tiempos de transmisión (que en este tipo de aplicaciones suelen ser importantes). La figura (6) muestra una de las imágenes tomadas por la cámara derecha, en el acercamiento puede observarse el patrón de puntos característico de una imagen tomada empleando un filtro de Bayer.

El patrón de Bayer [10], debe interpretarse para generar la información de color de la imagen (RGB) y posteriormente se transforman las imágenes de color a escala de grises, que es el formato utilizado durante el proceso de rectificación que corrige los efectos de la distorsión (radial y tangencial) y alinear las líneas de búsqueda con el eje x (véase figura 7).

La figura 8 muestra la precisión ΔZ con respecto a la profundidad Z (4) para el caso en el cual el mapa de disparidad es calculado sin hacer la estimación a nivel subpixel; los valores de b y f son característicos de la cámara y se determinan durante el proceso de calibración.

Fig. 5. Imagen original como es tomada por la cámara estéreo, observe que la imagen izquierda y derecha están entrelazadas.

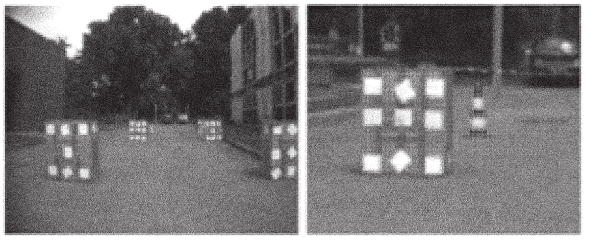

Fig. 6. Imagen izquierda separada, observe en la derecha el patrón de puntos representativo de una imagen tomada con un filtro Bayer.

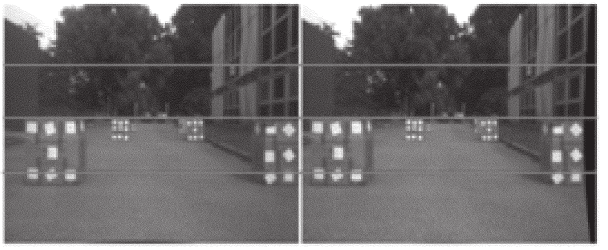

Fig. 7. Par de imágenes rectificadas y sin distorsión, observe que las líneas de búsqueda (líneas marcadas) son paralelas al eje x.

En la figura 8 se muestra, cómo es mejorada considerablemente la resolución en la estimación de profundidades, sobre todo para valores grandes de profundidad cuando se utiliza el refinamiento a nivel subpixel en el cálculo del mapa de disparidad.

Podemos observar que la estimación realizada será mejor para objetos de la escena que se encuentren cercanos a la cámara,

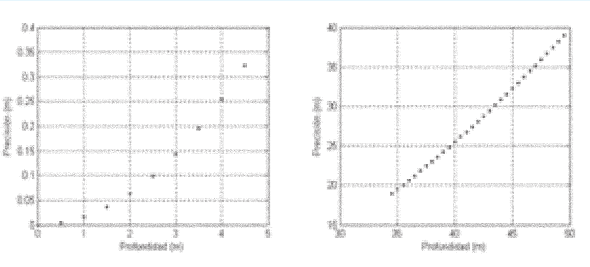

Fig. 8. Precisión en la estimación de profundidades para el sistema implementado sin utilizar refinamiento a nivel subpixel del mapa de disparidad. Izquierda, precisión a corta distancia. Derecha, precisión a larga distancia.

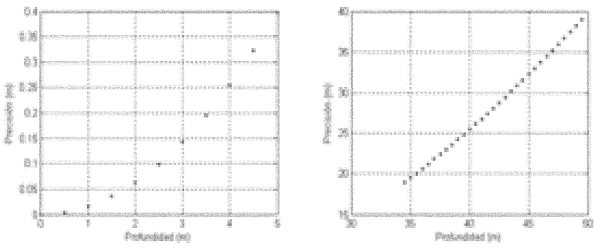

Fig. 9. Precisión en la estimación de profundidades para el sistema implementado utilizando refinamiento a nivel subpixel del mapa de disparidad. Izquierda, precisión a corta distancia. Derecha, precisión a larga distancia.

esto es debido a la relación no lineal existente entre la profundidad y la resolución.

J. Experimentos

El ajuste de los parámetros utilizados por el algoritmo de *block matching* es presentado en [11], la secuencia de imágenes es tomada desde un vehículo moviéndose a velocidad constante de 3.6 km/h. Dentro de un escenario de dimensiones conocidas, con objetos estáticos, de tal forma que se conoce la posición inicial del vehículo con respecto a cada uno de los objetos de la escena.

Para obtener las mediciones de profundidad, se seleccionan algunos puntos de la escena los cuales se siguen usando técnicas de flujo óptico. Con el propósito de mejorar el desempeño del algoritmo de flujo óptico y de este modo aislar sus

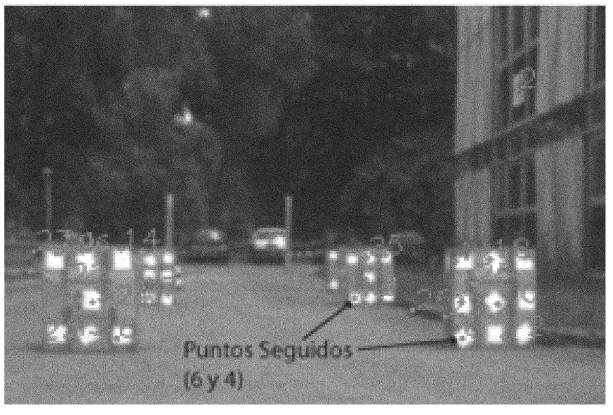

Fig. 10. Una de las imágenes de la secuencia, en ella pueden observarse los puntos que se siguen para calcular su profundidad.

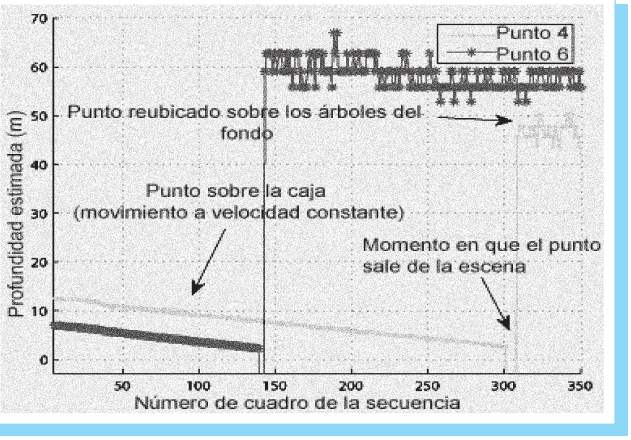

Fig. 11. Variación de la distancia del vehículo con respecto a los puntos objetivos, los cambios bruscos se presentan cuando los puntos se salen del campo visual de las cámaras.

efectos para poder analizar únicamente los efectos del algoritmo de cálculo de mapa de disparidad, se han colocado marcas (cuadros blancos) sobre los objetivos.

Cada uno de los cuadros es seguido durante toda la secuencia de imágenes y se calcula su profundidad en cada instante (cuadro de la secuencia) a partir del mapa de disparidad calculado.

En la figura 11 se muestran los resultados obtenidos para dos puntos de la imagen (puntos 4 y 6) el punto cuatro corresponde a la caja más próxima al vehículo del lado derecho, en tanto que el punto 4 corresponde a una caja situada 6 metros más atrás que la primera, también del lado derecho. El desplazamiento de 6 metros existente entre los dos objetivos puede verse en la Fig. 11, como la diferencia existente entre las dos rectas descendiendo con pendiente constante (cuadros 1 al 138). El error promedio al calcular la distancia entre cajas en este caso fue de 29 cm.

Otro experimento que se realizó consiste en comparar los resultados de profundidad obtenidos, con aquellos valores teóricos que deberían obtenerse considerando que el vehículo se mueve a una velocidad constante, conociendo su posición inicial con respecto a cada una de las cajas (véase figura 12). En este caso el error promedio fue de 11.3 cm, el hecho de que el error sea menor que el obtenido en el experimento anterior se debe a que en este caso sólo influye el error de la estimación para un objetivo, que además se trata de la caja más cercana situada inicialmente a 7 m del vehículo, en tanto que en el caso anterior influyen los errores de la estimación para el objetivo situado a 7 m y el siguiente, situado a 13 m del vehículo.

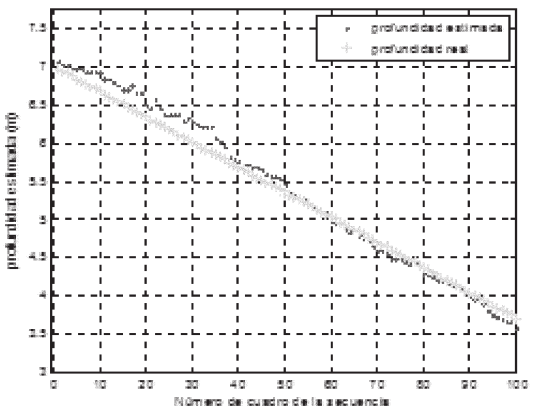

Fig. 12. Resultados de la estimación de profundidad con respecto a la disparidad ideal teórica que debería obtenerse al suponer una velocidad constante el vehículo.

5. Conclusiones

Los valores de profundidad 3D de una escena presentan una relación no lineal con respecto a los valores de disparidad calculados, como resultado los valores de profundidad que se calculan usando un sistema basado en visión estéreo como el que se muestra en este artículo, en general tendrán una mejor resolución para objetos cercanos a la cámara.

La estimación de la profundidad a nivel subpixel, mejora considerablemente la resolución de este sistema, haciendo posible su uso en aplicaciones tales como la robótica móvil, en el caso de robots que avancen a una velocidad de algunos metros por segundo.

Además de los errores inherentes al cálculo de disparidad, se deben considerar los errores existentes para seguir los puntos seleccionados a lo largo de toda la secuencia usando flujo óptico, así como los errores producidos en el control de la velocidad del vehículo debido a las irregularidades del suelo. Los errores promedio obtenidos son suficientemente buenos para vehículos desplazándose a bajas velocidades como en el caso presentado ya que nos permite estimar la posición de un objeto situado a 6 m de distancia con un error de tan sólo 11.3 cm.

Cuando se pretende implementar un sistema de búsqueda de correspondencias o cálculo de disparidad como parte del control de posición de un vehículo o robot móvil, es muy importante que la estimación de dicho algoritmo considere una etapa de refinamiento, para producir valores de disparidad a nivel subpixel.

Los resultados obtenidos ocupando el método de búsqueda de correspondientes mediante Block Matching presentado son alentadores para la utilización de sistemas estéreo de cálculo de disparidad, como herramienta para la estimación de distancias (profundidades) dentro de sistemas de control de navegación de robots móviles, o como en este caso, vehículos de baja velocidad, debido principalmente a que puede ser implementado de forma muy eficiente computacionalmente, siendo posible su utilización en tiempo real.

6. Referencias

[1] R. Shukla, H. Radha, and Martin Vetterli. 'Disparity dependent segmentation based stereo image coding'. *ICIP*, volume 1, pp.757-760, 2003.

[2] Matthies, L. 'Stereo vision for planetary rovers: stochastic modeling to near realtime implementation'. *IJCV* 8(1), (pp. 71-91), 1993.

[3] D. Marr and T. Poggio. 'A computational theory of human stereo vision'. *Proc R Soc Lond*, pp.301-328, Mayo 1979.

[4] Federico Lecumberry, 'Cálculo de disparidad en imágenes estéreo, una comparación', *III Workshop de Computación Gráfica, Imágenes y Visualización*.

[5] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge, UK: Cambridge University Press, 2006.

[6] R. Y. Tsai, 'A versatile camera calibration technique for high accuracy 3D machine vision metrology using off -the-shelf TV cameras and lenses', *IEEE Journal of Robotics and Automation* 3 (1987): 323-344.

[7] Olivier Faugeras, Quang-Tuan Luong y Theo Papadopoulo. The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications. MIT Press.

[8] K. Konolige, 'Small vision system: Hardware and implementation', *Proceedings of the International Symposium on Robotics Research* (pp. 111-116), Hayama, Japan, 1997.

[9] R. Szeliski and D. Scharstein, 'Symmetric Sub-Pixel Stereo Matching', *European Conference on Computer Vision*, 2002.

[10] A. Lukin and D. Kuvasov, 'High-Quality Algorithm for Bayer Pattern Interpolation', *Programming and Computer Software*, Vol. 30, No. 6, pp. 347-358, 2004.

[11] Jesús Arturo Escobedo Cabello, Volodymyr Ponomaryov, 'Evaluación y comparación del desempeño entre el algoritmo para el cálculo de Disparidad usando corte de Grafos propuesto por Vladimir Kolmogorov y el algoritmo usando Block Matching propuesto por Kurt Konolige', *SOMI-XXIV*.