Congreso Internacional en Ingeniería Electrónica. Mem. Electro 2015, Vol. 37, pp 187-192, Chihuahua, Chih. México
http://depi.itchihuahua.edu.mx/display/memorias_electro/MemoriaElectro2015
ISSN 1405-2172

COMPARACIÓN DE MEDICIONES DE PROFUNDIDAD ENTRE UN SISTEMA DE VISIÓN EN ESTEREO Y UN SENSOR RGBD

Víctor Hugo Velasco, Elisabet González Juárez, Isidro Robledo Vega, Alberto Pacheco González y Rogelio Baray Arana
Instituto Tecnológico de Chihuahua
División de Estudios de Posgrado e Investigación
Ave. Tecnológico #2909, Col. 10 de Mayo, Chihuahua, México
Tel. +52(614)201-2078 Ext. 30
{ vhvelasco, egonzalezj, irobledo, apacheco, rbaray}@itchihuahua.edu.mx

RESUMEN.

En este artículo se presenta una comparación de las mediciones de profundidad obtenidas por un sensor RGBD contra las calculadas por medio de la reconstrucción tridimensional de una escena capturada usando un sistema de visión en estéreo conformado por dos cámaras de video CCD. En nuestros experimentos pudimos observar que se requiere llevar a cabo un complicado proceso de calibración para obtener medidas de profundidad confiables con el sistema de visión en estéreo, mientras que los sensores RGBD proporcionan medidas mas exactas sin necesidad de ser calibrados, pero su rango de trabajo es mucho más limitado.
Palabras Clave: visión en estéreo, sensores RGBD, reconstrucción tridimensional.

ABSTRACT.

In this paper we present a comparison of depth measures obtained using an RGBD sensor against those calculated after a 3D reconstruction of a scene captured using a stereo vision system composed of two CCD video cameras. We could observe in our experiments that a complicated calibration process is required to obtain reliable depth measures with the stereo vision system, while RGBD sensors provide more precise measures without the need of being calibrated, but its working range is much more limited.
Keywords: stereo vision, RGBD sensor, 3D reconstruction.

1. INTRODUCCIÓN

Las aplicaciones de visión por computadora desarrolladas por un gran número de ingenieros e investigadores tienen como meta imitar la visión humana. La capacidad innata del ser humano de apreciar e identificar la diferencia de distancia que existe entre él y los objetos que lo rodean ha sido la inspiración para crear los sistemas de visión en estéreo [1]. Hace años era muy complicada su implantación ya que los sensores eran muy limitados y costosos. Lo mismo ocurría con los equipos de cómputo utilizados para el desarrollo de los algoritmos de captura y procesamiento de las imágenes. Aún así, la importancia de las aplicaciones potenciales de esta tecnología fue impulsando el desarrollo de métodos de captura más eficientes y cada vez menos costosos. En la actualidad, los sistemas de visión con múltiples cámaras son usados comúnmente para capturar escenas y realizar reconstrucciones tridimensionales en computadoras. Aunque el tiempo de procesamiento y el costo computacional sigue siendo una limitante.

Por otro lado tenemos los sensores RGBD, son un tipo de sensores que, tras tener pocos años en el mercado, se han convertido en una alternativa económica, que brinda información de profundidad sin necesidad de realizar mucho procesamiento. El interés de la industria de los videojuegos en interfaces naturales basadas en los movimientos de los jugadores para sustituir los controles manuales tradicionales impulsó el desarrollo de dispositivos para el control de los videojuegos mas conscientes del entorno que denominaron sensores de movimiento, con los que básicamente se detecta a un jugador y se reconoce gestos predefinidos realizados con sus extremidades para el desarrollo de las interfaces de control de los videojuegos. La compañía PrimeSense fue contratada por Microsoft para el desarrollo del sensor Kinect para la consola Xbox 360 y del software para su manejo. Después Microsoft puso disponible la tecnología de forma abierta para que los desarrolladores de videojuegos pudieran trabajar con ellos y crear nuevas aplicaciones. Lo interesante de esto es que su bajo

costo y disponibilidad ha permitido que la comunidad dedicada a hacer investigación en visón por computadora haya adoptado este tipo de sensores como uno de los métodos de captura más frecuentemente utilizado. Aparte del sensor Kinect de Microsoft, la tecnología desarrollada por PrimeSense ha sido implementada en otros sensores como el sensor Xtion Pro Live de la compañía Asus y los sensores Carmine y Capri de PrimeSense, aunque es más complicado obtenerlos.

El experimento realizado compara los procedimientos para obtener las medidas de profundidad usando un sistema de visión en estéreo y un sensor RGBD, los detalles técnicos se exponen en las siguientes secciones así como los resultados obtenidos. Nuestra intensión es mostrar diferentes procedimientos para obtener medidas de profundidad, el primero por medio de un sistema de visión en estéreo que es mucho mas caro y complejo en su calibración y el segundo usando un sensor RGBD que es barato y simple de usar, pero que tiene un rango de trabajo demasiado limitado.

2. VISIÓN EN ESTÉREO

Dentro de la percepción visual existe un fenómeno en el cual una persona puede percibir la profundidad basada en las diferencias, en apariencia, de una escena vista con el ojo izquierdo respecto de la vista con el ojo derecho. A esta diferencia de localización retinal se le conoce como disparidad [2]. Con base en este principio se han desarrollado los sistemas de visión estéreo.

2.1. Componentes del Sistema de Visión en Estéreo

En la configuración canónica de un sistema de visión en estéreo se usan dos cámaras con características similares y se colocan una a un lado de la otra sobre una barra montada en un trípode apuntando en la misma dirección de modo que sus ejes ópticos sean paralelos. Con esta configuración se proporciona la habilidad de inferir información de la estructura y distancia de una escena 3D a partir de 2 imágenes tomadas desde diferentes puntos de vista. La disparidad se calcula como la diferencia de la localización de un punto en ambas imágenes y es inversamente proporcional a su profundidad, que puede ser calculada por triangulación [3]. En la figura 1 se muestra la configuración del sistema de visión en estéreo, se utilizaron dos cámaras Sony modelo XCD-X710CR alineadas y separadas una distancia *Tx*=10cms. Ambas cámaras cuentan con lentes ópticos de 8.5mm de longitud focal y con sensores CCD a color con formato Bayer de 1077x788 elementos de forma cuadrada y su tamaño físico es de 4.65µm por lado. Las cámaras tienen interfaz Firewire (IEEE 1394) y son conectadas a una estación de trabajo con Matlab para llevar a cabo el proceso de adquisición por medio del Toolbox de Adquisición de Imágenes, fijando la resolución en 1024x768 pixeles. La conversión de Bayer a RGB se realiza por software utilizando interpolación bilineal.

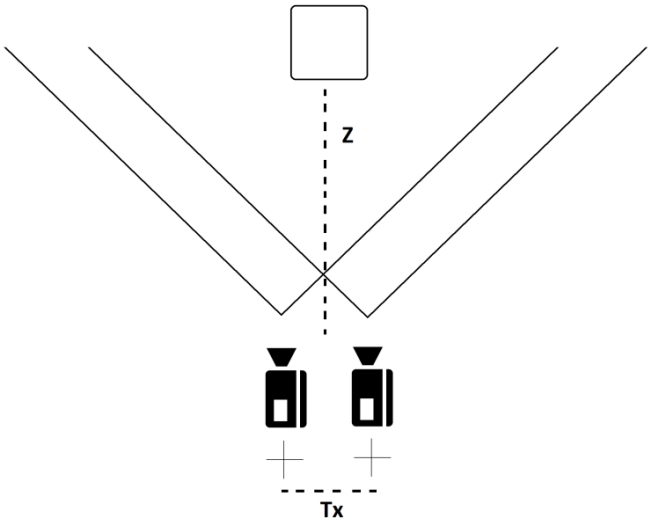

Figura 1.- Configuración del sistema de visión en estéreo.

2.2. Calibración

La calibración estéreo es el proceso en el cual se calcula la relación geométrica entre dos cámaras en el espacio, esto permite obtener los parámetros necesarios para adquirir la información en 3D de una escena. La calibración estéreo se basa encontrar la rotación de la matriz *R* y el vector de traslación *T* entre las dos cámaras, a estos se les conoce como

parámetros extrínsecos del sistema de visión en estéreo y se aprecian en la figura 2 [4].

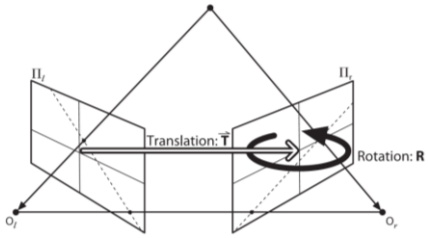

Figura 2. Calibración en estéreo.

Para calibrar el sistema de visión en estéreo se utilizó un procedimiento en base la las funciones de OpenCV [5]. Se utilizó un patrón de 9x6 cuadros blancos y negros de 2.5cms por lado y se adquirieron 15 pares de imágenes en estéreo donde se muestra el patrón de cuadros en diferentes posiciones y distancias con respecto a las cámaras del sistema. El patrón, parecido a un tablero de ajedrez, es útil ya que las esquinas son fáciles de detectar utilizando algoritmos de visión por computadora y su geometría es bastante simple.

En el siguiente paso del procedimiento de calibración se toma cada par de imágenes en estéreo y se realiza una detección de esquinas con el fin de segmentar el patrón de cuadros y utilizar estas esquinas como puntos de correspondencia entre las dos imágenes. El resultado de la detección de esquinas del patrón de cuadros sobre una par de imágenes en estéreo se muestra en la figura 3.

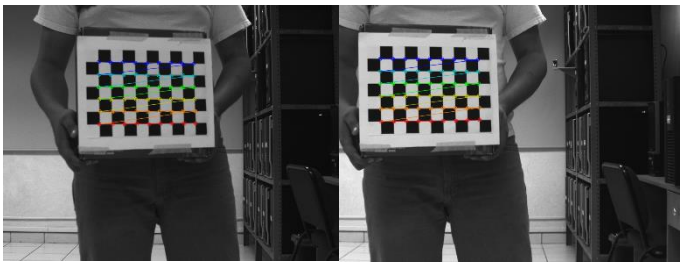

Figura 3. Detección de esquinas del patrón de cuadros.

Con los datos de correspondencia de las esquinas se realiza el proceso de rectificación de Bouguet [6] en el cual se realiza un deformación geométrica de las imágenes para que los puntos correspondientes queden alineados horizontalmente y facilitar el proceso de calibración.

2.3. Reconstrucción en 3D

Como resultado del proceso de calibración se obtiene la matriz de reproyección *Q*. Esta matriz contiene los parámetros intrínseco y extrínsecos del sistema y con ellos se puede llevar a cabo una reconstrucción tridimensional para obtener medidas de profundidad. La matriz *Q* tiene la siguiente forma:

$$Q = \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -1/T_x & (c_x - c_x')/T_x \end{bmatrix}$$

Donde c_x y c_y son las coordenadas del punto principal de la imagen izquierda, f es la longitud focal, T_x es la distancia de la línea base que existe entre las dos cámaras y c_x' la coordenada en x del punto principal de la imagen derecha. La matriz *Q* obtenida a partir de los pares de imágenes en estéreo adquiridos con nuestro sistema de visión en estéreo tiene los siguientes valores:

$$Q = \begin{bmatrix} 1 & 0 & 0 & -603.92 \\ 0 & 1 & 0 & -380.4 \\ 0 & 0 & 0 & 1303.08 \\ 0 & 0 & 0.1 & -13.4769 \end{bmatrix}$$

Con la matriz de reproyección *Q* es posible determinar la posición en 3D de un punto dado en la imagen izquierda y su punto correspondiente en la imagen derecha. Por lo tanto, dado un punto en coordenadas homogeneas en dos dimensiones *(x,y)* y su disparidad asociada *d*, se puede proyectar este punto hacia un espacio en 3D por medio de (1).

$$Q \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} (1)$$

Las coordenadas en 3D serán entonces $\left(\frac{X}{W}, \frac{Y}{W}, \frac{Z}{W}\right)$.

2.4. Cálculo de profundidades

Realizamos la adquisición de un nuevo par de imágenes en estéreo para probar la eficiencia de nuestro sistema de visión en estéreo. La escena capturada se muestra en la figura 4, donde se puede observar una mesa con objetos con superficies planas que nos permitirán obtener medidas reales de profundidad de manera controlada y así poder compararlas con las obtenidas utilizando la matriz de reproyección Q calculada con el algoritmo de Bouguet. Utilizando la matriz Q se puede realizar la rectificación de este par de imágenes.

Figura 4. Par de imágenes en estéreo para el cálculo de profundidades.

Tomando como referencia un punto específico de la imagen izquierda (x_l, y_l) y su punto correspondiente en la imagen derecha (x_r, y_r) en las imágenes rectificadas, se puede realizar la reconstrucción en 3D por medio de las siguientes ecuaciones:

$$d = x_l - x_r \quad (2)$$
$$X = x_l * Q[0,0] + Q[0,3] \quad (3)$$
$$Y = y_l * Q[1,1] + Q[1,3] \quad (4)$$
$$Z = Q[2,3] \quad (5)$$
$$W = d * Q[3,2] + Q[3,3] \quad (6)$$

Y así obtener la medida de profundidad de la coordenada Z en 3D calculando Z/W.
Se seleccionó el punto con coordenadas (291,584) de la imagen izquierda y su punto correspondiente en la imagen derecha con coordenadas (48,584), la coordenada en el eje y es la misma ya que las imágenes están rectificadas. Este punto se encuentra sobre la superficie de la caja más pequeña que se encuentra sobre la mesa. Entonces calculamos la profundidad de la siguiente forma:

$$x_l = 291 \quad x_r = 48$$
$$d = 291 - 48 = 243$$
$$W = (243 * 0.1) - 13.4769 = 10.8231$$
$$z = 1303.08$$
$$Z = \frac{1303.08}{10.8231} = 120.39$$

La medida real al punto seleccionado es de 160cms por lo que podemos observar un error relativo porcentual verdadero del 24.75%. De la misma forma se obtuvieron medidas de profundidad para otros dos puntos, uno sobre la caja del sensor Kinect y otro sobre el bloque de madera que se encuentra en la parte posterior de la mesa. Además se realizaron mediciones sobre los mismos tres puntos pero alejando la mesa del sistema de visión en estéreo a 2, 3 y 4 metros, en la figura 5 se pueden ver los pares de imágenes en estéreo adquiridas.

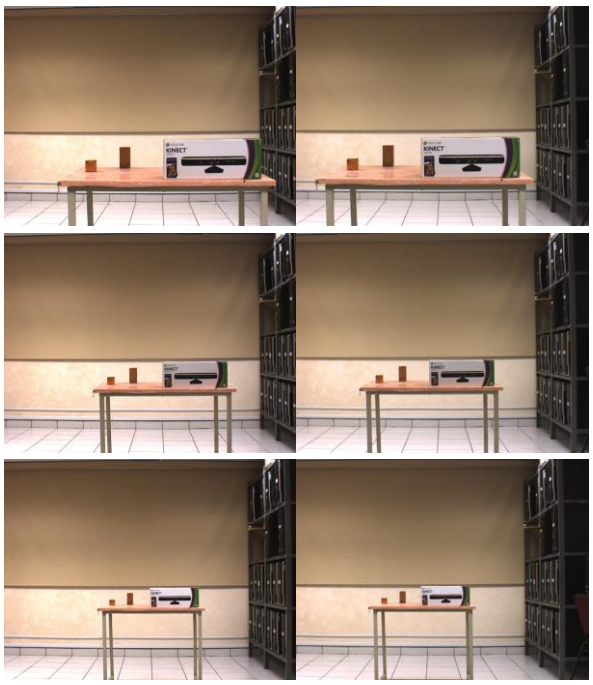
(2m)
(3m)
(4m)

Figura 5. Pares de imágenes en estéreo colocando la mesa con los objetos a 2, 3 y 4 metros del sistema de visión en estéreo.

En la Tabla 1 se muestran los resultados del experimento con las medidas calculadas por medio

de la reconstrucción tridimensional de los puntos y las medidas reales tomadas desde el sensor hasta las superficies de las cajas en cada diferente ubicación de la mesa y su correspondiente error relativo porcentual verdadero. Todas las medidas de profundidad están expresadas en metros. El objeto 1 se refiere a la caja del sensor Kinect, el objeto 2 se refiere a la pequeña caja de color naranja que se encuentra del lado izquierdo de la mesa y el objeto 3 se refiere al bloque de madera que se encuentra en el centro de los otros dos objetos.

Tabla 1. Resultados del experimento del cálculo de medidas de profundidad con el sistema de visión en estéreo.

Objeto	Profundidad Calculada	Profundidad Real	% Error
1.3 Metros			
1	0.9759	1.3	24.93
2	1.2039	1.6	24.75
3	1.3580	1.8	24.53
2 Metros			
1	1.5017	2.0	24.91
2	1.7298	2.3	24.79
3	1.8916	2.5	24.33
3 Metros			
1	2.3152	3.03	23.59
2	2.5465	3.33	23.52
3	2.7061	3.53	23.34
4 Metros			
1	3.1232	4.03	22.45
2	3.3406	4.33	22.84
3	3.4890	4.53	22.98

3. SENSORES RGBD

Los sensores RGBD son denominados de esta forma ya que proporcionan imágenes con datos del color (RGB) e imágenes de profundidad (D – *Depth* en inglés). En nuestras pruebas utilizamos el sensor Kinect de Microsoft, es una barra de 23 centímetros que cuenta con una cámara de video RGB, un proyector de rayos infrarrojos, una cámara sensible a los rayos infrarrojos, un arreglo de micrófonos y un procesador interno (figura 5) [7]. El sensor proporciona tres tipos de datos: color, profundidad e infrarrojo. La resolución de las imágenes es de 680x480 pixeles a 30 cuadros por segundo.

Figura 5. Sensor Kinect de Microsoft.

Con este sensor se adquirieron imágenes de la misma escena de la figura 4 y desde la misma distancia a los objetos que el sistema de visión en estéreo. La figura 6 muestra las imágenes adquiridas con los datos a color y de profundidad del sensor RGBD. Primero se seleccionó un punto sobre la misma superficie plana del experimento anterior, el dato proporcionado por el sensor nos da la medida de profundidad en milímetros y esta fue de 1600, comparada con los 1605mms de la medida real nos da un error relativo porcentual verdadero de 0.31%.

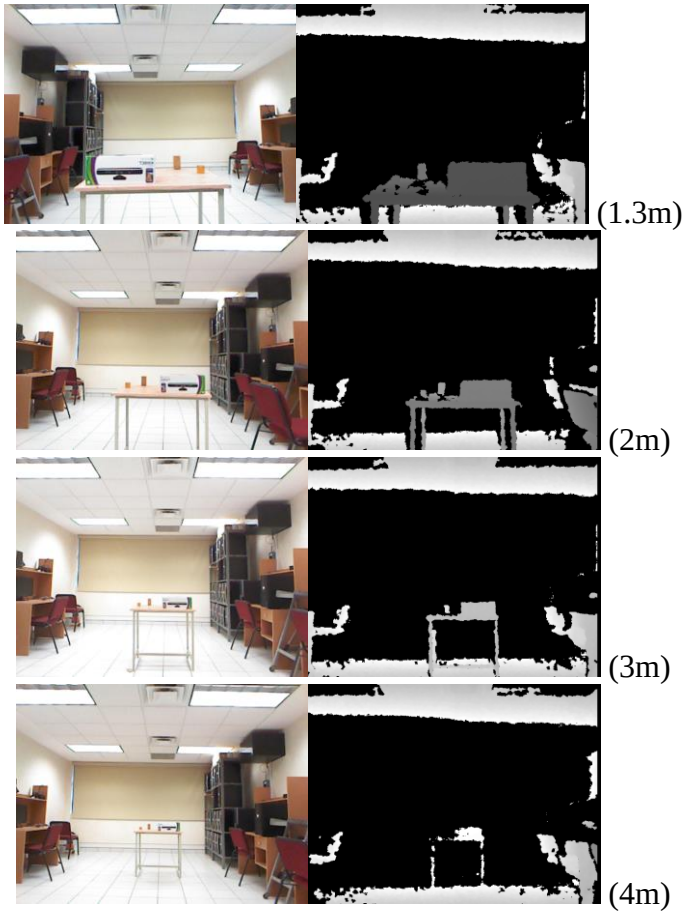

Figura 6.- Imágenes de los datos de color y profundidad del sensor RGBD.

Se adquirieron imágenes con la mesa a diferentes distancias al igual que en la sección anterior, se determinó la profundidad de los objetos y se realizó la comparación con respecto a las medidas reales como se muestra en la Tabla 2.

Tabla 2. Resultados del experimento de adquisición medidas de profundidad con el sensor RGBD.

Objeto	Profundidad Medida	Profundidad Real	% Error
1.3 Metros			
1	1.29	1.305	1.15
2	1.6	1.605	0.31
3	1.79	1.805	0.83
2 Metros			
1	1.952	2.0	2.4
2	2.289	2.3	0.47
3	2.487	2.5	0.52
3 Metros			
1	3.006	3.03	0.79
2	3.115	3.33	6.45
3	3.528	3.53	0.05
4 Metros			
1	3.975	4.03	1.36
2	0	4.33	100
3	0	4.53	100

4. OBSERVACIONES Y CONCLUSIONES

Al realizar las mediciones de profundidad con el sistema de visión en estéreo pudimos observar que el error es muy constante, por lo que nos dimos cuenta que utilizando un factor de escala por el cual multiplicamos todos nuestros resultados nos permite reducir el error de las mediciones a un promedio de 0.26%. Este ajuste de escala es necesario debido al proceso de calibración en estéreo y la determinación de los parámetros del sistema. Conociendo la dimensión real de un solo punto se pueden escalar todos los demás. Con respecto a las mediciones de profundidad obtenidas por medio del sensor RGBD observamos que si se trabaja dentro de las especificaciones de espacio definidas por el fabricante se pueden obtener resultados confiables, aunque la resolución del sensor es muy baja, además para obtener algunas de las mediciones se tuvo que determinar el valor a partir del análisis de varios puntos sobre la superficie del objeto ya que no se obtiene una medida válida en todos ellos. También observamos que cuando los objetos están fuera del rango de trabajo determinado, ya no son detectados por el sensor, como se aprecia en la imagen de profundidad de la última fila de la figura 6 donde los objetos 1 y 2, que son los más pequeños, ya no aparecen. De los experimentos llevados a cabo podemos concluir que se pueden obtener mediciones de profundidad confiables con el sistema de visión pero se debe de llevar a cabo un cuidadoso proceso de calibración y reconstrucción tridimensional. Se recomienda el uso del sensor RGBD para distancias no mayores a 3 metros y donde no se requiera una medición en puntos muy específicos, sino que pueda determinarse la profundidad en base a pequeñas superficies. Estamos planeando realizar estas mismas pruebas con la nueva versión del sensor Kinect de Microsoft que tiene una mejor resolución y un rango de trabajo más amplio.

5. REFERENCIAS

[1] Szeliski, R., *Computer Vision: Algorithms and Applications*, 1st ed., Springer-Verlag, 2011.

[2] Trucco, E. and Verri, A., *Introductory Techniques for 3D Computer Vision*, Prentice Hall, 1998.

[3] Forsyth, D. and Ponce, J., *Computer Vision: A Modern Approach*, 1st ed., Prentice Hall, 2002.

[4] Bradski, G. and Kaehler, A., *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed., O'Reilly Media, Inc., 2008.

[5] "OpenCV 2.4.11.0 documentation", OpenCV [en línea], disponible: http://docs.opencv.org/index.html, visitado: Junio 2015.

[6] J.Y.Bouguet. *MATLAB calibration tool*, [en línea], disponible: http://www.vision.caltech.edu/bouguetj/calib_doc/, visitado: Junio 2015.

[7] "Kinect Sensor", Microsoft MSDN Library [en línea], disp: https://msdn.microsoft.com/en-us/library/hh438998.aspx , visitado: Junio, 2015.