

Implementando matlab con visión estereoscópica para detección de distancia de una persona

Rodrigo Hinojosa Cruz
Universidad Autónoma del Estado de México

Resumen—La visión estereoscópica permite obtener una perspectiva 3D del mundo, mediante un procesamiento de dos imágenes con diferente perspectiva (generalmente izquierda y derecha). De lo anterior, ocurre algo similar en los humanos, ya que también contamos con esa capacidad, y en nuestro caso, es el cerebro el encargado de realizar esa mezcla de imágenes para poder obtener una sola imagen que integre la profundidad de los elementos que componen una escena tridimensional. Desde esa perspectiva, en la actualidad la visión artificial hace posible reconocer personas dentro de imágenes, y si a ello se le suma un uso adecuado de visión estereoscópica, se pueden detectar distancias de personas dentro de un vídeo. Es por ello que, en este proyecto se describe la metodología que se utiliza para realizar la detección de una persona dentro de un espacio 3D mediante el uso de las herramientas de visión estéreo de Matlab.

I. INTRODUCCIÓN

LA visión humana muestra el mundo de una forma diferente a la de otros seres vivos. En ese sentido, en los humanos cada ojo proporciona una visión 2D perspectiva del mundo que nos rodea, es decir, cada ojo percibe imágenes diferentes individuales. Pero eso no es lo que nos hace diferentes, lo que verdaderamente nos permite obtener la percepción del mundo que tenemos es la combinación de estas dos perspectivas, en cuyo caso, el cerebro es el encargado de hacer una mezcla entre esas dos perspectivas para formar un par estereoscópico que va a contener información acerca de las tres dimensiones (Percepción 2D más una profundidad Z) para así poder observar el mundo de la forma en que lo hacemos normalmente.

De lo anterior, el uso de visión artificial (VA) permite replicar esta forma de percepción al llevar a cabo un procesamiento de imágenes con el fin de detectar objetos, formas, personas y colores dentro de un espacio de trabajo, conocido también como campo de visión, el cual depende directamente de las propiedades de cámara que se use en los diferentes proyectos así como la ubicación desde donde se realiza la captura de la imagen.

Es por ello que actualmente el uso de VA nos permite incluso obtener la profundidad de diferentes objetivos ya sea implementando un sistema de visión estereoscópica u otros métodos como son los sistemas embebidos. Matlab es un software que desde el punto de vista de control, permite modelar y analizar sistemas, y adicional a esto, también permite el procesamiento de imágenes para usar VA.

Con respecto a lo anterior, este proyecto se enfoca en implementar matlab con visión estereoscópica para detectar la distancia a la que se encuentra una persona dentro de un espacio tridimensional.

Se dice que los seres humanos tenemos una percepción única del mundo, lo que en realidad es permitido por la visión binocular, con la que se mezcla en una sola percepción las sensaciones recibidas por cada una de las retinas [1].

I-A. Condiciones de la visión binocular

En caso de que se quiera obtener una visión binocular es necesario que:

- Los campos visuales monoculars se superpongan en todas las direcciones de mirada.
- Los campos de fijación individuales se superpongan debido a los movimientos coordinados de los dos ojos.
- La transmisión neuronal desde los dos ojos alcance la misma área cerebral para que se produzca la fusión de las imágenes y sé de la coordinación de la percepción.

I-B. Campo visual común

Para que este tipo de visión sea posible, las áreas en que están las dos órbitas y la estructura de los ojos deben de estar dispuestos para que se sobrepongan. En teoría, para que eso se cumpla, los ojos deben de divergir en 45° ver Figura 1. Sin embargo, en la realidad estos están situados de tal forma que sus ejes son paralelos.

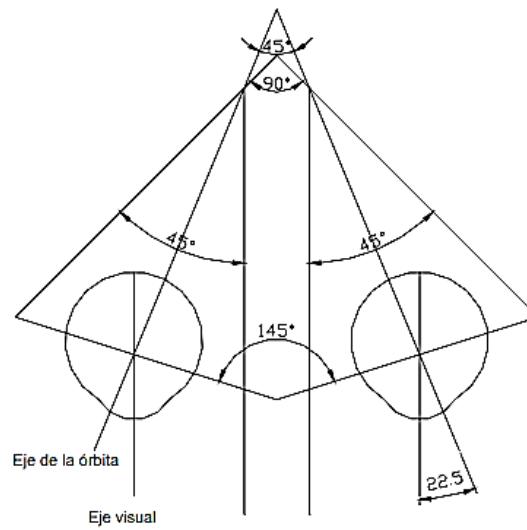


Figura 1. Esquema de las órbitas, orientación y relación entre sus ejes y sus paredes y los ejes visuales [1].

I-C. Disparidad binocular

La percepción de la profundidad en Imágenes Generadas por Ordenador (IGO) está considerablemente mejorada por la visualización estereoscópica. Así, los indicios asociados a la estereoscópica son más eficaces para evaluar la profundidad que el uso único de indicios monoculares. La proyección de IGO sobre la pantalla debe simular los indicios requeridos [2].

De lo anterior, la disparidad binocular se puede observar de dos formas: Cruzada y Paralela, se va a percibir una distancia específica para el uso que se quiera dar (Figura 2). Para poder simular un ambiente 3D que se aproxime al espacio en el mundo real, es necesario proporcionar tantas señales de profundidad como sea posible, o al menos las más significativas.

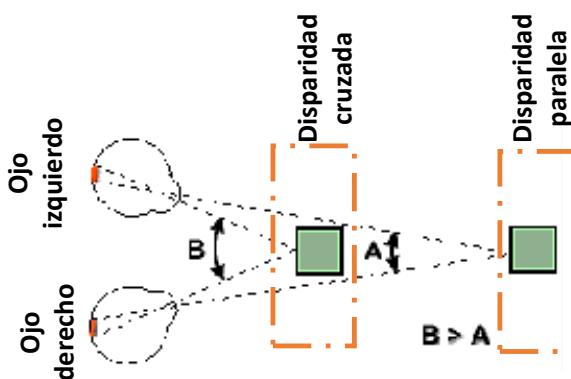


Figura 2. Disparidad binocular cruzada y paralela (editado de [3]).

I-D. Parámetros intrínsecos y extrínsecos de una cámara

Los patrones de calibración estiman los parámetros de lentes, el sensor de imagen para cámara o una videocámara. Y nosotros podemos usar estos parámetros para estimar las estructuras de una escena y remover la distorsión de los lentes [4]. Los parámetros de una cámara incluyen:

- **Intrínsecos.**- Estos están relacionados a las características internas de la cámara, como la distancia focal, el centro óptico (conocido como punto focal) y el coeficiente de inclinación.
- **Extrínsecos.**- Dentro de estos se describe la localización (posición y orientación) de la cámara en la escena 3D.

Para estimar los parámetros mencionados antes, se necesitan una nube de puntos de la escena en 3D y su correspondientes puntos de la imagen 2D. Estos se pueden obtener mediante el uso de múltiples imágenes de un patrón de calibración que puede ser representado en tablero o una cuadrícula circular [5].

II. MATERIALES Y METODOS

Usando dos cámaras (web logitech c920) y de las aplicaciones que nos brinda matlab para implementar visión

stereo, es posible reconstruir un escenario 3D mediante un procesamiento de vídeo estereoscópico y de esa forma detectar la distancia a la que se encuentra una persona dentro del campo de visión de los instrumentos de captura de imagen mencionados antes.

III. PROCEDIMIENTO

III-A. Capturas de imágenes stereo

Un sistema stereo consiste en dos cámaras con las mismas propiedades: cámara 1 y cámara 2, las cuales pueden estar separadas a cierta distancia pero la relación entre sus ejes debe ser paralela o cruzada para que puedan tener un campo de visión monocular cada una con la posibilidad de converger posteriormente, ver Figura 3.



Figura 3. Posición frontal de las cámaras Logitech Hd Pro C920

Se deben capturar mínimo 20 imágenes en cada una de las cámaras con perspectiva izquierda y derecha para poder proceder a la calibración stereo. La calibración stereo permite iniciar la calibración con solo 2 pares de imágenes pero en el proceso se eliminaran algunas y por ello es necesario tener un número entre 10 y 20 pares. En esta práctica se generaron 30 pares de imágenes (Figura 4).



Figura 4. Captura de imágenes izquierda y derecha con tablero de calibración en diferentes posiciones.

III-B. Calibración Setero Camera

En este paso se obtienen los parámetros intrínsecos y extrínsecos de las cámaras que se usan para realizar la captura de vídeo (Matriz "stereoParams.m"). Por lo que, de esa forma, en la Figura 5 se muestra el proceso que se sigue para ingresar las imágenes capturadas a la aplicación "Stereo Camera Calibrator App", en la cual se procesaran las imágenes estereoscópicas.



Figura 5. Proceso para el uso de la aplicación "Stereo Camera Calibrator" de matlab para obtener parámetros stereo.

De lo anterior, para el proyecto se cargaron 30 pares de imágenes, de las cuales, fueron descartados 7 pares. Una vez cargadas las imágenes y tras eliminar las que representaban un error de proyección muy alto, se generaron los parámetros stereo con un error absoluto de reprojeción medio de 0.24px, ver Figura 6.

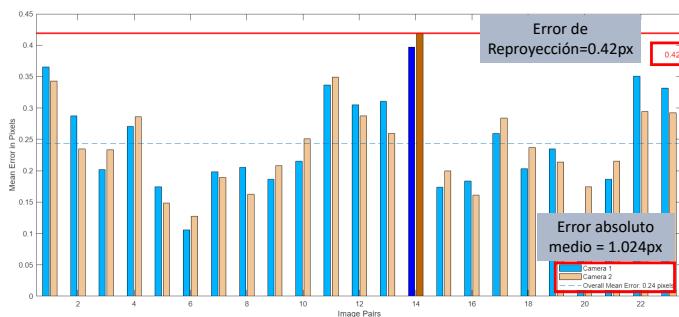


Figura 6. Errores de reproyección y mean error.

De lo anterior, en la Figura 7 se observan la posición de las cámaras, las distancias entre estas y del tablero, así como también las posiciones que se realizaron para la caracterización de rotaciones e inclinaciones. Así entonces, es como se obtienen los parámetros estéreo.

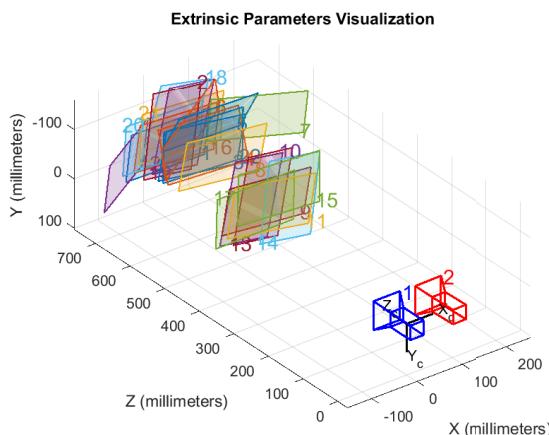


Figura 7. Visualización de los parámetros extrínsecos de las cámaras.

III-C. Grabación de video L&R

Una vez obtenidos los parámetros stereo, y sin cambiar de posición las cámaras, se debe grabar un vídeo para cada perspectiva, es decir, un vídeo para la cámara izquierda y un vídeo para la derecha.

Matlab no tiene como tal una aplicación o un script para captura de vídeo en proyectos de visión stereo, por lo que, para este proyecto se creó un script para grabar el vídeo a 25 frames por segundo, este parámetro corresponde directamente al modelo de cámara que se implemento en este proyecto y a las características de procesamiento de la computadora usada.

Por lo anterior, los requerimientos para obtener un procesamiento óptimo de una captura de vídeo por cámara fija, son: zona de visión, parámetros de visión y condiciones del ambiente. De esa forma, en este proyecto únicamente se mencionan los requerimientos a tomar en cuenta para los parámetros de vídeo.

Aunado a lo anterior, para obtener un mejor procesamiento de vídeo, hay tres factores que se deben tomar en cuenta:

- La *resolución* de una cámara tiene una alta influencia en la calidad de la grabación, así como también del procesamiento de visión por computadora. En ese sentido, mientras más datos sean proporcionados al sistema, mejor será el reconocimiento de objetos en este. Resoluciones de 1280px x 720px, 1920px x 1080px (FULL HD), o superiores garantizarán una mejor adquisición. Es por ello que, para este caso se reviso cual era la mejor resolución a la que nos permitían grabar las cámaras, con un valor de 1280px x 720px y se configuro la captura de vídeo con esta.
- Los *cuadros por segundo o frame rate (FPS)*, son una característica para definir la fluidez del movimiento, es por ello que, el FPS ideal para una buena visión es entre 25 y 30 fotogramas por segundo.
- Y por ultimo, la *velocidad de obturación* afecta la claridad de los contornos del objeto en movimiento, algunas cámaras cambian esta velocidad dependiendo de la iluminación con la finalidad de mantener el mismo brillo.

III-D. Procesamiento de vídeo para identificar distancia de personas

Una vez grabados los vídeos izquierdo y derecho, se debe proceder a incluir los archivos necesarios (StereoParams.m, Video_Left y Video_Right) al programa de captura de distancia. Es importante mencionar que para evitar un error en la comparación de parámetros es necesario verificar que la resolución usada al principio para la captura de imágenes izquierda y derecha con el tablero debe ser la misma que se usa en la captura de vídeo.

Para terminar, y una vez ejecutado el script de matlab mencionado antes, de acuerdo al proceso mostrado en la Figura 8, se irán abriendo ventanas emergentes mostrando los filtros que se van aplicando a los vídeos y se realizará la convergencia de las dos percepciones para poder calcular la distancia a la que se encuentra la persona en el vídeo.



Figura 8. Bloques para el procesamiento de vídeo.

De acuerdo a lo anterior, el primer paso es la rectificación de los fotogramas de las cámaras izquierda y derecha, realizado con el fin de calcular la disparidad y reconstruir la escena 3D. Las imágenes rectificadas se pueden combinar en un anaglifo y en ese caso, se puede apreciar al efecto 3D utilizando gafas estéreo rojo-cian.



Figura 9. Rectificación de los fotogramas.

Una vez realizada la rectificación, se hace un escalado a gris de las escenas para tratar de mejorar los datos que se obtienen 10. Continuando con el procesamiento, cualquier par de puntos correspondientes se ubican en la misma fila de píxeles. Lo anterior quiere decir que, para cada píxel de la imagen de la izquierda se calcula la distancia al píxel correspondiente de la imagen de la derecha. En ese sentido, esa distancia denominada disparidad, es proporcional a la distancia del punto del mundo desde la cámara, ver Figura 10.

Posteriormente, como se observa en la Figura 11, se realiza una reconstrucción del mundo tridimensional de acuerdo a los puntos 3D correspondientes a cada píxel del mapa de disparidad (nube de puntos).

Como último bloque del procesamiento, se reconoce o detecta a la persona en la imagen de la izquierda empleando la herramienta *vision.PeopleDetector* de matlab. Así como también, mediante la recolección de coordenadas tridimensionales del centroide de la persona detectada se calcula la distancia desde este hasta la cámara con una unidad de medida en metros. Por lo que, en términos generales, se determina la distancia de cada persona a la cámara en las diferentes escenas, ver Figura 12.

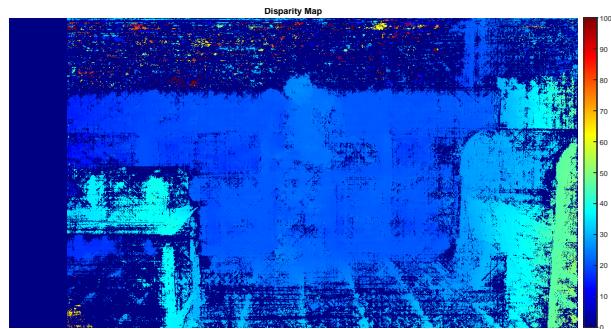


Figura 10. Obtención del mapa de disparidad.

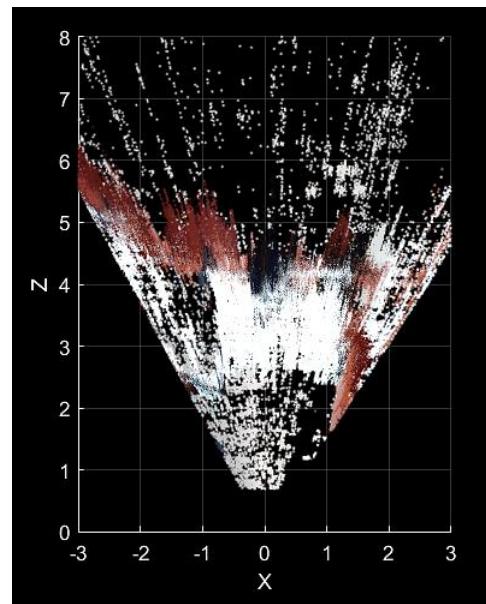


Figura 11. Nube de puntos para reconstrucción de escena 3D, vista desde los ejes X, Z.

IV. RESULTADOS

Se logró realizar la captura de los parámetros intrínsecos y extrínsecos de las cámaras logitech c920 usadas para esta práctica, con esos parámetros y con la captura de un vídeo estereoscópico, se detectó la distancia a la que se encuentra la persona dentro del espacio escaneado durante todo el vídeo analizado.

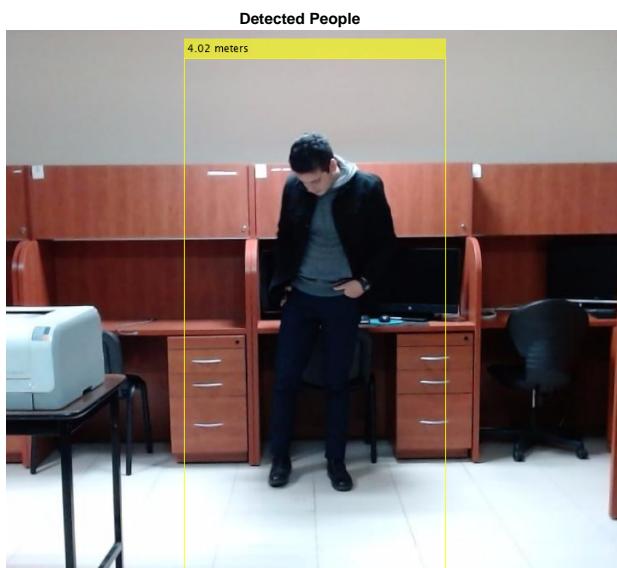


Figura 12. Detección de profundidad de una persona.

V. CONCLUSIONES

Al momento de realizar el procesamiento del vídeo, se logró detectar la distancia de una persona durante todos los cuadros generados en el vídeo. Por ello, el objetivo de utilizar matlab y sus herramientas de visión estereoscópica para la detección de distancia de una persona fue cumplido totalmente.

Durante la realización de pruebas con diferentes parámetros de visión, se encontró que si no se utilizan los adecuados para el tipo de cámara y las capacidades de procesamiento de vídeo es posible obtener distintos tipos de errores.

En suma, con el desarrollo de este proyecto se obtuvo un nuevo enfoque del uso de visión estereoscópica, ya que, integrando un sistema de procesamiento por computadora es posible analizar nubes de puntos de espacios 3D que nos brinda la posibilidad de detectar la profundidad a la que se encuentra alguno de los elementos que componen la escena.

REFERENCIAS

- [1] M. C. Puell Marín, “Óptica fisiológica: el sistema óptico del ojo y la visión binocular,” 2006.
- [2] M. Poyade and A. R. Lecuona, “Análisis de la disparidad binocular en la percepción de la profundidad en realidad virtual,” in *Actas del VIII Congreso Internacional de Interacción Persona Ordenador (INTERACCIÓN 2007)*, 2007, pp. 213–220.
- [3] C. Orona. (16/08/2011) Aplicaciones de gráficos de computadora. [Online]. Available: <http://graficacion08110935.blogspot.com/2011/08/actividad-1.html>
- [4] MathWorks. (2015) Using the stereo camera calibrator app. [Online]. Available: https://la.mathworks.com/help/vision/ug/using-the-stereo-camera-calibrator-app.html?searchHighlight=Stereo%20Calibration%20app%2C&s_tid=srchtitle_Stereo%20Calibration%20app%252C_2
- [5] ——. (2015) Camera calibration with matlab. [Online]. Available: <https://la.mathworks.com/videos/camera-calibration-with-matlab-81233.html>