

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE VISIÓN ESTEREOSCÓPICO PARA LA RECONSTRUCCIÓN TRIDIMENSIONAL DE MATERIALES

Cornejo Guerra Paulo Sergio^{1,2}, Traslosheros Michel Alberto², Chagolla Gaona Hernando¹, Castellanos Galindo Joaquín¹

¹Departamento de Ingeniería Eléctrica y Electrónica
Instituto Tecnológico de Querétaro

Av. Tecnológico S/N, Centro Histórico, Santiago de Querétaro, Querétaro, México, Tel: (442) 2274400
e- mail: paulo_guerra2000@yahoo.com, hchagollag@hotmail.com, jcastell_7700@yahoo.com.

²Universidad Aeronáutica en Querétaro

Carretera Estatal Querétaro-Tequisquiapan 22154, Colón. Querétaro, México, Tel: (442) 1016600
e-mail: alberto.traslosheros@unaq.edu.mx

RESUMEN.

Se presenta el desarrollo de un sistema de visión estereoscópica diseñado para generar, apartir de materiales u objetos, modelos tridimensionales que puedan ser empleados para su posterior maquinado o fabricación. El sistema utiliza un par de cámaras y un haz incidente de luz láser para escanear el objeto, produciendo dos imágenes correlacionadas. Cada par de imágenes se procesa digitalmente para determinar, mediante triangulación, una nube de puntos; la cual a la postre definirá el modelo tridimensional. La programación de reconstrucción y modelado fué desarrollada empleando OpenCV bajo Windows.

Para validar la precisión del modelo se empleó Solidworks, en donde se comparó el modelo tridimensional contra el objeto real.

Clave: visión estereoscópica, estéreo correspondencia, Rectificación estereoscópica, nube de puntos.

ABSTRACT.

In this work the development of a vision system, designed to generate stereoscopic three-dimensional models of materials or objects for subsequent machining or manufacture, is presented. The system uses a pair of cameras and an incident laser beam to scan the object to produce two correlated images. Each pair of images are digitally processed to determine by triangulation, a particular point cloud that eventually will define the three-dimensional model. The reconstruction and modeling programs were developed using OpenCV under Windows.

To validate the accuracy of the model, a comparison between the three-dimensional model and the real object was performed under SolidWorks.

Keywords: stereoscopic vision, stereo correspondence, stereoscopic rectification, cloud points.

1. INTRODUCCIÓN

La visión artificial es una de las disciplinas científicas y tecnológicas que más desarrollo ha tenido en las últimas décadas, siendo la reconstrucción tridimensional una de las actividades más relevantes en este campo, en gran parte debido al desarrollo de sistemas computacionales cada vez más poderosos y compactos. La robótica, reconocimiento de imágenes, topografía, arquitectura, realidad virtual, medicina

y escultura son algunas aplicaciones donde se han obtenido modelos tridimensionales (3D) cuya exactitud es igual a los objetos reales [1-3]. Uno de los dispositivos capaces de realizar la reconstrucción 3D recibe el nombre de escáner 3D.

La reconstrucción de los modelos se pueden realizar por contacto o sin contacto. Dentro de este último grupo destacan los basados en luz estructurada en los cuales un patrón de luz es proyectado sobre el objeto con la finalidad de analizar digitalmente la deformación generada debido a la geometría de la superficie en la que el patrón es reflejado. Con esta información es posible obtener, mediante el procesamiento computacional, un modelo 3D del objeto analizado.

Un escáner puede tener una o más cámaras dependiendo de la aplicación que se va a realizar, pero por lo general se emplea sólo un par, dando lugar a la técnica de reconstrucción estereoscópica [1], [4].

La visión estereoscópica es la forma en como los humanos vemos, a partir de nuestros ojos obtenemos imágenes del mundo que nos rodea y debido a su separación lo que vemos con el ojo derecho está ligeramente desplazado con respecto a lo que vemos con el izquierdo, esta diferencia se llama disparidad y nos permite determinar la profundidad o la distancia a la que se encuentran los cuerpos que nos rodean.

Un sistema de visión estereoscópico actúa en forma parecida [1], a partir de dos imágenes capturadas en el mismo instante por medio de dos cámaras separadas entre sí por una distancia dada es posible determinar la profundidad de los objetos que se encuentran en la escena. Es decir a partir de dos imágenes en dos dimensiones (2D) se obtiene una representación en 3D de la escena capturada. Este proceso se conoce como reconstrucción tridimensional 3D. Una de las técnicas fundamentales con los que se logra lo anterior se llama triangulación [5] este proceso permite determinar las coordenadas tridimensionales de un punto común en ambas imágenes a partir de la traslación (separación) entre las cámaras y su rotación. A estos dos últimos conceptos se les conoce

como parámetros intrínsecos y extrínsecos del sistema estereoscópico.

2. SISTEMA DE VISION ESTEREOSCOPICO DISEÑADO

Se propuso la implementación de un sistema de visión estereoscópica de bajo costo el cual se utilizará como herramienta didáctica y desarrollo. A continuación se describen los elementos que lo constituyen.

a). Par de cámaras Logitech c525 con una resolución de 1280x720 pixeles.

b). Apuntador laser (fuente de luz estructurada). Dispositivo montado sobre una base que lo mantiene fijo y en posición tal que el haz incide en un trocito de vidrio de forma cilíndrica con el que el haz cambia de dirección adquiriendo una forma vertical. La base se mueve en dirección paralela al soporte de las cámaras mediante un motor de pasos el cual es controlado por un microcontrolador de tal forma que su velocidad de movimiento ha sido calculada previamente y esté sincronizada con la captura de las imágenes. Se cuenta con un sistema optoelectrónico de control de disparo del haz para que el barrido se realice sobre el objeto a escanear.

c).- Escenario. Plataforma graduada donde es colocado el objeto a escanear de tal forma que su posición pueda ser determinada con facilidad para validación del modelo 3D obtenido.

d).- Computadora. Permite la ejecución del programa diseñado para la reconstrucción 3D del objeto o material de trabajo. En la figura 1 se muestra el sistema diseñado.

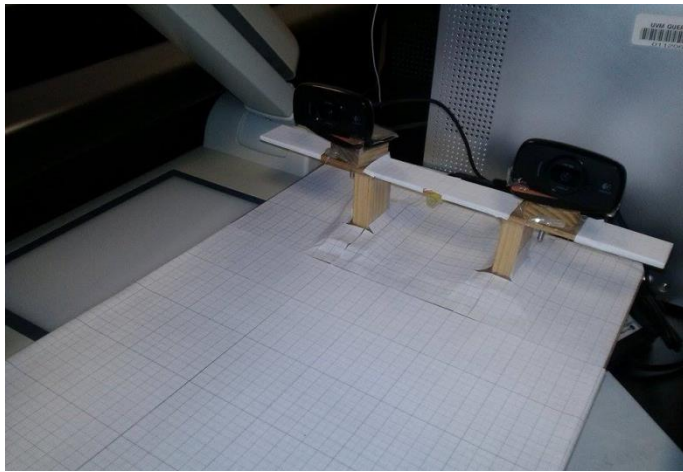


Figura 1. Disposición del par estéreo de cámaras del sistema de visión diseñado.

El programa desarrollado realiza las funciones descritas en el diagrama mostrado en la figura 2 las cuales serán explicadas por etapas.

2.1. Etapas del proceso de captura, procesamiento y reconstrucción 3D

Se muestra un diagrama de flujo (figura 2) en donde se describen las etapas que el programa diseñado realiza, este se desarrolló con OpenCV y C++, utilizando como base en su implementación Matlab básicamente con el toolbox calib [5].

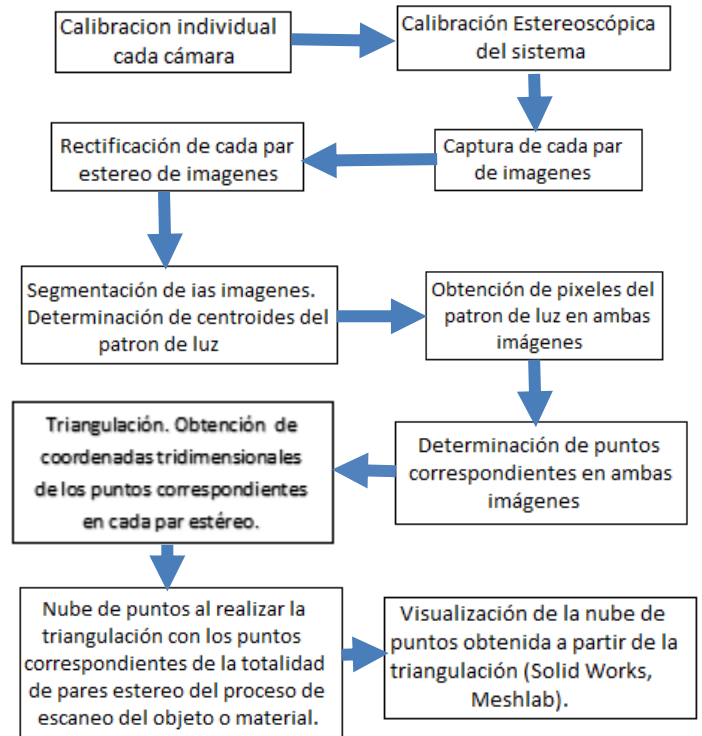


Figura 2. Diagrama general del programa diseñado para el sistema de visión estereoscópica de reconstrucción 3D.

A continuación se desarrollan cada uno de los bloques anteriores con mayor detalle.

2.2. Calibración de las cámaras

Las cámaras digitales son dispositivos que transforman la luz en impulsos eléctricos los cuales con el procesamiento adecuado se realiza la captura de los pares de imágenes del objeto o material que será reconstruido tridimensionalmente, es necesario conocer su funcionamiento y constitución física para obtener fotografías de buena calidad que permitan lograr un modelo 3D adecuado. La forma común de describir las cámaras se basa en el modelo *pinhole* en el cual el lente de la cámara actúa como un punto por el cual pasan los rayos de luz de la escena sin ser alterados en su dirección. En la figura 3 se muestra un esquema del modelo, en el cual un punto en el espacio 3D $A(X, Y, Z)$ corresponde a un punto del plano imagen de la cámara $O(x, y)$, es decir, se realiza una transformación lineal de $R^3 \rightarrow R^2$; del espacio 3D al plano 2D. Las cámaras tienen una matriz de sensores CCD (Charged Coupled Device)

de forma generalmente cuadrada los cuales convierten la luz incidente en señales eléctricas que al ser procesadas por un programa se obtienen las imágenes capturadas y que se representan en forma de pixeles como se muestra en la figura 4.

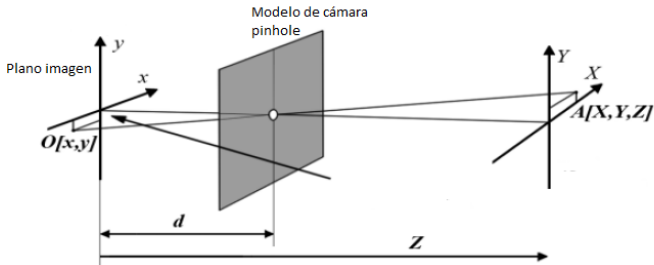


Figura 3. Modelo pinole para las cámaras utilizadas.

La información sobre los pixeles de una imagen dependen de las características internas de cada cámara y estas pueden variar sin importar si son del mismo tipo y marca.

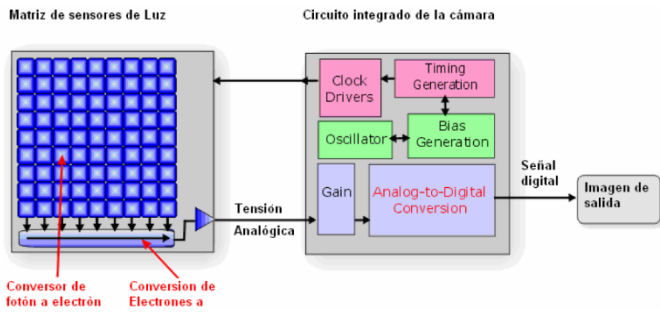


Figura 4. Esquema general de una cámara digital común con sensores CCD.

Esto se debe principalmente a que en su proceso de fabricación los lentes de la cámara generalmente presentan pequeñas imperfecciones que provocan distorsión (radial y tangencial) en las imágenes que no son deseables si se desea lograr fotografías de buena calidad. De igual forma es importante conocer las coordenadas del punto principal y la distancia focal d como se muestra en la figura 2 llamados *parámetros intrínsecos* así como la posición (translación) y orientación (rotación) de la cámara con respecto al objeto en el espacio real los cuales constituyen los *parámetros extrínsecos* de la cámara los cuales determinan por medio de su calibración. Para realizarla existen varias técnicas:

- Métodos basados en modelos 3D [Tsai 1987].
- Métodos basados en patrones 2D [Zhang 2000]. Es menos precisa que la anterior.
- Autocalibración [Harley 1994]. El planteamiento matemático es complejo.

Dada su facilidad se decidió utilizar el método de Zhang [6], para ello se capturó un número adecuado de imágenes (20 en total) de un patrón de calibración de tablero de ajedrez con

ambas cámaras en el mismo instante al cual se hará referencia como par estéreo de imágenes. Posteriormente por medio de OpenCV se realizó la calibración de cada cámara, en la figura 5 se muestra un par estéreo del patrón utilizado.

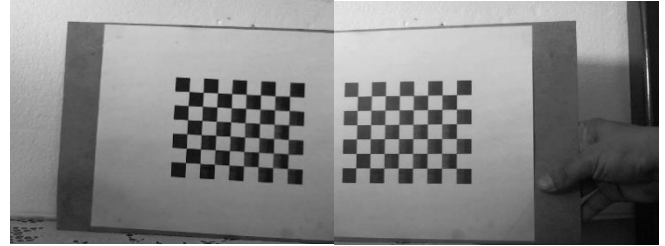


Figura 5 Ejemplo de par de imágenes estéreo capturadas para la calibración de las cámaras.

Es importante mencionar que la calibración se realizó también con Matlab obteniendo las siguientes conclusiones:

- OpenCV realiza la calibración más rápida.
- Matlab ofrece varias opciones de calibración de forma más interactiva con el usuario, el análisis es más simple y los resultados son visualizados gráficamente.
- En ambos casos los resultados fueron adecuados dado el valor de error obtenido (error ≈ 0.23).

Al finalizar se obtuvieron los parámetros intrínsecos (distancia focal, parámetros de distorsión y las coordenadas del punto principal), con los cuales se determina una matriz A 3×3 llamada matriz de parámetros intrínsecos ecuación (1) y los parámetros extrínsecos (vectores de rotación R y translación T) representados en una matriz compuesta en la ecuación (2).

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 0 \end{bmatrix} \quad (1)$$

$$[R \ T] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \quad (2)$$

En (1) f_x y f_y son las distancias focales con respecto a los eje x y y , c_x y c_y son las coordenadas del punto principal, los valores r_{ij} y t_i para $i = j = 1, 2, 3$ en (2) son los parámetros de rotación de las cámaras con respecto a los ejes X, Y, Z y las componentes de translación del origen del sistema de coordenadas del objeto en el espacio al sistema de coordenadas de la cámara.

Con esta información la relación de un punto (x, y, z) del mundo real (3D) con el punto correspondiente en cada imagen

(X, Y) (2D) se representa por medio de la transformación lineal representada en la ecuación 3.

$$\lambda \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = A [R \ T] \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (3)$$

2.3. Estéreo calibración del sistema

Ya que el objeto a reconstruir se realizará a partir de los pares de estéreo imágenes, se deben determinar los vectores que describen la posición o translación T y rotación R de las cámaras con respecto al objeto, estos son diferentes a los obtenidos en la calibración individual realizada en 2.2 ya que proporcionan la translación (separación) entre las dos cámaras y su rotación con respecto al objeto. La estéreo calibración es el procedimiento que permite realizar lo anterior y fue implementada por medio de OpenCV [6-8], la función “cvStereoCalibrate()” se ejecuta en esta parte del proceso y proporciona los parámetros intrínsecos y extrínsecos del sistema estereoscópico, los primeros son los calculados en la calibración individual y los últimos se refieren fundamentalmente a la transformación rígida (rotación y translación) que relaciona la posición y orientación entre las dos cámaras del sistema. En la figura 5 se muestra el resultado de la estéreo calibración del sistema realizada con Matlab solo para motivos de demostración.

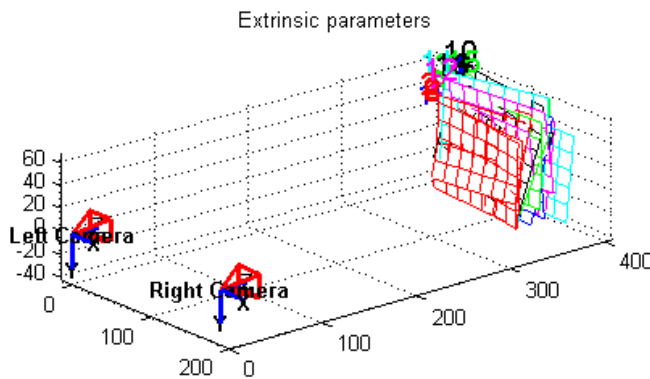


Figura 5. Resultado de la estéreo calibración del sistema.

Con los parámetros obtenidos de la estéreo calibración las matrices de la cámara derecha e izquierda permiten determinar un punto $P(x, y, z)$ en el espacio a partir de sus coordenadas $P_D(X_D, Y_D)$ en la imagen derecha y $P_I(X_I, Y_I)$ en la izquierda por medio de triangulación [8], como veremos más adelante, estos nos proporcionaran la nube de puntos para realizar la reconstrucción 3D del objeto escaneado y reciben el nombre de puntos de interés.

Los puntos de interés deben estar presentes en un mismo par de estéreo imágenes esta propiedad se llama correspondencia

estéreo, la búsqueda de estos puntos (búsqueda de correspondencias) es en general compleja dependiendo las características del par estéreo de imágenes y del objeto en el par que se desea modelar, en nuestro caso, el haz de luz incidente en el objeto a escanear. Como se describirá posteriormente, esto se facilita con la rectificación estéreo.

2.4. Captura de pares de estéreo imágenes del objeto

El objeto o material que se va a reconstruir tridimensionalmente debe colocarse en la base del escenario del escáner de modo que su posición se encuentre en el eje del sistema de visión y pueda ser captado por ambas cámaras. La iluminación debe ser adecuada (no muy intensa) para evitar el ruido ya que las imágenes se trabajaron en escala de grises como se explicará más adelante. El objeto que se escaneó es un recipiente de plástico cuya geometría y forma se adaptaron bien para la evaluación experimental del sistema.

Al iniciar el proceso de captura de pares estéreo de imágenes la plataforma que soporta el láser se mueve paralelamente con respecto a la base de las cámaras a velocidad constante.

Se dispuso en el escenario un par de leds infrarrojos de forma que se encuentran detrás del recipiente, la distancia entre ellos es ajustable al ancho del objeto y centrada con el eje de la base del escenario, estos activan un par de fototransistores infrarrojos en el soporte del láser encendiéndolo para enviar el haz de luz iniciando la captura de los pares de estéreo imágenes de forma que el movimiento del haz este sincronizado con el retardo en la captura de cada par. Con esto se obtuvo un conjunto de imágenes estéreo donde el objeto es barrido completamente por el haz láser en su parte frontal. En la figura 6 se muestra una secuencia de pares de imágenes capturadas.



Figura 6. Secuencia de barrido del objeto escaneado. Las imágenes superiores corresponden a la captura con la cámara izquierda y las inferiores con la derecha

2.5. Rectificación estéreo

La posición del sistema estéreo debe ser tal que los ejes ópticos de las cámaras sean paralelos para cumplir con la geometría estándar, generalmente esto no garantiza que los pares estéreo de imágenes estén alineados lo que complicaría la búsqueda de correspondencias. El problema se resuelve mediante la estéreo rectificación de cada par permitiendo que los puntos de interés (píxeles) en cada imagen del par estéreo estén proyectados en la misma fila mediante el cálculo de sus nuevas coordenadas, de este modo su búsqueda se reduce a solamente a las columnas

en el plano de las imágenes. En OpenCV se implementó el algoritmo basado en el método de Bouguet [8] cuya finalidad consiste en determinar una matriz de rectificación R_{rect} con la cual se realiza la alineación de cada par de imágenes definida por la relación (4):

$$R_{rect} = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ \theta_3^T \end{bmatrix}$$

(4)

Donde los vectores unitarios expresados por las fórmulas (5)

$$\theta_1 = \frac{T}{\|T\|}, \theta_2 = \frac{[-T_y T_x 0]^T}{\sqrt{T_x^2 + T_y^2}}, \theta_3 = \theta_1 \times \theta_2$$

(5)

son ortogonales entre sí. Al realizar la calibración se calculan las matrices de rotación para cada cámara R_D y R_I , derecha e izquierda respectivamente, dadas por las fórmulas (6) y (7).

$$R_D = R_{rect} r_D$$

(6)

$$R_I = R_{rect} r_I$$

(7)

El siguiente paso es calcular las matrices de proyección de cada cámara para transformar los puntos del espacio tridimensional en puntos de sus planos imagen las cuales corresponden a las relaciones (8) y (9).

$$P_D = \begin{bmatrix} f & 0 & c_{xD} & T_x * f \\ 0 & f & c_{yD} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(8)

$$P_I = \begin{bmatrix} f & 0 & c_{xI} & 0 \\ 0 & f & c_{yI} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(9)

Donde la componente T_x es la distancia o desplazamiento horizontal entre las cámaras, f la distancia focal de las cámaras que hemos supuesto iguales y c_{xD} , c_{yD} , c_{xI} , c_{yI} , son las coordenadas de los puntos principales de las cámaras derecha e izquierda respectivamente.

En OpenCV lo anterior se realizó al implementar la función “cvStereoRectify()”, por medio de la función “cvInitUndistortRectifyMap()” se realiza el mapeo de los pixeles de los pares estero a pares estéreo rectificados. Las imágenes mostradas en la figura 6 has corresponde a una secuencia de pares estéreo rectificados.

2.6. Correspondencia estéreo

Como fue mencionado en 2.3 el problema de correspondencia es tal vez el proceso más complicado y que más costo

computacional requiere en la reconstrucción 3D del modelo. Es necesario determinar la mayor cantidad de puntos comunes de interés en el par estéreo para obtener una nube de puntos lo suficientemente densa para lograr, mediante triangulación, un modelo tridimensional lo más exacto posible al objeto real. Existen varios algoritmos que solucionan el problema destacando el método basado en intensidades (Intensity-based Matching), basado en características (Feature-based Method) y los basados en correlación (Correlation-based Method) [6-10]. De igual forma, son relevantes los trabajos realizados por Barnard (1982), Lane (1994), Faugeras (1993), mencionando el método de flujo óptico de Lucas-Kanade (1981). En este proyecto se propuso un método alternativo de acuerdo a los resultados deseados y que consistió en la conversión de cada par de imágenes estéreo rectificadas obtenidas a escala de grises con un valor de umbral adecuado obteniendo un nuevo par de imágenes donde solo fué visible el patrón de luz (haz laser). Posteriormente mediante segmentación se determinaron los centroides del haz al incidir en el objeto. Al realizar esta etapa generalmente se obtienen cantidades diferentes de centroides en cada par debido a la diferencia de posición de las cámaras en el proceso de captura, el programa diseñado permitió determinar los centroides correspondientes en ambas imágenes para ello se estableció aquella que tiene la menor cantidad, asumiendo que debido al proceso de rectificación la búsqueda se limita a determinar aquellos cuya coordenada Y (fila) es la misma en la otra imagen, el procedimiento realizado se muestra en el pseudocódigo mostrado (ver figura 7) el cual hace referencia al caso donde la menor cantidad de centroides se encuentra en la imagen derecha, en caso contrario solo se intercambian las sentencias de control de los ciclos anidados.

```
D=vector centroides imagen derecha.
I=vector centroides imagen izquierda.
d=elementos de D.; i=elementos de I.
If d<i
    for n=1:d
        for m=1:i
            if coordenada pixel vertical de Di=coordenada vertical
de Ii.
                Se almacenan ambos pixeles en vector Centroides derecha y
vector Centroides izquierda.
            end
        end
    end
end
```

Figura 7. Pseudocódigo implementado para determinar los centroides correspondientes en el par estéreo de imágenes.

Al finalizar se obtuvo un vector de centroides correspondientes en cada par estéreo. Posteriormente el programa toma cada uno de ellos como una semilla y establece regiones de crecimiento vertical y horizontal a partir de él con las restricciones de que el valor del pixel debe ser blanco y sus coordenadas deben estar limitadas al tamaño de la imagen. De este modo se determinan

dos vectores de pixeles uno para la imagen derecha y otro para la izquierda. En general la dimensión de estos vectores no es la misma por lo que es necesario determinar que pixeles son correspondientes en ambas imágenes. Se realiza una búsqueda de pixeles mediante un algoritmo semejante al presentado asumiendo que debido a la estéreo rectificación realizada se limitara a determinar nuevamente aquellos cuya coordenada Y es la misma. El resultado final es un par de vectores de igual dimensión para cada imagen cuyos elementos son los pixeles del haz de luz comunes (correspondientes).

2.7. Triangulación

En esta etapa se calcularon las coordenadas tridimensionales (x,y,z) asociadas a las coordenadas (X,Y) de los pixeles obtenidos. La triangulación se basa en la tesis de que las coordenadas del pixel tendrán la misma coordenada vertical Y y diferente coordenada horizontal X , esta diferencia se conoce como disparidad y a partir de ella es posible determinar la profundidad a la que se encuentra el punto correspondiente al pixel. [5-6], el programa guarda las coordenadas 3D en un vector el cual constituye la nube de puntos para un par de estéreo imágenes dado, lo que significa que todas las etapas descritas anteriormente se realizan para la totalidad de las estéreo imágenes adquiridas. El vector con la nube de puntos final fué suficientemente densa con lo que se logró obtener un modelo 3D muy preciso, el programa guarda estos datos en un archivo cuyo formato pueda ser visualizado en un software conveniente como Meshlab ó Solidworks, en las pruebas mencionadas se utilizó este último y el archivo fue txt.

3. RESULTADOS OBTENIDOS

En la figura 8 se muestra el modelo 3D del bote mediante Solidworks.

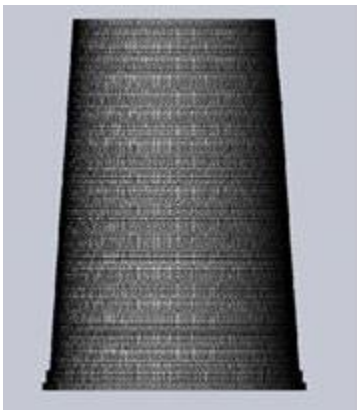


Figura 8. Modelo 3D del objeto visualizado con Solidworks.

La tabla 1 resume las dimensiones del modelo 3D y del material estudiado. Las dimensiones están dadas en cm .

Tabla 1. Resultados obtenidos con el sistema de Visión Estereoscópico de Reconstrucción 3D.

Dimensión	Modelo 3D	Objeto real	Error
Base Inferior	8.68	8.7	0.00230415
Base Superior	11.47	11.5	0.00261552
Altura	14.47	14.5	0.00207326

4. CONCLUSIONES

Dentro de las conclusiones cabe destacar las siguientes:

- El modelo 3D tiene una aproximación aceptable con respecto al objeto que se estudió.
- Se contempla mejorar el proceso de reconstrucción mediante la rotación del material mediante una base rotatoria y modificar el programa para que realice la conversión de coordenadas.

5. REFERENCIAS

- David 3D scanner, [en línea], disponible: <http://www.david-3d.com/en>. Sitio visitado 2014.
- C. Davatzikos, "Spatial Transformation and Registration of Brain Images Using Elastically Deformable Models". Science Direct, May 1997, volume 66, Issue 2 pp 207-222.
- R. Usamentiaga, J. Molleda, D. F. García. "Fast and Robust Laser stripe extraction for 3D reconstruction in industrial environments, Machine Vision and Applications, January 2012, Volume 23 Issue 1 pp 179-196.
- Siggraph 2009 and Siggraph Asia 2009 Courses [en línea] Disponible:<http://mesh.brown.edu/byo3d/Sitio> visitado 2014.
- "Simple triangulation with OpenCV from Harley & Zisserman", More Than Technical [en línea], disponible: <http://www.morethantechnical.com/2012/01/04/simple-triangulation-with-opencv-from-harley-zisserman-w-code/>
- OpenCv Camera Calibration and 3D reconstruction [en línea], disponible: http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html sitio visitado, 2014.
- R. Hartley y A. Zisserman, *Multiple View Geometry in Computer Vision* Cambridge University Press. Second Edition 2003 pp 405-458.
- G. Bradski y A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV library*. O'Reilly. First Edition September 2008. Pp 378-458.
- S. Brahmabhatt. *Practical OpenCV. Technology in action* pp. 95-119.
- S. Holzer, J. Shotton y P. Kohli. *Learning to Efficiently Reapetable Interest Points in Depth Data*. 12th Europeaqn Conference on Computer Vision. Firenze, Italy, 2012. Poster Session 1.