XVI SIMPOSIO INTERNACIONAL SELPER 2014

La Geoinformación al Servicio de la Sociedad

Memorias

Sociedad Latinoamericana en Percepción Remota y Sistemas de Información Espacial
Capítulo Colombia

30 AÑOS

Medellín, Colombia
29 de Septiembre al 3 de Octubre de 2014

APLICACIONES DE LA ESTEREOSCOPÍA EN LA RECONSTRUCCIÓN TRIDIMENSIONAL DE ZONAS DE TERRENOS

Jhonny Penna Salcedo[1], Elsa Adriana Cárdenas Quiroga[2], Luz Yolanda Morales[3]

Universidad Militar Nueva Granada, Bogotá Colombia, Grupo de Investigación Ingeniería, Geomática y Educación (IGE)

RESUMEN

Este trabajo presenta la evolución que han experimentado las técnicas de reconstrucción tridimensional de terrenos, desde las que involucraban el uso de imágenes y otros métodos limitados hasta las técnicas modernas que cuentan con equipos de última generación como láser escáner (2D y 3D) que permiten la reconstrucción precisa de áreas de interés. Además, se evalúa un algoritmo de mapeo y localización simultánea (RGB-D SLAM) y se analiza su aplicación en la generación de planos bidimensionales y tridimensionales de escenas que no cuentan con información previa.

Palabras claves: Estereoscopía, estereovisión, reconstrucción tridimensional, RGB-D SLAM, Kinect.

ABSTRACT

This paper presents the evolution undergone by three-dimensional land reconstruction techniques. It considers the initial methods that involved the use of paper-based images and other restricted techniques as well as modern techniques that apply cutting-edge technologies, such as laser scanner (2D and 3D), to make an accurate reconstruction of areas of interest. Moreover, an algorithm for simultaneous localization and mapping

[1] Ingeniero en Mecatrónica. jhonny.psalcedo@gmail.com
[2] Geóloga, Especialista en Docencia Universitaria. Magister en Gestión Ambiental
[3] Ingeniera Civil. Magister en Educación. Doctora en ciencias de la Ingeniaría

(RGB-D SLAM) is evaluated; its application to the generation of two- and three-dimensional indoor scenes that do not count on previous information is discussed.

Keywords: Stereoscopy, stereovision, three-dimensional reconstruction, RGB-D SLAM, Kinect.

INTRODUCCIÓN

En los últimos años, la reconstrucción tridimensional de terrenos ha tenido avances significativos en diferentes áreas del conocimiento (incluida la geomática) debido principalmente a la creciente necesidad de obtener modelos tridimensionales virtuales precisos que representen zonas de interés en las que se requiere aplicar diversos análisis para el estudio de fenómenos naturales y físicos, simulados o reales. La visualización y procesamiento de información de estas zonas exige el uso de algoritmos robustos que permitan el manejo de altos volúmenes de información que son normalmente obtenidos a través de dispositivos de captura de datos como cámaras y láser escáner. A pesar de que se trata de un proceso complejo e intrincado, los avances en visión computacional han permitido que este procedimiento sea cada vez más flexible y preciso. Esto ha generado gran cantidad de enfoques y propuestas de reconstrucción tridimensional con las que se busca garantizar la calidad, disminuir la complejidad de uso y lograr reducción en los costos.

En el siglo XXI, se crearon los primeros mapas tridimensionales interactivos para complementar la reconstrucción de espacios haciendo uso de imágenes adquiridas de la zona de interés (escena) sin importar sus propiedades geométricas, su resolución y, en general, su calidad. Un ejemplo es presentado por Radoui & Roman (2000), los autores diseñaron un sistema de visualización para la reconstrucción 3D de terrenos basándose

en el algoritmo de la interpolación de Shepard, utilizado para representar datos 2D en mapas interactivos 3D. El sistema propuesto utiliza como datos de entrada nodos no conectados con sus respectivas alturas asociadas, la no uniformidad de los datos es corregida a través de un proceso de remuestreo con el que se obtiene una distribución uniforme. La reconstrucción del terreno se realiza al relacionar la altitud de cada uno de los nodos con los puntos homólogos de la superficie contenidos en las respectivas imágenes 2D a través de la normal del vector proyectado.

Mckinley *et al.* (2001) muestran el uso de un estéreo par ignorando los parámetros intrínsecos del modelo estereoscópico, los extrínsecos de la escena o de los puntos especificados previamente al procesamiento. Para su aplicación, se debe asegurar que los ejes ópticos de las dos imágenes se intersectan y que los ejes *Y* son paralelos para mejorar el emparejamiento de puntos, y por consiguiente, la precisión del sistema. Se realiza una interpolación de un dieciseisavo sobre el arreglo de píxeles de las imágenes, y finalmente, una verificación para comprobar que los puntos calculados de la escena son confiables. Para realizar la reconstrucción final del modelo se necesita por lo menos un punto coordinado de la escena o las dimensiones reales de un objeto.

Tang *et al.* (2003) proponen un algoritmo que integra apropiadamente datos visibles y desconocidos para la realización de la reconstrucción, un ejemplo del método se presenta en la Figura 1. El método tiene como principal característica la aplicación de una etapa de optimización en donde se minimizan los errores de proyección de retorno de cada imagen, y se adicionan diferentes pesos a los puntos para distinguirlos entre visibles y ocluidos. Así, se inicia el proceso de reconstrucción uniendo las estructuras parciales existentes para obtener un modelo 3D inicial. Luego se aplicaba el proceso de optimización y posteriormente se proyectan los puntos característicos del espacio hacía

todas las imágenes dependiendo de las matrices calculadas sin importar si éstas eran o no visibles.

Figura 1. Reconstrucción 3D de un templo.

La integración de información manual y automática para la generación de terrenos se observa en proyectos como el de Shao *et al.* (2003), en este trabajo se muestra una estrategia topológica de reconstrucción 3D semiautomática de edificios y vías desde un estéreo par aéreo. El método se realizó inicialmente a través de operadores humanos para su validación, y posteriormente se complementó a través de una combinación de procesos en los que se mezclan diferentes niveles de geometrías junto con diferentes niveles de abstracción topológica. Los vértices, superficies y cuerpos se definen explícitamente por el operador, definiendo para esto cuatro niveles en el proceso. El primer nivel contiene vértices con la más sencilla extensión espacial y sus tipos se describen por sus atributos. El segundo nivel contiene líneas de juntas, las cuales se construyen de segmentos de línea conectados por puntos 3D. El tercer nivel contiene las superficies de los edificios mostrando la relación con sus vecinos, que pueden ser edificios o vías. El cuarto nivel consiste en la reconstrucción total de edificios o vías.

Otra aplicación es la propuesta por Brandou *et al.* (2007). Los autores diseñaron un método de adquisición de imágenes para la reconstrucción tridimensional de escenas

submarinas, de tal manera que se pueda complementar el análisis de muestras y medidas fisicoquímicas gracias a las medidas cuantitativas en 3D. Se utilizó un sistema de estereovisión para capturar diferentes tomas de un objeto, a intervalos fijos de acuerdo a una trayectoria predefinida. El sistema estaba junto al manipulador de un robot submarino que era preprogramado con la trayectoria a recorrer y a la vez ofrecía una imagen en vivo del escenario capturado. Este es un aspecto muy importante del sistema debido a que genera y sigue la trayectoria generada para conocer las posiciones exactas de la cámara en todo el transcurso del recorrido, y posteriormente lograr un modelo 3D de alta precisión de la escena.

Endres *et al.* (2012) evalúan el sistema RGB-D SLAM (Simultaneous Localization and mapping) para la reconstrucción del entorno y la estimación de la posición de la cámara usando estrategias de visión de máquina (Davidson *et al.* 2007, Newcombe *et al*. 2010). El algoritmo utilizado permite adquirir rápidamente los modelos a color en 3D de objetos y escenas con un sensor que proporciona una imagen RGB y la información de profundidad (D). Internamente usa características visuales SURF o SIFT para detectar las coincidencias entre pares de imágenes obtenidas y utiliza RANSAC para estimar la transformación 3D entre ellas. Para lograr el procesamiento en línea, la imagen actual se compara con un subconjunto de las imágenes anteriores. Posteriormente, se construye un grafo cuyos nodos corresponden a las vistas de cámara y cuyos bordes se corresponden con las transformaciones 3D estimados. La gráfica se optimiza con el algoritmo HOG-MAN de Grisetti *et al.* (2010) para reducir los errores acumulados. El proceso se observa en la Figura 2.

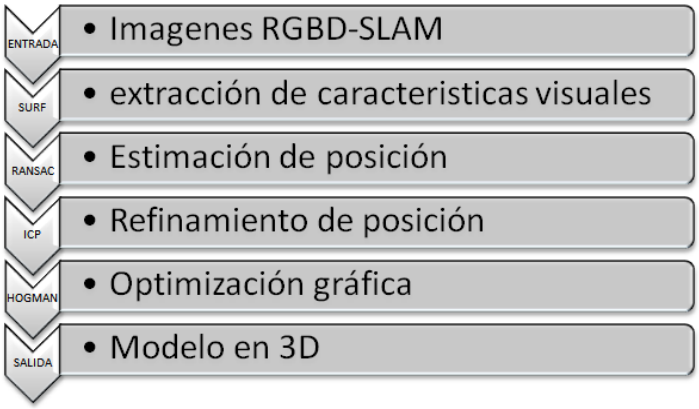

Figura 2. Etapas de procesamiento del algoritmo RGB-D SLAM

METODOLOGÍA

Después de efectuar la revisión de algunos de los trabajos relacionados con la reconstrucción tridimensional de escenas y terrenos, se decidió implementar y evaluar el algoritmo de RGB-D SLAM en razón a que se quería reconstruir un escenario que no contara con información base. Este trabajo se dividió en dos fases:

Implementación. El algoritmo seleccionado se implementó en el sistema operativo para robots (ROS), que funciona bajo la licencia BSD de software libre. Como dispositivo de entrada de datos se usó el sensor Kinect que proporciona información de las imágenes en RGB y mide de la profundidad de los elementos capturados ofreciendo una escena a color tridimensional de la realidad capturada.

Evaluación. Se realizó la evaluación de la precisión en la reconstrucción que realiza el algoritmo seleccionado y se comparó con mediciones directas tomadas sobre una escena real. Posteriormente, los datos capturados y las mediciones adquiridas se usaron para obtener un cálculo del porcentaje de error.

Para asegurar la fidelidad de los datos se construyó un banco de pruebas que permitió controlar variables tales como la ubicación del sensor, el tamaño y orientación de los objetos en la escena.

RESULTADOS

Diseño del banco de pruebas.

Se diseñó un banco de pruebas con el objeto de realizar medidas de profundidad con valores máximos de 1900 mm. Considerando los ángulos de apertura horizontal y vertical del sensor (ver Figuras 3 y 4), se determinó el ancho (2.172 m) y alto (1.576 m) de la pared de fondo para un área total de 3,42 m^2 y para el piso un área de 4,344 m^2. Sobre la superficie del piso se realizaron marcaciones con una separación de 100 mm desde la pared de fondo hasta los 2100 mm.

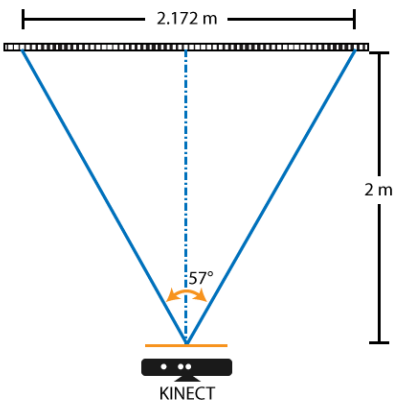

Figura 3. Vista de techo, espacio de prueba y ubicación del sensor.

Para la evaluación horizontal y vertical de la reconstrucción se crearon dos cuadrículas, una de 4 cuadros y otra de 9 cuadros.

Evaluación de profundidad

Toma de datos. Se capturan imágenes con el Kinect usando la cámara RGB y el sensor de profundidad a diferentes distancias: 500 mm, 1000 mm, 1500 mm, 2000 mm, 2500 mm y 3000 mm. La altura a la que se ubicó el sensor fue de 750 mm del suelo para garantizar la cobertura de la pared de fondo.

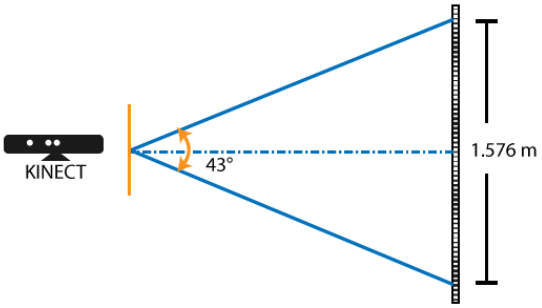

Figura 4. Vista lateral, espacio de prueba y ubicación del sensor.

Filtro de área de interés. Debido a que se capturan datos en tres ubicaciones que exceden la longitud máxima del banco de pruebas, se filtran las imágenes por área, de forma tal que solo se consideran 1000 puntos. Esto permite obtener un volumen menor de datos que facilitan los cálculos y elimina zonas de la imagen que no contienen información útil para la prueba. Posteriormente se calcula la media de los 1000 puntos de interés para obtener un promedio de la medición del sensor. Con la media obtenida y la medida real se calcula el error porcentual (ver tabla 1).

Tabla 1. Datos reales y medición de profundidad

Real (m)	0.5	1	1.5	2	2.5	3
Sensor (m)	0.52	1.04	1.55	2.06	2.57	3.09
Error %	4.04	4.35	3.4	3.45	2.80	3.21

Evaluación de las medidas horizontales y verticales

Toma de datos. Se capturan datos a diferentes distancias (1400 mm y 1900 mm) y alturas (50 mm, 350 mm y 1500 mm) usando dos configuraciones de marcadores (4 cuadros y 9 cuadros). La imagen RGB-D entregada tiene en promedio 250.000 puntos en el espacio (ver Figura 5).

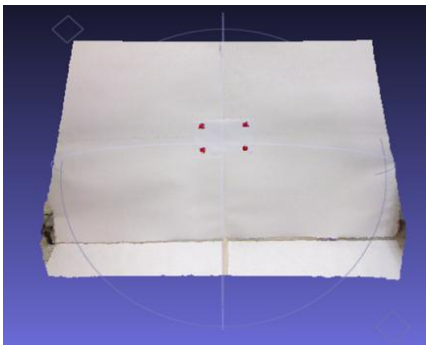

Figura 5. Captura inicial, datos entregados por el sensor.

Filtro por área de interés. Debido a la gran cantidad de puntos entregados en la captura de datos, es necesario realizar un filtrado por área de interés que coincida con la cuadrícula presente en la imagen de prueba, la imagen resultante tiene aproximadamente 8000 puntos (ver Figura 6).

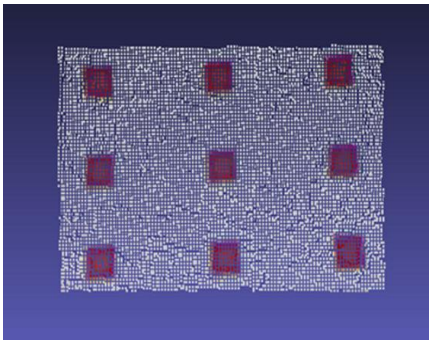

Figura 6. Resultado del filtro área de interés.

Filtro de color. Posterior a la extracción del área de interés, se aplica el filtro de color teniendo en cuenta que los marcadores utilizados son de color rojo y el fondo blanco. Se realiza un programa en el cual se hace la conversión de RGB a HSV para extraer únicamente los puntos rojos de la imagen, la imagen resultante posee aproximadamente 300 puntos (ver Figura 7).

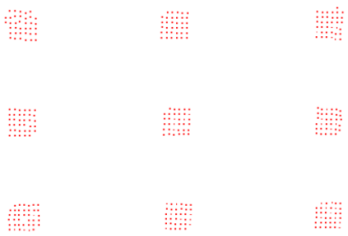

Figura 7. Resultado filtro de color.

Los puntos que componen la imagen resultante del filtro de color presentan varias tonalidades de rojo, para lo cual se realiza una estandarización del color asignando un valor RGB correspondiente a rojo puro.

Selección de puntos. Por cada cuadro presente en la cuadrícula original, se ubican dentro de la nube de puntos las esquinas de cada cuadro teniendo en cuenta las coordenadas del plano de la imagen (X, Y).

Medición de distancia. Teniendo las coordenadas (X, Y, Z) de las esquinas de los cuadros se calcula, mediante la distancia Euclidiana, el alto y ancho de cada cuadro.

Reconstrucción 3D

Para observar el comportamiento del algoritmo se capturaron dos escenas de interiores. La primera escena corresponde al laboratorio de robótica de la Universidad Militar Nueva Granada (ver Figura 8). Debido a la complejidad de esta escena, se generan varios artefactos en la reconstrucción de los objetos pequeños, observándose la falta de definición de los bordes, presencia de un elevado número de puntos muertos y sobre posición de objetos. La siguiente escena corresponde a un pasillo (ver Figura 9) que contiene un menor número de objetos, que adicionalmente poseen formas más regulares. En la reconstrucción se pudo observar una mejor definición en los bordes y contornos de los objetos, apreciándose también una mejor relación espacial entre ellos representando mejor las dimensiones de la escena real.

Figura 8. Reconstrucción del laboratorio de robótica de la UMNG.

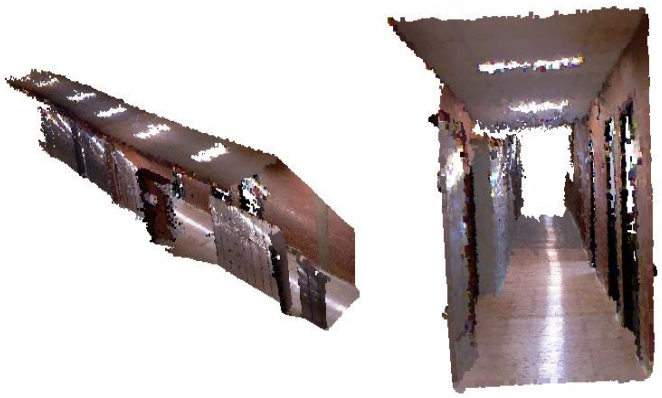

Figura 9. Reconstrucción de un pasillo de la UMNG.

DISCUSIÓN DE RESULTADOS Y CONCLUSIONES

Los resultados obtenidos en la implementación del método evaluado permiten considerar su uso en diferentes áreas de la geomática. Con un error medio inferior al 4% en entornos generales, y al 1% en objetos de gran tamaño ($x \geq 1.8$m), la precisión del método es adecuada para levantamientos fotogramétricos terrestres en los que no se requiere gran detalle y alta precisión sino la reconstrucción de escenarios generales. Considerando que la construcción del entorno se fundamenta en la captura de imágenes con condiciones geométricas y espaciales controladas, y que dichas condiciones pueden ser simuladas, las posibilidades de aplicación, tal y como se verá a continuación, son variadas.

El hecho de que la captura de datos (nube de puntos) sea controlada en los tres ejes permite un mayor control de la precisión según las necesidades de la escena. En el caso de barridos manuales, la calidad de esta escena estará dada por el sistema de referencia predefinido antes de la captura; todos los sensores y escenas (frames) capturadas estarán ligados a este sistema de referencia. Por ejemplo, en escenarios en donde la medida

relevante sea la altura de los elementos que componen la escena (i.e., MDE) se puede pensar en la localización del sensor no ortogonal a la superficie sino al plano X-Z de la escena, de este modo, los valores en el eje Z de la escena corresponderán a los capturados en Y, de este modo el control de la profundidad será más eficiente. En este caso, será necesario contar con sistemas de transformación y/o rotación de ejes.

Para escenas de gran extensión se deben considerar las restricciones del sensor. Conociendo la distancia máxima de captura (3m) y los ángulos máximos horizontal y vertical (57° y 43°), se requiere el desarrollo de métodos adicionales para la integración de escenas que permitan la reconstrucción de entornos que superen el volumen máximo de captura, e.g., 2.1x1.5x2.0 m^3 para una profundidad de 2m. Es posible considerar una integración de las escenas considerando un sistema de referencia simétrico lineal en donde puntos equidistantes sirvan de ancla a los datos capturados en sus sistemas relativos. En general, es posible explorar los métodos geométricos y de calibración de sensores aerotransportados avanzados y validar su aplicación en la tomas del Kinect para mejorar la integración de las escenas, y si es el caso, realizar correcciones geométricas. Esta restricción es eliminada si el sensor es transportado por una plataforma móvil que entregue información de odometría (velocidad, aceleración, orientación), el algoritmo RGB-D SLAM se encarga en tiempo real del cálculo para la unificación de escenas.

Los casos de estudio están ligados a la fotogrametría terrestre, todos aplicados a la reconstrucción de escenas. El sistema propuesto es rápido, económico y fácil de implementar; si bien no reproduce los datos con precisión cartográfica, si permite capturar entornos y escenarios geográficos de forma inmediata y entregar datos fáciles de procesar y con las calidades mínimas requeridas para solucionar problemas

inmediatos en situaciones en las cuales variables como el presupuesto o la localización restringen el uso de sistemas de alta precisión cómo el LIDAR o métodos de escaneo tradicionales. Entre las posibles aplicaciones del método presentado se encuentran los levantamientos forenses, la digitalización 3D de paisajes urbanos, la reconstrucción de escenarios arqueológicos, el levantamiento superficial de minas, siempre considerando las precisiones expuestas. Además, se puede considerar como complemento de escenarios en los que los sistemas convencionales pierden la señal o presenten fallos, por ejemplo, aeronaves o automotores que pierdan la señal GPS al entrar en túneles o depresiones geográficas.

En el caso de reemplazar el sensor kinect por otro de mejores especificaciones las limitaciones de captura y resolución de datos mejorarían de acuerdo a las especificaciones del nuevo sensor, el método y el algoritmo presentado en este artículo sigue siendo efectivo sin importar el tipo de sensor utilizado.

REFERENCIAS

Brandou, V., Allais, A., Perrier, M., Malis, E., Rives, P., Sarrazin, J. & Sarradin, P. (2007) 3D Reconstruction of Natural Underwater Scenes Using the Stereovision System IRIS, *OCEANS 2007 – Europe*, 1-6.

Davison, A., Reid, I., Molton, N. & Stasse, O. (2007) MonoSLAM: Real-Time Single Camera SLAM. *Pattern Analysis and Machine Intelligence,* 29(6), 1052-1067.

Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., & Burgard, W. (2012) An evaluation of the RGB-D slam system. *IEEE International Conference on Robotics and Automation.*

Grisetti, G., Kummerle, R., Stachniss, C., Frese, U., Hertzberg, C. (2010) Hierarchical Optimization on Manifolds for Online 2D and 3D Mapping. *Robotics and Automation (ICRA),* 273-278.

Mckinley, T., McWaters, M., & Jain, V. (2001) 3D reconstruction from a stereo pair without the knowledge of intrinsic or extrinsic parameters'. *Digital and Computational Video*, 148-155.

Newcombe, R., & Davison, A. (2010) Live Dense Reconstruction with a Single Moving Camera. *Computer Vision and Pattern Recognition (CVPR)*, 1498-1505.

Radoui, D., & Roman, A. (2000) 3D terrain reconstruction using scaterred data sets. *Universitatea Babeş-Bolyai*, 45(2), 3-10.

Shao. Z., Li, D., Cheng, Q. (2003) A topological 3D reconstruction of complicated buildings and crossroads. *Geoscience and Remote Sensing Symposium*, 1(1), 56-58.

Tang, L., Wu, C., Tsui, H., & Liu, S. (2003) Algorithm for 3D reconstruction with both visible and missing data. *Electronics Letters*. 39(23).