ScienceDirect
Disponible en www.sciencedirect.com
Revista Iberoamericana de Automática e Informática industrial 12 (2015) 189–198
www.elsevier.es/RIAI

Detección de Automóviles en Escenarios Urbanos Escaneados por un Lidar

Alfonso Ramírez-Pedraza[a,*], José-Joel González-Barbosa[a,*], Francisco-Javier Ornelas-Rodríguez[a], Angel-Iván García-Moreno[a], Adan Salazar-Garibay[b], Erick-Alejandro González-Barbosa[c]

[a]*Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada, Instituto Politécnico Nacional, Querétaro, México.*
[b]*Quantificare S.A., Sophia-Antipolis, Francia.*
[c]*Instituto Tecnológico Superior de Irapuato, Guanajuato, México.*

Resumen

En este trabajo se desarrollaron algoritmos computacionales los cuales permiten segmentar y detectar los automóviles que se encuentran dentro de una nube de puntos 3D pertenecientes a un entorno urbano. Las nubes son adquiridas con el sistema LIDAR velodyne HDL-64E, el cual es capaz de reconstruir objetos que se encuentran a una distancia de 3 a 120 metros. El desarrollo del trabajo está dividido en dos etapas: la segmentación de objetos 3D y la indexación. Se considera que el ambiente es semi estructurado debido a que objetos como las fachadas de los edificios y el suelo pueden ser modelados por planos. La segmentación está compuesta por tres módulos: el primero consiste en la extracción del plano principal o suelo; en el segundo módulo se extraen los planos perpendiculares al suelo, los cuales son segmentados por una adaptación al método de Hough; y en el último módulo se aplica una segmentación a los objetos restantes usando el método MeanShift. La indexación está dividida en dos módulos: en el primero, los objetos segmentados son modelados por histogramas usando las direcciones de las normales de los puntos 3D de los objetos; en el segundo módulo, los histogramas son usados para comparar con una biblioteca de datos previamente construida. En este trabajo hemos usado el análisis ROC para optimizar los umbrales empleados. Se utilizan dos bases de datos: la primera es de entrenamiento y está compuesta por 4500 objetos; mientras que la segunda corresponde a la de prueba, la cual está compuesta por 3000 objetos y fue usada para verificar los resultados de entrenamiento. La aportación de este trabajo es presentar un nuevo sistema para detectar automóviles utilizando puntos 3D. *Copyright © 2015 CEA. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.*

Palabras Clave: Nube de Puntos 3D, LIDAR, Segmentación 3D.

1. Introducción

Las reconstrucciones realizadas con sensores láser como el LIDAR, están siendo utilizadas para resolver problemas en una amplia gama de sectores, por ejemplo Zhou y Neumann (2008) mencionan que el obtener el modelo 3D de ciudades pueden resolver problemas de tráfico, detectar y prevenir desastres en minas, obtener la topología y cartografía de la tierra para diseñar el crecimiento organizado de las ciudades. Sun y Salvaggio (2013) calculan la calidad de la superficie superior de los edificios. En Sohn et al. (2012) utilizan un LIDAR aéreo para la reconstrucción del ambiente, ellos describen que la seguridad de la infraestructura de la red eléctrica puede llegar afectar significativamente nuestra vida diaria, así como las actividades industriales. Describen que existen diversos factores que pueden amenazar la seguridad de la red eléctrica como lo son: la vegetación, la temperatura del ambiente y fallas estructurales, mencionan que un seguimiento puntual es clave para evitar posibles apagones.

Las reconstrucciones de ambientes urbanos 3D van tomando mayor auge dentro de la segmentación y clasificación de objetos. El trabajo de Himmelsbach et al. (2009) permite detectar, clasificar y seguir objetos 3D, con un vehículo terrestre autónomo, el cual navega de manera segura dentro del tráfico. La eficiencia que aporta la metodología se deriva de la combinación 2D y técnicas de procesamiento de datos 3D. En la segmentación de objetos de las nubes de puntos 3D se utiliza el mapa de ocupación de rejilla. Posteriormente se extraen las características de los objetos, utilizando la distribución de la normal para calcular sus histogramas. Finalmente utilizan SVM (*Support Vector Machine*) como clasificador de automóviles para discriminar dos clases de interés: automóviles y otros objetos.

El trabajo presentado en Goron et al. (2010) resuelve problemas específicos de automatización. La primera parte describe los detalles de su sistema de escaneo basado en un láser 2D;

*Autor en correspondencia.
Correos electrónicos: ponchorp.1985@gmail.com (Alfonso Ramírez-Pedraza), jgonzalezba@ipn.mx (José-Joel González-Barbosa)

© 2014 CEA. Publicado por Elsevier España, S.L.U. Todos los derechos reservados
http://dx.doi.org/10.1016/j.riai.2015.03.003

la segunda parte presenta una técnica de segmentación específica para entornos urbanos, que se divide en cuatro módulos. En el primero se elimina el ruido, calculan y analizan la densidad de la nube, eliminan puntos poco densos por ser considerados puntos dispersos, utilizan un umbral para decidir si los conjuntos de puntos son dispersos o densos. El segundo módulo, aplican RANSAC (ver Schnabel et al. (2007)) para obtener segmentaciones razonables de la escena urbana, inician con un punto de consulta y búsqueda de vecinos más cercanos. En el tercer módulo, separan el fondo del primer plano, es decir, almacenan los puntos que corresponden a las paredes de edificios y el suelo en una nube de puntos, y en otra nube de puntos se almacena los objetos de pequeña escala como personas, coches, vallas, arboles. Y finalmente, segmentan los objetos de pequeña escala utilizando RANSAC que calcula y agrupa por medio de un cuadro delimitador.

En Pulido Fentanes Jaime (2012) presenta una plataforma robótica para reconstruir y modelar ambientes de exteriores 3D en forma automática. Mencionan que su robot es flexible para ser utilizado en investigaciones de diferentes ámbitos como en misiones de búsqueda y rescate en situaciones de riesgo para humanos, como incendios, derrumbes, manipulación de sustancias peligrosas o incidentes nucleares. Ellos describen la estructura mecánica, la arquitectura en hardware y software, así como la comunicación de la plataforma robótica. Sus aportaciones son el desarrollo de un algoritmo diseñado para reconstruir automáticamente el entorno y el diseño de la plataforma todo terreno. En este artículo categorizan los puntos capturados en tres grupos: obstáculos, suelo u objetos; y procesan los datos 3D para extraer la información relevante de la nube de puntos.

El trabajo de Brostow et al. (2008) propone un algoritmo para la segmentación semántica utilizando nubes de puntos 3D derivados de ego-movimiento. El método consiste en combinar las características con el movimiento de un automóvil. Las características utilizadas son una proyección de puntos 3D a un espacio bidimensional considerando la información espacial y contextual, para lograr una segmentación coherente y reconocer las categorías de los objetos presentes. La segmentación semántica se basa en el movimiento 3D; el método funciona en reconstrucciones parciales y con ruido.

Douillard et al. (2011), presentan un conjunto de métodos para segmentar diferentes nubes de puntos 3D. Empíricamente demuestran que la extracción del suelo mejora significativamente el resultado de la segmentación. La segmentación de datos 3D dispersos se realiza usando modelos de baja resolución y proporcionando una malla del área reconstruida a partir de la estructura de una imagen. Utilizan el proceso gaussiano (GP) del modelo de regresión y consenso de muestreo incremental (INSAC), la segmentación basada en el mallado, la convexidad. Todos los algoritmos se prueban con varios grupos de datos y se evalúa la segmentación. Los resultados muestran la adaptación a los diferentes tipos de datos como el seguimiento dinámico y la predicción de la trayectoria de los objetos.

La reconstrucción de objetos 3D a partir de nubes de puntos o imágenes son una tarea importante en muchos campos de aplicación, como lo mencionan Truong et al. (2010), el enfoque toma la estrategia cognitiva del ser humano para resolver el problema de extracción e identificación de objetos. Ellos mencionan que en la actualidad los métodos robustos y rápidos para extracción de objetos siguen siendo un tema de investigación en curso y sufren de la compleja estructura de los datos. El proceso se basa en seis etapas: en la primera etapa, reducen información irrelevante o ruido, utilizan los algoritmos realce de la imagen y reducción del ruido en las nubes de puntos; en la segunda etapa, reducen el peso de los datos tales como el número de puntos en áreas de alta densidad aplicando la reducción de resolución uniforme; en la tercera etapa, se aplica la segmentación basada en color y recuadro delimitador; la etapa cuatro y cinco, detectan y reconocen objetos respectivamente. Para detectar y reconocer los objetos, proponen algoritmos de ajuste del plano por mínimos cuadrados, región creciente y los eigen valores de estimación para obtener las estructuras de los objetos. Finalmente en la sexta etapa, la reconstrucción o representación toman los datos que fueron transformados a una representación basada en mallas triangulares e histogramas por cada punto de la base de datos

Liang-Chia Chen y Wu (2013) mencionan que el reconocimiento de objetos 3D es un problema complejo de visión por computadora debido al ruido y características que presentan las escenas. Presentan un nuevo algoritmo de reconocimiento de objetos empleando histogramas basados en la curvatura. Su sistema de detección de objetos se basa en dos principales etapas: la segmentación y reconocimiento de objetos. La segmentación subdivide la información en grupos más pequeños, utilizan el método de Xu y Fujimura (2003), donde ellos aplican elipses que se contraen de manera iterativa, las iteraciones terminan cuando ningún punto se encuentra fuera de la elipse. En la segunda etapa, el reconocimiento de objetos utiliza una combinación de diversas características como formas geométricas y curvatura. Se desarrolla en dos módulos: el primer módulo determina los modelos de los objetos y son almacenados en una base de datos, lo realizan fuera de línea, en el segundo módulo generan los histogramas según las dimensiones y curvaturas de los objetos, finalmente se comparan los modelos. El reconocimiento de objetos se logra mediante la búsqueda de correlación de dimensiones de los histograma contra los objetos que se encuentran almacenados en la base de datos de entrenamiento. Describen los resultados con una tasa de efectividad del 90,4 % en el reconocimiento de seres humanos, autos y motocicletas.

Huber (2002) presenta un sistema que permite la automatización del modelado a partir de un conjunto de vistas 3D. El sistema se divide en tres etapas: la primera etapa, presenta un método orientado a datos que permite modelar implícitamente los efectos del ruido; en la segunda, desarrollan un algoritmo que permite obtener una hipótesis correcta del modelo y finalmente en la tercera etapa presentan gráficamente el modelo. Muestran el modelado de una gran variedad de escenas, incluyendo objetos pequeños, zonas de interiores y exteriores. En nuestro trabajo, el modelado de los objetos se obtienen a partir de objetos que se encuentran solo en exteriores, ya sea estáticos o en movimiento. Los objetos son totalmente reconstruidos en Huber (2002), sin embargo, en nuestra aplicación los objetos reconstruidos presentan oclusiones.

En el artículo de Barnea y Filin (2008) presentan un enfoque

computacional para resolver el problema del registro, explotan las características invariantes de la transformación del cuerpo rígido 3D para reducir la carga computacional en la correspondencia entre características. La metodología se divide en tres etapas: en la primera etapa, se obtiene una aproximación de coincidencia de las nubes de puntos. En la segunda etapa, se extraen las características clave, y son usadas en la tercera etapa, donde se aplica un refinamiento. Los resultados muestran la eficiencia de los algoritmos para segmentación por el uso de características locales. Nuestra segmentación está basada en MeanShift, la cual agrupa los objetos aleatoriamente según sea la media encontrada en un grupo de datos, sin importar la distancia a la que se encuentren del LIDAR.

Wang et al. (2003) presenta la detección y seguimiento de objetos en movimiento, dando solución a la localización y mapeo simultáneo (SLAM), para estimar con exactitud la posición de los objetos en movimiento, utilizan el mapa del área y la estimación de SLAM, los objetos son detectados al encontrar inconsistencias entre el nuevo escaneo y el mapa. Para modelar el movimiento de un objeto y predecir su posición futura usan el algoritmo modelo de interacción múltiple (IMM). Posteriormente aplican la hipótesis del seguimiento múltiple (MHT) para perfeccionar la detección y asociar los datos. Los resultados demuestran que el algoritmo es fiable y robusto para detectar y realizar seguimiento de peatones y diferentes tipos de vehículos en movimiento en zonas de exteriores. A diferencia de Wang et al. (2003), nosotros asociamos los datos y conocemos la posición de los objetos aplicando MeanShift, para posteriormente indexarlos a una base de datos conocida y suprimir de nuestro mapa los automóviles que son considerados como objetos no pertenecientes al ambiente

En el trabajo de Steder et al. (2010) presentan un enfoque novedoso de reconocimiento de sitios basados en la indexación de puntos de interés 3D. Cada sitio es representado por un conjunto de puntos de interés, donde cada punto contiene un vector descriptor. Para decidir la similitud entre dos lugares, se comparan los dos conjuntos de puntos. Primeramente es calculada una transformación entre ambos conjuntos utilizando la distancia Euclidiana seguida de una comparación de los vectores descriptores. Nuestro método de indexación representa los objetos 3D como histogramas de los direcciones de los vectores normales a la superficie. Estos histogramas necesitan poco espacio para su almacenaje.

El trabajo desarrollado por Golovinskiy et al. (2009), propone el diseño de un sistema de reconocimiento de objetos en nubes de puntos 3D escaneadas en entornos urbanos. El sistema consta de cuatro etapas: localización, segmentación, caracterización y clasificación de clústers. En la primera etapa se realiza una agrupación jerárquica de un primer clúster, donde éste se toma como base para realizar la localización de los demás clúster. Comparan el método Image Filters, MeanShift, CC Clustering y NC Clustering. En la segunda etapa, segmentan los clúster localizados. En la tercera etapa, se construye un vector de características mediante un clasificador entrenado manualmente, el cual contiene autos, postes de luz, semáforos, basureros, cajas de control de tráfico, anuncios espectaculares, entre otros. Finalmente, se etiquetan las características de los vectores mediante el clasificador vecinos más cercanos (k-nearest neighbors). La ubicación y segmentación propuesta brinda una mayor localización de objetos con un alto grado de precisión.

La detección de objetos en datos 3D se ha convertido en un campo cada vez más activo en los últimos años. Como lo describen Lai y Fox (2010), el problema principal en la detección de objetos es la necesidad de etiquetar los datos de entrenamiento para el aprendizaje de los clasificadores. Los autores reducen la necesidad de etiquetar los datos de entrenamiento manualmente mediante el aprovechamiento de los datos disponibles en internet, la finalidad es hacer frente a las diferentes características de los objetos. Adquieren información con un sistema formado por cuatro cámaras que cubren en gran parte los 360^o de visibilidad del robot. Los resultados obtenidos demuestran que los datos adicionales 3D tomados de internet mejoran la precisión de la clasificación en los entornos del mundo real, ya sea en interiores o exteriores.

Hoy en día existen pocos métodos con los cuales se puedan segmentar y reconocer objetos captados con un LIDAR, estos objetos se encuentran almacenados como una nube de puntos en tres dimensiones (x, y, z) como lo muestra Lai y Fox (2010). El sistema LIDAR gira 360^o sobre su propio eje, el número de puntos reconstruidos varía entre 0.1 y 1 millón de puntos dependiendo de la velocidad de rotación. Los puntos adquiridos por el LIDAR pertenecen a diversos objetos reconstruidos parcialmente. Para solucionar el problema de la reconstrucción parcial de los objetos, se unieron varias nubes de puntos adquiridas en diferentes posiciones García-Moreno et al. (2013). Los resultados obtenidos fueron objetos reconstruidos casi en su totalidad.

Nuestra plataforma de adquisición de datos se muestra en la Figura1. La plataforma es capaz de adquirir datos estáticamente o en movimiento. Adquiere información de tres sistemas diferentes: el LIDAR, una cámara panorámica (LadyBug) y un sistema inercial. El mapeo móvil se refiere a la colección de datos provenientes de múltiples sensores móviles geo-referenciados. Las aplicaciones son numerosas, tales como cartografía de infraestructura vial, la adquisición de inventarios, la teledetección láser en la caracterización y gestión del medio fluvial, arqueología, geografía, geomorfología, sismología y física atmosférica, etc.

La plataforma es utilizada para la digitalización de ambientes urbanos, con el objetivo de tener un registro tridimensional del centro histórico de Querétaro, el cual es patrimonio cultural de la humanidad. Con el propósito de adquirir la menor información posible de objetos que no pertenecen al ambiente, las adquisidores son hechas de manera nocturna. El adquirir en este horario no se obstaculiza el tránsito. Sin embargo, las adquisiciones muestran que existen automóviles estacionados. En este trabajo se propone un método para segmentar y detectar autos a partir de la nube de puntos 3D adquirida por el LIDAR. Como los autos son objetos que no pertenecen al medio ambiente reconstruidos, el objetivo es identificarlos y suprimirlos de la nube de puntos 3D. La principal aportación de este trabajo es presentar un nuevo procedimiento para detectar automóviles en nubes de puntos 3D, mediante el uso de técnicas como lo son: Hough, MeanShift y Delaunay. En los resultados de la sección 4 se muestra la eficiencia al detectar automóviles dentro de una

gran variedad de objetos 3D, lo que demuestra la robustez de nuestro trabajo.

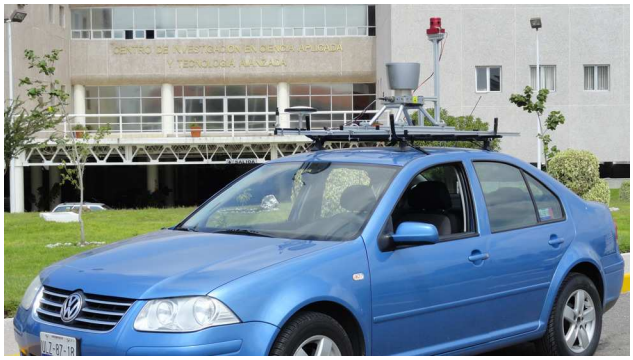

Figura 1: Sistema de Reconstrucción 3D. Nuestro sistema se encuentra montado en un vehículo, para realizar el levantamiento de los datos circulamos en áreas urbanas. La LadyBug cuenta con seis cámaras que capturan la escena en sus 360°, el GPS obtiene la posición en longitud y latitud del área en la que circulamos y el LIDAR reconstruye los 360° de la escena.

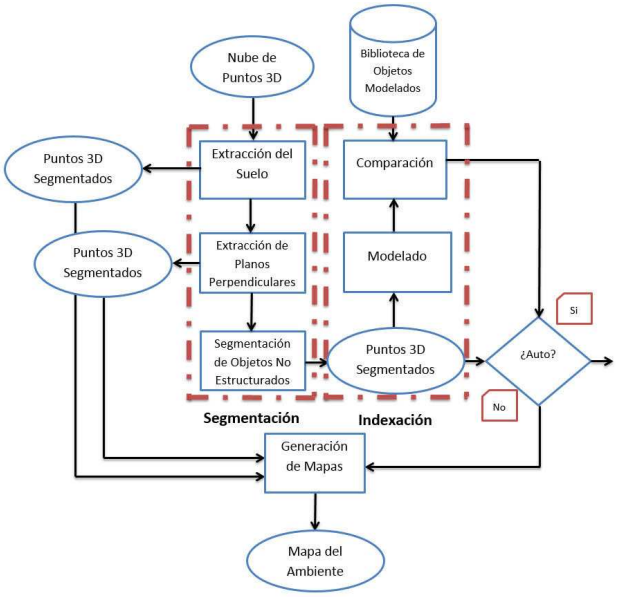

Figura 2: Metodología. El desarrollo de nuestro trabajo está dividido en dos etapas. La etapa de segmentación está compuesta por la extracción de objetos estructurados como el suelo, planos perpendiculares al suelo y objetos no estructurados. La indexación utiliza objetos segmentados, se calcula su modelo utilizando Delaunay. Usando la dirección de las normales de cada triangulo se calcula un histograma; y finalmente se compara con una biblioteca utilizando las técnicas descritas en la sección 4.4.

2. Segmentación de Objetos

El medio ambiente digitalizado contiene objetos estructurados y no estructurados. En primer lugar, los objetos estructurados que corresponden al suelo y fachadas son detectados y extraídos de la nube de puntos. Los puntos restantes son segmentados usando MeanShift, este algoritmo agrupa un conjunto de datos de dimensión d asociando a cada punto con la moda o el pico de la función de densidad de probabilidad del conjunto de datos. La Figura 2 muestra la secuencia de las etapas utilizadas en la segmentación.

2.1. *Extracción de Planos*

Los puntos que pertenecen al suelo son los primeros en ser detectados y extraídos de la nube de puntos, se utilizan los algoritmos que se describen en el trabajo de Hernández-García et al. (2011). En segundo lugar las fachadas son detectadas y extraídas.

La normal del suelo sirve para definir un nuevo sistema de coordenadas; los puntos se representan en este nuevo sistema. Partiendo del conocimiento de que las fachadas son planos perpendiculares al suelo, solo para esta etapa de la segmentación, la tercera coordenada de los puntos no se utiliza y por medio de la transformada Hough se busca el conjunto mayor de puntos que modelan un plano. El algoritmo 1 describe la forma de extracción de los planos.

En el algoritmo 1 γ es un umbral que define el mínimo número de puntos que forman un plano, $\theta \in [0, \Delta\theta, 2\Delta\theta, 3\Delta\theta, \cdot\pi]$ y $\rho \in [\rho_{min}, \rho_{min}, \rho_{min}+\Delta\rho, \rho_{min}+2\Delta\rho+3\Delta\rho, \cdot\rho_{max}]$, $acum$ es una matriz de contadores inicializada en 0 a la cual en cada iteración se le asigna la unidad en el espacio matricial correspondiente, θ es el ángulo de inclinación, $distStep = \Delta\rho$, n es la normal del plano encontrado con coeficientes A', B', C', D'. El factor de acumulación $exp(20 * \rho/T_x)$ es debido a que a mayor distancia se encuentren los planos el número de puntos que lo definen es menor.

Para extraer los puntos correspondientes al suelo es necesario establecer un umbral (α), el cual obtenemos aplicando la ecuación de incertidumbre del sistema de medición, que se presenta en (1) y demostrada en BIPM (2008). Esta expresión comúnmente es utilizada para calcular la calidad y confiabilidad de las medidas de un instrumento. Nosotros la adaptamos para obtener un punto de referencia que discrimine planos de otros objetos. Utilizamos los resultados obtenidos por Atanacio-Jiménez et al. (2011), ellos definen una varianza de calibración del sensor de $2{,}22cm^2$ y una media del error de $1{,}56cm$, estos datos se obtienen después de calibrar el LIDAR Velodyne 64E.

$$U = k * u_c(y) = k * \sqrt{U_{cal}^2 + U_p^2 + U_w^2} + |b| \qquad (1)$$

Donde U_{cal}^2 es la varianza de calibración del sensor igual a $2{,}22cm^2$, U_p^2 es la suma de los errores en el proceso de medición, en nuestro artículo es igual a 0, U_w^2 es el tamaño promedio de la banqueta igual a $13cm$, k es factor de cobertura igual a 2, para conseguir un nivel de confianza en la incertidumbre $p = 94{,}5\,\%$ y $|b|$ es la media del error igual a $1{,}56cm$. Por lo que un umbral α está definido como:

$$\alpha = U = 15{,}46cm$$

Algoritmo 1 Transformada de Hough Modificada

Entrada: Nube de Puntos Z_i, donde $Z_i \in \Re^d, i = 1...N$
Inicialización:
Discretiza el espacio paramétrico $[T_x, T_y] = acum[\rho, \theta]$
$acum[\rho_i, \theta_i] \leftarrow 0$
$k \leftarrow 0$
$h \leftarrow 0$
$max \leftarrow 0$
Hacer:
(1) **Para cada punto** $Z_i = (x_i, y_i, z_i)^T, i = 1, 2...\rho_{max}$
a) Obtener $\rho = x_i \cos\theta_i + y_i \sin\theta_i$
b) Encontrar posición con respecto θ y ρ
c) Incrementar matriz contadores $acum[k, h] = acum[k, h]$ $+exp(20 * \rho/T_x)$
(2) Encuentra el máximo local $max \leftarrow acum(k_{max}, h_{max})$
(3) $A' = \rho_{max} cos(\theta_{max})$
(4) $B' = \rho_{max} sin(\theta_{max})$
(5) $C' = 0$
(6) $n \leftarrow [A', B', C']D'_0 \leftarrow D'$ Valores Inicializados
(7) Aplicamos segmentación del suelo
(8) Obtener parámetros del plano $[n, D']$ y puntos que no pertenecen al plano encontrado
Mientras: $max > \gamma$
Salida: Planos encontrados

3. Indexación

En esta sección, se modelan globalmente los objetos sin estructura que no fueron extraídos de la nube, se analizan las características del objeto para formar sus firmas. Se generan los histogramas a partir de la dirección de la normal del modelo del objeto. Finalmente se utilizan las métricas de similitud en la comparación entre histogramas.

3.1. Modelado

Las características 3D de un objeto pueden ser representadas por el vector normal a la superficie. En este trabajo calculamos el vector normal a la superficie de manera local. Nosotros proponemos modelar el objeto 3D a partir de la distribución de la dirección de las normales calculadas. El desarrollo del modelado se realizó en tres pasos:

1. Normalización de la orientación. Cada objeto segmentado está formado por un conjunto de puntos Z_i, el cual es orientado respecto a la distribución de los datos. Se calculan los eigen vectores $V = [v_1, v_2, v_3]$ y eigen valores $\{e_1, e_2, e_3\}$ de cada conjunto de puntos, donde $e_1 < e_2 < e_3$, por lo tanto $Z'_i = V^T Z_i$.

2. Triangulación de Delaunay. Se calculan las normales de cada triángulo que forman el objeto segmentado. Un objeto contiene un conjunto de normales $N = \{n_1, n_2, ..., n_j\}$ donde $n_j = [P^j_x, P^j_y, P^j_z]^T$.

3. Finalmente, utilizamos las direcciones de las normales para construir los histogramas $H(\phi, \theta)$ de cada objeto.

Donde $\phi = \tan^{-1}(\frac{P^j_x}{P^j_y})$ y $\theta = \tan^{-1}(\frac{P^j_z}{\sqrt{P^2_x + P^2_y}})$

Modelados los objetos se calculan sus histogramas, se utilizan las características de la normal de cada triangulación, como son: su dirección y repeticiones, para calcular los histogramas esféricos completos. Se utiliza Delaunay por ser considerado un descriptor eficiente.

3.2. Comparación

Un objeto desconocido modelado por un histograma H_i es comparado con el histograma que modela un automóvil de la biblioteca de objetos H_b. Al igual que en Gonzalez-Barbosa y Lacroix (2002) y Luber (2013) se utilizan las técnicas χ^2, intersección de histogramas, distancia de Haussler, distancia Euclidiana y *Earth Mover´s Distance* para la comparación de los histogramas.

Método χ^2. Es un método estadístico y es definido como se muestra en la ecuación 2:

$$\chi^2_{H_b,H_i}(H_b, H_i) = \sum_{\phi} \sum_{\theta} \frac{(H_b(\phi, \theta) - H_i(\phi, \theta))^2}{H_b(\phi, \theta) + H_i(\phi, \theta)} \tag{2}$$

Distancia de Haussler. Es definida por:

$$\|H_b - H_i\| = \sum_{\phi} \sum_{\theta} \frac{H_b(\phi, \theta) - H_i(\phi, \theta)}{1 + H_b(\phi, \theta) + H_i(\phi, \theta)} \tag{3}$$

Distancia Euclidiana. La distancia euclidiana o euclidea se deduce a partir del teorema de Pitágoras. El resultado nos arroja la distancia euclidiana existente, como se demuestra en la ecuación 4:

$$\|H_b - H_i\| = \sqrt{\sum_{\phi} \sum_{\theta} (H_b(\phi, \theta) - H_i(\phi, \theta))^2} \tag{4}$$

Intersección de Histogramas. Es definida como:

$$\|H_b - H_i\| = \sum_{\phi} \sum_{\theta} \frac{H_b(\phi, \theta)}{\sum_{\phi} \sum_{\theta} H_{barea}(\phi, \theta)} - \frac{H_i(\phi, \theta)}{\sum_{\phi} \sum_{\theta} H_{iarea}(\phi, \theta)} \tag{5}$$

Earth Mover´s Distance. El método consiste en determinar el número de posiciones (bins) en el que dividiremos nuestros histogramas. Obteniendo de H_b y H_i el promedio de sus repeticiones, estas repeticiones son las características propias de cada histograma. Posteriormente, se obtiene la normal de las características y de los histogramas H_b y H_i. Finalmente se optimizan todos los datos aplicando la ecuación 6:

$$EMD(H_b, H_i) = \frac{\sum_\phi \sum_\theta f_{\phi\theta} d_{\phi\theta}}{\sum_\phi \sum_\theta f_{\phi\theta}} \quad (6)$$

Aquí $d_{\phi\theta}$ es la distancia existente entre los bins, $f_{\phi\theta}$ es la cantidad de movimientos entre los bin's. Y H_b es la frecuencia observada, H_i es la frecuencia esperada para todas las técnicas.

Se aplica el análisis ROC para optimizar los umbrales de MeanShift, Delaunay y el modelado por histogramas. Se utiliza una biblioteca de datos de entrenamiento de 4500 objetos para optimizar los umbrales y una biblioteca de datos de prueba de 3000 objetos para verificar los resultados del análisis ROC.

4. Resultados

En esta sección, se presentan los resultados obtenidos después de aplicar las técnicas descritas en la sección 2 y 3. Estos resultados reconocen particularmente automóviles 3D que han sido digitalizados en ambientes urbanos. La extracción de planos se aplica para generar los mapas del ambiente, permitiendo mayor rapidez en la segmentación y en la indexación de objetos.

4.1. Extracción de Planos

Aplicamos la trasformada de Hough en busca de un conjunto mayor de puntos que modelen una recta. Consideramos que los planos que pertenecen al ambiente son perpendiculares al suelo, por tal motivo, solo en este módulo reducimos el problema a las coordenadas (x, y) de los puntos. Con esta consideración los planos del ambiente son proyectados a líneas. Cada punto del borde tiene asociado una línea paramétrica.

La intersección de la línea paramétrica indica la existencia y la posición de puntos colineales. Cuanto mayor sea el número de puntos colineales, mayor será la posibilidad de encontrar el plano. En esta etapa las fachadas extraídas corresponden a posiciones paramétricas con número de puntos colineales mayor a 600 puntos. Este parámetro permite que la lateral de un carro no sea confundida como una fachada. Por otro lado, varias fachadas con número de puntos colineales menores a 600 puntos no son detectadas. Sin embargo estos segmentos de fachadas al modelarlas son descartadas como autos (observe la Figura 6 (B), (D) y (F)) que modelan un segmento de fachada que no ha sido extraído por Hough.

La Tabla 1 muestra las ecuaciones correspondientes a la extracción del suelo y planos perpendiculares, para lo cual fue necesario utilizar la metodología descrita por Hernández-García et al. (2011) y el algoritmo 1. En la Tabla 1 se puede ver la extracción de dos planos perpendiculares encontrados y el plano principal (suelo). La extracción de los planos perpendiculares depende del lugar donde se haya decidido realizar la reconstrucción. Así que pueden existir n extracciones de planos si así se requiere.

La extracción de puntos pertenecientes a objetos estructurados es fundamental en nuestro trabajo, ya que descartamos objetos que no corresponden a autos. Esto nos ayuda a disminuir el tiempo de procesamiento de la información.

Tabla 1: Extracción de objetos estructurados. La nube de puntos tiene un total de 62858 puntos, de los cuales se extrajeron 35257 puntos pertenecientes al suelo y fachadas. La ecuación del plano está definida por $AX + BY + CZ + D = 0$

Planos	*A*	*B*	*C*	*D*	**# Puntos**
Suelo	-0.0076	-0.0012	1.0000	184.02	21694
Plano 1	0.1439	0.9896	0.0068	-583.73	13563
Plano 2	0.1748	0.9845	-0.0127	141.48	764
No segmentados					26837
Total					62858

La lateral de un auto no es confundida con un plano, debido a que el número de puntos de la lateral del auto es menor al número de puntos de un plano. Si por alguna razón varias laterales de automóviles estuvieran alineadas, los umbrales utilizados para detectar las fachadas permiten no segmentar estas laterales como un plano.

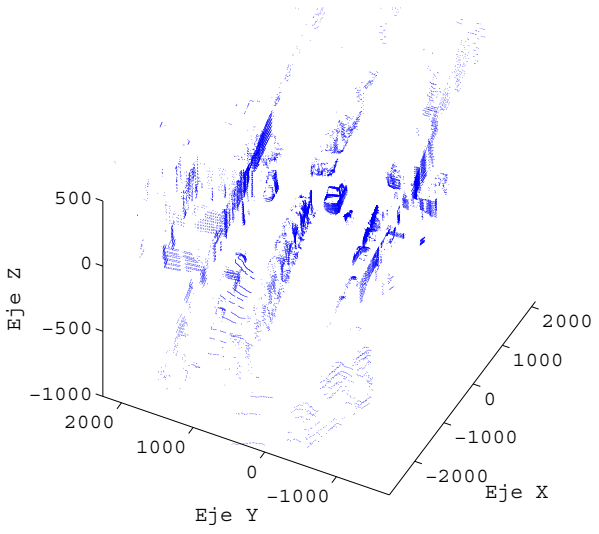

Figura 3: Nube típica sin objetos estructurados. Algunos puntos pertenecientes a planos no han sido segmentados debido al número de puntos que lo componen o al umbral usado para la segmentación.

Después de extraer el suelo y las fachadas de edificios, la nube de puntos restante contiene objetos no estructurados, como se muestra en la Figura 3. De esta manera dejamos de considerar una gran cantidad de puntos, para posteriormente aplicar la segmentación de objetos no estructurados. Los segmentos de planos que se observan en la Figura 3 no fueron segmentados por completo debido al número de puntos que lo componen y al umbral usado en Hough.

4.2. *Segmentación de Objetos No Estructurados*

La Figura 4 muestra los objetos segmentados por el método MeanShift, los puntos que pertenecen al mismo objeto son pintados en color y etiquetados automáticamente con la finalidad de conocer la posición de ese objeto dentro de la nube de puntos. Los clúster de colores corresponden a diferentes objetos. Se han usado colores repetidos debido a las limitantes gráficas. El uso de un umbral con valor a 150 puntos elimina objetos más pequeños, así como ruido.

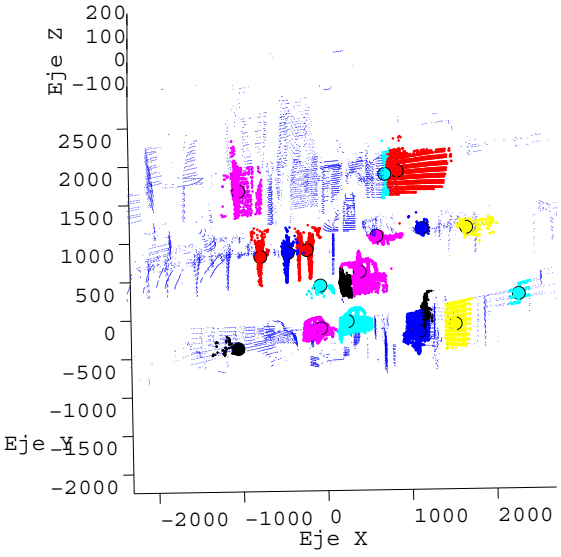

Figura 4: Segmentación MeanShift. Se muestran los objetos segmentados por MeanShift, los puntos que pertenecen a un mismo objetos son pintados de un mismo color, los clúster de colores corresponden a diferentes objetos. Se han usado colores repetidos debido a las limitantes gráficas, se desprecian objetos formados por menos de 150 puntos.

La segmentación con MeanShift es optimizada para cada nube de puntos utilizando análisis ROC. En la Figura 4 se puede apreciar la segmentación de la mayor parte de los automóviles, árboles y segmentos de planos. El resultado de la segmentación es utilizada en la etapa de indexación.

4.3. *Modelado*

Una vez aplicado el método MeanShift se tienen segmentados diversos objetos, como personas, casetas telefónicas, árboles, segmentos de pared, pilares y automóviles. Estos objetos segmentados son modelados por histogramas de la dirección normal. La indexación nos ayuda a identificar los objetos en base a su histograma o firma particular. El obtener un buen modelado de los objetos es fundamental para obtener una buena detección.

La triangulación de Delaunay modela los objetos utilizando triángulos, donde la normal de cada triángulo es utilizada para construir los histogramas de cada objeto. Se optimizaron los umbrales utilizando análisis ROC. Para el método de MeanShift se optimizó el radio del perfil del kernel. En Delaunay se optimizaron dos variables: la primera, define el máximo nivel de tolerancia alcanzable para evitar errores de propagación cuando un tetraedro está mal marcado y la segunda tolerancia decrece a razón de colocar en el interior o exterior los triángulos del tetraedro. Finalmente en el modelado por histogramas se optimizó el número de bin's de los histogramas. Las Figuras 5 (A) y 5 (B) muestran la nube de puntos de un automóvil y un árbol, respectivamente. En las Figuras 5 (C) y 5 (D) se muestra el mismo automóvil y el mismo árbol modelado con Delaunay. Computacionalmente tiene una gran ventaja obtener el modelo debido a que podemos definir las características propias de cada objeto, esto en base a su forma. Y finalmente la Figura 5 (E) y 5 (F) muestran los histogramas particulares del automóvil y el árbol. En las Figuras 6 (A) y (B) se muestra un contenedor de basura y un fragmento de muro, estos objetos frecuentemente los podemos encontrar en áreas urbanas, en las Figuras 6 (C) y (D) podemos observar el modelo de los objetos, y en las Figuras 6 (E) y (F) se muestran los histogramas de los objetos, se puede observar en la Figura 6 (E) y en la Figura 5 (E) histogramas con distribución de frecuencias diferentes por lo cual no se puede detectar un contenedor de basura como automóvil aunque cuentan con características semejantes como el tamaño y la forma, esto debido a las curvas más pronunciadas en el auto. En estas dos últimas figuras se utilizan las métricas descritas en la sección 3.2 para cuantificar la diferencia de los histogramas.

Utilizamos cuatro modelos de automóviles diferentes para detectar con mayor eficiencia los automóviles presentes en nuestros escaneos. Los automóviles tomados como muestra son los más comunes en México y son: el sedán, el compacto que son autos no mayores a 4,7m y no menores a 4m de largo, los utilitarios mejor conocidos como pick-up o SUV por sus siglas en inglés (*Sport Utility Vehicle*) y los denominados hatchbacks que consisten en una cabina para pasajeros y un área de carga. El utilizar estos diferentes tipos de automóviles ayudo a optimizar el algoritmo y detectar de forma correcta la mayor parte de los vehículos reconstruidos.

4.4. *Comparación*

Nuestras bases de datos de entrenamiento y prueba fueron formadas con 4500 y 3000 objetos respectivamente. Estas bases de datos contienen objetos como: automóviles, postes, árboles, segmentos de planos o suelo, personas entre otros. En la Tabla 3 se muestran las matrices de confusión de la base de datos de prueba obtenidas después de aplicar los cinco métodos que se describieron en la sección 3.2.

En las Figuras 7 y 8 se muestran las curvas ROC de los cinco métodos descritos en la sección . El mejor método para detectar objetos utilizando esta técnica es χ^2 con un 81,30 % de probabilidad de que las detecciones sean correctas (vea la Figura 7). Estos resultados concuerdan con Gonzalez-Barbosa y Lacroix (2002) quien utiliza estas mismas medidas de comparación para imágenes panorámicas que han sido representadas como histogramas de apariencias locales. Luber (2013) utiliza las mismas métricas en el seguimiento de personas bajo diferentes restricciones sociales. Las cinco maneras de comparar se

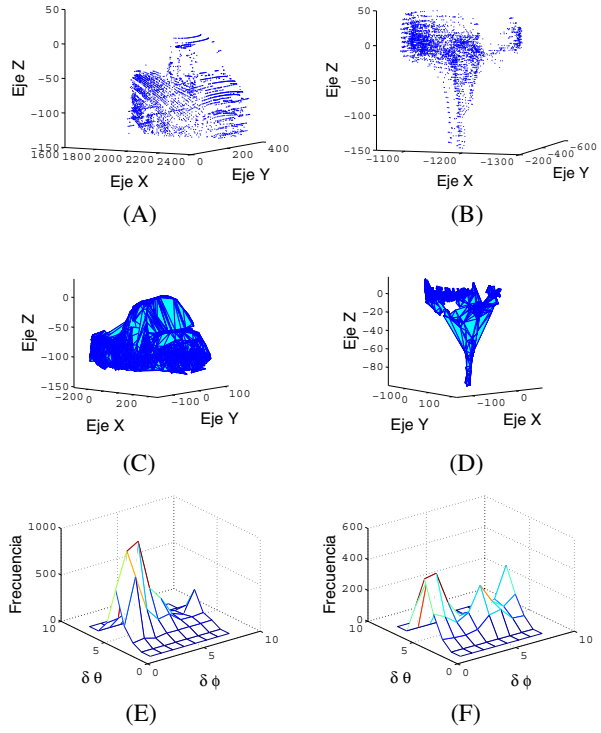

Figura 5: Modelado. En (A) y (B) se muestran los puntos correspondientes a un auto y un árbol respectivamente, segmentados con MeanShift. Mientras que en (C) y (D) se observa el modelo aplicando Delaunay. Y finalmente (E) y (F) muestran los histogramas de los objetos construidos en base a la normal de cada triangulación.

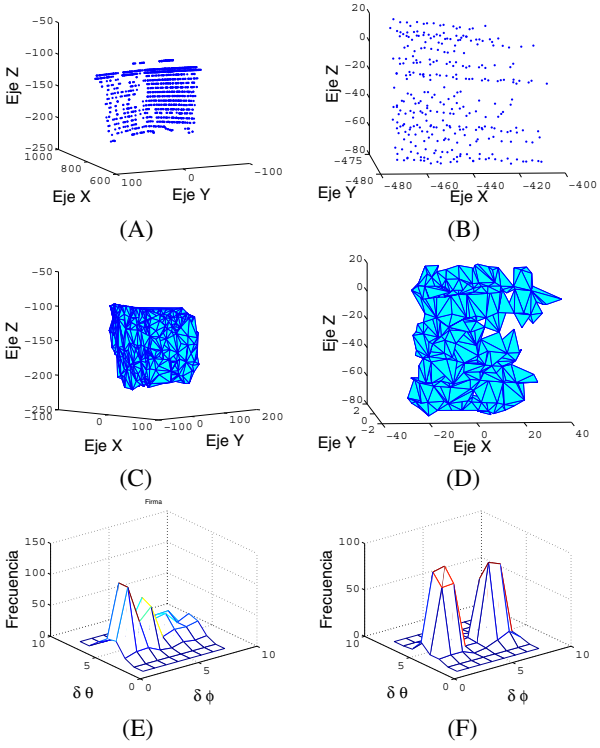

Figura 6: Modelado. En (A) y (B) se observan un contenedor de basura y un fragmento de muro. Se muestran en (C) y (D) los modelos de los objetos. En (E) y (F) se presentan los histogramas.

han separado en dos gráficas con el propósito de apreciar mejor los resultados.

La Tabla 2 ha sido calculada utilizando la base de datos de prueba, la cual está compuesta de 3000 objetos. En esta tabla se presentan las cinco métricas utilizadas y el área bajo la curva de cada métrica. Siendo Λ el área bajo la curva $\{x \in \mathbb{R} \mid b \leq \Lambda \leq c\}$ los valores de b y c corresponde a 1 y 0.5 respectivamente. Si Λ se acercan más a la unidad los métodos son mejores, sin embargo, entre más se acerquen a 0.5 los métodos son considerados ineficientes Park et al. (2004).

La Figura 6 (B), (D) y (F) muestra una fachada extraída por MeanShift mientras que la Figura 9 (A), (B) y (C) muestra una fachada extraída con Hough. La densidad de puntos de los planos segmentados por Hough es mucho mayor a los planos segmentados por MeanShift. Sin embargo, los histogramas representados en ambas figuras visualmente son muy similares.

Tabla 2: Comparación cuantitativa de las diferentes métricas de similitud.

Métrica	**Área bajo la curva Λ**
χ^2	0.8130
Intersección	0.8060
Haussler	0.8114
Euclidiana	0.8045
EMD	0.7798

5. Conclusión

En este trabajo se desarrolló una nueva metodología para la detección de automóviles. La propuesta consta de dos partes: segmentación e indexación. La segmentación descarta los puntos que pertenece al suelo y fachadas. El algoritmo 1 extrae fachadas en cualquier tipo de reconstrucción 3D de ambientes estructurados. La técnica no paramétrica MeanShift es aplicada a los puntos restantes para segmentar planos con pocos puntos y objetos no estructurados presentes en el medio ambiente, tales como árboles, personas, autos, postes, etc. En la etapa de indexación, los objetos son representados con una firma basada en histogramas, la cual es construida usando las direcciones de las normales del objeto segmentado. La metodología fue aplicada a dos bases de datos. La primera corresponde a la base de datos de entrenamiento y está compuesta por 4500 objetos. En esta parte se calcularon los umbrales óptimos usados en el sistema. Una vez conocidos los umbrales óptimos se repitió el proceso en una base de prueba con el objetivo de evaluar el sistema. La base de prueba consta de 3000 objetos, la metodología muestra una correcta detección de autos del 81.30 % cuando se aplica la métrica χ^2. El detectar objetos de nubes de puntos puede tener diferentes beneficios; como por ejemplo en visitas 3D guiadas por computadora de un área de interés, como pudiera ser el cen-

tro histórico de una ciudad importante. Al usuario le interesaría observar la zona de interés sin que exista información no pertenece al área como pudieran ser los automóviles. En visitas guiadas, muchos de los escenarios de interés mundial están llenos de información que no pertenece al área y en muchos de los casos estos objetos obstaculizan la visión de las zonas de interés. Lo que se pretende con nuestro trabajo es proponer una metodología que permita detectar y eliminar esos puntos que no son parte del área reconstruida, para posteriormente generar una reconstrucción limpia. La metodología desarrollada en este trabajo está ampliamente estudiada para cualquier tipo de reconstrucción sin importar la topología del terreno donde fue levantada la información. El levantamiento de los datos se realizó en movimiento y estáticamente. En la segmentación, algunos objetos pierden parte de sus características, debido a que son segmentados parcialmente o un mismo objeto es segmentado en dos secciones diferentes. Este tipo de segmentación no permite obtener el modelo completo del objeto y por consecuente obtener histogramas poco reales, lo cual afecta la correcta detección de los objetos. Por lo tanto, los trabajos futuros estarán orientados a mejorar la segmentación, y estudiar la información de los objetos que puedan ser usados para la detección, como pudieran ser el tamaño, número de superficies u otras características.

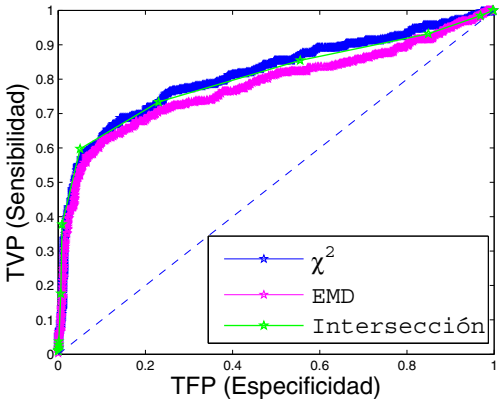

Figura 7: Comparación de Métodos. Se presentan los tres mejores métodos de comparación: χ^2, EMD e Intersección. El área bajo la curva nos indica que la distancia χ^2 es el mejor método de comparación de histogramas.

English Summary

Vehicle Detection in Urban Environments Scanned by a Lidar.

Abstract

Detection of vehicles on 3D point clouds is performed by using the algorithm presented in this work. Point clouds correspond to urban environments and were acquired with the LIDAR Velodyne HDL-64E. The environment is considered semi-structured so that can be modeled using planes.

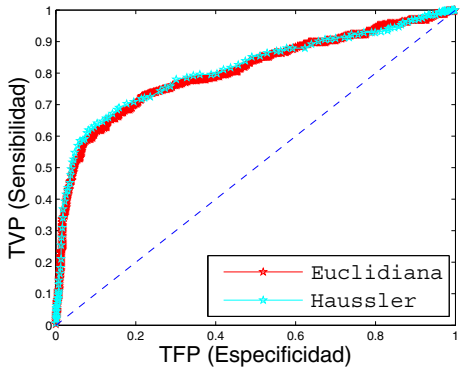

Figura 8: Comparación de Métodos. Se presentan las distancias Euclidiana y Haussler.

Vehicle detection is carried out on to stages, segmentation and indexation. First stage is at the same time composed of three sub-stages. In the first one the principal plane (in this case the floor) is extracted, in the second sub-stage secondary planes are extracted using a tailored version of Hough's method, secondary planes are those perpendicular to the main plane. Finally in the third sub-stage and using MeanShift method, the remaining objects are segmented.

Indexation on its side is divided into two sub-stages, in the first one, last segmented objects using MeanShift method are modeled using histograms according to the direction of the object's 3D points normal; in the second stage histograms are compared to those previously stored on a database of object's histograms. Optimizing of detection thresholds was carried out through ROC analysis. Two databases were used during the experiments, the first DB have 4500 objects and was used for ROC analysis training; the second one contained 3000 objects and was used for verification.

Keywords:

3D point cloud, Lidar, 3D Segmentation.

Agradecimientos

Los autores desean agradecer el apoyo financiero recibido durante el desarrollo de este trabajo al Consejo Nacional de Ciencia y Tecnología (CONACYT) a través del proyecto SEP-2005-O1-51004 / 25293 y el número de CVU 348861 y al Instituto Politécnico Nacional a través del proyecto SIP-20144623.

Referencias

Atanacio-Jiménez, G., González-Barbosa, J.-J., Hurtado-Ramos, J. B., Ornelas-Rodríguez, F. J., Jiménez-Hernández, H., García-Ramírez, T., González-Barbosa, R., 2011. Lidar velodyne hdl-64e calibration using pattern planes. International Journal of Advanced Robotic Systems Vol. 8. (Núm. 5), pp. 70–82.

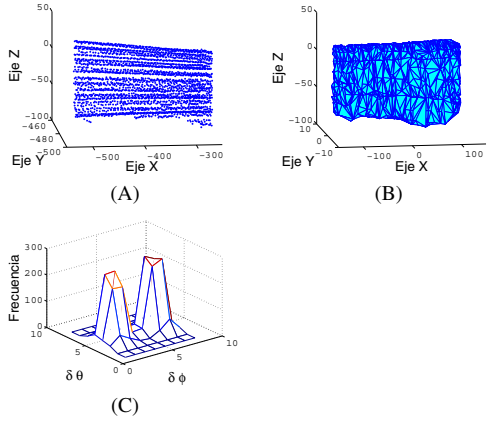

Figura 9: Modelado. En (A) y (B) se observa un fragmento de muro detectado por Hough (ver el algoritmo 1). Se presenta en (B) el modelo y en (C) su histograma.

Tabla 3: Matriz de Confusión. Se usa un automóvil modelado como objeto base y es comparado con los restantes objetos modelados. Se utiliza la base de datos de prueba de 3000 objetos.

χ^2				Predicción
	85.82 %	31.27 %	117.09 %	Autos (P')
	14.18 %	68.73 %	82.91 %	Otros (N')
	100 %	100 %		
Valor real	Autos (P)	Otros(N)		
Intersección de Histogramas				Predicción
	76.96 %	26.63 %	103.58 %	Autos (P')
	23.04 %	73.37 %	96.42 %	Otros (N')
	100 %	100 %		
Valor real	Autos (P)	Otros(N)		
Distancia de Haussler				Predicción
	82.27 %	29.72 %	111.99 %	Autos (P')
	17.73 %	70.28 %	88.01 %	Otros (N')
	100 %	100 %		
Valor real	Autos (P)	Otros(N)		
Distancia Euclidiana				Predicción
	78.43 %	27.24 %	105.67 %	Autos (P')
	21.57 %	72.76 %	94.33 %	Otros (N')
	100 %	100 %		
Valor real	Autos (P)	Otros(N)		
Earth Mover Distance				Predicción
	82.13 %	32.20 %	114.33 %	Autos (P')
	17.87 %	67.80 %	85.67 %	Otros (N')
	100 %	100 %		
Valor real	Autos (P)	Otros(N)		

Barnea, S., Filin, S., 2008. Keypoint based autonomous registration of terrestrial laser point-clouds. ISPRS Journal of Photogrammetry and Remote Sensing Vol. 63. (Núm. 1), pp. 19–35.

BIPM, 2008. Evaluation of measurement dataguide to the expression of uncertainty in measurement. JCGM 100.

Brostow, G. J., Shotton, J., Fauqueur, J., Cipolla, R., 2008. Segmentation and recognition using structure from motion point clouds. En: European Conference on Computer Vision (ECCV).

Douillard, B., Underwood, J. P., Kuntz, N., Vlaskine, V., Quadros, A. J., Morton, P., Frenkel, A., 2011. On the segmentation of 3d lidar point clouds. En: International Conference on Robotics and Automation (ICRA). IEEE, pp. 2798–2805.

García-Moreno, A. ., Gonzalez-Barbosa, J. ., Ornelas-Rodríguez, F. ., Hurtado-Ramos, J. ., Ramirez-Pedraza, A., González-Barbosa, E. ., 2013. Automatic 3d city reconstruction platform using a lidar and dgps. Vol. 7629. LNAI of Lecture Notes in Computer Science. pp. 285–297.

Golovinskiy, A., Kim, V. G., Funkhouser, T., 2009. Shape-based recognition of 3D point clouds in urban environments. En: International Conference on Computer Vision (ICCV).

Gonzalez-Barbosa, J.-J., Lacroix, S., 2002. Rover localization in natural environments by indexing panoramic images. En: International Conference on Robotics and Automation (ICRA). IEEE, Vol. 2. (pp. 1365 - 1370).

Goron, L., Tamas, L., Reti, I., Lazea, G., 2010. 3d laser scanning system and 3d segmentation of urban scenes. En: International Conference on Automation, Quality and Testing, Robotics (AQTR), IEEE. Vol. 1. pp. 1–5.

Hernández-García, D.-E., Gonzalez-Barbosa, J.-J., Hurtado-Ramos, J.-B., Ornelas-Rodríguez, F.-J., Castillo Castaneda, E., Ramírez, A., García, A., Gonzalez-Barbosa, R., Aviña-Cervantez, J., 2011. 3d city models: Mapping approach using lidar technology. En: 21st International Conference on Electronics, Communications and Computers (CONIELECOMP). pp. 206–211.

Himmelsbach, M., Luettel, T., Wuensche, H.-J., 2009. Real-Time Object Classification in 3D Point Clouds Using Point Feature Histograms. En: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS. St. Louis, MO, USA, pp. 994–1000.

Huber, D., 2002. Automatic three-dimensional modeling from reality. Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

Lai, K., Fox, D., 2010. Object recognition in 3d point clouds using web data and domain adaptation. International Journal of Robotics Research Vol. 29. (Núm. 8), pp. 1019–1037.

Liang-Chia Chen, Hoang Hong Hai, X.-L. N., Wu, H.-W., 2013. Novel 3-d object recognition methodology employing a curvature-based histogram. International Journal of Advanced Robotic Systems Vol. 10. (Núm. 294).

Luber, M., 2013. People tracking under social constraints. Ph.D. thesis, Universitätsbibliothek Freiburg.

Park, S. H., Goo, J. M., Jo, C.-H., 2004. Receiver operating characteristic (roc) curve: practical review for radiologists. Korean Journal of Radiology Vol. 5 (Núm. 1), pp. 11–18.

Pulido Fentanes Jaime, Zalama Eduardo, G. G.-B. J., 2012. Plataforma robótica para tareas de reconstrucción tridimensional de entornos exteriores. RIAI Vol. 46. (Núm. 01), pp. 81–92.

Schnabel, R., Wahl, R., Klein, R., 2007. Efficient ransac for point-cloud shape detection. En: Computer graphics forum. Vol. 26. Wiley Online Library, pp. 214–226.

Sohn, G., Jwa, Y., Kim, H. B., 2012. Automatic powerline scene classification and reconstruction using airborne lidar data. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. I-3, pp. 167–172.

Steder, B., Grisetti, G., Burgard, W., 2010. Robust place recognition for 3d range data based on point features. En: International Conference on Robotics and Automation (ICRA), IEEE. pp. 1400–1405.

Sun, S., Salvaggio, C., 2013. Aerial 3d building detection and modeling from airborne lidar point clouds. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing Vol. 6. (Núm. 3), pp. 1440–1449.

Truong, H. Q., Hmida, H. B., Marbs, A., Boochs, F., 2010. Integration of knowledge into the detection of objects in point clouds. En: Paparoditis, N., Pierrot-Deseilligny, M., Mallet, C., Tournaire, O. (Eds.), Proceedings of the ISPRS Commission III symposium - PCV. Vol. 38. of International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 143–148.

Wang, C.-C., Thorpe, C., Suppe, A., 2003. LADAR-based detection and tracking of moving objects from a ground vehicle at high speeds. En: Intelligent Vehicle Symposium IEEE.

Xu, F., Fujimura, K., 2003. Human detection using depth and gray images. En: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 115–121.

Zhou, Q.-Y., Neumann, U., 2008. Fast and extensible building modeling from airborne lidar data. En: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems (GIS). pp. 7:1–7:8.