PRELIMINARY MACHINE VISION SYSTEM FOR GLASS SEGMENTATION ON SKYSCRAPER WINDOWS

SISTEMA DE VISIÓN ARTIFICIAL PRELIMINAR PARA LA SEGMENTACIÓN DE VIDRIO EN VENTANAS DE RASCACIELOS

Ing. Artur Mosquera Mykh

Universidad de Pamplona, Facultad de Ingenierías y Arquitectura,
Maestría en Controles Industriales,
Pamplona – Colombia, Km 1 vía Bucaramanga.
Tel.: (+57 7) 568 5303, Fax: (+57 7) 568 5303, Ext. 144.
E-mail: artur_mosquera@hotmail.com

Abstract: A robotic arm prototype designed to be mounted on a drone and do the window cleaning on skyscraper windows, requires a machine vision system able to obtain the values of the height of the window, width of the window and the position of the window's center in relation to the position of the camera on the X, Y and Z coordinates. This paper presents a preliminary machine vision system for glass segmentation on skyscraper windows, using a database compiled from stock images websites and applying on them edge detection algorithms such as Canny, Sobel, Prewitt, Roberts and logarithmic, Hough transform for line detection, and region properties to detect and segment rectangular areas. Lastly, the steps to optimize the program using support vector machines, neural networks and a multi-agent system using a decision tree are explored.

Keywords: Edge detection, skyscraper, window, Hough transform.

Resumen: Un prototipo de brazo robótico diseñado para ser montado sobre un dron y realizar el trabajo de limpieza de ventanas de rascacielos, requiere de un sistema de visión artificial que obtenga los parámetros de la altura de la ventana, el ancho de la ventana y la posición del centro de la ventana con respecto a la cámara en los ejes X, Y y Z. Este articulo presenta un sistema de visión artificial preliminar para segmentar el vidrio de las ventanas de rascacielos, empleando una base de datos compilada de páginas de imágenes stock, y aplicando en ellas técnicas de detección de bordes como Canny, Sobel, Prewitt, Roberts y logarítmico, transformada de Hough para la detección de líneas y propiedades de regiones para detectar y segmentar áreas rectangulares. Por último, se exploran los pasos a tomar para optimizar el programa resultante mediante maquinas de soporte vectorial, redes neuronales y un sistema multiagente empleando un árbol de decisiones.

Palabras clave: Detección de bordes, rascacielo, ventana, transformada de Hough.

1. INTRODUCION

Aprovechando los avances tecnológicos de los drones, el prototipo de brazo robótico que se está desarrollando con la compañía J&D Trading Group Inc, está diseñado para ser montado sobre un dron y realizar el trabajo de limpieza de ventanas de rascacielos. Este prototipo inicial requiere de un operario para realizar las tareas de limpieza, adicional al requerido por la Federal Aviation Administration (FAA) para volar el dron seleccionado. Debido a esto, luego de realizar el análisis de los costos de operación del modelo propuesto, se llegó a la conclusión de que es indispensable que este sea automatizado. Debido a la variedad de condiciones y características físicas de las ventanas de rascacielos, la parte más importante y extensa de esta automatización es el requerimiento de un sistema de visión artificial robusto y preciso, que

sea capaz de indicarle al brazo robótico el lugar donde se encuentra el vidrio de la ventana del rascacielo con respecto al dron, y el tamaño de esta, sin importar el tipo de ventana, condiciones de iluminación u otros factores que influyan en la obtención de las variables requeridas.

En este artículo se presenta una metodología para la segmentación de vidrio en ventanas de rascacielos a partir de detectores de bordes, filtros, técnicas de detección de líneas y parámetros de regiones. Los resultados obtenidos por este programa preliminar demuestran que es posible obtener las variables requeridas, y exponen las estrategias futuras que optimizarán la confiabilidad y precisión de la segmentación.

2. ANTECEDENTES

El número de rascacielos en el mundo está en constante incremento, sin embargo con los recientes avances en la optimización de sus estructuras (Aldwaik & Adeli, 2014), se puede apreciar un aumento considerable con respecto a años anteriores. El año 2017 fue el cuarto año consecutivo en el que se supera el record anterior de número de rascacielos mayores a 200m construidos, con 144 nuevos edificios completados en 2017 (Fig. 1), comparado al record anterior de 127 completados en 2016. Esto llevó el número total de edificios de más de 200 metros de altura en el mundo a 1319, mostrando un incremento de 12.3% desde 2016 y marcando un incremento de 402% desde el año 2010. (Council on Tall Buildings and Urban Habitat, 2018).

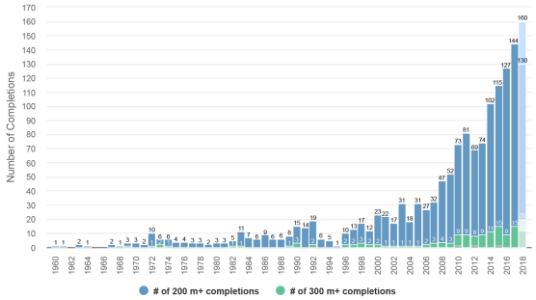


Fig. 1. Edificios de más de 200m y 300m completados por año entre 1960 y 2017.
Fuente: Council on Tall Buildings and Urban Habitat, 2018.

Al caracterizar la dinámica de crecimiento de los rascacielos, evaluar los requerimientos para efectuar la limpieza de sus ventanas e identificar las tecnológicas empleadas en los últimos años para su limpieza, se hizo evidente la necesidad de desarrollar nuevas alternativas que realicen esta tarea de forma más segura, rápida y precisa. Una posible alternativa involucra el aprovechamiento de los avances en el tiempo de vuelo y capacidad de carga de los drones para integrarlos con brazos robóticos diseñados con las limitantes de los drones en mente. Estas dos industrias paralelas de desarrollo rápido combinan sus fortalezas para lograr tareas que antes no eran posibles; los drones pueden llegar a varios lugares, pero están limitados en cuanto a las operaciones que pueden hacer, mientras que los brazos robóticos tienen un área de trabajo y capacidades extensas, pero están limitados a un espacio en particular (Mosquera, 2017).

En cuanto al procesamiento digital de la imagen, este se puede dividir en tres categorías, procesos de nivel bajo, medio y alto. Se consideran procesos de bajo nivel a operaciones como un preprocesamiento para la reducción de ruido o aumentar el contraste, y se caracterizan por el hecho de que tanto la entrada como la salida resultante son una imagen. El procesamiento de nivel medio involucra tareas como partición de una imagen en regiones u objetos, descripción de esos objetos para reducirlos a una forma apropiada para que el computador los procese, y clasificación y reconocimiento de objetos individuales. Este tipo de procesamiento se caracteriza por el hecho que, aunque la entrada en una imagen, la salida son atributos extraídos de esas imágenes, como por ejemplo bordes o contornos. Finalmente, un procesamiento de nivel alto involucra un sistema que realice funciones cognitivas normalmente asociadas con la visión humana (Gonzalez, Woods & Eddins, 2004).

El procesamiento de alto nivel se basa en tener el conocimiento de cómo se pretende lograr la tarea, partiendo de un modelo formal del entorno, y luego, como la realidad percibida por la imagen digital se compara al modelo. A partir de las diferencias que se presenten, se añaden nuevas tareas que permitan sobreponerse a estas diferencias. Con el fin de lograr lo anterior, se hace un cambio a procesamiento de bajo nivel para encontrar información que permita actualizar el modelo, y se itera el proceso resultante en un bucle donde el procesamiento de alto nivel genera tareas parciales para el procesamiento de bajo nivel (Sonka, Hlavac & Boyle, 2014).

Cuando se presenta un nuevo problema es necesario entender como los métodos y algoritmos funcionan para encontrar la metodología que lleve a una posible solución. Por esta razón, la visión artificial se puede considerar una disciplina que da solución al cómo combinar imágenes derivadas de sensores, modelos previamente obtenidos de la escena, y el conocimiento del procesamiento de imágenes pata construir una descripción explicita del

modelo del entorno. Este modelo permite resolver una imagen que no ha sido interpretada, empleando una descomposición de la imagen en partes independientes y coherentes, para que el subsecuente análisis pueda dar solución a una colección de problemas pequeños en lugar de un solo problema complejo (Fischler & Firschein, 2014).

La detección de ventanas en la visión artificial se utiliza en el campo de la fotogrametría, centrándose en las diferentes características de la imagen como color, histograma, texturas, para clasificar los elementos de una fachada y saber que parte es una ventana, puerta, muro, cielo, vegetación u otro elemento. Esto se realiza con el fin de mejorar los modelos 3D de los edificios, reconstruir edificios en 3D o realizar análisis de deformaciones a partir de imágenes en alta resolución (Yang, Förstner & Chai, 2012).

Existen bases de datos que ayudan en esta tarea, como la de eTRIMS (Fig. 2), la cual proporciona varias muestras de fachadas juntos con imágenes que indican la segmentación de objetos, la clase a la que pertenece cada objeto y los bordes de los objetos detectados (Korč and Förstner, 2009).

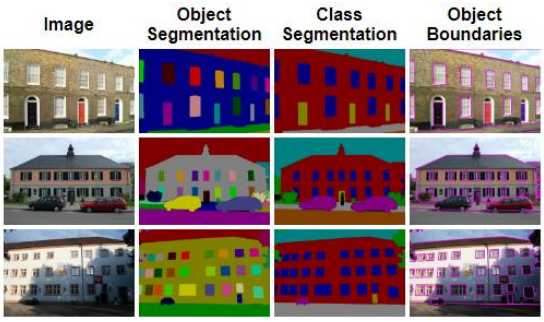


Fig. 2. Ejemplos de la base de datos eTRIMS.
Fuente: Korč and Förstner, 2009.

Generalmente para la detección de ventanas se emplean características de los bordes de los marcos, puesto que estas tienen una falta de características propias, cambiando según reflejos u objetos que se ven a través de ellas. Otro factor importante, es que en fachadas completas pueden presentarse objetos que generan respuestas en los detectores de bordes pero que no son ventanas, por lo cual es necesario conocer a cual elemento pertenece cada borde o sección detectada (Neuhausen, Koch & König, 2016).

En el caso de este proyecto, no es necesaria la detección de la ventana sino la segmentación de la parte de la ventana que debe ser limpiada (vidrio), para lo cual no ha habido un enfoque específico. Esto puede deberse principalmente a la falta de necesidad o aplicación de este resultado. Sin embargo, el enfoque para lograr esto comparte bases teóricas con otras aplicaciones como el reconocimiento de placas de autos (Azad & Baghdadi, 2014), la detección de elementos en fachadas mencionada anteriormente y la detección de señales de tráfico (Berkaya, Gunduz, Ozsen, Akinlar & Gunal, 2016), en especial las rectangulares (Pink & Eickeler, 2016).

3. METODOLOGIA

El documento corresponde a un trabajo investigativo de carácter descriptivo enfocándose en los resultados del proceso de creación de un sistema de visión artificial preliminar para la segmentación de vidrio en ventanas de rascacielos.

El proyecto inicia con la búsqueda de distintas fotos de ventanas de rascacielos que más se asemejan a las que tomaría una cámara montada sobre un dron, es decir, directamente en frente de la ventana con el vidrio a segmentar. En este caso, fotos tomadas por quadcopteros volando cerca de ventanas de rascacielos o fotos de las fachadas tomadas desde edificios opuestos. Estas fueron encontradas en páginas web de imágenes y videos de stock completamente gratuitos o con vista previa gratuita como GettyImages, ShutterStock, iStock, entre otros.

La prioridad fue conseguir la mayor variedad posible en cuanto a condiciones de iluminación, marcos de distintos tamaños, grosores y tonalidades, variación entre marcos en una sola ventana, vidrio que permiten ver dentro del edificio y los objetos que se ven (mesas, sillas, columnas, persianas, escritorios), vidrio completamente reflectante y los objetos que se reflejan (otros edificios, grúas de construcción, cielo despejado o nublado, sol tanto en los marcos como en el vidrio), entre otros. Mientras más difícil pareciese la segmentación mejor, puesto que garantiza un programa final más robusto.

Partiendo de la base de datos se evaluaron las necesidades de detección y segmentación, analizando las características de diferentes ventanas de rascacielos actuales y la variedad de condiciones a las que estas pueden ser expuestas.

Con los resultados obtenidos se seleccionaron los métodos o técnicas de visión artificial que permitían la detección y segmentación de vidrio en ventanas de rascacielos. Esto abarcó técnicas de pre-procesamiento como la corrección de defectos en la imagen, ruido, contraste, iluminación no uniforme, color y distorsiones geométricas, así como algoritmos y técnicas de detección de bordes como Canny, Sobel, Prewitt,

Roberts y logarítmico, técnicas de detección de líneas, y propiedades de áreas de regiones, especialmente el de áreas rectangulares debido a que es la forma predominante de ventanas de rascacielos.

Luego de seleccionar los métodos que más se adecuaban a los requerimientos del proyecto se realizó un programa preliminar de segmentación de vidrio en ventanas de rascacielos. Este programa cuenta con una comparación visual de los métodos que presentaron resultados positivos en algunas de las ventanas de la base de datos, teniendo en cuenta las fortalezas y debilidades de cada método.

4. RESULTADOS

Siguiendo los parámetros establecidos en la metodología, se compilaron más de cien fotos de rascacielos encontradas en páginas web de imágenes y videos de stock como GettyImages, iStock y ShutterStock. A partir de estas se seleccionaron y se recortaron 400 imágenes de ventanas (Fig. 3), siendo estas la base de datos para las pruebas del sistema de visión artificial (Fig. 4).



Fig. 3. Base de datos de 400 ventanas de rascacielos.



Fig. 4. Ejemplos de la base de datos de ventanas de rascacielos.

Como las ventanas de rascacielos tienen una falta de características propias a excepción de los marcos, el programa preliminar para lograr la segmentación del vidrio consiste en detectar los bordes de los marcos mediante los algoritmos de detección de bordes Canny, Sobel, Prewitt, Roberts y Logarítmico, dilatación de los bordes detectados, creación de áreas a partir de los bordes detectados, filtros para eliminar áreas que estén en contacto con los márgenes de la imagen (otras ventanas incompletas) y por último, propiedades de región para detectar formas rectangulares. Lo anterior se programó en Matlab junto con la creación de una interfaz gráfica de usuario hecha con el GUIDE de Matlab (Fig. 5).

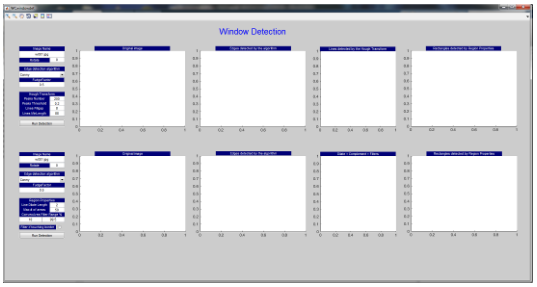


Fig.5. Interfaz gráfica del programa preliminar para la prueba de distintos algoritmos.

Las opciones del programa preliminar permiten seleccionar el algoritmo de bordes, cambio de umbrales, y otros filtros para las propiedades de regiones. Empleando este programa, se compararon los algoritmos de bordes mencionados anteriormente (Fig. 6), donde el de Sobel mostró mejores resultados en cuanto a la precisión de los bordes encontrados y la sensibilidad de la detección en cuanto a las características buscadas de las ventanas de rascacielos.



Fig. 6. Ejemplo de la segmentación empleando distintos detectores de bordes.

Se realizó un programa que aplica el código de segmentado preliminar a cada una de las imágenes de la base de datos, manteniendo un valor fijo en los umbrales de detección. Empleando el detector de bordes Sobel (Fig. 7) y Prewitt, fueron 175 de 400 las segmentaciones correctas, mientras que Roberts (Fig. 8) tuvo el mayor número de segmentaciones exitosas con 186 de 400, pero estas son un poco imprecisas.



Fig. 7. Ejemplos de segmentación exitosa con el algoritmo de detección de bordes Sobel.



Fig. 8. Ejemplos de segmentación exitosa con el algoritmo de detección de bordes Roberts.

En el caso del detector de bordes de Canny, este tuvo el peor número de segmentaciones correctas con 105 de 400, seguido por el detector logarítmico con 139 de 400.

A partir de estos resultados se seleccionaron los algoritmos de detección de bordes Sobel y Roberts como los principales a emplear en pruebas consiguientes de optimización.

A pesar de la segmentación exitosa en varios casos, estos fallaron principalmente en las ventanas con marcos poco definidos, cuando hay objetos como otros edificios, grúas, la línea del horizonte o la ciudad en el reflejo del vidrio, o cuando hay demasiados objetos presentes dentro del edificio que se pueden ver a través del vidrio transparente.

Aumentar los umbrales de los algoritmos de bordes solo aumenta la sensibilidad de la detección, solucionando problemas de bordes poco definidos pero afectando la segmentación del área con los nuevos bordes detectados. La cantidad de secciones de bordes detectados resultan en varias áreas segmentadas pequeñas en lugar del vidrio deseado (Fig. 9).

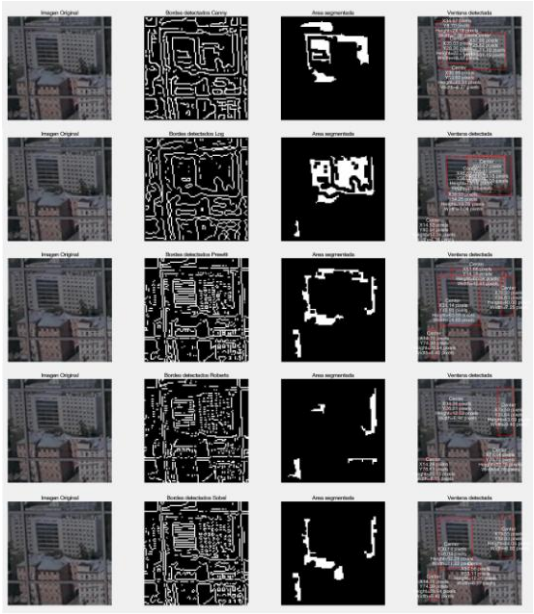


Fig. 9. Problema en la segmentación en las ventanas con reflejos de edificios.

Para solucionar el problema expuesto, se empleó la transformada de Hough para detectar líneas entre los bordes, usando parámetros que permiten obviar discontinuidades entre líneas, y filtrar por la longitud de las líneas detectadas, así como el ángulo de las líneas buscadas (Fig. 10).

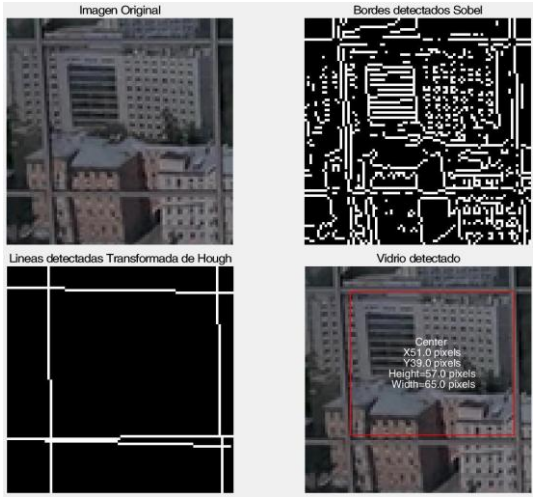


Fig. 10. Segmentación exitosa empleando transformada de Hough.

La transformada de Hough con parámetros fijos y empleando el detector de bordes Sobel logró la segmentación exitosa en 192 de 400 imágenes. Sin embargo, al realizar un ajuste manual de los parámetros, se logró una segmentación exitosa en 364 de las 400 ventanas (Fig, 11). Esto se realizó para demostrar el potencial de este método, así como planear los ajustes automáticos que se realizarán en el programa multiagente final con las distintas características encontradas.



Fig. 11. Segmentación exitosa con transformada de Hough con umbrales variados manualmente.

Con el fin de mejorar la detección de bordes, se decidió cambiar los algoritmos de detección de bordes, a un método de inteligencia artificial para entrenar un programa que sea capaz de diferenciar los bordes de los marcos, del de los bordes en los reflejos y transparencias del vidrio.

El primer método de inteligencia artificial implementado es el de las maquinas de soporte vectorial, para el cual se partió la imagen de la base de datos de imágenes en celdas de 3x3 pixeles (Fig. 12). Se separaron las distintas celdas en categorías de bordes (horizontal, vertical y no borde) y se acomodaron las matrices de características para el entrenamiento del clasificador. Otro método que se probará en el futuro es el de las redes neuronales.

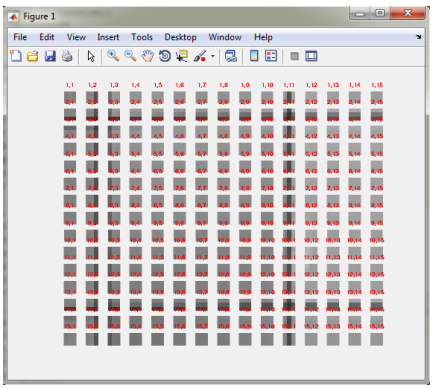


Fig.12. Partición de una imagen de la base de datos en celdas de 3x3 pixeles.

5. CONCLUSIONES

El programa preliminar demuestra que es posible segmentar el vidrio en las ventanas de rascacielos empleando métodos de visión artificial. Sin embargo, el porcentaje de la tasa de fallos es demasiado alto para considerar confiable el programa preliminar.

Quedó en evidencia que mejorando el detector de bordes, es posible mejorar todo el programa de detección preliminar, y el hecho de que hay casos en los que un método es mejor que otro, y al mismo tiempo el método empeora otros que fueron exitosos, demuestra el requerimiento de un árbol de decisiones que seleccione el método y los parámetros según distintos casos y condiciones en ventanas de rascacielos.

El programa preliminar consigue la altura y ancho de la ventana, así como dos de los parámetros de posicionamiento, pero los valores están dados en pixeles. Para conseguir el parámetro faltante es necesario encontrar la mejor forma para calcular la distancia entre la cámara y la vidrio de la ventana, ya sea empleando solo la cámara, sensores de distancia o un hibrido de ambos si es necesario. Al mismo tiempo, esto permitirá hacer una relación entre los pixeles de la imagen y el tamaño en centímetros de la ventana.

Existe la posibilidad de optimizar la precisión de la segmentación y la confiabilidad frente a distintos tipos de ventanas y condiciones de operación, empleando técnicas de inteligencia artificial como las maquinas de soporte vectorial y las redes neuronales, para enseñarle al programa lo que se considera un borde de los marcos de las ventanas.

REFERENCIAS

Aldwaik, M. & Adeli, H. (2014). Advances in optimization of highrise building structures. *Structural and Multidisciplinary Optimization*, *50*(6), 899-919.

Azad, R. & Baghdadi, M. (2014). Novel and Fast Algorithm for Extracting License Plate Location Based on Edge Analysis. *arXiv preprint arXiv:1407.6496*.

Berkaya, S. K., Gunduz, H., Ozsen, O., Akinlar, C. & Gunal, S. (2016). On circular traffic sign detection and recognition. *Expert Systems with Applications*, *48*, 67-75.

Council on Tall Buildings and Urban Habitat (2018). Another Record-Breaker for Skyscraper Completions. Recuperado de: http://www.ctbuh.org/

Fischler, M. A., & Firschein, O. (Eds.). (2014). *Readings in computer vision: issues, problem, principles, and paradigms*. Elsevier.

Gonzalez, R. C., & Woods, R. E. (2002). Digital image processing. Upper Saddle River, New Jersey: Pearson-Prentice-Hall.

Korč, F., Förstner, W. (2009). eTRIMS Image Database for interpreting images of man-made scenes. Technical report, Department of Photogrammetry, University of Bonn.

Mosquera, A. (2017). Limpieza de Ventanas de Rascacielos y Alternativas Tecnológicas Emergentes. Revista Colombiana de Tecnologías de Avanzada ISSN: 1692-7257 - Volumen 2 – Número 30. Universidad de Pamplona. Pamplona, (pp.109-118).

Neuhausen, M., Koch, C., & König, M. (2016). Image-based window detection: an overview.

Pink, L. & Eickeler, S. (2016). Performance Enhancements for the Detection of Rectangular Traffic Signs. In *Advanced Microsystems for Automotive Applications 2016* (pp. 113-123). Springer International Publishing.

Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.

Yang, M. Y., Förstner, W., & Chai, D. (2012). Feature evaluation for building facade images-an empirical study. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences:[XXII ISPRS Congress, Technical Commission I] 39 (2012), Nr. B3* (Vol. 39, No. B3, pp. 513-518). Göttingen: Copernicus GmbH.