

Reconocimiento y segmentación de la mano con una cámara infrarroja

Recognition and segmentation of the hand with an infrared camera

LÓPEZ-DÍAZ, Roberto Enrique †*, NIETO-YÁÑEZ, Alma Delia, HERNÁNDEZ-BÁEZ, Irma Yazmín y VELASCO-CASTILLO, Miguel Ángel

Universidad Politécnica del Estado de Morelos, Dirección de Ingeniería en Informática

ID 1^{er} Autor: Roberto Enrique, López-Díaz / **ORC ID:** 0000-0003-1716-6704, **CVU CONACYT ID:** 212691

ID 1^{er} Coautor: Alma Delia, Nieto-Yáñez / **ORC ID:** 0000-0001-8554-8152, **CVU CONACYT ID:** 212730

ID 2^{do} Coautor: Irma Yazmín, Hernández-Báez / **ORC ID:** 0000-0002-3078-9618, **CVU CONACYT ID:** 212658

ID 3^{er} Coautor: Miguel Ángel, Velasco-Castillo / **ORC ID:** 0000-0003-0787-1872, **CVU CONACYT ID:** 582790

DOI: 10.35429/JID.2019.9.3.1.7

Recibido 09 Octubre, 2019; Aceptado 03 de Diciembre, 2019

Resumen

En este artículo se presenta como a través de una cámara infrarroja se obtienen imágenes bajo diferentes condiciones de luz que permiten segmentar, clasificar y reconocer gestos realizados con una mano. La cámara infrarroja utilizada se construyó haciendo pequeñas modificaciones físicas a una cámara web convencional y agregando un filtro infrarrojo. Adicionalmente, se diseñaron estructuras de LEDs infrarrojos para la iluminación de la escena, logrando que pueda ser utilizada en diferentes condiciones de luz

Lógica difusa, Vhdl, Temperatura

Abstract

In this article we present how, through an infrared camera, images are obtained under different light conditions that allow us to segment, classify and recognize gestures made with one hand. The infrared camera used was built by making small physical modifications to a conventional webcam and adding an infrared filter. Additionally, infrared LED structures were designed for the lighting of the scene, making it possible to be used in different light conditions.

Fuzzy logic, Vhdl, Temperature

Citación: LÓPEZ-DÍAZ, Roberto Enrique, NIETO-YÁÑEZ, Alma Delia, HERNÁNDEZ-BÁEZ, Irma Yazmín y VELASCO-CASTILLO, Miguel Ángel. Reconocimiento y segmentación de la mano con una cámara infrarroja. Revista del Diseño Innovativo. 2019 3-9: 1-7

* Correspondencia del Autor (Correo electrónico: rlopezd@upemor.edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

Actualmente la Interacción Humano Computadora (HCI por sus siglas en inglés, Human Computer Interaction) es un concepto muy nombrado entre grandes empresas, pues se busca una relación más intuitiva entre el hombre y la computadora para aplicarla en diferentes dispositivos tecnológicos. Considerando la satisfacción del usuario final al reducir su esfuerzo para realizar tareas. Se han trabajado proyectos donde la comunicación se realiza mediante gestos corporales; Microsoft¹ es una de las empresas que hizo esto posible con uno de sus proyectos nombrado Kinect², otro caso es el de Nintendo³ con la consola Wii⁴. Es por eso que contribuyendo a la mejora de HCI y las nuevas tendencias de crear interacción con gestos corporales se plantea la búsqueda de interacción con el usuario donde se requiere identificar una mano utilizando luz infrarroja como herramienta clave.

La luz infrarroja brinda información que no se puede obtener de la luz visible. Todos los cuerpos tienden a emitir radiaciones infrarrojas; las cuales dependen directamente de la temperatura en que se encuentra el cuerpo y esta es una de las ventajas que se aprovechan dentro de los desarrollos de visión computacional ya que la captura de imagen proporciona imágenes nítidas, con una iluminación realmente llamativa que permite identificar claramente cuerpos en la escena. Una cámara infrarroja es un dispositivo que, a partir de las emisiones de infrarrojos medios del espectro electromagnético de los cuerpos detectados, forma imágenes luminosas visibles por el ojo humano y el hecho de que cualquier cuerpo con temperatura, emana cierta cantidad de radiación infrarroja hace que este tipo de toma infrarroja esté asociada a los disparos nocturnos o la posibilidad de ver en situaciones de mucha oscuridad (Benitez Restrepo, 2007).

En este proyecto se construye una cámara infrarroja a partir de pequeñas modificaciones a nivel de hardware de una cámara web convencional, logrando una imagen limpia, y mediante la segmentación de la escena se identifica la mano, llevando finalmente a un proceso de reconocimiento de gestos.

El aporte del presente trabajo se puede determinar como la obtención de un dispositivo que permita capturar imágenes con las características mencionadas anteriormente y con lo cual se abre la posibilidad de ser utilizado en proyectos que requieren capturar escenas con condiciones de desventaja de iluminación que favorezca la segmentación e identificación de elementos en ella.

Trabajos relacionados

En esta sección se presenta parte de la revisión bibliográfica realizada para verificar el estado del arte respecto al reconocimiento de gestos.

Los autores en (Ghotkar, 2012) se propone el desarrollo de un sistema de reconocimiento de gestos para comunicarse a través de una interacción humano-computadora más natural. El objetivo principal es desarrollar un algoritmo de segmentación sólido y eficiente basándose en espacios de color y procesamiento morfológico necesario para la detección del color de piel, la eliminación del fondo de la imagen y con condiciones de iluminación variable. En el trabajo se usó la biblioteca OpenCV para realizar el seguimiento y un algoritmo de recorrido perimetral para la detección del contorno de la mano.

Para (García Cortés, 2014), la detección de gestos es una de las partes importantes en HCI por lo cual propone un trabajo en el que se haga la detección de gestos a través de pre-procesamiento de imágenes para disminuir el ruido además de usar máquinas de soporte vectorial (SVM por sus siglas en inglés) para la detección y extracción de la región donde se encuentran las manos utilizando características evidentes con un enfoque basado en apariencia.

También se hace uso de un clasificador AdaBoost para la detección de los objetos de interés. Como se conclusión se menciona que las características del ambiente no restrictivo influyen en el desempeño del sistema.

En (S. E. Ghobadi, 2007) se describe una técnica de segmentación robusta y rápida, usando la fusión de imágenes 2D y 3D con el objetivo de reconocimiento de gestos.

¹ <https://www.microsoft.com/es-mx>

² <http://www.xbox.com/es-MX/Kinect>

³ <https://www.nintendo.es/>

⁴ <https://www.nintendo.es/Wii/Wii-94559.html>

Esta técnica de segmentación se basa en la combinación de dos enfoques de clustering no supervisado como son K-medias y Maximización de la expectativa. Los resultados experimentales mostraron que esta técnica logra segmentar correctamente la mano del cuerpo, la cara, el brazo de una persona y de otros elementos encontrados en la imagen la cual fue tomada con condiciones de iluminación variables en tiempo real.

En (Tanibata, 2002) se propone un método de obtención de características de la mano a partir de secuencias de imágenes, específicamente cuando una persona realiza las señas del lenguaje de signos japonesa para reconocer las palabras de dicho lenguaje en un fondo complejo.

Para el reconocimiento se usa la técnica de modelos ocultos de Markov (HMM por sus siglas en inglés) utilizando seis rasgos de la cara y manos. Los resultados presentados indican que se cumple el objetivo de hacer el seguimiento de la cara y las manos.

Por otro lado, en (Byungkyu K., 2013) se presenta un método de predicción de las poses de una mano articulada en tiempo real con una cámara de profundidad (Kinect, Xtion Pro) con el propósito de realizar la interacción en un ambiente de realidad mixta y para realizar el estudio de efectos de modelos de manos articuladas reales y virtuales en un simulador. Para realizar el reconocimiento se usaron bosques de decisión aleatorios, que demostraron tener mejores resultados en aplicaciones de tiempo real evitando los errores típicos en el uso de estas tecnologías, además de mostrar bajo consumo de recursos computacionales y alta precisión.

Metodología de desarrollo

Para el desarrollo de este proyecto se decidió adaptar la metodología cascada mejorada empleada en el desarrollo de software (Cabrera, 2009) debido a que esta metodología establece las etapas del ciclo de vida del software, de tal forma que el inicio de cada etapa debe esperar a la finalización de la anterior, sin embargo, permite el mejoramiento de etapas anteriores, si es necesario. En Figura 2, se muestra el esquema de la metodología que incluye tres etapas.

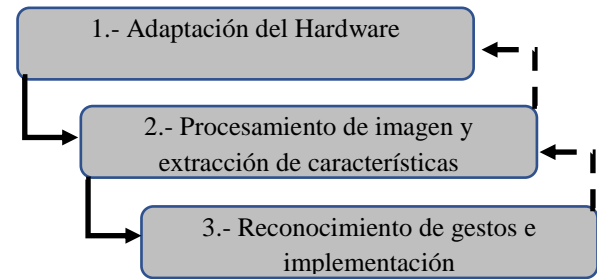


Figura 1 Esquema de metodología propuesta
Fuente: Elaboración Propia

La primera etapa nombrada “Adaptación del hardware” se refiere a todas las actividades para obtener la cámara con las modificaciones necesarias, así como la mejor alternativa para la captura de la imagen deseada para continuar con el proceso. En la segunda etapa “Procesamiento de imagen y extracción de características” se definen aspectos importantes para las características a comparar, así como la estructura de la biblioteca a construir. En la tercera etapa “Reconocimiento de gestos e Implementación” que se refiere a cómo se logra el entrenamiento para llegar a la detección de gestos y a implementar la biblioteca en una aplicación.

Adaptación del hardware

En esta sección se presenta la estructura del hardware y las modificaciones realizadas al mismo para obtener una cámara infrarroja con las características necesarias para el proyecto. En la Figura 2 se presenta un esquema de la cámara y las adiciones necesarias para convertirla en cámara infrarroja.

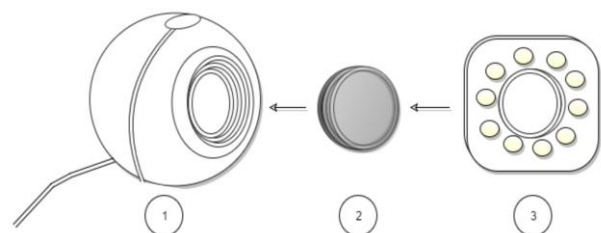


Figura 2 Esquema de cámara y adiciones
Fuente: Elaboración Propia

Primeramente se modifica la cámara web (Número 1 en la imagen) modelo COM-105 con sensor CMOS de 480 pixeles, de tal manera que se elimine el filtro ICR siguiendo el procedimiento explicado en (Instructables, 2008) (wikiHow, 2016). Posteriormente, con el objetivo de realizar una iluminación adecuada de la escena que se evaluaría durante las pruebas, es necesario construir un arreglo de LEDs (Número 3 de la imagen).

En una segunda fase, esta estructura se construye usando un arreglo de 10 LEDs modelo TN130BF cuyo mayor ángulo de amplitud de iluminación permiten iluminar los objetos a segmentar de más forma más completa a diferencia de otros modelos de diodos. En la Figura 3 se presenta la estructura terminada.



Figura 3 Estructura con arreglo de LEDs terminada
Fuente: Elaboración Propia

Finalmente, se utiliza algún material que permita el filtrado de la luz infrarroja para determinar los elementos presentes en la imagen (Número 2 de la imagen). Los materiales usados fueron seleccionados considerando la longitud de onda captada por cada uno de los materiales, realizando pruebas experimentales para determinar aquel que ofreciera la captura adecuada de la mano y los gestos realizados. Al finalizar dichas pruebas, se seleccionó como material el negativo de una película fotográfica en rollo (Figura 4) el cual permite observar elementos con una longitud de onda entre 800nm y 1000 nm.

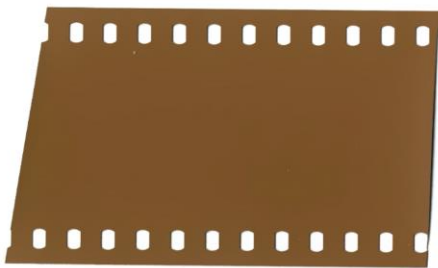


Figura 4 Tira de negativo de película fotográfica

Procesamiento de la imagen y extracción de características

Una vez que se terminan las modificaciones a la cámara, debía comprobarse que las imágenes que se capturan contienen información útil para el reconocimiento de gestos efectuados con una mano. Para realizar la comprobación, se construyó una biblioteca de funciones basadas en lenguaje C++ y la biblioteca OpenCV (OpenCV, 2015).

El primer paso a realizar es la segmentación de una mano en primer plano del resto o fondo de la imagen capturada. La forma seleccionada para realizar esta segmentación se basa en la segmentación por umbral en donde basándose en valores que se establecen por una persona, una imagen en colores obtenida por la cámara construida mediante una máscara que verifica pixel a pixel, si el valor numérico de cada uno de estos pixeles se encuentra en el rango que define una mano y en caso contrario, se determina que ese pixel no pertenece a la mano. En la figura 5 se presenta el proceso para el establecimiento del umbral para un canal de color encontrado en las imágenes.

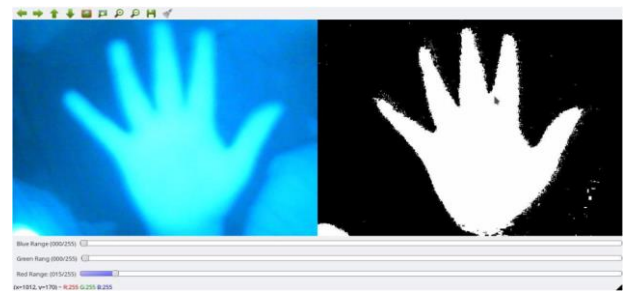


Figura 5 Asignación de umbral para el canal con el color rojo y segmentado con esa información
Fuente: Elaboración Propia

Debido a que una de las características de las imágenes capturadas mediante una cámara infrarroja es que dicho tipo de luz es reflejado por los objetos, se analizó el uso de dicho canal para que un programa transformara la información en él, en otra imagen a escala de grises con el objetivo de apoyar a una mejor segmentación de la mano en la imagen. Esta transformación ofreció una mejora en la segmentación, razón por la cual se seleccionó continuar usando la escala de grises. En la imagen 6 se puede observar la imagen resultante de dicho proceso.



Figura 6 Asignación de umbral para escala de grises y segmentado con dicha información

Ya que la mano se ha segmentado, se deben extraer las características que la definen. Dichas características se obtendrán del análisis del contorno de la mano y de la ubicación del centroide de la misma.

El contorno de la mano se obtiene como una secuencia de puntos utilizando el algoritmo desarrollado por Satoshi en 1985 (Satoshi Suzuki, 1985) y que se encuentra implementado en la biblioteca OpenCV. Una vez obtenido, se debe encontrar la envoltura convexa (polígono mayor que envuelve una figura en la imagen) y sus defectos (puntos profundos entre la envoltura convexa). Esto permite determinar el espacio ocupado por la mano y ubicar la punta de los dedos como puede observarse en la Figura 7. En la figura se pueden observar el inicio y el fin de un defecto los cuales corresponden a la unión del espacio entre los dedos de la mano y la punta de los mismos.

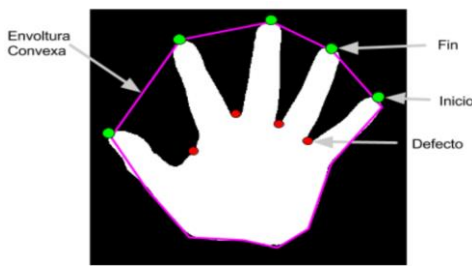


Figura 7 Ubicación y descripción de envoltura convexa y sus defectos

Posterior a la ubicación del polígono mayor, que corresponde a la mano con los dedos incluidos, se determina el centroide de la mano utilizando el polígono detectado (Figura 8).



Figura 8 Ubicación del centroide de la mano

Con lo anterior, se determina que las características a utilizar para identificar los gestos estarían definidos por un defecto y el centroide de la forma definida en la tabla 1.

Característica en la imagen	Característica extraída a partir de la imagen
Defecto	<ul style="list-style-type: none"> - Distancia de inicio a fin - Distancia de inicio al punto más profundo - Distancia del fin al punto más profundo
Centroide	<ul style="list-style-type: none"> - Distancia de inicio al centroide - Distancia de final al centroide

Tabla 1 Descripción de características extraídas Fuente Elaboración Propia

Con lo anterior se establece que se usan 20 características para representar una mano extendida a los que se agrega el ancho y la altura de la mano para hacer un total de 22 características a analizar (Figura 9).

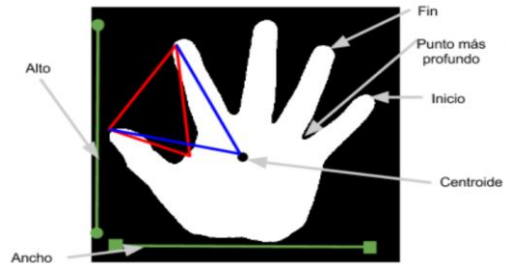


Figura 9 Descripción de características extraídas Fuente: Elaboración Propia

Resultados

Para comprobar el funcionamiento de identificación de la mano, se desarrollaron funciones para clasificar las características extraídas de las imágenes. Esto se logró usando la técnica de máquinas de soporte vectorial (SVM por sus siglas en inglés) (Morales España, 2005) la cuál aprende a clasificar datos en dos clases distintas. Una implementación de SVM está incluida en la biblioteca OpenCV con lo cual fue posible utilizarla.

Para realizar el entrenamiento de las máquinas, se ejecutó la extracción de características de diversas tomas realizadas con la librería desarrollada y etiquetando cada toma de gesto como cinco dedos y cuatro dedos. Ejemplos de las características se presentan en las figuras 10 y 11.

Para cada gesto de dedos extendidos, se usaron 55 muestras con las cuales se entrenó a las máquinas para posteriormente hacer la comprobación con imágenes capturadas directamente de la cámara encendida. Cada gesto se comprueba por separado con otro gesto en donde los dedos no se encuentran separados como se observa en la figura 12.



Figura 10 Identificación de mano con cinco dedos extendidos y características extraídas



Figura 11 Identificación de mano con cuatro dedos extendidos y características extraídas



Figura 12 Identificación de mano con cero dedos extendidos y características extraídas

Fuente: Elaboración Propia

Los resultados obtenidos permitieron observar que el desarrollo realizado identifica de mejor manera el gesto de cinco dedos contra el gesto sin dedos extendidos que el gesto de cuatro dedos contra el de sin dedos extendidos.

Conclusiones y trabajos futuros

Con el desarrollo del proyecto se puede concluir que tanto la cámara construida como la librería implementada pueden apoyar en la detección y reconocimiento de los gestos de una mano.

Es de resaltar que las tecnologías desarrolladas son relativamente de bajo costo lo que podría ser parte de consideración al momento de compararse con algunos de los trabajos realizados previamente y analizados en la sección de trabajos relacionados.

No obstante, los resultados obtenidos abren la posibilidad de mejorar las tecnologías desarrolladas después de hacer una revisión de los aspectos negativos que se encontraron.

Uno de ellos por ejemplo es la afectación de la identificación de los gestos debido a la distancia entre la cámara y la mano que hace los gestos, ya que al entrenar a las máquinas de soporte vectorial con una característica como la altura y la anchura, las muestras usadas en entrenamiento pueden no haber sido suficientes para evitar ser susceptible a errores por dicha distancia.

Adicionalmente, en cuanto al hardware modificado, se plantea el cambio en el tipo de diodos utilizados por otro con un ángulo de amplitud mayor para abarcar mayor espacio visible o también cambiar el material usado para filtrado de luz infrarroja de manera que se ajusten a los valores usados por los LEDs.

Por último, se puede buscar modificar la librería desarrollada para evitar que se considere una parte del brazo como parte de la mano lo cual también pudo ser razón por la que se tuvieron algunos problemas en la identificación del tamaño de la mano.

Referencias

Cabrera, A. S. (2009). Procesos de Ingeniería del Software. Universidad Técnica Particular de Loja.

Benitez Restrepo, H. D. (2007). Procesamiento de imágenes infrarrojas para la detección de defectos en materiales. Revista Tecnura, 40-51.

Byungkyu K., R. M. (2013). Real time hand pose recognition with depth sensors for mixed reality interfaces. 2013 IEEE Symposium on 3D User Interfaces (3DUI), 171-172.

García Cortés, D. C. (2014). Reconocimiento de gestos de manos como mecanismo de interacción humano-computadora, Trabajo de Grado. Bogotá: Universidad Nacional de Colombia, Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial.

Ghotkar, A. K. (2012). Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction. International Journal of Human-Computer Interaction. 3, 15-25.

Instructables. (2008). Infrared (IR) Webcam. Obtenido de Instructables: <https://www.instructables.com/id/Infrared-IR-Webcam/>

Morales España, G. y. (2005). Estudio e implementación de una herramienta basada en máquinas de soporte vectorial aplicada a la localización de fallas en sistemas de distribución. Trabajo de Grado. Bucaramanga: Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones, Universidad Industrial de Santander.

OpenCV. (18 de Diciembre de 2015). OpenCV Modules. Obtenido de OpenCV Open Source Computer Vision: <https://docs.opencv.org/3.1.0/index.html>

S. E. Ghobadi, O. E. (2007). Hand Segmentation using 2D/3D Images. Proceedings of Image and Vision Computing New Zealand, 64-69.

Satoshi Suzuki, K. b. (1985). Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, Volume 30, Issue 1, 32-46.

Tanibata, N. S. (2002). Extraction of Hand Features for Recognition of Sign Language Words. Osaka: Osaka University, Computer-Controlled Mechanical Systems, Graduate School of Engineering.

wikiHow. (2016). Cómo convertir una cámara web en una cámara infrarroja. Obtenido de wikiHow: <https://es.wikihow.com/convertir-una-c%C3%A1mara-web-en-una-c%C3%A1mara-infrarroja>