

Una aplicación interactiva parametrizable para la segmentación de la mano humana en imágenes RGB



A parameterizable interactive application for human hand segmentation in RGB images

Cattafi, Ricardo

Ricardo Cattafi

rcattafi@gmail.com

Universidad Católica Santa María la Antigua (USMA),
Panamá

Investigación y Pensamiento Crítico

Universidad Católica Santa María La Antigua, Panamá

ISSN: 1812-3864

ISSN-e: 2644-4119

Periodicidad: Cuatrimestral

vol. 11, núm. 3, 2023

investigacion@usma.ac.pa

Recepción: 10 Abril 2023

Aprobación: 20 Septiembre 2023

URL: <http://portal.amelica.org/ameli/journal/333/3334707004/>

DOI: <https://doi.org/10.37387/ipc.v11i3.366>



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Resumen: La segmentación es una técnica fundamental en el análisis de imágenes que permite detectar y clasificar objetos; entre otros, las manos humanas. En este trabajo se presentan los resultados del desarrollo y pruebas de una aplicación interactiva parametrizable, *Hands Segmentation*, para la segmentación de manos humanas en imágenes RGB. Los resultados muestran que la aplicación fue eficiente para segmentar un 100% de las regiones de las imágenes RGB donde existían manos humanas, aunque un 2% de las segmentaciones incluía otras partes del cuerpo humano, como por ejemplo los brazos. Excluyendo ese 2%, el promedio de la métrica de efectividad Índice de Jaccard fue de 0.73, considerándose un valor aceptable para este tipo de aplicaciones. Se utilizó una muestra de veintidós (22) manos humanas distribuidas en diez (10) imágenes obtenidas por muestreo sistemático de un banco de imágenes de uso libre. El procesamiento de las imágenes se realizó con la aplicación *Hands Segmentation* desarrollada en C++ utilizando técnicas de visión artificial y la librería especializada *OpenCV*. El método de procesamiento incluyó: a) transformación del espacio de color de RGB a YCrCb, b) segmentación de la imagen por simulación de análisis hidrográfico (*Watershed*), c) delimitación de las regiones basada en contornos y d) cálculo de la efectividad usando la métrica Índice de Jaccard (IoU) y de la precisión, usando una métrica de cualificación ad hoc. El uso sistemático de la aplicación muestra que una adecuada escogencia de los parámetros a) brillo, b) contraste, c) dimensión del operador pixel a pixel, d) dimensión del operador morfológico y e) umbrales de dimensión de las regiones, son esenciales para obtener una adecuada segmentación de una mano humana en una imagen RGB.

Palabras clave: visión artificial, técnicas de segmentación de imágenes, aplicaciones para segmentación de imágenes..

Abstract: Segmentation is a fundamental technique in image analysis that allows objects, including human hands, to be detected and classified. This work presents the results of developing and testing a parameterizable interactive application, *Hands Segmentation*, for the segmentation of human hands in RGB images. The results show that the application efficiently segmented 100% of the regions of the RGB images where human hands existed. However, 2% of the segmentations included other human body parts, such as the arms. Excluding that 2%, the average of the Jaccard Index effectiveness metric was 0.73, which is considered an acceptable value for this type of application. A sample of twenty-two (22) human hands distributed in ten (10) images obtained

by systematic sampling from a free-use image bank was used. Image processing was carried out with the Hands Segmentation application developed in C++ using artificial vision techniques and the specialized OpenCV library. The processing method included: a) transformation of the color space from RGB to YCrCb, b) image segmentation by simulation of hydrographic analysis (Watershed), c) delimitation of the regions based on contours and d) calculation of the effectiveness using the Jaccard Index (IoU) and precision metric, using an ad hoc qualification metric. The systematic use of the application shows that an adequate choice of the parameters a) brightness, b) contrast, c) dimension of the pixel-by-pixel operator, d) dimension of the morphological operator, and e) dimension thresholds of the regions, are essential to obtain an adequate segmentation of a human hand in an RGB image.

Keywords: computer vision, image segmentation techniques, applications for image segmentation..

INTRODUCCIÓN

La segmentación (clasificación en regiones según características comunes) en imágenes es una de las técnicas estudiadas por los investigadores de visión artificial (*computer vision* o CV) y que se está utilizando hoy en día en muchas aplicaciones útiles. Por ejemplo, actualmente se utiliza en el área de la salud, lenguaje de señas, control robótico, entorno virtual, domótica, aplicaciones móviles y juegos (Oudah, Al-Naji & Chahl, 2020).

Cada uno de esos campos de aplicación tiene sus propios objetivos y contexto. Si se considera el campo del lenguaje de señas, el objetivo es detectar gestos para comunicar la interacción humano-humano o humano-robot. En este caso, el contexto es un entorno controlado en el que además de las técnicas de CV, se utilizan sensores (como guantes o dedales) para obtener señales y movimientos en coordenadas espaciales.

La técnica de segmentación es fundamental en el análisis de imágenes debido a que permite detectar y clasificar objetos; entre otros, las manos humanas. En contraste con las técnicas de segmentación de rostros (una técnica madura en CV), hasta hoy no hay un consenso para un método efectivo para la segmentación de manos. Esto se debe a que las técnicas de segmentación de rostros se basan en la detección previa del área del rostro mediante patrones de Haar usando el método de Viola & Jones (Viola & Jones, 2001) que es suficientemente efectivo, al menos cuando no existe oclusión.

En los casos de segmentación de manos hay un gran conjunto de combinaciones de posiciones de manos, orientaciones y colores de piel, como se puede ver en la Figura 1. Además, las condiciones de la imagen como el brillo, el contraste, el ruido, la oclusión y el tamaño relativo deben considerarse para diseñar una técnica efectiva de detección de manos. Significa que la detección de manos es un problema complejo en comparación con el problema de detección de rostros; por lo tanto, se debe usar un enfoque diferente para resolver dicho problema (Shah, Hurmat et al, 2012).

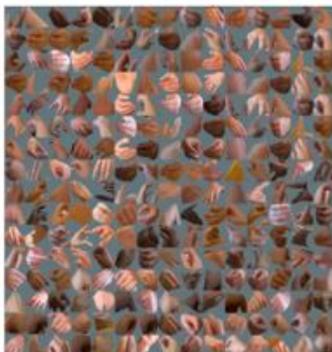


Figure 1. Un subconjunto de las posibles combinaciones de posiciones de manos, orientaciones y colores de piel. (Tomado de <http://vision.soic.indiana.edu/projects/egohands/>)

FIGURE 1

Un subconjunto de las posibles combinaciones de posiciones de manos, orientaciones y colores de piel
Tomado de <http://vision.soic.indiana.edu/projects/egohands/>

No obstante, dichas limitaciones, se han realizado avances en relación con las técnicas de segmentación de manos. Actualmente se han analizado algunas características de las manos como el color de la piel y la apariencia, y se han utilizado algoritmos para reconocer las manos (y posteriormente segmentarlas) a través algunas técnicas como las basadas en los movimientos de las manos, basadas en la posición de las articulaciones, basadas en profundidad de visión o basadas en modelos 3D. Incluso, se han desarrollado modelos basados en plantillas y algoritmos de aprendizaje de máquina (Adaboost) para reconocerlas (Oudah, Al-Naji, & Chahl, 2020).

Técnicas de segmentación de imágenes

La segmentación de imágenes se puede obtener por similitud de píxeles en la misma región o por discontinuidad entre píxeles en diferentes regiones. Existen varias técnicas para segmentar una imagen. Por ejemplo, técnicas de a) transformaciones basadas en umbrales, b) basadas en el crecimiento de regiones con similitud de color o, c) por simulación de análisis hidrográfico. La técnica basada en umbrales es útil cuando se utilizan imágenes en escala de grises, mientras que los basados en el crecimiento de regiones con similitud de color requieren una función de comparación para incrementar el área de la región a partir de píxeles “semillas” (píxeles tomados al azar o sistemáticamente en la imagen). En este caso, el número de regiones depende del número de semillas predefinidas; sin embargo, dado que no se conoce a priori cuántas regiones serán detectadas debe realizarse un preprocesamiento de la imagen.

Por otro lado, la idea intuitiva detrás del algoritmo de transformación por simulación de análisis hidrográfico es simular que se “inunda” cada región con un color. Cada región coincide con el área de una cuenca (definida por un mínimo de una función

que representa matemáticamente la imagen), lo que significa que la cuenca se convierte en una región de segmentación. Para implementar esta técnica comúnmente se utiliza el algoritmo *Watershed* (Rao, Kumar & Nagaraju, 2006) que es una técnica robusta para segmentar regiones densas incluso conteniendo problemas de oclusión de objetos.

Al igual que la técnica por crecimiento de regiones con similitud de color, dicho algoritmo también requiere seleccionar algunos puntos “semilla” para localizar los mínimos. Estos puntos se llaman “marcadores” y se pueden encontrar mediante una primera fase de preprocesamiento. Generalmente, se utiliza una transformación euclídea de distancias sobre una imagen binaria (blanco y negro) para encontrar los puntos

de marcadores para las regiones y un punto para marcar el área de fondo. Cada región detectada se identifica con una etiqueta (generalmente con un color) para diferenciarla.

La imagen transformada de la cuenca (imagen etiquetada) es útil para encontrar los contornos de cada región con el fin de obtener algunas características útiles de la región; por ejemplo, su área o puntos centroides de cada región. Utilizando las características del área de contorno podemos filtrar una región segmentada de acuerdo con:

$$A_{min} < A_{src} < A_{max}$$

Donde A_{min} y A_{max} son los máximos y mínimos de área para el contorno de la región segmentada; y A_{src} es el valor real del área del contorno de la región segmentada.

Técnicas de segmentación de manos por color de la piel

Una técnica frecuentemente utilizada para segmentar las manos en una imagen es usar las características de color y apariencia de la piel.

El uso de la función de detección de regiones según el color de piel para segmentar las manos en una imagen es una técnica que se ha utilizado en una variedad de aplicaciones. La detección del color de la piel se puede realizar utilizando operadores de píxel por píxel o de regiones espaciales locales (Oudah, Al-Naji, & Chahl, 2020).

Uno de los aspectos más desafiantes de esta técnica es tratar con los parámetros de brillo, contraste y ruido de la imagen, debido a sus características irregulares, especialmente en imágenes de espacios de color RGB (*Red, Green, Blue*). Debido a que las imágenes son generadas por cámaras en el espacio de color RGB, estas deben convertirse en otro espacio de color ya que la mezcla del canal de color y la intensidad en el espacio de color RGB no permiten separar la información de la luminancia. Este problema puede resolverse utilizando el espacio de color YCbCr (Rhyne, 2016). En el espacio de color YCbCr, la luminancia se guarda en el canal Y, mientras que los componentes de color se guardan en los canales Cb y Cr. Por lo tanto, los problemas de luminancia se pueden evitar simplemente considerando solo los canales Cb y Cr. Luego, se puede aplicar un operador píxel por píxel para generar una imagen resultante con solo píxeles de color de piel (Kolkur, et al S., 2017).

En este trabajo se utilizará el siguiente operador píxel por píxel para transformar la imagen original de espacio de color RGB a espacio de color YCrCb:

$$h(x,y) = \begin{cases} 132 < f^{'cr}(x,y) < 173 \rightarrow h_R(x,y) = 235 \\ 76 < f^{'cb}(x,y) < 127 \rightarrow h_G(x,y) = 128 \\ \text{otro} \rightarrow h_B(x,y) = 128 \end{cases}$$

Siendo $f^{'cr}(x,y)$ y $f^{'cb}(x,y)$ los canales Cb y Cr en el espacio de color YCrCb y $h_R(x,y)$, $h_G(x,y)$, $h_B(x,y)$ los canales RGB de la imagen resultante $h(x,y)$. Antes de la aplicación del operador, la imagen RGB original $f(x,y)$ se transforma en $f^{(x,y)}$ en el espacio de color YCbCr. Los rangos (132,173), (76,127) corresponden a los valores aceptables para $f^{'cr}(x,y)$ y $f^{'cb}(x,y)$ en los que existe una máxima probabilidad de coincidir con los colores de las manos humanas. Dichos rangos fueron tomados de Mohammed & Melhum (2020). Los valores 235,128 y 128 definidos para $h_R(x,y)$, $h_G(x,y)$ y $h_B(x,y)$ corresponden al color RGB(235,128,128).

Sin embargo, evitar la luminancia utilizando el espacio de color YCbCr no es suficiente para considerar la imagen adecuada para la segmentación de la región. El brillo y el contraste son dos parámetros que deben ajustarse para obtener una respuesta de señal efectiva. El brillo se puede controlar mediante la siguiente transformada lineal:

$$h(x, y) = a f(x, y) + \beta$$

Los parámetros a y β transforman lineal y respectivamente el contraste y el brillo de la imagen. Por otra parte, para imágenes con una alta densidad de objetos, una transformación de contraste adaptativa como la técnica de transformada Gamma es más apropiada de la siguiente manera:

$$h(x, y) = c(f(x, y))^{\gamma}$$

Donde c es una constante y γ es el parámetro Gamma.

Después de obtener la imagen resultante $h(x, y)$, el siguiente paso es segmentar esa imagen en dos regiones: regiones de manos y fondo.

Una técnica de segmentación interactiva para la segmentación de la mano humana

Para realizar la segmentación interactiva de manos humanas en imágenes RGB se presenta un método de cuatro fases (Figura 2): 1) transformación del color de piel pixel a pixel, 2) segmentación de regiones por medio de simulación de análisis hidrográfico, 3) delimitación de la región basado en sus contornos y 4) cálculo de la efectividad con la métrica IoU y la precisión con una escala de cualificación. La escogencia de este método es producto de una investigación documental sobre las técnicas actuales para segmentar manos humanas en imágenes RGB.



Figure 2. Fases del método de segmentación de manos humanas en imágenes RGB

FIGURE 2

Fases del método de segmentación de manos humanas en imágenes RGB

La Aplicación *Hands Segmentation*

Para alcanzar el objetivo de segmentación de manos humanas, se desarrolló una aplicación interactiva parametrizable (denominada *Hands Segmentation*) cuyos parámetros se ajustan a través de controles deslizantes en la interfaz de usuario (sección inferior de la Figura 3). Los parámetros de control son: brillo (*Brightness*), contraste (*Contrast*), dimensión del operador pixel a pixel (*Kernel Size*), dimensión del operador morfológico (*Struct Element Size*), umbral de área de región mínima (*Min Hands Region Area*) y umbral de área de región máxima (*Máx Hands Region Area*). Funcionalmente, cada vez que se cambia un valor de parámetro en la interfaz de usuario, se ejecutan iterativamente las cuatro fases del método anterior y se muestra un nuevo conjunto de regiones etiquetadas con colores aleatorios. El usuario detiene el proceso de ajuste de los parámetros cuando observa una o más manos humanas etiquetadas y enmarcadas dentro de un recuadro.

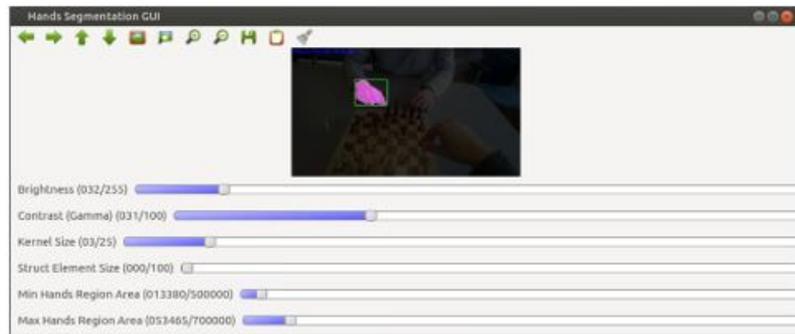


Figure 3. Interfaz de usuario de la aplicación *Hands Segmentation*

FIGURE 3
Interfaz de usuario de la aplicación Hands Segmentation

Por ejemplo, en la Figura 4 se muestra una imagen RGB de la escena de un juego de ajedrez (en modo transparencia) mostrando una región (en color rosa y enmarcada en un recuadro verde) que corresponde a una de las manos de un jugador. También se observa en la parte superior izquierda de la imagen el resultado de la métrica de IoU y el número de manos detectadas.

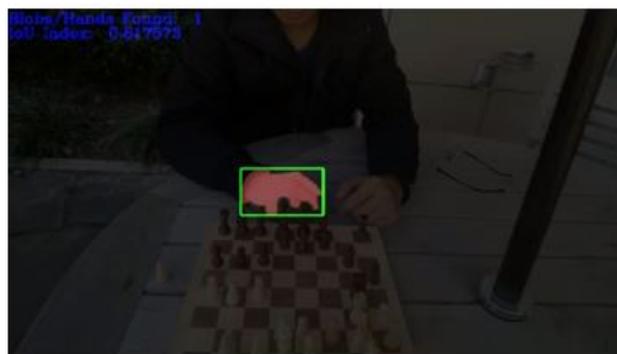


Figura 4. Imagen procesada que muestra la mano etiquetada en color rosa, el número de manos segmentadas y el valor de IoU

FIGURA 4
Imagen procesada que muestra la mano etiquetada en color rosa, el número de manos segmentadas y el valor de IoU

Para realizar las tareas de la segmentación de una imagen se codificaron las siguientes funciones:

- a. *skinFilter*: Ejecuta la transformación del espacio de color y el operador pixel a pixel de color de la piel.
- b. *watershedSegmenter*: Segmentación en regiones con el del algoritmo *Watershed*.
- c. *handDetectionProcess*: Ejecución de operadores morfológicos, delimitación del contorno de cada región y presentación en la interfaz.
- d. *IoUIndex()*: Cálculo del Índice de Jaccard (IoU).

La aplicación se implementó en C++; con OpenCV 4.5.5 como biblioteca especializada de funciones CV y sistema operativo Ubuntu 20.04. Las pruebas se ejecutaron en un equipo PCx64 con procesador Intel i5, 2.5GHz, 2 GPU, 3GB RAM.

MATERIALES Y MÉTODOS

El objetivo principal de este trabajo es utilizar la aplicación *Hands Segmentation* para segmentar regiones de manos humanas en un conjunto de diez imágenes RGB seleccionadas sistemáticamente del sitio web del Laboratorio de Visión por Computadora de la Universidad de Indiana (disponible en <http://vision.soic.indiana.edu/projects/egohands/>). El rendimiento del software se verifica utilizando la métrica de efectividad Índice de Jaccard (también denominada *Intersection Over Union* o IoU) y una métrica de escala cualitativa (0 a 5 puntos) para especificar el grado de precisión (cualificación) de la aplicación. La escala se construye de la siguiente forma:

0 puntos: No se segmentó ninguna región que incluyan manos.

1 punto: Se segmentó una o dos manos incluyendo otras partes del cuerpo.

2 puntos: Se segmentaron las dos manos en forma conjunta.

3 puntos: Se segmentaron dos o más manos indistintamente

4 puntos: Se segmentó cada mano por separado con diversos parámetros

5 puntos: Se segmentó cada mano por separado con los mismos parámetros

Como datos de entrada se utilizan (Figura 5):

1. Diez (10) imágenes de espacio de color RGB, resolución 720x1280, categoría 8UC3 (tres canales y una profundidad de 8 bits).

2. Diez (10) imágenes binarias (máscaras) segmentadas en dos regiones: fondo (región negra) y regiones de mano (regiones blancas). Las máscaras se utilizan para fines de cálculo de IoU. En este caso se denominan imágenes de referencia real (*Groundtrue*).

3. Veintidós 4-tuplas (x, y, ancho, alto) correspondientes a las coordenadas de las máscaras de cada mano.

| Image # | Image | Mask (Groundtrue) | Ground true coordinates |
|---------|--|--|---|
| 1 |  |  | 631,318,217,122 451,308,175,134 |
| 2 |  |  | 641,322,133,111 518,339,201,166 |
| 3 |  |  | 727,322,153,185 493,328,162,101 |
| 4 |  |  | 682,347,138,106 489,331,172,106 |
| 5 |  |  | 669,316,151,102 483,312,201,197 |
| 6 |  |  | 620,414,235,232 730,273,181,94 437,224,125,153 |
| 7 |  |  | 740,413,400,249 627,221,164,124 365,190,154,137 |
| 8 |  |  | 627,314,247,137 430,339,187,135 |
| 9 |  |  | 668,533,109,93 406,490,98,116 |
| 10 |  |  | 608,276,211,118 421,234,143,173 |

Figura 5. Imágenes utilizadas en las pruebas mostrando el número de la imagen (*Image #*), la imagen en espacio de color RGB (*Image*), la imagen de referencia (*Mask (Groundtrue)*) y los valores de las coordenadas de ubicación de las imágenes de referencia (*Ground True Coordinates*)

FIGURA 5

Imágenes utilizadas en las pruebas mostrando el número de la imagen (*Image #*), la imagen en espacio de color RGB (*Image*), la imagen de referencia (*Mask (Groundtrue)*) y los valores de las coordenadas de ubicación de las imágenes de referencia (*Ground True Coordinates*)

El método para alcanzar los objetivos de este trabajo consistió en un plan de tres pasos:

- 1) Analizar visualmente cada una de las imágenes y obtener sus principales características; es decir, identificar el brillo aparente, contraste, ruido, oclusiones, posición de las manos, entre otros.
- 2) Utilizar la aplicación *Hands Segmentation* para identificar patrones que permitan alcanzar un resultado preciso para cada imagen.
- 3) Recopilación de los valores de los parámetros ajustados para cada imagen y del correspondiente métrica IoU y cualificación de la efectividad de la aplicación.

RESULTADOS

Siguiendo el método antes mencionado se obtuvieron los siguientes resultados:

- 1) Análisis visual de cada imagen.

Una inspección visual de cada imagen de entrada indicó que no son uniformes en relación con los valores de brillo, contraste y ruido. Además, tienen regiones de oclusión y diversas apariencias de las manos. Por ejemplo, las manos están en posición lateral, tienen un tamaño relativamente pequeño o tienen vestimenta y objetos superpuestos que crean una oclusión de los dedos. Como apoyo para el análisis visual de cada imagen

se utilizaron histogramas de los canales RGB. Así, se puede ver en el histograma que se muestra en la Figura 6 (imagen #1) tienen un contraste más bajo que el resto de las imágenes.

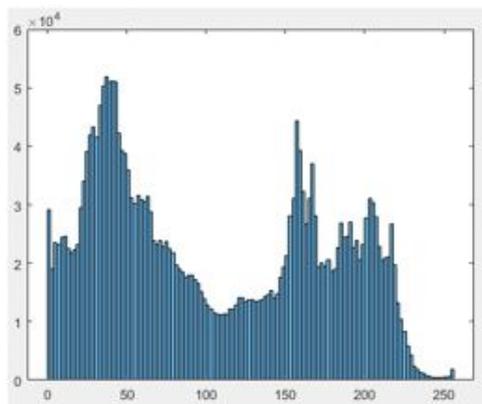


Figura 6. El histograma de la imagen #1 muestra la frecuencia de intensidad de los canales RGB

FIGURA 6

El histograma de la imagen #1 muestra la frecuencia de intensidad de los canales RGB

Otras, como las imágenes # 2 a # 4, tienen contenido similar, pero las condiciones de iluminación y contraste son diferentes. Este hecho se interpretó como que no existe una combinación única de valores de parámetros para encontrar cada región de la mano en una imagen.

2) Ejecución de la aplicación *Hand Segmentation*.

Las diez imágenes se procesaron con la aplicación *Hand Segmentation*.siguiendo las fases mostradas en la Figura 2 y los resultados se organizaron como se muestra en la Figura 7.

| Image # | Input image | Segmented output image | Binary output image | Best fitting set of parameter values | | | | | | | | | |
|---------|-------------|------------------------|---------------------|--------------------------------------|------|-----------|----------|-------------|--------------|--------|--------------|----------------|---------------|
| | | | | Region # | Area | Perimeter | Centroid | Orientation | Aspect Ratio | Volume | Surface Area | Mean Curvature | Quality Index |
| 1 | | | | 113 | 96 | 12 | 4 | 74953 | 107124 | 0.0027 | 0.091 | + | |
| 2 | | | | 5 | 23 | 2 | 5 | 24091 | 84613 | 0.2874 | 0.2874 | + | |
| 3 | | | | 31 | 26 | 6 | 7 | 44441 | 27367 | 0.8111 | 0.2334 | + | |
| 4 | | | | 16 | 17 | 3 | 5 | 9613 | 9324 | 0.8175 | + | | |
| 5 | | | | 48 | 100 | 5 | 4 | 9155 | 38961 | 0.7023 | + | | |
| 6 | | | | 21 | 0 | 8 | 4 | 13380 | 27723 | 0.6175 | + | | |
| 7 | | | | 46 | 2 | 5 | 0 | 12134 | 12867 | 0.6168 | + | | |
| 8 | | | | 58 | 70 | 12 | 0 | 7100 | 11251 | 0.6832 | + | | |
| 9 | | | | 36 | 50 | 6 | 13 | 38939 | 60267 | 0.1944 | + | | |
| 10 | | | | 9 | 3 | 0 | 5 | 8001 | 25309 | 0.5162 | + | | |
| 11 | | | | 12 | 31 | 3 | 0 | 11300 | 13663 | 0.7872 | + | | |
| 12 | | | | 60 | 21 | 0 | 6 | 51708 | 62376 | 0.6167 | + | | |
| 13 | | | | 35 | 93 | 15 | 0 | 7116 | 11989 | 0.6540 | + | | |
| 14 | | | | 122 | 97 | 4 | 11 | 49334 | 153851 | 0.1399 | - | | |
| 15 | | | | 136 | 86 | 9 | 1 | 7042 | 7326 | 0.7876 | + | | |
| 16 | | | | 97 | 48 | 0 | 5 | 7116 | 9992 | 0.3811 | + | | |
| 17 | | | | 91 | 18 | 1 | 1 | 28911 | 14982 | 0.1940 | - | | |
| 18 | | | | 62 | 32 | 6 | 7 | 11180 | 17663 | 0.6722 | + | | |

Figura 7. Se muestran las imágenes de entrada (*Input image*), de salida (*Segmented output image*) y de referencia (*Binary output image*). Además, se muestran los valores de parámetros (*Best fitting set of parameter values*) que mejor se ajustan a la detección manual y la segmentación seleccionadas de forma interactiva mediante la GUI. También se muestran los valores del índice de Jaccard (*IoU*) y el índice de cualificación (*Quality index*)

FIGURA 7

Se muestran las imágenes de entrada (*Input image*), de salida (*Segmented output image*) y de referencia (*Binary output image*). Además, se muestran los valores de parámetros (*Best fitting set of parameter values*) que mejor se ajustan a la detección manual y la segmentación seleccionadas de forma interactiva mediante la GUI. También se muestran los valores del índice de Jaccard (*IoU*) y el índice de cualificación (*Quality index*)

3) Recopilación de imágenes de salida y de la métrica IoU y la escala cualitativa

Como se puede observar en la Figura 7, cada fila de la tabla contiene toda la información de procesamiento sobre una imagen de entrada y salida. Por ejemplo, la fila #1 contiene, de izquierda a derecha, las imágenes de entrada (imagen RGB e imagen *Groundtrue*) y las imágenes de salida (imagen segmentada de la o las mano(s) etiquetadas en color aleatorio e imagen binaria obtenida).

En la mayoría de los casos (imágenes #4, #5, #6, #7, #9 y #10) se obtuvo más de una imagen de salida porque el proceso de segmentación manual no se realizó de forma única; lo cual significa que la parametrización para ajustar el algoritmo no se pudo hacer utilizando un subconjunto único de valores de los parámetros (*Brightness*, *Contrast*, *Kernel Size*, *Struct Element Size*, *Min Hands Region Area* y *Máx Hands Region Area*) para todas las manos en la imagen debido a los diferentes condiciones de iluminación, contraste y oclusión de cada una. Además, la misma fila de tabla contiene los valores para los parámetros que se seleccionaron de forma interactiva mediante la interfaz de usuario. Cuando fue necesario, se utilizaron dos o tres filas para mostrar esos valores. La penúltima columna contiene el valor de IoU; mientras la última columna contiene la cualificación de la efectividad.

DISCUSIÓN

El objetivo de este proyecto fue detectar un total de veintidós manos (22) distribuidas en diez imágenes de entrada y crear un conjunto de n regiones R_n tal que $R_1 \cup R_2 \dots \cup R_n = f'(x, y)$. Para alcanzar tales objetivos se siguió un plan de cuatro fases aplicado a cada imagen usando la aplicación *Hands Segmentation* obteniendo un conjunto de resultados que se muestran en la Figura 7.

Mediante la inspección visual de cada imagen de entrada, realizada como primer paso del método establecido, queda evidenciado que el conjunto de diez imágenes de entrada no es uniforme en relación con los valores de brillo, contraste y ruido. Este hecho se interpretó como que no hay una combinación única de valores de parámetros para segmentar manos humanas en una imagen.

Por otra parte, el diseño de la aplicación *Hand Segmentation* estuvo fundamentada en una investigación documental sobre las técnicas actuales para detectar manos en una imagen. En este sentido, la mayoría de los autores sugieren una doble estrategia: a) umbralizar la imagen de entrada utilizando un operador pixel a pixel basado en el color -estadísticamente, corresponde al rango de color más probable- de la piel de las manos y b) aplicar un algoritmo basado en la posición de las articulaciones de la mano, o una de las diferentes técnicas de segmentación. Dichos algoritmos de detección de manos basadas en la posición de las articulaciones utilizan la posición de las articulaciones de los dedos como una característica reconocible. Esta técnica funciona bien si los dedos son evidentes (extendidos). Sin embargo, basándonos en el análisis visual realizado sobre las imágenes de entrada, consideramos que los dedos incluidos en las imágenes de entrada utilizadas en este estudio no están lo suficientemente extendidos para aplicar esa técnica. Además, la apariencia de las manos tiene cambios sensibles entre las imágenes; Por ejemplo, están en posición lateral, tienen un tamaño relativo pequeño o tienen vestimenta y objetos superpuestos que crean una oclusión de los dedos.

Con respecto a la funcionalidad de la aplicación *Hands Segmentation*, se observó que la para una efectiva parametrización son útiles las siguientes heurísticas:

1. **Brillo y contraste:** se debe cambiar la luminancia de la imagen de entrada con el fin de mejorar la definición de los bordes usando una transformación lineal para el brillo, mientras para el contraste se utiliza una transformación Gamma. Para obtener una mejor precisión en la transformación de brillo y contraste, se aplicó la transformada Gamma sobre cada cambio en el parámetro de brillo.

2. **Tamaño del núcleo:** para controlar el radio del elemento estructural k_s ; es decir, la dimensión de la matriz de los operadores morfológicos (dilatación y erosión) aplicado en la generación de la región de marcadores (Szeliski, 2011), se utilizó calculan la siguiente ecuación:

$$k_s = 2P + 1$$

Donde P es el parámetro seleccionado por el usuario en la interfaz de la aplicación.

Los valores paramétricos obtenidos (Figura 7) no se interpretaron estadísticamente para determinar si existen patrones significativos que permitan mejorar la heurística. No obstante, se puede inferir que el uso de la aplicación *Hands Segmentation* sobre una muestra más numerosa podría ser útil para elaborar un conjunto de datos de entrenamiento para un algoritmo de aprendizaje de máquina como Adaboost (Freund & Schapire, 1997).

Por último, los resultados de la Figura 7 muestran la efectividad de la aplicación *Hands Segmentation*. Como se puede ver, el 73% (16 manos de 22 en las imágenes #3, #4, #5, #6, #7, #9 y #10) de las manos se detectaron de forma individual, mientras que un 9% (imágenes de entrada # 2 y # 4) se detectaron en forma conjunta. Dos de ellos (imágenes #1 y #8) no pudieron ser aislados de los brazos. Para estos casos, se podría considerar una estrategia basada en el cálculo de centroides (Dey & Anand, 2014). El algoritmo también fue capaz de reconocer el 100% de las regiones basado en el color de la piel de las manos, considerando diferentes colores de piel, oclusiones y posiciones de las mismas.

CONCLUSIONES

La detección y segmentación de manos es un campo de investigación útil en CV debido a su variedad de aplicaciones. El objetivo principal de este proyecto fue el desarrollo y pruebas de una solución de software para la detección y segmentación de manos humanas en imágenes RGB utilizando técnicas de CV y codificación en lenguaje C++ con la librería especializada OpenCV. Se utilizó una estrategia basada en cuatro fases consecutivas para resolver el problema y alcanzar los objetivos. Esa estrategia se implementó con un algoritmo interactivo parametrizable implementado en la aplicación *Hands Segmentation*. Los resultados muestran que el algoritmo fue eficiente para detectar manos humanas en un 100% de las imágenes analizadas, aunque un 2% de las segmentaciones obtenidas incluía otras partes del cuerpo humano como los brazos. Además, se midió el rendimiento de la detección manual utilizando la métrica IoU y la calidad de segmentación utilizando la métrica de precisión *ad hoc*. Excluyendo ese 2%, el promedio de la métrica IoU corresponde a 0.73 lo que se considera un algoritmo de buena calidad para detectar manos humanas.

RECONOCIMIENTOS

Este trabajo es parte del proyecto Nro. SRUI-CPEI-ID-2021-2022-009 subvencionado por la Universidad Católica Santa María la Antigua.

REFERENCIAS

- Dey, S. K., & Anand, S. (2014). Algorithm For Multi-Hand Finger Counting: An Easy Approach. doi:10.48550/ARXIV.1404.2742
- Freund, Y. and Schapire, R. (1997). A decision theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55 119–139. MR1473055
- Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaudoise Sci. Nat.* 44:223-270
- Jones, M. & Rehg, J. (1999). Statistical color models with application to skin detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 1, pp. 274-280.
- Kolkur, S., Kalbande, D., Shimpi, P., Bapat, C., & Jatakia, J. (2017). Human skin detection using RGB, HSV and YCbCr color models. arXiv preprint arXiv:1708.02694.
- Mohammed, M. G., & Melhum, A. I. (2020). Implementation of HOG feature extraction with tuned parameters for human face detection. *International Journal of Machine Learning and Computing*, 10(5), 654-661.

- Rao, A.N., Kumar, V., & Nagaraju, C. (2006). A New Segmentation Method using Watersheds on Grey Level Images. *2006 International Conference on Advanced Computing and Communications*, 275-278.
- Rehg, J. & Kanade, T. (1993). DigitEyes: Vision-Based Human Hand-Tracking. School of Computer Science Technical Report CMU-CS-93-220, Carnegie Mellon University.
- Rhyne, T.-M. (2016). *Applying Color Theory to Digital Media and Visualization* (1st ed.). CRC Press. <https://doi.org/10.1201/9781315380384>
- Sato, Y., Kobayashi, Y. & Koike, H. (2000). Fast tracking of hands and fingertips in infrared images for augmented desk interface. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 462-467.
- Shah, Hurmat & Khan, Rehanullah & Hasan, Laiq & Khan, Jebran & Khattak, A. & Jan, Sadaqat. (2012). On Experimental Comparison of Color and Supervised Face Detection. *SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)*. 44. 561-564.
- Szeliski, R. (2011). *Computer vision algorithms and applications*. London; New York: Springer. ISBN: 9781848829343 1848829345 9781848829350 1848829353
- Oudah, M.; Al-Naji, A.; Chahl, J. (2020) Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. *J. Imaging* 2020, ., 73. <https://doi.org/10.3390/jimaging6080073>
- Viola, P. and Jones, M. (2001) Rapid Object Detection Using a Boosted Cascade of Simple Features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, 8-14 December 2001*, I-511-I-518.