

Segmentación mediante Agregación Multiescala para Biometría basada en Geometría de Mano

Ángel García-Casarrubios Muñoz, Carmen Sánchez Ávila, Alberto de Santos Sierra, Javier Guerra Casanova, Gonzalo Bailador del Pozo, Vicente Jara Vera

Grupo de Biometría, Bioseñales y Seguridad (GB2S)
Centro de Domótica Integral (CeDInt)
Universidad Politécnica de Madrid
{agarcia,csa,alberto,vjara,jguerra,gbailador}@cedint.upm.es

Resumen La biometría basada en geometría de mano precisa de un paso inicial vital para asegurar una identificación válida del usuario: la segmentación. Dicho paso busca aislar completamente la mano del fondo, para una captura tomada mediante una fotografía. Hoy en día, los sistemas biométricos tienden a extenderse hacia aplicaciones diarias, y por ello es importante diseñar algoritmos que sean capaces de realizar las mismas funciones biométricas en dispositivos cotidianos. Por ello, este artículo presenta un algoritmo de segmentación basado en agregación multiescala para biometría de mano. Este algoritmo será validado mediante imágenes de mano adquiridas con un dispositivo móvil, presentando además un enfoque orientado a una posterior implementación en dichos dispositivos, y conservando la precisión necesaria de un algoritmo de segmentación.

1. Introducción

En un sistema biométrico no invasivo de reconocimiento de personas a través de una imagen de la palma de la mano, es requisito indispensable extraer correctamente el contorno de la misma. Para lograrlo, es necesario utilizar un algoritmo de segmentación de imágenes, que es la rama de la teoría de procesamiento de imágenes que se encarga de estudiar diversas técnicas que permitan separar de forma eficiente un objeto presente en una imagen [8]. En los últimos veinticinco años se han impulsado un gran número de algoritmos que abordan el problema de la segmentación desde un enfoque matemático distinto, cada uno con sus ventajas y sus inconvenientes. Uno de los enfoques que más evolución ha experimentado es el relativo a la agregación multiescala. Esta técnica plantea un algoritmo que somete la imagen a determinados procesos matemáticos, agrupando los píxeles en segmentos de propiedades similares, de forma que se reduce el número de segmentos en sucesivos pasos (escalas) [5]. Además, los resultados ofrecidos recientemente por estos métodos han mostrado una mejoría en comparación a otros métodos [1], como el método Normalized-Cuts [7] y el método Mean-Shift [2]. En este artículo se propone un algoritmo de segmentación basado en agregación multiescala que hace uso de técnicas de lógica difusa para evaluar la similitud entre los segmentos restantes en cada escala, incluyendo además algunas medidas de textura. Así, dichas medidas de textura ayudan a decidir si dos segmentos deben ser agrupados en un único segmento o no. Asimismo, teniendo en cuenta que este algoritmo está dirigido a la segmentación de imágenes de manos, se incorporan algunas medidas de color de la piel [9] para reafirmar el peso de los puentes existentes entre cada par de segmentos. A diferencia de otros métodos basados en agregación multiescala [5], nuestro algoritmo no hace uso de una matriz de interpolación para cada paso de

escala, reduciendo, por tanto, la complejidad de las operaciones. Sin embargo, sí que utiliza la triangulación Delaunay [3] para incluir información espacial, de forma que proporciona en cada escala los segmentos que son más probables de agruparse a través de las coordenadas de su centroide.

2. Enfoque multiescala

Considérese una imagen I en escala de grises de $M \times N$ píxeles como una red de nodos $G = (V, E, W)$, donde los nodos (Vertex, V) corresponden a los píxeles de la imagen y los puentes (Edge, E) conectan píxeles vecinos con unos pesos w_{ij} de acuerdo a una estructura de 4 vecinos por píxel. En este algoritmo dichos pesos w_{ij} son calculados mediante una función de lógica difusa. En lo que se refiere a la imagen I , ésta es obtenida por un dispositivo móvil según el modelo RGB [4]. Sin embargo, se obtienen mejores resultados mediante la transformación al espacio Lab [9]. Concretamente, se usa la capa b, por lo que de ahora en adelante $I = b$. Considerando $G^{[s]}$ la representación de la imagen en la escala s , $G^{[0]} = I$. En esta primera escala los pesos w_{ij} son calculados de acuerdo a una función de lógica difusa que sigue una distribución gaussiana (Ec. 1):

$$w_{ij} = \exp\left(-\frac{(I_i - I_j)^2}{2\sigma^2}\right) \quad (1)$$

donde I_i e I_j representan las intensidades de I en las posiciones i y j respectivamente, siendo $\sigma > 0$ un parámetro que influye en la precisión de la segmentación. En efecto, si $\sigma \rightarrow \infty$ entonces $w_{ij} \rightarrow 1$. De esta manera, se fija de forma empírica $\sigma = 0,01$ para que los valores de w_{ij} sean comparables a los valores de I_k .

Una vez calculados los pesos w_{ij} entre cada par de nodos de la imagen, éstos son ordenados de forma descendente, de tal manera que los pares de nodos más fuertemente unidos quedan dispuestos en las primeras posiciones de la lista. Seguidamente, se procede a efectuar el proceso de agregación de los nodos comenzando por la primera posición de la lista y continuando en sentido descendente. El proceso de agregación agrupará los $M \times N$ píxeles de la imagen en T segmentos, donde $s_t \subseteq V$ (con $t = 1, 2, \dots, T$) representa a un segmento que aglutina a un conjunto de píxeles con propiedades similares. Teniendo en cuenta que un segmento s_i sin asignar se denota por $s_i = 0$, el proceso de agregación se ejecutará hasta que todos los nodos en V queden asignados a algún segmento, de acuerdo a la siguiente relación (Ec. 2):

$$(s_i, s_j) = \begin{cases} (s_q, s_q) & s_i = s_j = 0, \\ (s_i, s_i) & s_i \neq 0, s_j = 0 \\ (s_j, s_j) & s_i = 0, s_j \neq 0 \end{cases} \quad (2)$$

donde las etiquetas s_i y s_j se asignan a los nodos v_i y v_j para indicar a qué segmento pertenecen en función de tres situaciones posibles: se asigna una etiqueta nueva, s_q , si los dos nodos no han sido asignados; y se asigna la correspondiente etiqueta s_i en un puente e_{ij} uno de los nodos ya ha sido asignado. Finalmente, se calculan los nodos representativos, que acumulan la información de cada segmento. Esta información se refiere a la posición del centroide, ξ , la intensidad media μ y la varianza σ^2 . La forma de calcular el centroide se basa en la posición física de los nodos que componen el segmento dentro de la imagen I . Así, si consideramos que un nodo v_i tiene dos componentes (x_{v_i}, y_{v_i}) y que x_{s_i} e y_{s_i} representan las coordenadas de los nodos pertenecientes al segmento s_i , el centroide será (Ec. 3):

$$\xi_i = (\bar{x}_{s_i}, \bar{y}_{s_i}) \quad (3)$$

Igualmente, si v_{s_i} representa el conjunto de nodos dentro del segmento s_i , la intensidad media de dicho conjunto será:

$$\mu_i = \bar{I}_{v_{s_i}} \quad (4)$$

En cuanto a la varianza, si N_{s_i} representa el número de nodos de V que componen el segmento s_i , y basándose en Ec. 4, la varianza será:

$$\sigma_i^2 = \frac{1}{N_{s_i}} \sum_{i=1}^{N_{s_i}} (I_{v_{s_i}} - \mu_i)^2 \quad (5)$$

Por lo tanto, a partir de la segunda escala, un segmento s_i vendrá definido por la tupla $s_i = (\xi_i, \mu_i, \sigma_i^2)$, facilitando así el proceso recursivo en las sucesivas escalas, donde los representantes de los segmentos de una escala s pasan a ser nodos en la escala $s + 1$. Debido al proceso de agregación anterior, la estructura de 4 vecinos por cada nodo no se conserva, por lo que es necesario crear una nueva estructura para cada nueva escala. Esta estructura es creada mediante una triangulación de Delaunay [3], que establece una red de vecinos a partir de los centroides ξ_i calculados en la escala anterior. Aunque el número de vecinos no puede ser fijado para cada nodo de la nueva escala, este algoritmo proporciona una forma rápida de asignar los vecinos de cada nodo en función de su proximidad.

3. Condición de agregación

Para evitar la agregación de segmentos unidos por puentes débiles se añade una condición de agregación, de forma que si no es cumplida por un par de nodos, éstos no serán agregados, aunque seguirán siendo asignados a algún segmento. Este requisito ofrece en realidad información de la textura de los segmentos, ya que se basa en la varianza de los nodos involucrados. De esta forma, dos nodos v_i y v_j de cualquier escala sólo serán agregados si se cumple la siguiente condición (Ec. 6):

$$\sigma_{ij}^2 \leq \sqrt{\sigma_i^2 \sigma_j^2} + k \quad (6)$$

donde σ_{ij}^2 representa la varianza resultante de la agregación de los nodos v_i y v_j , σ_i^2 y σ_j^2 representan las varianzas acumuladas de los nodos v_i y v_j , y k es una constante fijada empíricamente que controla el grado de agregación, es decir el número de segmentos restantes tras finalizar el proceso.

4. Resultados

El algoritmo ha sido evaluado con una base de datos propia [6], con fotos de usuarios de diferentes razas, género y edad. Una muestra de los resultados puede apreciarse visualmente en la Figura 1.

Debido a los buenos resultados que ofrece el uso del espacio de color Lab (capa b), se ha medido su eficiencia de acuerdo al factor F [1]:

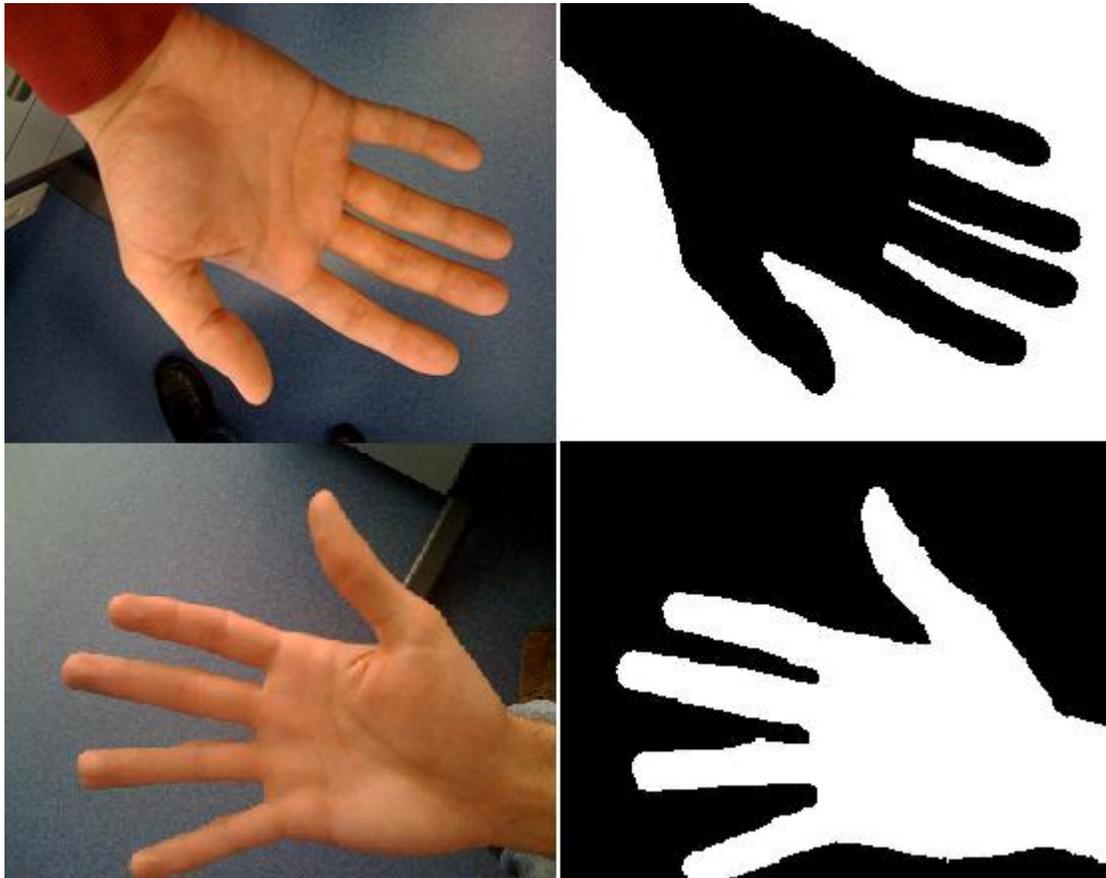


Figura 1. Ejemplo de resultados para diferentes manos. Obsérvese como es capaz de detectar la mano a pesar de presentar diferentes fondos.

$$F = \frac{2RP}{R + P} \quad (7)$$

donde P (Precision) se define como (Ec. 8):

$$P = \frac{\text{número de verdaderos positivos}}{\text{número de verdaderos positivos} + \text{número de falsos positivos}} \quad (8)$$

y R (Recall) se define como (Ec. 9):

$$R = \frac{\text{número de verdaderos positivos}}{\text{número de verdaderos positivos} + \text{número de falsos negativos}} \quad (9)$$

Para calcular los parámetros R y P hemos comparado la segmentación del algoritmo desarrollado en este proyecto con un algoritmo basado en segmentación en color que se considera como segmentación manual perfecta. Así, en primer lugar se incluye el Cuadro 2, donde se muestran los valores del factor F para imágenes de manos sin ruido pero con distorsión gaussiana. Seguidamente, en el Cuadro 3, se incluyen los valores del factor F para imágenes de manos con ruido gaussiano añadido.

En el Cuadro 2 se comprueba cómo el valor del factor F disminuye al introducir ruido gaussiano (de media μ y varianza σ^2) a la imagen, luego la calidad de la segmentación se deteriora, especialmente si se aumenta la varianza del ruido añadido. Igualmente, en el Cuadro 3 se muestra el deterioro que sufre la segmentación al aplicar a la imagen original un filtro gaussiano de varianza σ^2 , lo que provoca que se vuelva borrosa.

En ambos casos, a pesar de que la distorsión de la imagen es evidente, el resultado de la segmentación sigue siendo estable. Los resultados obtenidos en el Cuadro 3 simulan el comportamiento de fotografías que ha sufrido un emborronamiento por causa de una mala adquisición de la mano en movimiento, por ejemplo. Con lo cual, dichos resultados tienen aplicación directa en el sistema de identificación biométrica global.

5. Conclusiones y Trabajo Futuro

Los algoritmos de segmentación multiescala SWA (Segmentation by Weighted Aggregation) han demostrado en los últimos años proporcionar buenos resultados en diferentes ámbitos, por lo que en este proyecto se ha desarrollado e implementado un método de segmentación multiescala basado en la estructura de los algoritmos SWA que trata de ofrecer buenos resultados ante el uso de imágenes de manos tomadas por dispositivos móviles.

Una de las características fundamentales que presenta el algoritmo es la introducción de una condición de parada que recoge información de la textura de los segmentos y que controla el número de los mismos a través del parámetro k . De esta forma, se ha visto cómo al aumentar el valor de k , disminuye el número de segmentos restantes tras la ejecución del algoritmo. Dicho valor de k debe ser ajustado previamente a la ejecución del algoritmo, demostrándose que el comportamiento es similar ante el uso de imágenes de una misma base de datos.

Los resultados obtenidos reflejan la precisión de este algoritmo para entornos no controlados y para condiciones de captura de mano adversas (ruido gaussiano y filtro gaussiano).

En relación a posibles líneas de investigación futuras, una mejora relevante a desarrollar, aparte de la mencionada adaptación automática del parámetro k , puede girar en torno a la disminución de

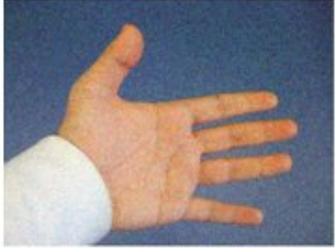
Imagen	Factor F	μ	σ^2	Parámetro k
	90.02%	0	0.01	0.003
	92.59%	0.1	0.01	0.003
	92.57%	0.2	0.01	0.003
	54.85%	0	0.05	0.005

Figura 2. Deterioro de la segmentación al introducir ruido gaussiano.

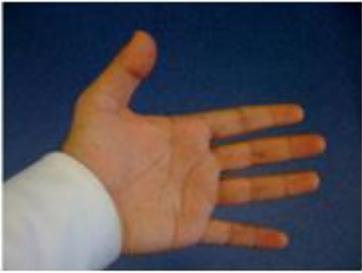
Imagen	Factor F	σ^2	Parámetro k
	99.11%	1	0.003
	89.04%	10	0.003
	93.81%	20	0.003

Figura 3. Deterioro de la segmentación al filtrar la imagen con un filtro gaussiano.

tiempos de ejecución. Dicha disminución puede ser lograda mediante la implementación de ciertas funciones de alta carga computacional (triangulación de Delaunay, ordenación de pesos) en otro lenguaje de programación más eficiente.

Por último, se estudiará la implementación de este algoritmo para dispositivos móviles y su inclusión en aplicaciones de tiempo real dentro de un sistema biométrico completo de mano.

Referencias

1. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07. pp. 1–8 (june 2007)
2. Comaniciu, D., Meer, P., Member, S.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 603–619 (2002)
3. Dyer, R., Zhang, H., Möller, T.: Delaunay mesh construction. In: SGP '07: Proceedings of the fifth Eurographics symposium on Geometry processing. pp. 273–282. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2007)
4. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1992)
5. Meirav, G., Eitan, S., Basri, R., Brandt, A.: Texture segmentation by multiscale aggregation of filter responses and shape elements. In: ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. p. 716. IEEE Computer Society, Washington, DC, USA (2003)
6. de Santos Sierra, A., Casanova, J., Avila, C., Vera, V.: Silhouette-based hand recognition on mobile devices. In: Security Technology, 2009. 43rd Annual 2009 International Carnahan Conference on. pp. 160–166 (5-8 2009)
7. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
8. Shirakawa, S., Nagao, T.: Evolutionary image segmentation based on multiobjective clustering. In: CEC'09: Proceedings of the Eleventh conference on Congress on Evolutionary Computation. pp. 2466–2473. IEEE Press, Piscataway, NJ, USA (2009)
9. Vassili, V.V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: in Proc. Graphicon-2003. pp. 85–92 (2003)