

# Identificación biométrica vascular del dorso de la mano mediante imágenes infrarrojas

Marco A. Mayén García<sup>1</sup>, Daniela Rodríguez García<sup>1</sup>,  
Benjamín Luna Benoso<sup>1</sup>, Uriel Corona Bermúdez<sup>2</sup>

<sup>1</sup> Instituto Politécnico Nacional,  
Escuela Superior de Cómputo,  
México

<sup>2</sup> Instituto Politécnico Nacional,  
Centro de Investigación en Computación  
México

{antoniomayen1997, mobius\_95}@hotmail.com,  
{danierodga, urielcoro}@gmail.com

**Resumen.** Actualmente uno de los mayores problemas de seguridad en las organizaciones es la falta de sistemas computacionales que integren tecnologías que permitan la identificación confiable de miembros y visitantes. En este trabajo se propone una arquitectura que permite la identificación de un individuo por medio de la red de venas en el dorso de la mano haciendo uso de imágenes infrarrojas obtenidas en un ambiente controlado a través de una cámara digital, aplicando técnicas de análisis de imágenes y reconocimiento de patrones.

**Palabras clave:** Biometría, patrón vascular, procesamiento de imágenes, reconocimiento de patrones, identificación biométrica.

## Vascular bBometric Identification of the Back of the Hand using Infrared Images

**Abstract.** Currently one of the biggest security problems in organizations is the lack of computer systems that integrate technologies that allow the reliable identification of members and visitors. In this work, an architecture is proposed that allows the identification of an individual through the network of veins on the back of the hand using infrared images obtained in a controlled environment through a digital camera, applying image analysis techniques and pattern recognition.

**Keywords:** Biometrics, vascular pattern, image processing, pattern recognition, biometric identification.

## 1. Introducción

En México, el delito de robo de identidad va en aumento día con día, según datos del Banco de México, el país ocupa el octavo lugar a nivel mundial en este delito. Confirmar la identidad de un individuo se ha convertido en la pieza clave para reducir el riesgo de fraudes por robo de identidad en las organizaciones alrededor del mundo [1, 2].

Los identificadores biométricos son mediciones del cuerpo humano vivo. El reconocimiento biométrico, o simplemente biometría, se refiere al uso de características o identificadores anatómicos y conductuales distintivos (por ejemplo, huellas dactilares, cara, iris, voz, geometría de la mano, etc.) para reconocer automáticamente a una persona [3].

Los identificadores biométricos no se pueden extraviar, falsificar o compartir fácilmente, se consideran más confiables para el reconocimiento de personas que los métodos tradicionales de tarjeta (tarjetas de identificación) o basados en conocimientos como contraseñas o número de identificación personal (PIN, por sus siglas en inglés).

Los objetivos del reconocimiento biométrico son los beneficios del usuario (retiro de dinero en un cajero automático sin tarjeta o PIN), mejor seguridad (solo la persona autorizada puede entrar en una instalación), una mayor responsabilidad (difícil de negar haber accedido a registros confidenciales), y una mayor eficiencia (menor sobrecarga que el mantenimiento de contraseñas en computadora) [3].

En este artículo se presenta una arquitectura que permite la identificación de una persona por medio del patrón biométrico vascular de las venas del dorso de la mano. La tecnología de patrón de venas utiliza la red vascular subcutánea en el dorso de la mano para verificar la identidad de los individuos en las aplicaciones biométricas. El principio de esta tecnología se basa en el hecho de que el patrón de los vasos sanguíneos es único para cada individuo, incluso entre gemelos idénticos [4].

## 2. Materiales y métodos

### 2.1. Ambiente controlado

Un ambiente controlado, es un ambiente en el que los factores de luz, temperatura, humedad relativa, etc., se establecen artificialmente [5]. En el presente trabajo, se ha implementado un ambiente que permite la entrada de luz natural o artificial únicamente por un cuadrado de lado de 11.5cm. En su interior contiene un aparato localizador de venas con un regulador manual de potencia que utiliza leds infrarrojos para transmitir luz infrarroja a través de la palma de la mano hasta el dorso, develando así la red vascular dorsal, como se puede visualizar en la fig. 1. También, contiene tres espejos en las caras inferior, posterior y una lateral para mayor reflexión de la luz infrarroja transmitida por el localizador de venas.



Fig. 1. Imagen ilustrativa de uno de los usos del localizador de venas [6].

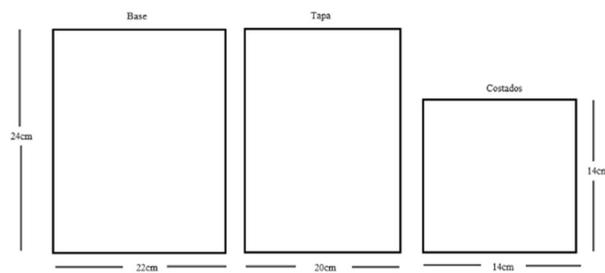


Fig. 2. Planos del prototipo de ambiente controlado.



Fig. 3. Imagen capturada en el ambiente controlado.

El ambiente controlado provee un espacio lo suficientemente amplio al usuario para introducir la palma de una mano y cerrar el puño tomando el localizador. Además, permite la entrada del lente de una cámara digital, por la tapa, para tomar fotografías de las venas visualizadas. En la fig. 2 se pueden apreciar las dimensiones del ambiente.

Posteriormente, se capturaron un conjunto de 10 imágenes, que correspondían al dorso de una mano de un sujeto, de un total de 18 sujetos. Dichas imágenes fueron tomadas con una cámara digital Canon PowerShot A470 de 7.2 megapíxeles, utilizando una configuración de ISO 200, que permitía que el sensor fuese menos sensible a la luz,

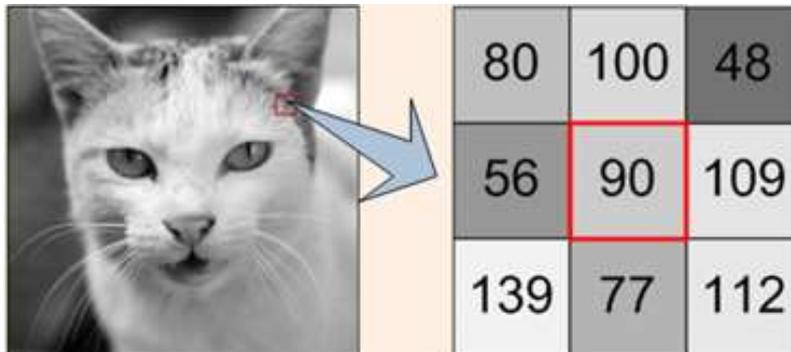


Fig. 4. Imagen en escala de grises [8].

esto con el fin de evitar obtener reflejos rojizos o blanquecinos en la imagen final, y el modo macro, que aseguraba el acercamiento a un objeto relativamente pequeño como lo era la superficie plana en el dorso de una mano. La intensidad del localizador de venas se colocó en la máxima potencia. La fig. 3 es una imagen tomada dentro del ambiente controlado.

Con el fin de incrementar el tamaño del banco de imágenes para el proceso de entrenamiento y prueba del clasificador, se realizaron 24 transformaciones geométricas que incluyen rotaciones en diferentes ángulos, traslaciones, sesgamientos, zoom y reflejos.

## 2.2. Mejoramiento de la imagen

El procesamiento digital de imágenes, incluye un conjunto de técnicas que operan sobre la representación digital de una imagen, a objeto de destacar algunos de los elementos que conforman la escena, de modo que se facilite su posterior análisis por parte de un sistema de visión artificial [7].

La técnica de suavizado, eliminación de ruido y la detección y realce de bordes se realizó haciendo uso de diferentes técnicas, descritas a continuación:

### 2.2.1. Modo de imagen escala de grises

En este modo de color la imagen está constituida por píxeles que pueden adoptar distintas tonalidades de un mismo color, por ejemplo, desde el blanco (0% de negro) hasta el negro (100% de negro). Este espectro de tono se gradúa, normalmente, en una escala que tiene 256 niveles [8]. La fig. 4 muestra un ejemplo de una imagen en escala de grises y su representación de niveles de intensidad luminosa.

### 2.2.2. Convolución

La convolución es un filtro de uso general que proporciona una forma de realizar un producto de dos matrices numéricas, normalmente de diferente tamaño, pero con el mismo número de dimensiones, para producir una tercera matriz numérica de la misma dimensión [9].

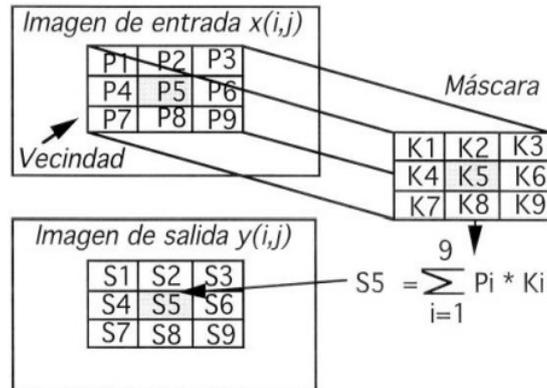


Fig. 5. Ilustración del proceso de convolución con una máscara [9].

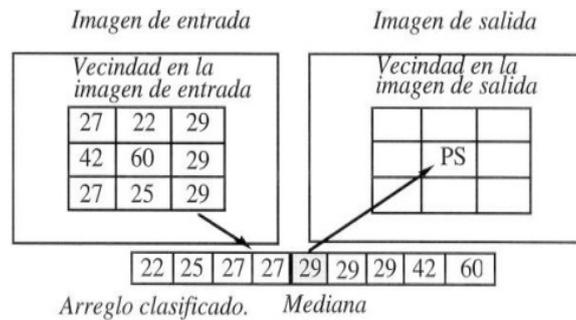


Fig. 6. Ilustración del procedimiento para implantar el filtro mediana [9].

Si consideramos una imagen como un arreglo bidimensional denotado por  $x(i,j)$  y el filtro (núcleo o máscara de convolución) con respuesta impulsiva  $h(i,j)$ , su convolución produce una imagen de salida  $y(i,j)$ , de acuerdo a la siguiente ecuación, en donde  $m$  y  $n$  definen la vecindad a considerar de acuerdo al tamaño del núcleo de convolución  $h(i,j)$  [9]:

$$y(i,j) = \sum_{m=-K1}^{K2} \sum_{n=-L1}^{L2} h(m,n)x(i-m,j-n). \quad (1)$$

La implementación de esta ecuación de convolución se hace de manera directa cuando el tamaño del filtro o máscara de convolución es pequeño (usualmente menor que 9x9 píxeles). Para la implementación directa de la ecuación de convolución es una matriz de tamaño  $(N_1 \times N_2)$ , usualmente de 3x3 píxeles, la cual adicionalmente define el tamaño de la vecindad dentro de la imagen, de manera que sea del mismo tamaño de la máscara [9].

Como se puede observar en la fig. 5, cada píxel en la imagen de salida es el resultado de la suma de los productos entre los píxeles de la máscara y los píxeles incluidos en la vecindad correspondiente en la imagen de entrada.

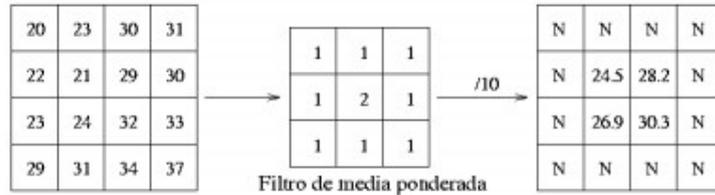


Fig. 7. Filtro media ponderada [10].

### 2.2.3. Filtro mediana

El filtraje mediana, es un procedimiento no lineal, útil para reducir el ruido impulsivo y del tipo “sal y pimienta”, presente en las imágenes. El filtro mediana utiliza los valores de los píxeles contenidos en una vecindad de tamaño impar, para determinar el nuevo valor de píxel de interés. El procedimiento para ello, consiste en ordenar todos los píxeles incluidos en la ventana en orden creciente y sustituir el píxel ubicado en el centro de la vecindad por el píxel mediano luego de la clasificación, es decir, si tenemos una secuencia discreta de tamaño  $N$  impar, entonces la mediana de tal secuencia, es aquel miembro de la secuencia, para el cual,  $(N-1)/2$  elementos son más pequeños o a lo sumo iguales y  $(N-1)/2$  elementos son más grandes. En la fig. 6 se muestra un ejemplo de la implantación del filtro mediana [9].

### 2.2.4. Filtro media ponderada

El filtro media ponderada es similar al filtro media, donde se le asigna a un píxel central la media de todos los píxeles incluidos en la ventana. Sin embargo, la matriz de filtrado está compuesta no únicamente por unos, sino que se le otorga más peso a uno de ellos (el central, habitualmente), para obtener un resultado similar a la imagen original y evitar que parezca borrosa [10]. La fig. 7 muestra una máscara media ponderada.

### 2.2.5. Modificaciones al brillo (aclorado)

El brillo aumenta la luminosidad total de una imagen, es decir, modificarlo permite aclarar los píxeles oscuros de una imagen y blanquear totalmente los claros [11].

### 2.2.6. Operación morfológica de dilatación

Sea  $A$  y  $B$  conjuntos de  $Z^2$ , la dilatación de  $A$  y  $B$ , denotada como  $A \oplus B$ , está definida como:

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\}. \quad (2)$$

Esta ecuación se basa en reflejar  $B$  sobre su origen, y moviendo esta reflexión de  $z$ . La dilatación de  $A$  y  $B$  es el conjunto de todos los desplazamientos,  $z$ , tal que  $B$  y  $A$  se superponen por al menos un elemento. Basado en esta interpretación, puede ser escrita una equivalencia como:

$$A \oplus B = \{z | [(\hat{B})_z \cap A] \subseteq A\}. \quad (3)$$

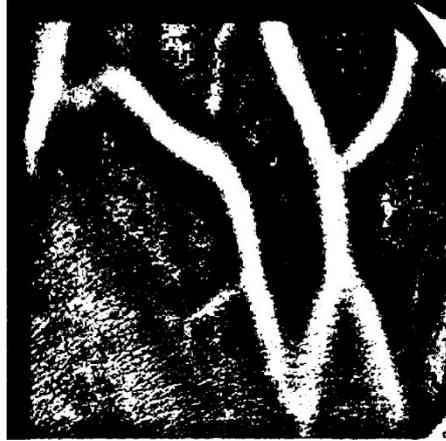


Fig. 8. Umbralado por método de Niblack.

Como antes, asumimos que B es un elemento estructural y A es el conjunto (objeto de la imagen) a ser dilatado [12].

## 2.3. Segmentación de la imagen

### 2.3.1. Segmentación por umbralado local

Este tipo de segmentación permite separar un objeto dentro de la imagen del fondo que lo circunda. La técnica se basa en comparar alguna propiedad de una imagen con un umbral fijo o variable, realizando tal comparación para cada uno de los píxeles que conforman la imagen, si el valor de la profundidad del píxel supera el valor del umbral, entonces el píxel pertenece al objeto, en caso contrario, el píxel pertenece al fondo [13].

El algoritmo de Niblack determina un valor umbral para cada píxel en sentido deslizando una ventana rectangular sobre la imagen en niveles de gris. El tamaño de la ventana rectangular puede diferir. El umbral se calcula en función de la media local  $m$  y la desviación estándar  $S$  de todos los píxeles en la ventana como se muestra en la ecuación 5:

$$T_{Niblack} = m + k * s , \quad (4)$$

$$T_{Niblack} = m + k \sqrt{\frac{1}{NP} \sum (p_i - m)^2} = m + k \sqrt{\frac{\sum p_i^2}{NP} - m^2} = m + k\sqrt{B} , \quad (5)$$

donde  $NP$  es el número total de píxeles presentes en la imagen en escala de grises,  $T_{Niblack}$  representa el valor de umbral,  $m$  es el valor promedio de los píxeles  $p_i$ , y  $k$  es fijado dependiendo del ruido que aún existe en el fondo, pudiendo ser -0.1 o -0.2 [14]. En la fig. 8 puede verse el algoritmo aplicado a la imagen capturada en el ambiente controlado.



Fig. 9. Imagen aplicando Niblack, reversión y componentes conexos.



Fig. 10. Imagen original contrastada con su esqueleto.

### 2.3.2. Componentes conexos

Sea  $G=(V, E)$  un grafo no conexo. Se denomina componente conexo de  $G$  a un subgrafo conexo maximal de  $G$  [15].

Aplicar componentes conexos permite el etiquetado de diferentes conjuntos de información conectados, y su diferenciación por el área que abarcan. Se colocó una restricción para que cualquier componente conexo menor de una dimensión de  $200 \times 200$ , fuese eliminado de la imagen, para únicamente quedarse con la mayor cantidad de información de la red, como se aprecia en la fig. 9.

### 2.3.3. Segmentación basada en esqueletizado

El esqueletizado es un proceso para reducir las regiones de primer plano en una imagen binaria a un remanente esquelético que conserva en gran medida el alcance y la conectividad de la región original mientras se descarta la mayoría de los píxeles de primer plano originales [16].

El esqueleto se puede producir de dos formas principales. El primero es usar algún tipo de adelgazamiento morfológico que erosione sucesivamente los píxeles del límite (al tiempo que conserva los puntos finales de los segmentos de línea) hasta que no sea

posible un adelgazamiento, en cuyo punto lo que queda se aproxima al esqueleto. El método alternativo es calcular primero la transformación de distancia de la imagen.

El esqueleto se encuentra a lo largo de las singularidades (es decir, pliegues o discontinuidades de curvatura) en la transformación de distancia [16]. A continuación, la fig. 10 muestra una comparación entre la imagen original capturada en el ambiente controlado y la misma imagen esqueletizada.

## 2.4. Extracción de características para el análisis de imágenes

La forma característica de un objeto puede cuantificarse mediante momentos, los cuales describen la manera en que se distribuyen los píxeles de un objeto sobre el plano de la imagen. Los momentos deben ser invariantes (es decir, valores similares para objetos del mismo tipo) a las transformaciones geométricas (traslación, rotación y escala) que pueden sufrir los objetos y al mismo tiempo deben ser discriminantes (es decir, valores distintos para objetos de diferente tipo). Estas características son deseables para poder reconocer los objetos con mayor facilidad [17].

Los momentos de Hu son un conjunto de siete descriptores invariantes que expresan numéricamente la forma de un objeto, donde los dos primeros momentos se computan como:

$$\phi_1 = \eta_{20} + \eta_{02}, \quad (6)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \quad (7)$$

donde los momentos centrales normalizados de segundo orden se calculan como:

$$\eta_{20} = \frac{m_{20} - \frac{m_{10}^2}{m_{00}}}{m_{00}^2}, \quad (8)$$

$$\eta_{02} = \frac{m_{02} - \frac{m_{01}^2}{m_{00}}}{m_{00}^2}, \quad (9)$$

$$\eta_{11} = \frac{m_{11} - \frac{m_{10}m_{01}}{m_{00}}}{m_{00}^2}, \quad (10)$$

y los momentos geométricos de orden  $p + q$  se calculan como:

$$m_{pq} = \sum_x \sum_y x^p y^q I(x, y), \quad (11)$$

donde  $I(x, y)$  es un píxel del objeto con las coordenadas  $(x, y)$  [17].

## 2.5. Entrenamiento del clasificador de patrones

Un patrón es un conjunto de características, mientras que una clase de patrones es un conjunto de patrones similares. El objetivo del reconocimiento de patrones es asignar un patrón a la clase a la que pertenece [18].

Random Forest es un método versátil de aprendizaje automático capacitado de efectuar tanto tareas de regresión como clasificación. Es un tipo de método de

**Tabla 1.** Comparativa de exactitud de algoritmos clasificadores.

Algoritmo	Exactitud	Imágenes clasificadas correctamente
SVM	83.00%	747/900
MLP	≈84.44%	760/900
kNN	≈84.44%	760/900
Árbol de decisión	≈84.44%	760/900
Random Forest	≈90.22%	812/900

**Tabla 2.** Comparativa de exactitud para diferentes valores de  $k$  utilizando validación cruzada.

$K$	Exactitud promedio
10	≈94.51%
30	≈97.15%
50	≈97.62%
100	≈98.19%

aprendizaje por conjuntos, donde un grupo de modelos débiles se combinan para formar un modelo robusto. Se generan múltiples árboles y cada árbol otorga una clasificación, es decir, vota por una clase y el resultado es la clase con mayor número de votos en todo el bosque [19].

Cada árbol se construye así:

- a) Dado que el número de casos en el conjunto de entrenamiento es  $N$ . Una muestra de esos  $N$  casos se toma aleatoriamente, pero con reemplazo. Esta muestra será el conjunto de entrenamiento para construir el árbol  $i$ .
- b) Si existen  $M$  variables de entrada, un número  $m < M$  se especifica tal que, para cada nodo,  $m$  variables se seleccionan aleatoriamente de  $M$ . La mejor división de estos  $m$  atributos es usado para ramificar el árbol. El valor  $m$  se mantiene constante durante la generación de todo el bosque.
- c) Cada árbol crece hasta su máxima extensión posible y no hay proceso de poda [20].

### 3. Los resultados

Después de aplicar 24 transformaciones geométricas, el banco de imágenes final se incrementó en un 2,500%, obteniendo 4,500 imágenes de cada una de las cuales fue sustraído un vector característico de tamaño 7.

Se destinó un 80% de estos vectores obtenidos para el conjunto de entrenamiento y el 20% restante para el conjunto de prueba, es decir, que se usaron 3,600 vectores característicos para el primer conjunto y 900 para el segundo.

Se realizaron pruebas con otros tipos de clasificadores de patrones, como Máquina de Soporte Vectorial (SVM), Red Neuronal Multicapa (MLP), kNN, y árbol de decisión, para comparar el desempeño de éstos con respecto al clasificador utilizado en este trabajo, Random Forest. Enseguida, en la tabla 1, se muestran los porcentajes de exactitud y el número de imágenes clasificadas correctamente alcanzados por cada uno de los algoritmos implementados haciendo uso de librerías del lenguaje de alto nivel Python en su versión 3.7.

Sin embargo, dada la naturaleza estocástica de la asignación de porcentajes a los conjuntos de entrenamiento y de prueba, no es posible asegurar que se obtendrán los mismos resultados y, en consecuencia, las mismas precisiones y errores en cada corrida del algoritmo. Por lo cual, se implementó un algoritmo de validación para observar el desempeño del clasificador de patrones en diferentes escenarios.

El método de validación cruzada consiste en tomar los datos originales, en este caso los 4,500 vectores característicos generados, para elaborar a partir de ellos los conjuntos separados de entrenamiento y prueba, para luego dividir en  $k$  subconjuntos el conjunto de entrenamiento, y en el momento del entrenamiento del algoritmo, tomará cada  $k$  subconjunto como conjunto de prueba del modelo, mientras que el resto será utilizado como conjunto de entrenamiento. El proceso está pensado para repetirse  $k$  veces, y en cada repetición, se elegirá un conjunto de prueba diferente y nuevamente los restantes serán parte del conjunto de entrenamiento. Cuando se realicen en su totalidad las  $k$  iteraciones, se calcula la precisión y el error para cada uno de los modelos producidos y se calcula el promedio de los  $k$  modelos para obtener la precisión y error finales [21].

Para diferentes valores de  $k$ , se muestra a continuación la exactitud promedio de los modelos generados en la tabla 2.

#### **4. Conclusiones y trabajos futuros**

Se ha presentado una arquitectura que permite la identificación de un sujeto, con una exactitud promedio mayor al 90%. Esto nos hace capaces de afirmar que, será capaz de identificar correctamente una imagen correspondiente al dorso de una mano de una persona, 9 de cada 10 veces que esa imagen sea ingresada.

Se logró implementar un ambiente controlado, que permite la captura de imágenes infrarrojas a través de la técnica de transiluminación.

También, se consiguió segmentar la red vascular subcutánea del dorso de una mano, manteniendo una abstracción de la información original.

Con todo esto dicho, las implicaciones de esta arquitectura podrían llevar a implementar la arquitectura en el ámbito público o privado, permitiendo la identificación de individuos de organizaciones pequeñas y medianas, dados los tiempos de procesamiento.

Como trabajo futuro debería considerarse la reducción de tiempos de procesamiento y optimización de los algoritmos implementados para realizar la identificación, así como la automatización de los procesos de captura de imágenes infrarrojas y transformar una sola toma del dorso de una mano, en un barrido fotográfico del dorso de una mano.

## Referencias

1. Barrera-Rubio, P.: Realidad y prevención: Robo de identidad en México. *El Economista* (2020)
2. Cero Papel.: El 26% de las empresas en el mundo utilizan biometría para combatir el fraude (2019)
3. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: *Handbook of fingerprint recognition*. Springer-Verlag, pp. 2–3 (2009)
4. Barrales-López, A.L.: Identificación biométrica por medio de la detección de venas de la mano en imágenes infrarrojas. Tesis de maestría CIDETEC (2015)
5. Real Academia de Ingeniería: Ambiente controlado. *Diccionario de la Real Academia de Ingeniería* (2015)
6. Sonomedical: Localizador de venas (2019)
7. Passariello, G., Mora, F.: *Imágenes médicas: Adquisición, análisis, procesamiento e interpretación*. Ediciones de la Universidad Simón Bolívar, pp. 61–98 (1995)
8. Passariello, G., Mora, F.: *Imágenes médicas: Adquisición, análisis, procesamiento e interpretación*. Ediciones de la Universidad Simón Bolívar, pp. 36–38 (1995)
9. González, R., Woods, R.: Image segmentation. *Digital Image Processing* (pp. 778-780). Ed. Pearson-Prentice Hall (2007)
10. Alonso, F.: Tema 6 técnicas de filtrado. <https://um.es/geograf/sigmur/teledet/tema06.pdf> (2005)
11. Arcgis.com: Función de brillo y contraste. ArcGIS for Desktop <https://desktop.arcgis.com/es/arcmap/10.3/manage-data/raster-and-images/contrast-and-brightness-function.htm> (2020)
12. González, R., Woods, R.: Morphological image processing. *Digital Image Processing*. Pearson-Prentice Hall, pp. 653 (2007)
13. Gómez, W.: Análisis de imágenes digitales, segmentación por umbralado. Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV), pp. 35 (2015)
14. Ramudu, K., Krishna-Reddy, V., Abdul-Rahim, B.: Niblack method based segmentation for microscopic imagery. *International Journal of Electrical and Electronics Engineers*, 7 (1), pp. 26–27 (2015)
15. UDEA: 1.6. Conectividad, componentes conexas, vértice y aristas de corte. Universidad de Antioquia. [http://docencia.udea.edu.co/regionalizacion/teoriaderedes/informaci%F3n/C1\\_Conectividad.pdf](http://docencia.udea.edu.co/regionalizacion/teoriaderedes/informaci%F3n/C1_Conectividad.pdf) (2005)
16. Homepages: Morphology-skeletonization/medial axis transform. <https://homepages.inf.ed.ac.uk/rbf/HIPR2/skeleton.htm> (2020)
17. Gómez, W.: Reconocimiento de objetos en fotografías. Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV), <https://tamps.cinvestav.mx/~wgomez/toptamps/presentacion.pdf> (2020)
18. Grupo de Topología Computacional y Matemática Aplicada: Introducción al reconocimiento de objetos. <http://grupo.us.es/gtocoma/pid/tema7.pdf> (2020)
19. González, L.: Aprendizaje supervisado: Random forest classification - Ligdi González <https://ligdigonzalez.com/aprendizaje-supervisado-random-forest-classification/> (2020)
20. Orellana, J.: Árboles de decisión y Random Forest. <https://bookdown.org/content/2031/ensambladores-random-forest-parte-i.html> (2020)
21. Delgado, R.: Introducción a la validación cruzada (k-fold cross validation) en R. [http://rstudio-pubs-static.s3.amazonaws.com/405322\\_6d94d05e54b24ba99438f49a6f8662a9.html](http://rstudio-pubs-static.s3.amazonaws.com/405322_6d94d05e54b24ba99438f49a6f8662a9.html) (2018)