

# Large-scale palm vein recognition on synthetic datasets

Ruber Hernández-García<sup>\*†</sup>, José I. Santamaría<sup>\*</sup>, Ricardo J. Barrientos<sup>\*</sup>, Edwin H. Salazar Jurado<sup>\*</sup>,  
Francisco Manuel Castro<sup>‡</sup>, Julián Ramos-Cózar<sup>‡</sup>, and Nicolás Guil<sup>‡</sup>

<sup>\*</sup>Laboratorio LITRP, Depto. DCI, Facultad de Ciencias de la Ingeniería, Universidad Católica del Maule, Talca, Chile

<sup>†</sup>Centro de Investigación de Estudios Avanzados del Maule (CIEAM), Universidad Católica del Maule, Talca, Chile

<sup>‡</sup>Depto. Arquitectura de Computadores, ETSI Informática, Universidad de Málaga, Málaga, España

Email: rbarrientos@ucm.cl

**Abstract**—During the last decade, palm vein recognition has gained the attention of the research community on biometric systems since it presents high security. However, its applications on massive individuals identification are limited mainly because publicly available datasets have very small numbers of subjects. In this context, synthetic datasets are helpful to evaluate the performance and scalability of biometric systems on large-scale databases. Thus, the present work evaluates CNN-based models on two self-created synthetic datasets. For this purpose, we implemented two end-to-end CNN architectures based on AlexNet and Resnet32. Besides, we created two large-scale synthetic datasets by using a StyleGAN-based model and a specific method based on biological transport networks, which are comprised of 10,000 and 2,000 individuals, respectively. The generated datasets are the largest of the state-of-the-art and were validated by using different quantitative metrics in order to measure their visual quality and realism comparing to real images. The experimental results show the applicability and quality of the proposed synthetic databases in order to evaluate the efficiency and scalability of palm vein recognition methods.

**Index Terms**—Convolutional neural networks, Palm vein recognition, Large-scale recognition, Synthetic datasets.

## I. INTRODUCCIÓN

Durante la última década, el reconocimiento de venas de la palma de la mano ha ganado la atención de la comunidad científica en el área biométrica ya que presenta una alta seguridad. Un sistema de reconocimiento basado en las venas de la palma se puede resumir en cuatro procesos: la adquisición de la imagen, el preprocesamiento, la extracción de características y el reconocimiento [1]. En comparación con otras técnicas biométricas tradicionales, los patrones vasculares como rasgo biométrico no está completamente estudiado, por lo que existen varios desafíos en cada uno de los procesos. En particular, sus aplicaciones en la identificación masiva de personas son muy limitadas, principalmente porque las bases de datos disponibles públicamente tienen un número muy reducido de sujetos.

Para hacer frente a las limitaciones mencionadas, en algunas técnicas biométricas (p.e. huellas dactilares, iris o rostro) se ha adoptado la generación de bases de datos sintéticas. En particular, las imágenes sintéticas son muy eficaces para evaluar el rendimiento de diferentes algoritmos de procesamiento de imágenes, pero su uso con fines biométricos es una cuestión polémica. Sin embargo, las bases de datos sintéticas tienen la

ventaja de evitar un proceso de recopilación de muestras que requiere mucho tiempo, y además no comprometen en absoluto la seguridad de los usuarios, que suele ser una preocupación personal en cuanto a la violación de la privacidad [2]. Además, aunque este tipo de conjunto de datos no sustituye totalmente la validación con imágenes reales, sí facilita la validación de los enfoques propuestos en bases de datos a gran escala.

En este trabajo, se evalúan dos modelos de redes convolucionales (CNN) en dos conjuntos de datos de imágenes sintéticas de venas de la palma a gran escala. Para ello, se implementaron dos arquitecturas CNN de extremo a extremo (end-to-end) basadas en AlexNet [3] y Resnet32 [4]. Ambos conjuntos de datos sintéticos son de creación propia, para lo que se desarrolló un esquema de generación de imágenes sintéticas de patrones de venas de la palma que puede ser adaptado a diferentes modelos generativos para la simulación de las imágenes reales. La primera base de datos llamada *Synthetic-PVDB* está conformada por 10,000 individuos y sus imágenes fueron generadas mediante un modelo StyleGAN2 [5]. El otro conjunto de datos, llamado *NS-PVDB*, fue creado usando un método basado en redes de transporte biológico [6] y se compone de 2,000 sujetos. Los conjuntos de datos generados son los más grandes del estado del estado del arte y fueron validados usando diferentes métricas cuantitativas para medir sus similitudes en comparación con las imágenes reales.

## II. ESTADO ACTUAL DE LAS BASES DE DATOS PÚBLICAS PARA RECONOCIMIENTO DE VENAS DE LA PALMA

Entre las aplicaciones reales del reconocimiento de venas de la palma, se debe destacar su adopción para fines de autenticación en varios tipos de servicios públicos [7]. Fujitsu Labs ofrece un escáner de venas de la palma y un kit de desarrollo de software para la autenticación mediante venas de la palma, que presenta una tasa de falsa aceptación (FAR) igual a 0.00008%, y una tasa de falso rechazo del 0.01% [8]. Para esto, Fujitsu realizó el mayor estudio conocido sobre reconocimiento de venas de la palma con una base de datos compuesta por 75,000 sujetos y un total de 150,000 muestras. Sin embargo, su tecnología es privada y está patentada, lo que no permite el uso de las imágenes de venas capturadas por su dispositivo para evaluar diferentes enfoques o crear bases de datos con fines de investigación. Además, el estudio se realizó

con fines comerciales y la base de datos no está disponible para reproducir los experimentos o comparar nuevos enfoques.

Por estas razones, la comunidad científica ha introducido algunas bases de datos de imágenes de venas de la palma de creación propia [9], que se han recogido utilizando dispositivos de bajo coste. Como resultado de lo anterior, hay muchos enfoques no estandarizados mediante el uso de diferentes protocolos de adquisición y dispositivos. Además, estas condiciones producen diferentes parámetros de imagen, variaciones en la iluminación y el contraste, y también variados resultados de rendimiento. En la Tabla I se ofrece una visión general de las bases de datos disponibles públicamente, donde se compara el número de sujetos, las manos y las muestras capturadas, las sesiones de adquisición, el total de imágenes y su resolución.

Como puede observarse en la Tabla I, las bases de datos públicas generalmente están compuestas por un número limitado de personas con no más de 300 individuos. Por lo general, los enfoques propuestos para el reconocimiento de venas de la palma siguen el protocolo *nom L & R* de [11], donde ambas manos del mismo sujeto se consideran sujetos diferentes, duplicando el número de sujetos en el conjunto de datos. Sin embargo, todavía se requiere un número mayor de individuos que permitan estudios de escalabilidad de los métodos propuestos para la identificación masiva de personas. Una posible solución para abordar este inconveniente podría ser utilizar imágenes de varias bases de datos [17], pero esto no resolvería el problema debido a dos razones principales. En primer lugar, con esto solo se obtendría un total de 2,510 sujetos, lo que aún sería insuficiente para evaluar métodos de reconocimiento masivo de personas. En segundo lugar, y muy notable, las variaciones de las imágenes resultantes de diferentes sistemas de adquisición (p.e. resolución, iluminación, contraste, etc.) pueden causar problemas en el entrenamiento y ajuste de los parámetros de los algoritmos propuestos.

En este sentido, el presente estudio pretende analizar el rendimiento de modelos CNN sobre conjuntos de datos de creación propia compuestos por imágenes sintéticas de patrones de venas. Dentro del contexto biométrico y según la revisión bibliográfica, existen pocos trabajos enfocados a proporcionar metodologías para la generación de imágenes sintéticas para el reconocimiento basado en las venas, como son las venas de los dedos [18], el dorso de la mano [19], y la esclera [20]. Por lo tanto, en la siguiente sección, se presenta un nuevo esquema para la generación de bases de datos de imágenes sintéticas de venas de la palma con fines biométricos.

### III. RECONOCIMIENTO DE VENAS DE LA PALMA SOBRE BASES DE DATOS SINTÉTICAS A GRAN ESCALA USANDO MODELOS CNN

#### A. Bases de datos de imágenes sintéticas de venas de la palma

En la Fig. 1, se formaliza un esquema general del modelo propuesto para la generación de bases de datos sintéticas a gran escala. Este enfoque se centra en la generación de muestras de la región de interés (ROI) de las imágenes de venas de la palma de la mano, ya que normalmente los

métodos propuestos en la literatura utilizan solamente el ROI de la mano. El diagrama de flujo propuesto consta de tres procesos principales, mientras que el primero varía en función del método de generación implementado, los dos últimos son comunes a ambos enfoques.

- 1) **Generador de imágenes de venas (VIG):** El proceso VIG tiene como objetivo la generación de imágenes sintéticas aleatorias que simulan tanto los patrones venosos (i.e. patrones grises oscuros de las estructuras vasculares) como las características visuales (i.e. textura, contraste e iluminación) de las imágenes de las venas de la palma. Es importante señalar que existen diferentes métodos para producir estructuras vasculares sintéticas. Por lo tanto, el modelo presentado puede ser implementado utilizando una variedad de enfoques mediante la modificación del generador en correspondencia con los diferentes parámetros involucrados en el proceso de adquisición de imágenes reales. Así, se evalúan dos enfoques de propósito general para implementar este proceso, que son explicados en la Sección IV-A. En primer lugar, se estudia una novedosa aplicación de dominio específico para la arquitectura de Redes Neuronales Generativas basada en estilos (StyleGAN) [21] para generar imágenes sintéticas de venas de la palma, bajo el fundamento de que el proceso de adquisición produce imágenes con características muy particulares en términos de textura, iluminación y contraste. Este enfoque pretende generar imágenes sintéticas de venas de la palma simulando el mismo estilo de las imágenes reales. Por otra parte, se examina un enfoque para simular el plexo venoso de la palma de la mano, que se inspira en la red de transporte natural del crecimiento del *Physarum polycephalum* que forma patrones de trayectoria similares a las estructuras vasculares.
- 2) **Discriminador de unicidad de individuos (IUD):** El proceso anterior genera aleatoriamente imágenes sintéticas de venas de la palma, pero estas deben satisfacer el requisito de unicidad para formar parte de una base de datos biométrica. Por esto, el proceso IUD compara cada imagen recién generada con todas las imágenes de la base de datos sintética, garantizando este requisito. Dado que el proceso de creación de la base de datos sintética es incremental, se adopta un procedimiento de comparación basado en un umbral de unicidad mediante el cálculo de la medida de similitud propuesta en [22]. De acuerdo con los experimentos de [22], se fija el umbral de unicidad en 0.1 como un criterio aceptable de separabilidad entre muestras genuinas e impostoras. Por lo tanto, si la similitud con cualquier muestra de la base de datos es superior al umbral de unicidad, la imagen generada se descarta. En caso contrario, si no se supera el umbral para ninguna de las muestras de la base de datos, se considera que la imagen generada puede incorporarse a la base de datos como un nuevo sujeto. Este umbral garantiza además la dificultad de reconocimiento de individuos en

TABLA I: Resumen general de bases de datos de venas de la palma disponibles publicamente. El número de sujetos y el número de muestras son expresados como Sujetos  $\times$  Manos, y Muestras  $\times$  Sesiones de adquisición, respectivamente.

Base de datos	Sujetos	Muestras	Total de Imágenes	Tamaño de Imágenes	Intervalos de Sesiones	Etiquetado
CASIA [10]	100 $\times$ 2	3 $\times$ 2	7,200	768 $\times$ 576	30 días	Id
VERA [11]	110 $\times$ 2	5 $\times$ 2	2,200	480 $\times$ 680	5 minutos	Id, Edad, Sexo
PUT [12]	50 $\times$ 2	4 $\times$ 3	1,200	1280 $\times$ 960	7 días	Id
PolyU [13]	250 $\times$ 2	6 $\times$ 2	6,000	352 $\times$ 288	9 días	Id
Tongji* [14]	300 $\times$ 2	10 $\times$ 2	12,000	800 $\times$ 600	61 días	Id
IITI* [15]	185 $\times$ 2	6 $\times$ 1	2,220	2,592 $\times$ 1944	N/A	Id
FYO [16]	160 $\times$ 2	1 $\times$ 2	640	800 $\times$ 600	10 minutos	Id

\*Estas bases de datos están disponibles mediante solicitud a los autores.

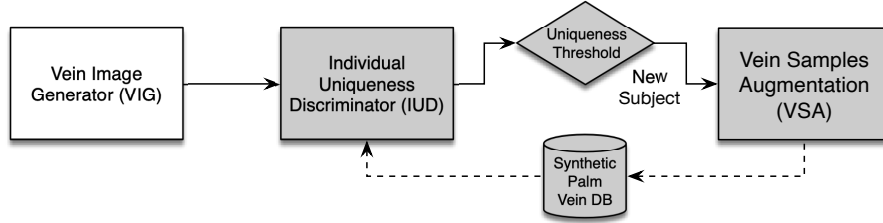


Fig. 1: Diagrama de flujo de los enfoques propuestos para la generación de imágenes sintéticas de venas de la mano para la creación de bases de datos sintéticas. Los procesos resaltados en color gris son comunes a ambos enfoques, mientras que el proceso VIG varía en función del método de generación implementado.

el conjunto de datos al preservar individuos con cierta similitud, ya que de acuerdo a los experimentos previos, en el rango de  $[\sim 0.03..0.1]$  existen ciertas muestras de individuos diferentes con alta similitud.

- 3) **Aumento de Muestras (VSA):** Dado que una base de datos biométrica debe contener más de una muestra por individuo (p.e. muestras de galería y de prueba), con el fin de aumentar el número de imágenes por sujeto, se realiza un proceso de aumento de muestras para cada nueva imagen validada por el proceso IUD. Así, el proceso VSA tiene como objetivo obtener diferentes muestras para cada sujeto aplicando transformaciones aleatorias sobre la imagen generada, simulando las variaciones naturales que ocurren durante un procedimiento de adquisición sin contacto en el mundo real. Para ello, se aplican aleatoriamente dos tipos de transformaciones sobre la muestra principal: variaciones de textura/iluminación (contraste, iluminación y desenfoque) para simular el efecto de la iluminación NIR, y transformaciones afines del ROI (traslación, rotación, recorte y acercamiento) para replicar los cambios de posición de la mano en el dispositivo de adquisición. Como resultado, se obtienen seis muestras por cada nuevo sujeto de la base de datos.

### B. Proceso de aumentación de muestras de ROI de la palma

Debido a que los conjuntos de datos públicos de imágenes reales tienen un número limitado de muestras, esta es una limitación a solventar para el entranamiento de modelos CNN. Así, a diferencia de otros enfoques que aplican el aumento de la imagen a partir de las muestras de ROI durante el

proceso de entrenamiento, se desarrolló un procedimiento de aumento de muestras durante la etapa de preprocesamiento. De esta manera, se logran obtener 87x muestras de ROI de cada imagen a partir de la traslación y rotación de un ROI central (cuadrado azul) dentro de toda la palma (cuadrado rojo), como se muestra en la Fig. 2. El ROI central está centrado en la región de interés de la palma, y su lado está determinado por los puntos medios entre los dedos índice y medio, y el anular y el meñique, respectivamente. Para aumentar el número de muestras de ROI, este ROI central se desplaza repetidamente en ambos ejes con  $\Delta(X, Y) = 5px$ , y también se rota con  $\Delta\theta = 5^\circ$  hasta  $\pm 15^\circ$ .

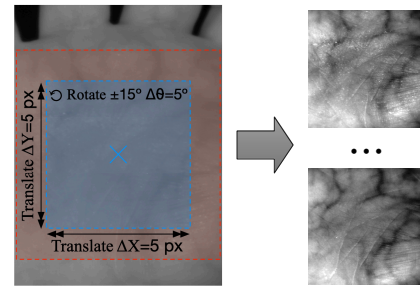


Fig. 2: Proceso de aumentación de muestras de ROI.

Este procedimiento de aumentación de muestras ROI fue realizado mediante un script de MATLAB, que a partir de la detección del ROI, de forma automática permitió aumentar los conjuntos de imágenes de las bases de datos empleadas. Se utilizó el mismo proceso en todas las implementaciones de la metodología propuesta. En primer lugar, se aumentaron las muestras de referencia para el entrenamiento de los mo-

delos generativos implementados, tanto para la generación de imágenes sintéticas de venas como para la simulación de la textura de la palma de la mano. Luego, el mismo proceso fue ejecutado para evaluar la efectividad de los modelos CNN en el reconocimiento de personas.

### C. Modelos CNN evaluados para el reconocimiento de personas mediante las venas de la palma

En la metodología propuesta, se evaluaron dos arquitecturas CNN de principio a fin (end-to-end) que varían en función de su profundidad. El primer modelo se basó en la arquitectura AlexNet [3], la cual se denominó como SingleNet y fue ajustada en función de la tarea de clasificación. La segunda arquitectura fue seleccionada teniendo en cuenta la aplicación de un esquema de clasificación más complejo que incluyera operaciones residuales, para esto se implementó un modelo basado en la arquitectura ResNet32 [4]. A continuación, se describen los detalles de las implementaciones de cada modelo evaluado. Ambos modelos fueron seleccionados con el objetivo de validar el rendimiento de modelos CNN de principio a fin de poca complejidad para el reconocimiento de personas a gran escala, sin la necesidad de usar modelos pre-entrenados.

1) *Modelo SingleNet*: Las especificaciones de la arquitectura implementada se puede detallar en la Tabla II. El modelo está compuesto por cinco capas convolucionales seguidas de tres capas densas (*fully-connected*). Todas las capas convolucionales usan activaciones *ReLU*, además están seguidas de operaciones de normalización (*BatchNorm*) y agrupamiento (*max-pooling*). Además de las operaciones de normalización anteriores, las dos primeras capas *fully-connected* se acompañaron de operaciones de abandono (*dropout*) con el objetivo de reducir el sobreajuste de la red. Finalmente, la última capa densa está compuesta por tantas neuronas como clases (individuos) del conjunto de datos y utiliza la función de pérdida *softmax*. Para el entrenamiento, el modelo CNN fue compilado usando la función de pérdida *categorical\_crossentropy* y la función de optimización *SGD*.

TABLA II: Especificaciones de la arquitectura SingleNet. Nomenclatura: P=tamaño de la agrupación; BatchNorm=normalización por lotes; Dr=tasa de abandono; C=número de salidas (clases).

Capas	Especificaciones	Post-procesamiento
Input	$64 \times 64$	-
Conv01	$7 \times 7 \times 96$	BatchNorm; P: $2 \times 2$
Conv02	$5 \times 5 \times 192$	BatchNorm; P: $2 \times 2$
Conv03	$3 \times 3 \times 284$	BatchNorm; P: $2 \times 2$
Conv04	$3 \times 3 \times 512$	BatchNorm; P: $2 \times 2$
Conv05	$2 \times 2 \times 1024$	BatchNorm; P: $2 \times 2$
Full01	1024	Dr: 0.65
Full02	512	Dr: 0.65
Full03	C	-

2) *Modelo ResNet32*: La arquitectura ResNet [4] se basa en una estructura de conexión de acceso directo llamada módulo residual, donde se apilan varios módulos residuales. La conexión residual realiza un mapeo de identidad y sus salidas se agregan a la salida de la siguiente capa, alterando ligeramente la entrada en función de la entrada original o una

anterior. Este tipo de operación ayuda a prevenir la pérdida de información a lo largo del flujo de procesamiento sin aumentar el número de parámetros ni la complejidad computacional del modelo, y puede mejorar el rendimiento del modelo.

En la Tabla III se muestran las especificaciones de la arquitectura, que está compuesta por tres grupos de módulos residuales (ResStack). Al principio de cada grupo, el tamaño del mapa de características se reduce a la mitad mediante una capa convolucional con *stride* = 2, mientras que el número de filtros se duplica. Dentro de cada grupo, todas las capas tienen la misma cantidad de filtros y la última activación ReLU se realiza después de la conexión residual. Después de los grupos residuales se computa una agrupación promedio (*AvgPooling*) y se añade la capa densa de salida para la clasificación. El modelo se compiló con la función de pérdida *categorical\_crossentropy* y una optimización *Adam*.

TABLA III: Especificaciones de la arquitectura ResNet32, el número de salidas de la capa Full01 se corresponde con el número de clases C.

Capas/Bloques	Especificaciones	Post-procesamiento
Input	$64 \times 64$	-
Conv2D-BN-ReLU	$3 \times 3 \times 16$	BatchNorm
ResStack01	$32 \times 32$ conv; 16	-
ResStack02	$16 \times 16$ conv; 32	-
ResStack03	$8 \times 8$ conv; 64	-
AvgPooling	$8 \times 8$	-
Full01	C	-

## IV. RESULTADOS EXPERIMENTALES

Los experimentos se realizaron en un servidor dedicado compuesto por dos CPUs Intel Xeon Gold 6140 (36 núcleos físicos), 126 GB de RAM y cuatro GPUs NVIDIA GeForce GTX 1080Ti. Los modelos se implementaron utilizando Python 3.7 y OpenCV 4.1, mientras que la implementación de StyleGAN2 y ambos modelos CNN requirió TensorFlow 1.15, CUDA 10.1 y cuDNN 7.5.

### A. Implementación del proceso VIG para la generación de imágenes sintéticas

1) *Generación de imágenes sintéticas basada en estilo*: Diferentes enfoques del estado del arte han adoptado variantes GAN para fines de aumento de muestras en el reconocimiento de las venas [23]. En particular, el modelo StyleGAN2 [5] ha demostrado obtener excelentes resultados en términos de calidad de imagen, variabilidad, eficiencia y generando imágenes más realistas. Por esto, se implementó el proceso VIG usando el modelo StyleGAN2, con el objetivo de generar muestras aleatorias de las venas de la palma de la mano y reproducir el *estilo* de las imágenes reales de venas de la palma de la mano con un alto nivel de realismo, simulando las características visuales de las imágenes en términos de textura, iluminación y contraste. Este proceso se ha llamado *Style-VIG* y la Fig. 3(a) muestra el diagrama de flujo del proceso.

En nuestros experimentos, se entrenó la red StyleGAN2 usando imágenes de la banda 850nm del conjunto de datos CA-SIA [10], que se utiliza habitualmente para el reconocimiento

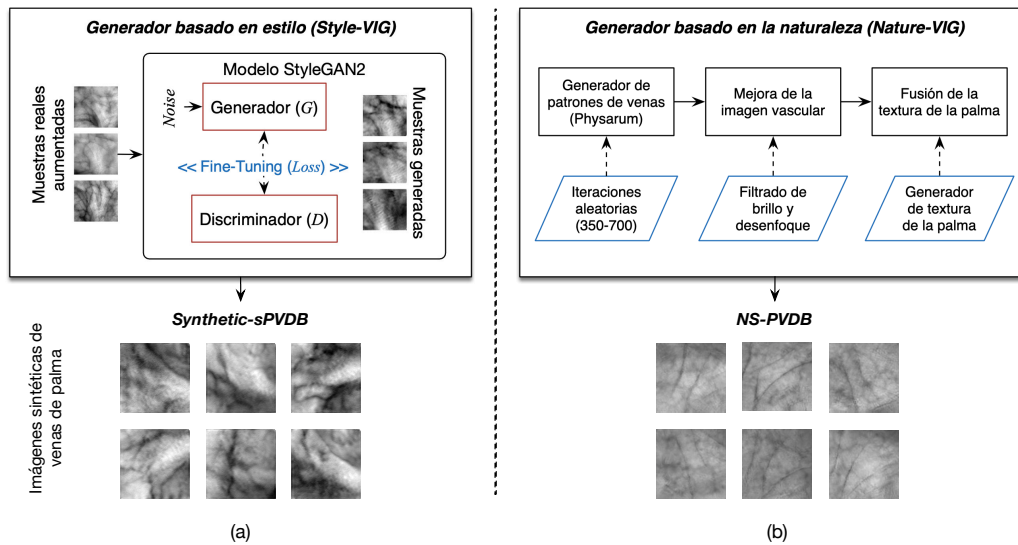


Fig. 3: Representación del diagrama de flujo de las variantes implementadas del proceso VIG: (a) Style-VIG, y (b) Nature-VIG.

de venas de la palma. Los datos de entrenamiento originalmente estaban compuestos por 1,200 imágenes reales, que se incrementaron hasta 104,400 muestras de entrenamiento después de realizar el procedimiento de aumentación de ROI. La implementación se basó en el código fuente oficial de StyleGAN2 de NVIDIA Research Projects<sup>1</sup>, utilizando la configuración original (config-e) de la arquitectura para 25,000 k-imágenes. Finalmente, el proceso de entrenamiento reportó una medida FID igual a 63.89. Luego de entrenado el modelo, el proceso de generación de una nueva imagen sintética demora en promedio 58.1ms ( $\pm 259.69$ ms).

2) **Generación de imágenes sintéticas basada en la naturaleza**: Como segundo enfoque para la implementación del proceso VIG, se desarrolló un método algo más específico basado en las redes de transporte biológico. En este caso, se aplicó un modelo matemático para simular el plexo venoso de la mano. Este enfoque se basa en el principio de que la red vascular adapta su estructura de forma óptima al movimiento de la mano y, por tanto, a sus músculos. El modelo matemático se basa en una búsqueda de caminos con dinámica adaptativa, que se ha utilizado para simular el crecimiento del *Physarum polycephalum* [6]. En particular, este organismo forma dinámicamente patrones de trayectorias regulares similares a la red del plexo venoso, ya que sus patrones de crecimiento establecen estructuras organizativas óptimas con alta eficiencia. Este variante del proceso VIG se denominó *Nature-VIG* y se representa en la Fig. 3(b).

El modelo matemático implementado para simular una estructura aproximada de los patrones de las venas de la palma fue el propuesto en [6] y se basa en un sistema multi-agente autoorganizado. Inicialmente, los agentes se mueven aleatoriamente dentro del cuadrado que representa el ROI de la palma y luego convergen a la red más óptima para conectar las fuentes de alimento situadas en los puntos medios de la

parte superior e inferior de la región, que análogamente corresponden a la parte proximal y distal de la palma. Este proceso de búsqueda aleatoria se realiza de forma iterativa, según los experimentos realizados en [6], para más de 200 iteraciones el cambio dinámico de los agentes tiende a converger a un patrón estable. Por lo tanto, se estableció este parámetro por un número aleatorio de 350 a 700 iteraciones, que conjuntamente con el movimiento aleatorio inicial de los agentes aseguran las variaciones entre clases de las muestras generadas.

El segundo procedimiento del proceso *Nature-VIG* tiene como objetivo mejorar los patrones vasculares producidos por el generador, aplicando filtros de brillo y desenfoque sobre la imagen vascular para reducir las discontinuidades formadas por la huella de los agentes. Así se logra obtener una estructura vascular continua, y también se simulan los efectos de iluminación NIR que se producen en la adquisición de imágenes reales. Las variaciones de brillo se aplican de forma aleatoria aumentando la iluminación global entre  $[1.2..1.8]$  como factor del brillo original. Posteriormente, se computan una secuencia de filtros de desenfoque sobre las imágenes obtenidas, que utiliza un kernel gaussiano (con  $radio = 3 \sim 5$ ).

Por último, para obtener la imagen final se fusionan la imagen resultante del procedimiento anterior con una imagen de la huella de la palma generada aleatoriamente. Esto se realiza con el fin de añadir los detalles y textura de la mano para realzar el realismo de las imágenes generadas. Para ello, se entrenó igualmente un modelo StyleGAN2 para generar aleatoriamente imágenes de huellas palmares a partir de las imágenes de la base de datos CASIA [10] del espectro visible (WHT). Se decidió utilizar estas imágenes debido a que fueron capturadas con el mismo protocolo de adquisición de las imágenes de venas del mismo conjunto de datos. Las muestras originales fueran aumentadas usando el procedimiento propuesto, obteniendo 104,400 muestras a partir de 1,200 imágenes originales de huellas palmares. Igualmente, se utilizó la implementación de NVIDIA Labs, entrenando con

<sup>1</sup><https://github.com/NVLabs/stylegan2>

la configuración original para 25,000 k-imágenes con un FID final igual a 77.28. Luego de esto, se obtiene una imagen sintética de patrones de venas de la palma en un tiempo promedio de 179.32 ( $\pm$  37.29) segundos.

### B. Validación de las bases de datos sintéticas

Utilizando el método propuesto, se crearon dos conjuntos de datos de imágenes sintéticas de venas de la palma. El primer conjunto de datos llamado Synthetic-sPVDB fue generado mediante la variante Style-VIG y está compuesto por 10,000 individuos con seis muestras por cada uno, sumando un total de 60,000 imágenes. Este conjunto de datos es el más grande del estado del arte, siendo 16 veces mayor en el número de individuos y 5 veces mayor en el número total de muestras en comparación con Tongji [14]. El segundo conjunto de datos fue generado usando la variante Nature-VIG y se ha nombrado NS-PVDB, estando conformado por 12,000 imágenes correspondientes a 2,000 sujetos diferentes. Cabe mencionar que como el proceso de creación de las bases de datos es incremental e iterativo, es factible seguir aumentando el número de individuos que componen ambas bases de datos, algo que sin dudas se continuará realizando.

Con el objetivo de medir cuantitativamente la similitud de la distribución entre las imágenes reales y las generadas, se utilizaron tres métricas en el espacio de características, el cual es más robusto que el espacio de píxeles para medir la similitud. Con este fin, se calcularon representaciones de características de las imágenes utilizando las activaciones de un modelo InceptionV3 [24]. En primer lugar, se obtuvo la medida FID para evaluar las estadísticas de dos subconjuntos de imágenes sintéticas y reales. Adicionalmente, se implementó un clasificador binario basado en el clasificador KNN ( $k=1$ ) para predecir dos clases: imágenes reales e imágenes sintéticas. De este modo, se puede estimar la similitud de ambos conjuntos de datos mediante la exactitud de la clasificación (Accuracy) en un esquema Leave-One-Out (LOO). Cuanto más cerca esté la exactitud del 50%, se supone que los dos conjuntos de datos son más similares, ya que el clasificador los clasifica erróneamente. Para complementar el análisis, finalmente se evaluaron ambas distribuciones calculando las métricas *Precision* ( $F_{1/8}$ ) y *Recall* ( $F_8$ ) propuestas en [25]. La clave de este método es que cuantifica qué parte de los datos sintéticos puede ser generada por una parte de la distribución real y viceversa.

La Tabla IV resume los resultados de la validación cuantitativa entre las imágenes reales y las generadas, para ambos conjuntos de datos generados. Para esta comparación solo se tuvieron en cuenta las bases de datos públicas más representativas del estado del arte de acuerdo con su uso en la literatura. Se puede observar que de acuerdo a las métricas evaluadas el conjunto Synthetic-sPVDB tiene mayores similitudes con las bases de datos CASIA, VERA, Tongji e IITI; resultados que son equivalentes a cuando se realiza una inspección visual de las imágenes. En particular, la mayor similitud se aprecia con la CASIA teniendo en cuenta que el modelo generativo fue entrenado con la CASIA-850. Por otra parte, se puede

notar que el conjunto NS-PVDB presenta menor similitud con las bases de datos de referencia, exceptuando la PUT y la VERA en las que se visualizan de mejor manera los patrones correspondientes al plexus venoso de la palma.

TABLA IV: Resultados de la validación cuantitativa de ambos conjuntos de datos generados (a) Synthetic-sPVDB y (b) NS-PVDB, frente a los conjuntos de datos públicos más representativos del estado del arte.

BD pública	FID	Accuracy	$F_8$	$F_{1/8}$
CASIA-940 [10]	67.54	0.604	0.897	0.898
PolyU [13]	103.60	0.665	0.743	0.834
PUT [12]	209.11	0.896	0.368	0.309
VERA [11]	93.01	0.637	0.750	0.844
Tongji [14]	88.91	0.621	0.853	0.862
IITI [15]	97.18	0.664	0.727	0.814

(a)

BD pública	FID	Accuracy	$F_8$	$F_{1/8}$
CASIA-940 [10]	227.41	0.841	0.295	0.318
PolyU [13]	203.60	0.815	0.348	0.322
PUT [12]	71.92	0.719	0.641	0.737
VERA [11]	93.01	0.634	0.825	0.865
Tongji [14]	153.01	0.714	0.403	0.447
IITI [15]	137.01	0.672	0.371	0.386

(b)

### C. Rendimiento de modelos CNN evaluados

Para el entrenamiento de ambos modelos se seleccionó el 80% de las muestras de cada individuo, de los cuales el 10% fue separado para validación, mientras se reservó el restante 20% para la evaluación. Todas las imágenes fueron pre-procesadas para normalizar los valores de intensidad de grises y luego se le aplicó una Ecuación del Histograma de Contraste (CLAHE) para realzar los detalles de los patrones de venas. Ambos modelos fueron entrenados sobre las dos bases de datos sintéticas generadas usando las muestras originales sin el proceso de aumentación de muestras y además aplicando este solo al conjunto de entrenamiento. Debido a limitaciones de memoria, para la base de datos Synthetic-sPVDB solo se entrenó con 2,000 sujetos. Esto permitió evaluar el rendimiento sobre los dos conjuntos de datos en condiciones similares y además se pudo comprobar cómo generalizan los modelos ante un incremento del número de sujetos en la etapa prueba mediante un clasificador KNN. Los tiempos que necesitaron los modelos evaluados para el entrenamiento estuvieron en el orden promedio de 6 minutos y 8 horas para SingleNet, y para Resnet de 1 hora y 27 horas; sin muestras aumentadas y con muestras aumentadas, respectivamente. Si bien estos tiempos son relativamente altos (para solo 2,000 individuos) cuando se usa el proceso de aumentación, este es un proceso fuera de línea que no tiene una incidencia directa en la eficiencia de tiempo de ejecución de ambos modelos.

Se establecieron 100 épocas de entrenamiento con un tamaño de lote  $batch\_size = 32$  y una tasa de aprendizaje  $lr = 0.01$ . Además, se configuró una reducción de la tasa de aprendizaje con un factor igual a 0.2 cuando la función de pérdida se mantiene invariable. Además, se estableció un

criterio de parada temprana en caso que el error de validación no mejore durante siete épocas. Para la optimización SGD, el  $momentum = 0.9$  y el  $weight\ decay = 1 \times 10^{-5}$ .

La Tabla V muestra los resultados finales del entrenamiento y pruebas de cada modelo en ambos conjuntos de datos. Se puede apreciar las altas tasas de precisión y bajos valores de pérdida obtenidos tanto en los conjuntos de entrenamiento como validación. Debido a la configuración de parada temprana, solo para la base de datos Synthetic-sPVDB se alcanzaron el máximo de épocas de entrenamiento. Comparando los dos modelos, se obtuvieron resultados superiores para el modelo SingleNet (Tabla V(a)). En comparación al uso de aumentación de muestras, la tasa de acierto fue superior al usar el procedimiento propuesto. De cierta manera, se puede decir que los resultados obtenidos eran esperados teniendo en cuenta que las bases de datos creadas no simulan muestras de distintas sesiones de adquisición de forma que aumenta la variabilidad inter-clase. Por esto son determinantes los resultados mostrados a continuación, en virtud de confirmar la validez del método propuesto.

Una vez entrenados los modelos, de forma adicional, se utilizaron ambos como codificadores de características para evaluar la eficacia de un clasificador KNN. Esta prueba permitió conocer el grado de generalización de los modelos obtenidos para la representación de descriptores de características. Además, se pudo comprobar la escalabilidad de los modelos sobre el conjunto de datos Synthetic-sPVDB ante un incremento del número de individuos sin necesidad de reentrenarlos. Al mismo tiempo, se validaron las bases de datos sintéticas haciendo pruebas cruzadas de los modelos entrenados sobre estas y evaluando sobre las bases de datos reales más representativas del estado del arte. Para la ejecución de estas pruebas, se usó el 50% del conjunto de prueba como galería biométrica y el otro 50% como muestras de prueba. Los descriptores de características se obtuvieron a partir de la salida de la última capa convolucional, eliminando todas las capas densas de la red. El clasificador KNN se implementó usando la distancia euclidiana y un  $k = 1$ .

Las tablas VI y VII resumen los resultados obtenidos en ambas pruebas usando el clasificador KNN, en las bases de datos sintéticas y las de referencia, respectivamente. Al usar el modelo entrenado sin aumentación de muestras, se aprecia una degradación de los resultados que es más significativa en el modelo ResNet y el conjunto Synthetic-sPVDB. Por otra parte, usando los modelos entrenados sobre las muestras aumentadas aunque los resultados son ligeramente inferiores no se aprecian diferencias significativas. Se debe resaltar que al aumentar la cantidad de individuos utilizados en las pruebas sobre Synthetic-sPVDB no disminuyó la tasa de reconocimiento, lo que evidencia en primera instancia el grado de generalización de los modelos propuestos. Al realizar pruebas cruzadas en las bases de datos de referencia (Tabla VII), los resultados son superiores al 90% para el modelo SingleNet entrenado en ambos conjuntos sintéticos, con excepción de la base de datos CASIA. En el caso del modelo ResNet, los resultados son comparables con los resultados reportados en la literatura

sobre estas bases de datos, igualmente quedando por debajo la base de datos CASIA. Es importante destacar que estos resultados fueron obtenidos entrenando sobre un conjunto de datos completamente sintético sin ninguna información previa acerca de las bases de datos reales en las que se probaron. Estos resultados podrían ser mejorados mucho más realizando un proceso de refinamiento (*fine-tuning*) de los filtros convolucionales de los modelos propuestos.

## V. CONCLUSIONES

El presente trabajo evaluó por primera vez modelos CNN sobre bases de datos a gran escala compuestas por imágenes sintéticas de venas de la palma. Como parte del trabajo, se presentaron y validaron dos enfoques para la generación de imágenes sintéticas de venas del ROI de la palma de la mano. El objetivo del modelo y bases de datos propuestas no es sustituir a los conjuntos de datos de imágenes del mundo real, sino poner a disposición de la comunidad científica conjuntos de datos mucho más grandes que los disponibles en el estado del arte que pueden ser alternativas para estimar rápidamente la eficiencia, la escalabilidad y el rendimiento general de nuevos métodos de reconocimiento biométrico. Además, los conjuntos de datos creados quedaron disponibles en línea en: <http://www.litrp.cl/repository.html>.

Para la evaluación de los modelos CNN, se entrenaron dos arquitecturas de extremo a extremo sobre las bases de datos generadas usando un procedimiento propio de aumentación de datos durante el pre-procesamiento de muestras. Los resultados experimentales validaron la aplicabilidad y calidad de las bases de datos sintéticas propuestas para evaluar la eficacia y escalabilidad de métodos de reconocimiento de venas de la palma basados en CNN. En las pruebas realizadas, se pudo comprobar la necesidad de variantes de implementaciones que hagan un buen uso de la memoria para manejar grandes volúmenes de muestras por los modelos CNN. Además, para la generación de las bases de datos se requiere la introducción de procesamiento paralelo para el proceso de verificación de unicidad. Como trabajo futuro, se planea ampliar este estudio teniendo en cuenta las limitaciones detectadas, así como al empleo de otros modelos de reconocimiento y extensión de las pruebas de bases de datos cruzadas.

## REFERENCIAS

- [1] W. Wu, S. Elliott, S. Lin, S. Sun, and Y. Tang, "Review of palm vein recognition," *IET Biometrics*, vol. 9, no. 1, pp. 1–10, 2019.
- [2] J. Zuo, N. A. Schmid, and X. Chen, "On generation and analysis of synthetic iris images," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 77–90, 2007.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [5] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110–8119.
- [6] Y. Liu, C. Gao, Z. Zhang, Y. Wu, M. Liang, L. Tao, and Y. Lu, "A new multi-agent system to simulate the foraging behaviors of physarum," *Natural computing*, vol. 16, no. 1, pp. 15–29, 2017.

TABLA V: Resultados de métricas de rendimiento de ambos modelos CNN evaluados sobre las bases de datos sintéticas generadas, aplicando el proceso de aumentación de muestras y sin aplicarlo: (a) SingleNet y (b) ResNet. La métrica de Exactitud (Accuracy) se reporta en formato porcentual (%).

Base de datos	Aumentación de muestras	Épocas	Entrenamiento		Validación		Prueba
			Exactitud	Pérdida	Exactitud	Pérdida	Exactitud
Synthetic-sPVDB	No	54	98.93	0.0607	100	0.0210	95.38
NS-PVDB	No	58	99.33	0.0533	99.80	0.0260	99.37
Synthetic-sPVDB	Sí	124	99.90	0.0479	99.94	0.0464	99.92
NS-PVDB	Sí	19	99.48	0.1341	99.85	0.1499	99.80

(a)

Base de datos	Aumentación de muestras	Épocas	Entrenamiento		Validación		Prueba
			Exactitud	Pérdida	Exactitud	Pérdida	Exactitud
Synthetic-sPVDB	No	59	98.92	0.0321	99.95	0.0291	80.30
NS-PVDB	No	65	99.80	0.0321	98.65	0.0586	96.32
Synthetic-sPVDB	Sí	27	99.71	0.0316	99.63	0.0352	99.90
NS-PVDB	Sí	46	99.87	0.0039	99.88	0.0036	99.80

(b)

TABLA VI: Resultados de exactitud (accuracy) del clasificador KNN usando los modelos CNN entrenados como codificadores de características sobre las bases de datos sintéticas generadas, aplicando el proceso de aumentación de muestras y sin aplicarlo: (a) SingleNet y (b) ResNet.

Base de datos	Modelo sin Aumentación	Modelo con Aumentación
Synthetic-sPVDB (2,000)	85.20	99.85
Synthetic-sPVDB (10,000)	71.39	99.84
NS-PVDB	98.65	99.60

(a)

Base de datos	Modelo sin Aumentación	Modelo con Aumentación
Synthetic-sPVDB (2,000)	52.50	99.85
Synthetic-sPVDB (10,000)	44.05	99.86
NS-PVDB	86.65	99.60

(b)

TABLA VII: Resultados de exactitud (accuracy) de pruebas cruzadas entre los modelos entrenados en las bases de datos sintéticas y bases de datos reales usando un clasificador KNN.

Base de datos	Synthetic-sPVDB		NS-PVDB	
	Modelo SingleNet	Modelo ResNet	Modelo SingleNet	Modelo ResNet
CASIA-940 [10]	84.00	96.50	84.50	88.50
PolyU [13]	96.00	99.90	98.10	99.30
PUT [12]	93.00	99.00	93.00	98.00
VERA [11]	94.09	96.82	94.09	95.45
Tongji [14]	92.00	97.67	90.00	94.00
IITI [15]	93.78	98.92	94.86	95.95

- [7] T. Shinzaki, "Use case of palm vein authentication," in *Handbook of Vascular Biometrics*. Springer, 2020, pp. 145–158.
- [8] M. Watanabe, "Palm vein authentication," in *Advances in biometrics*. Springer, 2008, pp. 75–88.
- [9] A. Uhl, C. Busch, S. Marcel, and R. Veldhuis, *Handbook of vascular biometrics*. Springer Nature, 2020.
- [10] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 281–284.
- [11] P. Tome and S. Marcel, "Palm vein database and experimental frame-

- work for reproducible research," in *2015 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2015, pp. 1–7.
- [12] R. Kabaciński and M. Kowalski, "Vein pattern database and benchmark results," *Electronics Letters*, vol. 47, no. 20, pp. 1127–1128, 2011.
- [13] D. Zhang, Z. Guo, G. Lu, L. Zhang, and W. Zuo, "An online system of multispectral palmprint verification," *IEEE transactions on instrumentation and measurement*, vol. 59, no. 2, pp. 480–490, 2009.
- [14] L. Zhang, Z. Cheng, Y. Shen, and D. Wang, "Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset," *Symmetry*, vol. 10, no. 4, p. 78, 2018.
- [15] S. Bhilare, G. Jaswal, V. Kanhangad, and A. Nigam, "Single-sensor hand-vein multimodal biometric recognition using multiscale deep pyramidal approach," *Machine Vision and Applications*, vol. 29, no. 8, pp. 1269–1286, 2018.
- [16] Ö. Toygar, F. O. Babalola, and Y. Bitirim, "FYO: a novel multimodal vein database with palmar, dorsal and wrist biometrics," *IEEE Access*, vol. 8, pp. 82461–82470, 2020.
- [17] R. Hernández-García, S. Guidet, R. J. Barrientos, and F. E. Frati, "Massive finger-vein identification based on local line binary pattern under parallel and distributed systems," in *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*. IEEE, 2019, pp. 1–7.
- [18] F. Hillerström, A. Kumar, and R. Veldhuis, "Generating and analyzing synthetic finger vein images," in *2014 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2014, pp. 1–9.
- [19] S. Crisan, I. G. Târnovan, and T. Crisan, "A hand vein structure simulation platform for algorithm testing and biometric identification," in *16th IMEKO TC4 Symposium, Florence, Italy*, 2008.
- [20] A. Das, P. Mondal, U. Pal, M. Blumenstein, and M. A. Ferrer, "Sclera vessel pattern synthesis based on a non-parametric texture synthesis technique," in *Proceedings of international conference on computer vision and image processing*. Springer, 2017, pp. 241–250.
- [21] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.
- [22] R. Hernández-García, R. J. Barrientos, C. Rojas, and M. Mora, "Individuals Identification Based on Palm Vein Matching under a Parallel Environment," *Applied Sciences*, vol. 9, no. 14, p. 2805, 2019.
- [23] J. Zhang, Z. Lu, M. Li, and H. Wu, "GAN-Based Image Augmentation for Finger-Vein Biometric Recognition," *IEEE Access*, vol. 7, pp. 183 118–183 132, 2019.
- [24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [25] M. S. Sajjadi, O. Bachem, M. Lucic, O. Bousquet, and S. Gelly, "Assessing Generative Models via Precision and Recall," in *Advances in Neural Information Processing Systems*, 2018, pp. 5228–5237.