



PONTIFICIA
UNIVERSIDAD
CATÓLICA DE
VALPARAÍSO



Mauricio Alejandro Rojas Carrasco

Fusión de imágenes visibles y térmicas aplicando técnicas metaheurísticas para sistemas de reconocimiento de rostros

Informe Proyecto de Título de Ingeniero Electrónico



Escuela de Ingeniería Eléctrica



Fusión de imágenes visibles y térmicas aplicando técnicas metaheurísticas para sistemas de reconocimiento de rostros

Mauricio Alejandro Rojas Carrasco

Informe Final para optar al título de Ingeniero Electrónico,
aprobada por la comisión de la
Escuela de Ingeniería Eléctrica de la
Pontificia Universidad Católica de Valparaíso
conformada por

Sr. Gabriel Hermosilla Vigneau
Profesor Guía

Sr. Jorge Mendoza Baeza
Segundo Revisor

Sr. Sebastián Fingerhuth Massman
Secretario Académico

Valparaíso, Mayo de 2017

Para mi tata, Juan Castillo Quiroga.

Agradecimientos

Quiero agradecer a toda mi familia por su apoyo, sus palabras de aliento y su constante preocupación a lo largo de este proceso. De manera especial agradezco a mi madre, quien siempre ha estado para mí y que a pesar de estar lejos físicamente, la siento muy cerca, gracias por tanto.

También quiero agradecer a mis amigos, “los cabros” que fui conociendo durante esta etapa que con su simpatía, energía y motivación, logramos avanzar en nuestras respectivas carreras, gracias por los días de estudio, las risas, los cafés y las cervezas. A mi amigo de la juventud, “el chato” Andrés por estar constantemente preocupado por mí, por las largas conversaciones que a veces eran necesarias, gracias por ser parte de mi segunda familia. A Gino por aportar en cada situación con algún chiste absurdo o fuera de lugar, que hicieron que cada día en la universidad sean más acogedores. A Miguel, una persona que aprendí a conocer, gracias por escucharme y tener una palabra motivadora para cada instante. Y en fin, a todos mis compañeros que conocí a lo largo de la carrera, cada uno con una historia, todos aportando alguna enseñanza, gracias.

De igual manera quiero agradecer a mi profesor guía quien con su exigencia, paciencia y enseñanzas pude cerrar de la mejor manera esta etapa de pregrado.

Y finalmente agradezco a mi banda “Último Andén”, que con las horas de ensayo, reuniones, y las presentaciones que tuvimos, logran hacer un paréntesis en la rutina universitaria. Gracias por la música y las risas, ¡Vamos Arriba!

Este proyecto es financiado gracias al proyecto FONDECYT N°11130466.

Valparaíso, Marzo de 2017

M.R.

Resumen

En este trabajo se presenta un nuevo método para el reconocimiento de rostros que utiliza una fusión de descriptores de imágenes visibles y térmicas, aplicando técnicas metaheurísticas para encontrar un vector de pesos que permita la fusión óptima de los descriptores. Las técnicas presentes en este estudio son: algoritmo PSO y el Algoritmo Genético (GA).

Este nuevo sistema de reconocimiento de rostros consta de dos etapas fundamentales: Etapa de entrenamiento y etapa de validación. En el proceso de entrenamiento el sistema obtiene los pesos multiplicativos óptimos mediante técnicas metaheurísticas las cuales aumentan la significancia de un espectro por sobre otro, esto se lleva a cabo utilizando diferentes combinaciones de conjuntos de la base de datos estándar Equinox, la cual posee imágenes en el espectro visible y en el espectro térmico. En la etapa de validación los pesos óptimos obtenidos se aplican a la base de datos Equinox completa y a la base de datos PUCV-VTF, entregando como resultado sus respectivas tasas de reconocimiento.

Los métodos bajo estudio utilizan descriptores basados en patrones locales de una vecindad de un pixel en la imagen, los cuales son “Local Binary Pattern” (LBP), “Local Derivative Pattern” (LDP) y “Histograms of Oriented Gradients” (HOG), que se apoyan en el uso de histogramas regionalizados, y mediante la medida de similitud de intersección de histogramas (HI) se evalúa su desempeño.

Se comprueba que al aplicar la fusión de ambos espectros utilizando técnicas metaheurísticas como métodos de optimización aumentan el rendimiento en el sistema de reconocimiento de rostros, permitiendo así sistemas más eficientes, fiables y robustos ante variaciones reales. El desempeño de este nuevo método consigue tasas que superan los 98% y 99% de tasa de reconocimiento, según la base de datos utilizadas. Este método supera el desempeño de la aplicación de reconocimiento de rostros en el espectro visible y térmico por separado.

Palabras claves: Reconocimiento de rostros, fusión de imágenes, técnicas metaheurísticas.

Abstract

This work presents a new method for face recognition that uses a fusion of descriptors of visible and thermal images, applying metaheuristic techniques to find a vector of weights that allows the optimal fusion of the descriptors. The techniques present in this study are: PSO algorithm and Genetic Algorithm (GA).

This new face recognition system consists of two fundamental stages: Training stage and validation stage. In the training process the system obtains the optimal multiplicative weights by means of metaheuristic techniques which increase the significance of one spectrum over another, this is carried out using different combinations of sets of the standard database Equinox, which has images in the visible spectrum and thermal spectrum. In the validation stage, the obtained weights are applied to the complete Equinox database and the PUCV-VTF database, resulting in their respective recognition rates.

The methods under study use descriptors based on local patterns of a neighborhood of a pixel in the image, which are "Local Binary Pattern" (LBP), "Local Derivative Pattern" (LDP) and "Histograms of Oriented Gradients" (HOG), based on the use of regionalized histograms and by the measure of similarity of histograms intersection (HI), its performance is obtained.

It is verified that applying the fusion of both spectra using metaheuristic techniques as optimization methods increases the performance in the face recognition system, thus allowing systems more efficient, reliable and robust to real variations. The performance of this new method achieves rates that surpass the 98% and 99% of rate of recognition, according to the database used. This method overcomes the performance of the face recognition application in the visible and thermal spectrum separately.

Keywords: Face recognition, image fusion, and metaheuristics techniques.

Índice general

Introducción	1
Objetivos generales	3
Objetivos específicos	3
1 Antecedentes generales.....	4
1.1 Reconocimiento de rostros.....	4
1.2 Reconocimiento de rostros en el espectro IR	6
1.2.1 Ventajas del espectro infrarrojo	6
1.2.2 Desventajas del espectro infrarrojo	7
1.3 Estado del arte.....	8
1.4 Conclusiones del capítulo.....	10
2 Descripción de metodologías y algoritmos	11
2.1 Técnicas de Fusión de imágenes.....	11
2.1.1 Fusión a nivel de características (Features)	12
2.1.2 Fusión a nivel de puntaje (Scores)	12
2.1.3 Fusión a nivel de decisión	13
2.2 Métodos de reconocimiento de rostros	13
2.2.1 Local Binary Pattern (LBP)	13
2.2.2 Histogram of Oriented Gradients (HOG)	14
2.2.3 Local Derivative Pattern (LDP)	16
2.2.4 Medidas de similitud	18
2.3 Conclusiones del capítulo.....	19
3 Técnicas de optimización	20
3.1 Definición de metaheurística.....	20
3.2 Técnicas metehurísticas	21
3.2.1 Metaheurísticas basadas en trayectoria	22
3.2.2 Metaheurísticas basadas en población	23
3.3 Algoritmo basado en cúmulos de partículas (PSO)	24
3.3.1 Descripción del algoritmo PSO	24
3.3.2 Tipos de PSO	26

3.4 Algoritmo Genético	27
3.5 Conclusiones de capítulo.....	28
4 Sistema propuesto	29
4.1 Fusión de descriptores usando el algoritmo PSO	29
4.1.1 Etapa de entrenamiento.....	30
4.1.2 Etapa de validación.....	34
4.2 Fusión de descriptores usando algoritmo genético.....	34
4.3 Conclusiones del capítulo.....	36
5 Experimentos y resultados	37
5.1 Bases de datos utilizadas	37
5.1.1 Base de datos Equinox.....	37
5.1.2 Base de datos PUCV-VTF	39
5.2 Experimentos para la etapa de entrenamiento	40
5.2.1 Ajuste de parámetros del algoritmo PSO.....	41
5.2.2 Entrenamiento utilizando distintos conjuntos	44
5.2.3 Experimentos para la etapa de validación	47
5.2.4 Validacion con la base de datos Equinox	47
5.2.5 Validacion con PUCV-VTF	55
5.3 Conclusiones del capítulo.....	60
Discusión y conclusiones.....	62
Bibliografía.....	65

Introducción

En la actualidad podemos encontrar variados sistemas biométricos, éstos se basan en el estudio del reconocimiento de personas, utilizando uno o varios rasgos físicos o de conducta. Entre ellos podemos encontrar, sistemas de reconocimiento por la información de la retina, iris, huellas dactilares, forma de la mano, la forma de caminar y el reconocimiento de rostros. Ésta última ha tomado bastante interés en el último tiempo, ya que, a diferencia de las otras es una técnica no invasiva por lo que facilita la obtención de las muestras que conforman el sistema biométrico, además de que es un tema desafiante para los investigadores ya que permite entender el cómo identificamos, reconocemos objetos y rostros de las personas. El reconocimiento de rostros es un área estudiada por la visión por computador, reconocimiento de patrones e inteligencia artificial [1].

El reconocimiento de rostros ha sido investigado por la visión por computador desde sus inicios, y toma como premisa el emular, como las personas pueden ver y reconocer los objetos que se encuentran a su alrededor, principalmente en cómo se identifican y reconocen las personas entre sí a lo largo de su vida. Esto se debe a que cada persona posee ciertas características únicas e irrepetibles que hacen que se diferencien unas a otras. El poder “ver” es un acto biológico propio de los seres humanos que se produce gracias a células ubicadas en la retina del ojo humano que se denominan bastones y conos.

Tomando en cuenta lo anterior, la visión por computador desde sus inicios se ha desempeñado utilizando imágenes que trabajan en el espectro visible del espectro electromagnético, la cual cuenta con longitudes de onda desde los $0,38(\mu\text{m})$ a los $0,75(\mu\text{m})$, que con años de desarrollo ha logrado un gran nivel de madurez y buenos resultados prácticos, aplicando métodos holísticos y de calce local. Sin embargo, el trabajo con imágenes visibles cuenta con factores que dificultan el desempeño de estos sistemas como por ejemplo, las variaciones de iluminación, variaciones en la posición, cambios la expresión facial en el rostro de los individuos y la oclusión en las imágenes del rostro. Por estas razones los investigadores de la visión por computador se ven obligados a buscar distintas soluciones para aumentar el desempeño, dado que es importante generar sistemas que cuenten con altas tasas de reconocimiento, de manera que logren ser robustos a distintas situaciones [2].

Como solución a los problemas encontrados en espectro visible se comienzan a utilizar imágenes que trabajan en el espectro infrarrojo, las cuales son invariantes a la iluminación y distintas

expresiones del rostro, a esto se suman los recientes avances tecnológicos en este campo, que ha producido una importante disminución en el costo de las cámaras infrarrojas, en especial a las que trabajan en el rango de 8 a 12(μm), que es llamado espectro térmico. Al utilizar el espectro térmico las investigaciones resultan más atractivas, y dado a esto se genera un aumento en estudios y aplicaciones, aprovechando los métodos aplicados en el espectro visible [2].

Las cámaras térmicas a diferencia de las visibles, captan la radiación electromagnética emitida por la superficie de los objetos en el rango de 8 a 12(μm), haciéndolos independiente de la iluminación del objeto. A pesar de esto, la energía captada por la cámara no solo depende del cuerpo en observación, sino que también de los objetos presentes en la escena. Debido a esto la estimación de la temperatura de un rostro puede variar, además se presentan otras desventajas en la aplicación del espectro térmico, como los cambios metabólicos de los individuos, como fiebre, esfuerzo físico y problemas propios de las cámaras como la elevación de la temperatura del sensor tras largos periodos de trabajo. Por lo cual se buscan nuevas formas abordar el problema, creando variados sistemas de reconocimiento de rostros que usan la fusión de imágenes y la aplicación de técnicas de optimización, con el fin de aumentar el rendimiento de los sistemas [2] [3].

En el reconocimiento de rostros, la fusión de imágenes representa una alternativa interesante para la solución de los problemas que pueden presentarse en las imágenes térmicas y visibles. La fusión de imágenes visibles y térmicas aplicado al reconocimiento de rostros pretende generar sistemas que sean capaces de que cada espectro supere las deficiencias del otro, es decir que con una buena combinación de las imágenes, los sistemas pasan a ser más estables y robustos. En el tema de tesis se busca crear un sistema de reconocimiento de rostros que fusione imágenes térmicas y visibles, aplicando técnicas metaheurísticas con el fin de maximizar las tasas de reconocimiento del sistema.

Las metaheurísticas son estrategias inteligentes para mejorar los procedimientos heurísticos muy generales con un alto rendimiento [4] [5]. La idea de usar el calificativo heurístico es definir un procedimiento al cual se le tiene un gran grado de confianza en que encuentra soluciones de alta calidad, utilizando una cantidad de recursos razonables, aunque no se garantice que dichas soluciones sean las óptimas o las más factibles del problema, incluso puede existir el caso en el que no se llegue a una solución. La heurística se usa en contraposición a lo exacto. En la actualidad es posible encontrar una gran variedad de técnicas heurísticas, las cuales es posible clasificar de diversas maneras, pero en este trabajo nos centraremos en el estudio de metaheurísticas basadas en el comportamiento social de una población, donde encontramos métodos como el Ant Colony Optimization (ACO), Algoritmos Evolutivos (EA) que se puede subdividir en Computación Evolutiva (CE), Estrategias Evolutivas (ES) y Algoritmos Genéticos (GA) , y los algoritmos basados en cúmulos de partículas (Swarm Particle Optimization) [6]. En este trabajo se estudiará particularmente el algoritmo PSO, pero de acuerdo a las investigaciones planteadas en [7] se compara su desempeño con el algoritmo genético (GA).

PSO fue propuesto por Kennedy y por Eberhard en 1995 [8] este algoritmo está basado en el comportamiento de las aves o cardúmenes de peces en busca de comida. Este algoritmo se ha

empleado en una gran variedad de campos de investigación. En el reconocimiento de rostros ha sido utilizado con el fin de aumentar las características específicas de ciertos descriptores o bien aplicando la transformada de Wavelet o de Fourier [9] [10].

Este trabajo se basa en la investigación realizada en [7] donde se usa el algoritmo genético como herramienta de fusión de descriptores aplicando pesos multiplicativos que buscan maximizar la tasa de reconocimiento del sistema. Estos pesos aumentan la relevancia de un espectro sobre otro en ciertas áreas de la cara, por lo que resulta un método apropiado para el estudio. De esta forma se propone la aplicación del algoritmo PSO para la búsqueda de los pesos multiplicativos óptimos que maximicen las tasas de reconocimiento, utilizando combinaciones de distintos descriptores basados en características locales del rostro como histogramas LBP [11][29], histogramas LDP [12] e histogramas HOG [13]. Además se usan distintos conjuntos de entrenamiento tomados de la base de datos Equinox [14], y finalmente se comparan los resultados con la aplicación del algoritmo genético.

El sistema propuesto consta de dos etapas, la primera es llamada etapa de entrenamiento en la cual se usará la base de datos Equinox para encontrar los pesos óptimos que resalten las áreas del rostro más relevante de ambos espectros. La siguiente etapa es llamada etapa de validación, donde los pesos óptimos encontrados son aplicados a la base de datos Equinox completa y a la base de datos PUCV-VTF, dándonos así el real desempeño del sistema.

Objetivos generales

El objetivo general de este proyecto es implementar un sistema de reconocimiento de rostros en el cual se aplica la fusión de imágenes visibles y térmicas, con el fin de aumentar las tasas de reconocimiento del sistema. La fusión se realiza utilizando distintos descriptores de las imágenes visibles y térmicas del rostro.

El método de fusión aplica el algoritmo de optimización PSO [8] [15], el cual se encarga de buscar valores óptimos, que permiten que se destaquen ciertas áreas del rostro en el cual predomine la información de los descriptores aplicados en cada uno de los espectros, obteniendo una combinación adecuada de ellos y de esta forma maximizar las tasas de reconocimiento del sistema. Finalmente se comparan los resultados con los obtenidos en [7] donde se utiliza el algoritmo genético.

Objetivos específicos

- Estudio del estado del arte de reconocimiento de rostros visibles y térmicos,
- Estudio del estado del arte de la fusión de imágenes visibles y térmicas.
- Desarrollar algoritmos de fusión de imágenes aplicados al reconocimiento de rostros.
- Analizar la aplicación de técnicas metaheurísticas en sistemas de reconocimiento de rostros.
- Estudiar y analizar el comportamiento de los parámetros del algoritmo PSO y como se desempeña en reconocimiento de rostros.

1 Antecedentes generales

El presente capítulo se centra en el estudio de las bases que componen el trabajo de tesis. El reconocimiento de rostros como principal foco de este trabajo, ha desarrollado un gran aumento en cuanto a investigaciones y aplicaciones, destacándose principalmente en área de la visión por computador tanto en el espectro visible como en el espectro infrarrojo (térmico), desarrollando algoritmos capaces de obtener buenos resultados en sus tasas de reconocimiento. A pesar de esto no se logra obtener un sistema lo suficientemente robusto utilizando los espectros por separado por lo cual, se comienzan a desarrollar algoritmos que buscan fusionar la información de ambos y lograr sistemas de reconocimiento de rostros estables y robustos.

1.1 Reconocimiento de rostros

El reconocimiento de rostros es una tarea que los seres humanos realizan de manera rutinaria y sin esfuerzo alguno a lo largo de sus vidas. Es impresionante como el cerebro puede memorizar una gran cantidad de rostros sin importar el ángulo de visión y sin la necesidad de alguna información adicional. Esto ha despertado el interés de investigadores dentro de áreas como: la visión por computador, reconocimiento de patrones y la inteligencia artificial, esta motivación no nace solamente por los desafíos que posee, sino que también las numerosas aplicaciones que este entrega en cualquier área donde se requiera identificar correctamente a las personas [16].

Un sistema de reconocimiento de rostros, forma parte también del área denominada biometría, que consiste en la elaboración de métodos automatizados para la identificación o verificación de personas mediante el uso de sus características físicas o de comportamiento únicos e irrepetibles de cada persona. Las características físicas tienen la cualidad de ser relativamente estables en el tiempo, como por ejemplo: la huella dactilar, la silueta de la mano, patrones de la retina de iris, el sistema cardiovascular y el rostro humano. Los rasgos basados en el comportamiento son menos estables, ya que van a depender netamente del estado psicológico de la persona y el estado de salud, como por ejemplo la firma y la forma de caminar.

El reconocimiento facial se ha popularizado rápidamente como una de las tecnologías biométricas dominantes debido a sus grandes ventajas por sobre el resto: es intuitiva, fácil de utilizar por parte del sujeto, no es invasiva y además permite hacer reconocimiento de individuos en masa y a distancia. Este tipo de sistemas se pueden aplicar sistemas de seguridad y de acceso, con lo que se espera que identifique y verifique la identidad de los sujetos utilizando imágenes o

videos de entrada. También puede ser utilizado de forma exclusiva para identificar personas sospechosas, o bien clasificar a las personas que no pueden tener acceso a lugares específicos, esto se aplica con un criterio de uno contra pocos [16]. Desde el primer sistema de reconocimiento de rostros que fue desarrollado por Kanade en 1974 [17], se han desarrollado una gran variedad de algoritmos y metodologías, permitiendo incluso múltiples reconocimientos simultáneos o reconocimientos en tiempo real. A pesar de estos avances estos siguen un proceso estructurado y definido que se compone de cuatro etapas fundamentales.

- **Adquisición:** Primeramente se debe detectar el rostro en una imagen de entrada o de prueba la cual posee los datos biométricos del sujeto hasta ahora desconocido.
- **Normalización:** Se localizan componentes del rostro como la nariz, ojos o labios y se alinean, siguiendo reglas específicas como distancia entre ojos, posición de la nariz o labios.
- **Extracción de características:** En esta etapa se toma la imagen normalizada, la cual mediante un algoritmo extrae las características que pueden ser de patrones locales, globales o ambas. Estas posteriormente son guardadas en un vector de características, la cual contiene información que distingue las distintas secciones de la cara.
- **Reconocimiento:** Las características extraídas en los pasos anteriores son comparadas contra los vectores almacenados en el sistema que mediante medidas de similitud generan un puntaje de correspondencia, con lo que dependiendo de ese puntaje se acepta o se rechaza al sujeto.

El proceso explicado anteriormente correspondiente a un sistema de reconocimiento de rostros básico, mostrando cada una de sus etapas, es posible apreciarlo de manera esquemática en la Figura 1-1.

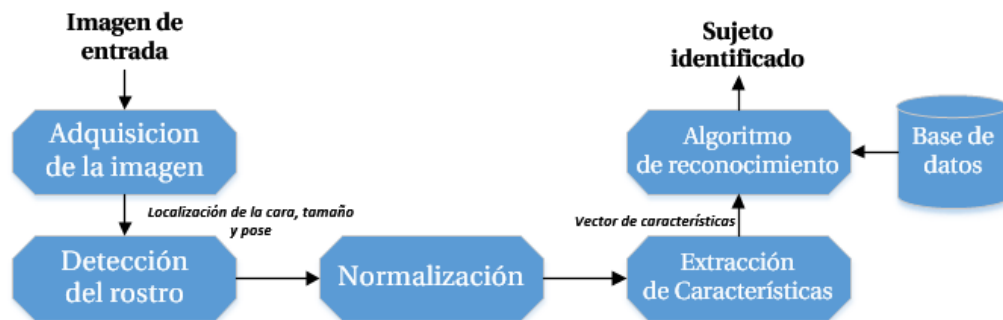


Figura 1-1: Esquema básico de un sistema de reconocimiento de rostros.

Todo lo mencionado anteriormente se ha desempeñado durante años en el espectro de luz visible del espectro electromagnético obteniendo muy buenos resultados, pero tiene ciertas limitaciones que hacen que los sistemas no sean los suficientemente robustos. Entre estas limitaciones se encuentra la dependencia a la iluminación, la oclusión por la utilización de máscaras o disfraces, los cambios de pose de los sujetos y la variación de las expresiones de los sujetos.

Una solución a los problemas que presentan las imágenes visibles es el uso de imágenes térmicas, las cuales son invariantes a la iluminación, cambio de pose y distintas expresiones del rostro, esto será explicado detalladamente en la siguiente sección.

1.2 Reconocimiento de rostros en el espectro IR

Hace algunos años era muy complejo trabajar en el espectro infrarrojo, ya que las cámaras que se usan para capturar esta radiación eran costosas, además eran utilizadas fundamentalmente en aplicaciones militares y médicas. Con el paso de los años y debido a avances tecnológicos las cámaras térmicas han bajado sus costos y aumentado su fiabilidad, mejorando sus limitaciones en resolución y ruido inherente, abriendo nuevos campos de estudio e investigación.

El espectro infrarrojo perteneciente al espectro electromagnético posee mayor longitud de onda que la luz visible, pero menor que las microondas. Su longitud de onda se expande entre los 0.75 μm y los 1000 μm aproximadamente. El infrarrojo se divide en 3 bandas espectrales, estas son: Infrarrojo cercano (0.75 μm – 1.4 μm), el infrarrojo mediano (1.4 μm -8 μm) y el infrarrojo lejano (8 μm – 100 μm) [6]. Éstas son llamadas así teniendo como referencia el espectro visible y sus divergencias, se diferencian entre ellas por su contraste, transmisión atmosférica bajo distintas condiciones climáticas, la señal de fondo y las limitaciones de apertura. La emisión infrarroja que entrega la zona del infrarrojo mediano y lejano es en forma de calor o radiación térmica y es particularmente donde trabajan las cámaras térmicas, que es específicamente entre 8 μm y los 12 μm , es por esto que se le llaman imágenes térmicas. En la Figura 1-2 se pueden observar imágenes del rostro de una persona tomadas en distintos rangos del espectro infrarrojo [1] [2] [18] [19][20].

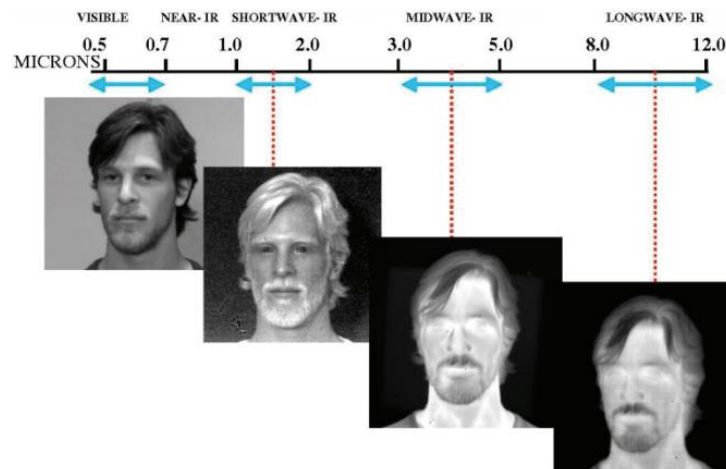


Figura 1-2: Imágenes de rostros capturadas en distintos rangos del espectro infrarrojo [2].

1.2.1 Ventajas del espectro infrarrojo

Como es mencionado en el apartado anterior, las imágenes visibles cuentan con algunas limitaciones a la hora de hablar de sistemas de reconocimiento de rostros. Sin embargo, las imágenes térmicas solucionan varias de éstas:

- Son invariantes a los cambios de luminosidad, esto se puede ver en la Figura.1-3, en la cual se visualizan las imágenes de un hombre en el espectro visible con distintos niveles de luminosidad y su comparación con la imagen térmica, donde esta no se ve afectada por esto.
- Las imágenes térmicas poseen robustez ante los cambios de pose y expresiones del rostro.
- La piel humana emite gran cantidad de radiación infrarroja que se encuentra entre los 8 a los 12 μ m, por lo que es un ser de naturaleza térmica.
- Es posible que obtener la información del sistema vascular. Esto es único, irrepetible, además de ser inmune al envejecimiento.

1.2.2 Desventajas del espectro infrarrojo

En el espectro IR también encontramos ciertas limitaciones o desventajas que hacen que los sistemas de reconocimientos de rostros se vean afectados en su eficiencia y tasas de reconocimiento como las que se describen a continuación:

- Las imágenes de rostros térmicas son afectadas por cambios en el metabolismo del sujeto, la actividad física, estado emocional y también por el consumo de alcohol, ya que afecta la distribución térmica.
- Los anteojos son opacos, por lo que no se puede captar su radiación.
- La temperatura de la piel puede cambiar de acuerdo a la temperatura ambiente.
- Ruidos de patrón fijo: Esto se debe a que los detectores de arreglo de patrón fijo captan porciones de radiación no uniforme por lo que la imagen pareciera que fue tomada con una cámara con el lente sucio. Esto es posible de solucionar con métodos de corrección de la no uniformidad (NUC).
- Las imágenes térmicas de un mismo individuo sufren cambios si son tomadas en distintos periodos de tiempo, ya que el metabolismo cambia constantemente y puede que las condiciones ambientales no sean las mismas.



Figura 1-3: Arriba se observan imágenes en el espectro visible con distinto tipo de iluminación, mientras que abajo se observan las imágenes térmicas con la misma condición [19].

1.3 Estado del arte

En la actualidad es posible encontrar una gran variedad de sistemas de reconocimiento de rostros, aunque todos operan acorde al sistema planteado en la Figura 1-1, la mayoría de estos métodos usan técnicas de extracción de características del rostro, estos datos posteriormente se almacenan en vectores de características que se compara con una galería de imágenes o base de datos del sistema. Estas bases de datos o galerías de imágenes representan el vocabulario que posee el sistema que representa las imágenes o los descriptores de imagen, el cual es consultado para obtener la identificación de los sujetos.

Entre los métodos de extracción de características más destacados se encuentra el Local Binary Pattern (LBP) [11], Weber Law Descriptor (WLD) [21], Principal Component Analysis (PCA) [22], Independent Component Analysis (ICA) [23], Scale-Invariant Feature Transform (SIFT) [24], Speeded-Up Robust Features (SURF) [25], Local Derivative Descriptor (LDP) [26], Histograms of Oriented Gradients (HOG) [13], Gabor Jet Descriptor (GJD) [27], entre otros. La aplicación de cada uno de estos métodos puede encontrarse en las investigaciones [7] [11] [2] [20] [25] [26].

Cada uno de los métodos mencionados anteriormente son aplicados a sistemas de reconocimiento de rostros tanto visibles como a térmicos, pero dado que estos poseen limitaciones se ha intentado crear nuevas formas de aumentar el rendimiento de los sistemas de reconocimiento de rostros, como la aplicación de sistemas multibiométricos que combinan uno o más sistemas biométricos como: rostros con la zona peri ocular, rostros con patrones del iris, etc. Otra forma de solucionar estos problemas son la fusión de imágenes, la cual es el método utilizado a lo largo de este trabajo, los métodos utilizados serán profundizados la siguiente sección.

La fusión de imágenes, presenta una alternativa para estudiar y evaluar, pues se espera disminuir las debilidades del sistema. La motivación principal para fusionar imágenes es mejorar la calidad de información que entregada a la salida de un sistema. Generalmente, las imágenes son capturadas con la misma cámara en distintos periodos de tiempo o por distintas cámaras simultáneamente, que pueden ser de dos espectros distintos. Existen muchos estudios donde se prueban y comparan algoritmos para fusionar imágenes [28], principalmente con el objetivo de mejorar la percepción visual del ojo humano, en [19] [28] se habla de esta teoría y aplicación de algoritmos de fusión estándar. Algunas aplicaciones de la fusión de imágenes se encuentran en el área de diagnóstico médico [26], imágenes satelitales, vigilancia militar, navegación guiada, detección de objetos y reconocimiento.

El proceso de fusionar imágenes es posible dividirlo en cuatro niveles: 1) fusión a nivel de píxeles, 2) fusión a nivel de características, 3) fusión a nivel de puntaje y 4) fusión a nivel de decisión. La fusión a nivel de píxel trabaja directamente con los píxeles de las imágenes, en cambio la fusión a nivel de características trabaja sobre los vectores de características de las imágenes. Para el nivel de puntaje se le asignan puntajes a cada una de las fuentes por separado para luego mediante un criterio establecido por el usuario asignar un valor final de "score", y finalmente para el nivel de decisión es posible aplicar los tres métodos anteriores, generalmente se define un criterio de

decisión común que permite combinar los resultados. En este último se utiliza para optimizar el sistema y elegir el mejor método o una combinación de ellos.

En este trabajo se busca utilizar técnicas metaheurísticas como herramientas para la fusión de imágenes. Principalmente se centra en la aplicación del algoritmo PSO (Particle Swarm Optimization) pero comparándose directamente con lo realizado en [7] donde se aplica el Algoritmo Genético (GA) para realizar la fusión. Estos algoritmos han sido utilizados en una gran cantidad de aplicaciones en distintas áreas, donde realiza soluciones exhaustivas a distintos problemas. También se han utilizado en problemas relacionados con el reconocimiento de patrones, tal como el reconocimiento de gestos en el espectro visible y en los que respecta a este trabajo en la fusión de imágenes térmicas y visibles [10].

El algoritmo genético y el algoritmo PSO han sido utilizados en reconocimiento de rostros principalmente para mejorar la selección de las características para mejorar el reconocimiento como por ejemplo PCA [9][20] [29, 30]. Sin embargo poco se ha estudiado sobre la aplicación de estos algoritmos directamente al reconocimiento de rostros. En la literatura se pueden encontrar métodos de fusión de imágenes visibles y térmicas donde se aplica PSO y GA, donde se utilizan distintos algoritmos de extracción de características en [7] [31] se usa LBP y LDP, mientras que en [29][32] se usa la Transformada Discreta de Wavelet (DWT).

De igual manera existen estudios comparativos de PSO y GA aplicado a sistemas de reconocimiento de rostros ya que ambos son técnicas metaheurísticas que están bioinspiradas y se basan el comportamiento de una población de individuos para encontrar el óptimo, GA se inspira en la evolución de las especies según el trabajo de Darwin, mientras PSO se inspira en el comportamiento de las aves en busca de comida. En reconocimiento de rostros se comparan resultados de estos algoritmos con el fin de ver quién es que tiene un mejor desempeño en estos estudios. En [29] [33] se propone un estudio comparativo de estos algoritmos para un sistema de reconocimiento que usa fusión de imágenes a nivel de puntajes, además se plantea un híbrido entre ambos algoritmos.

Con los avances de la tecnología y de la inteligencia computacional se han creado algoritmos inteligentes de reconocimiento de rostros basados en redes neuronales, estos algoritmos tienen una gran eficiencia pero significan un coste computacional mayor, y entrenamientos que pueden tardar hasta semanas. Entre estos trabajos se puede mencionar la aplicación de redes neuronales y la transformada de wavelet [34], como también donde se crea una interfaz de reconocimiento para la siguiente generación de dispositivos móviles aplicando redes neuronales [35].

El método de fusión propuesto en este trabajo de tesis se basa en lo expuesto en [7], donde se aplica la fusión de descriptores visibles y térmicos usando GA. El objetivo de esto es que las deficiencias de un espectro lo compense el otro, complementándose de manera que las tasas de reconocimiento se maximicen. En este trabajo se utilizará la misma metodología pero aplicando el algoritmo PSO, además se realizan distintos un descriptor que fue utilizado en dicha investigación que es HOG. De todos los descriptores involucrados, se plantean distintas combinaciones de descriptores para realizar la fusión. Sumado a esto se utilizan distintos conjuntos de entrenamiento y aplicación de pesos complementarios durante todo el proceso de

fusión. Finalmente se comparan los resultados obtenidos aplicando PSO con los resultados GA en el estudio plantado en [7].

1.4 Conclusiones del capítulo

En el capítulo se presenta el funcionamiento de los sistemas de reconocimiento de rostros y en virtud de sus avances se crean nuevas alternativas para mejorar su rendimiento. La aplicación de las imágenes térmicas fue un gran paso para el desarrollo de las investigaciones ya que logran compensar las dificultades del espectro visible. Posteriormente se encuentran con desventajas que afectan el correcto rendimiento de los sistemas en distintas condiciones. Ante estas limitaciones se crean distintos algoritmos de fusión que buscan utilizar la información de ambos espectros para mejorar los sistemas.

La fusión de imágenes se aplica en una gran variedad de áreas, con esto se espera tener mayor información y mejores resultados. Se debe tener presente que la metodología de fusión a utilizar dependerá del problema que se quiere resolver, por ejemplo para la mejora de visualización de una imagen se utilizan transformada de wavelet o calcular el promedio pixel a pixel, mientras que en el caso de la fusión generada en este proyecto donde la fusión se aplica al reconocimiento de rostros no es necesario obtener la visualización de la imagen fusionada.

Los técnicas metaheurísticas buscan la mejor solución entre una serie de posibles resultados a un problema específico, es por esto que es una herramienta útil en el reconocimiento de rostros y en la fusión de imágenes. Ya que estos son utilizados para seleccionar la información más adecuada con tal de aumentar el desempeño de un sistema de reconocimiento o bien seleccionar la información más relevante a la hora de fusionar las imágenes.

2 Descripción de metodologías y algoritmos

A lo largo de este trabajo se utilizan distintas metodologías y algoritmos. Primeramente se debe tener claro los niveles de fusión que se pueden utilizar y aplicar a sistemas de reconocimiento. Siguiendo con los métodos de reconocimiento aplicados al sistema propuesto y las medidas similitud las que al fin y al cabo desempeñan la labor de reconocimiento.

2.1 Técnicas de Fusión de imágenes

En el trabajo de tesis se busca fusionar imágenes térmicas y visibles. En la actualidad existen numerosas aplicaciones donde se requieren necesariamente imágenes procedentes de diferentes sensores, pero de una misma escena, tales como imágenes satelitales o imágenes médicas. Esto es posible porque cada imagen de distinto sensor proporciona un tipo de información distinta [10]. En consecuencia, el objetivo de realizar la fusión es combinar la información dispersa con el fin de obtener la máxima información posible en una sola imagen.

Esto puede ser abordado desde el punto de vista de un sistema biométrico, ya que tanto el reconocimiento de rostros en el espectro visible y el reconocimiento de rostros en espectro térmico, son sistemas biométricos. La fusión de ambos espectros será considerado entonces, como un sistema multibiométrico. Se pueden catalogar sistemas multibiométricos como: sistemas multi-sensor, multi-algorítmicos, multi-instancias y multimodales. Siendo este último el aplicado en el presente estudio [30] [30][36].

La fusión multimodal es un sistema que cuenta con distintas fuentes biométricas esto proporciona una mayor cantidad de información, por lo que se espera que sea más robusto y confiable. El término multimodal se refiere a la fusión de distintos tipos de información, por lo tanto es un enfoque más amplio que solo fusionar características directamente [30].

Se puede realizar fusión multimodal a partir de dos características unimodales las cuales pueden realizarse a distinto nivel: a nivel de sensor, a nivel de características (features), a nivel de puntaje (scores) o de decisión. Como la fusión a nivel de sensor no suele involucrar distintas modalidades biométricas, no será motivo de estudio, por lo que se centra el análisis en los otros niveles. Cabe

destacar que los distintos niveles de fusión no son excluyentes, por lo que es posible diseñar un sistema donde se apliquen varios de ellos.

2.1.1 Fusión a nivel de características (Features)

En la fusión a nivel de *features* se obtienen parámetros de los diferentes sistemas de extracción de características, formando un vector único con la información de ambos. Este tipo de fusión reúne lo más relevante de cada sistema, esto garantiza que no haya pérdidas de información, sin embargo esto puede contener algunas complicaciones si es que la dimensión de la fusión llega a ser tan grande que sea difícil de tratar [30]. Para estos casos es posible normalizar el valor resultante mediante técnicas como el PCA.

Dentro de la fusión a nivel de características se pueden aplicar diversas fusiones de imágenes convencionales como la fusión de imágenes usando distintos métodos como los son: fusión por la transformada discreta de wavelet y la transformada de Fourier [32].

2.1.2 Fusión a nivel de puntaje (Scores)

La fusión de puntajes es, en su expresión más simple, asociar niveles de certeza a cada sub-sistema, para luego fusionar estos resultados y asignar un puntaje final la salida del sistema.

Un ejemplo sencillo sería determinar si una persona es quien dice ser basándose en dos rasgos, imagen de rostro y los patrones de iris. El sistema de acceso indica positivo con un 90% de certeza, mientras que el iris indica positivo el 98% de certeza. Usando la regla del promedio simple, se puede estimar la certeza del sistema completo. En este caso la certeza del sistema sería del 94%, donde dependerá del desarrollador del sistema de reconocimiento si es un puntaje aceptable o no. La implementación de reglas de fusión de este estilo busca crear sistemas robustos ante cambios externos, donde, por ejemplo, una de las características podría no presentarse o podría presentar errores [30]. La fusión de scores se compone de dos elementos principales: normalización y la fusión propiamente tal [19] [29] [37].

Esquemas de normalización

El proceso de normalización transforma los diferentes resultados de los dos sistemas de tal manera que se encuentren en un rango equivalente o comparable. De no realizarse esta transformación, uno de los sistemas que tenga un amplio rango de valores podría eliminar la contribución del otro sistema con un rango menor. Por la naturaleza diversa de los puntajes, la normalización es crucial para el desempeño correcto del sistema [30] [36].

- **Normalización minMax:**

Corresponde al estiramiento lineal de los datos, normalizándolos ente 0 y 1. La ecuación 2-1 muestra la normalización donde \hat{S}_i corresponde al i -ésimo puntaje normalizado, S_i al i -ésimo puntaje sin normalizar, y donde $\min(S)$ y $\max(S)$ corresponden a los puntajes mínimos y máximos [37].

$$\hat{S}_i = \frac{S_i - \min(S)}{\max(S) - \min(S)} \quad (2-1)$$

- **Normalización z-score:**

Esta normalización modifica la media y la varianza globales de las puntuaciones del sistema para que valgan 0 y 1 respectivamente. Para esto se tiene que \hat{S}_i corresponde al i -ésimo puntaje normalizado, S_i al i -ésimo puntaje sin normalizar, \bar{S}_i corresponde a la media del conjunto de puntajes y $\sigma(S)$ a su desviación estándar [30] [37]. Esto se representa en la ecuación (2-2) .

$$\hat{S}_i = \frac{S_i - \bar{S}_i}{\sigma(S)} \quad (2-2)$$

2.1.3 Fusión a nivel de decisión

En la fusión a nivel de decisión las características o el puntaje de cada sistema son irrelevantes, por lo que los sistemas obtienen los parámetros necesarios para tomar su propia decisión. Luego el sistema debe tomar su decisión final basándose en las decisiones de ambos sub-sistemas.

Las técnicas de decisión final pueden ser de carácter lógico como OR y AND, como también se puede hacer mediante votación, en caso de haber varios sub-sistemas.

2.2 Métodos de reconocimiento de rostros

En esta sección se muestra el estudio y análisis de algoritmos usados para el reconocimiento de rostro. Estos métodos son algoritmos de extracción de características, los cuales están basados en patrones locales. En este estudio se utilizan tres, llamados: histograma de característica LBP (Local Binary Pattern), LDP (Local Derivative Pattern) y HOG (Histograms of Oriented Gradients). Siguiendo lo planteado en la Figura 1-1, una vez obtenido el vector de características se procede a obtener el valor de similitud o correspondencia el cual entrega un puntaje o score, que permite finalmente realizar el reconocimiento.

2.2.1 Local Binary Pattern (LBP)

El reconocimiento de rostros usando histogramas de características LBP fue propuesto originalmente en [11], y ha sido usado en variadas investigaciones desde ese entonces. LBP es uno de los mejores descriptores de textura [11], donde la imagen es dividida en regiones, en cada región el operador LBP asigna una etiqueta a cada pixel de la imagen donde mediante una ventana de 3x3 se compara el pixel central con sus vecinos, generando un código binario. Para cada región se genera un histograma que luego son concatenados, este histograma resultante o concatenado contiene información del rostro de características como, líneas, puntos o diferentes áreas del rostro. La imagen es dividida en regiones para mantener la información espacial del rostro, con esto cada región se toma como una imagen diferente y se calcula su histograma. Finalmente los histogramas de cada región son concatenados en una única imagen que representa la imagen del rostro. Su representación matemática se aprecia en la ecuación(2-3).

$$g(I(Z_o), I(Z_i)) = \begin{cases} 0, & \text{si } I(Z_o) > I(Z_i) \\ 1, & \text{si } I(Z_o) \leq I(Z_i) \end{cases}, \forall Z_i \in \Omega \subseteq I, i=0, \dots, 8 \quad (2-3)$$

Donde $g(I(Z_o), I(Z_i))$ es la función binaria que define el valor LBP correspondiente al central pixel Z_o de la imagen I comparado con sus pixeles vecinos Z_i . Esta representación se muestra en la en la Figura 2-1.

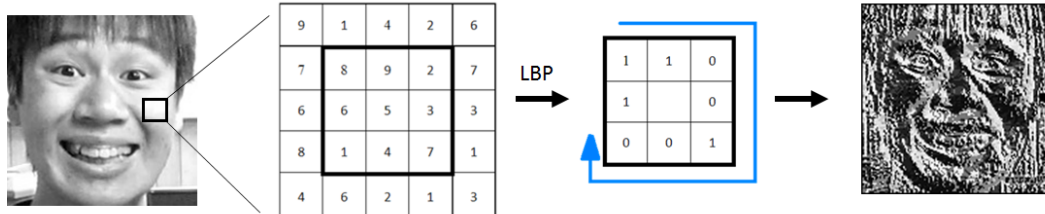


Figura 2-1: Ejemplo de codificación LBP [30].

A partir del código binario creado, se obtiene el histograma resultante o concatenado que contiene la información del rostro de características como: líneas, puntos o diferentes áreas del rostro. El uso de la división por regiones es para mantener la información espacial del rostro, con esto cada región se toma como una imagen diferente y a cada una se le calcula su histograma. Finalmente cada histograma de las distintas regiones es concatenado obteniendo una representación completa del rostro [11] [36]. En este trabajo las imágenes son divididas en $V \times U$ regiones, donde V representa la cantidad de divisiones verticales de la imagen y U la cantidad de divisiones horizontales de la imagen. Cada histograma es un vector que se calcula a partir de las intensidades de los pixeles que van de 0 a 255 (8 bits). En la Figura 2-2 se muestra el funcionamiento del histograma por regiones de las imágenes del rostro.

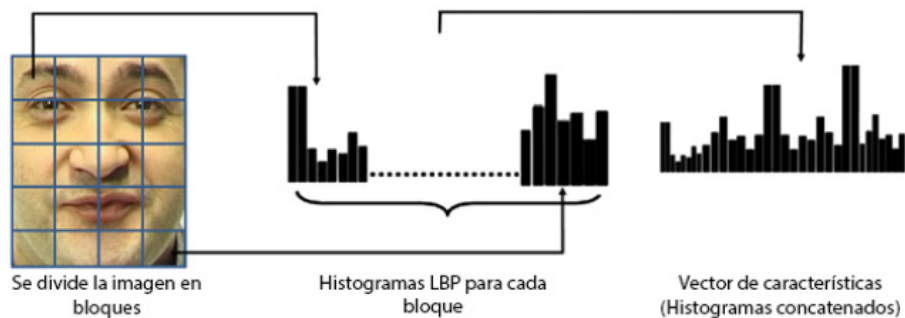


Figura 2-2: Histogramas generados por LBP [36].

2.2.2 Histogram of Oriented Gradients (HOG)

Los histogramas de gradientes orientados están inspirados en Scale-Invariant Features Transform (SIFT), y se proponen inicialmente en [13], es un descriptor invariante a las rotaciones en dos dimensiones. El descriptor HOG trabaja con los gradientes y las magnitudes de los pixeles de la imagen. Para obtener la gradientes y magnitud, se aplican dos máscaras observadas en la Figura 2-3, con esto se obtiene la información del competente horizontal y vertical.

0	0	0
-1	0	1
0	0	0

Vx

0	1	0
0	0	0
0	-1	0

Vy

Figura 2-3: Filtros para obtención de la componente Vx y componente Vy de los gradientes de la imagen

Con la información obtenida de las componentes Vx y Vy de los filtros, se puede obtener el ángulo y la gradiente de un pixel por medio de las ecuaciones (2-4) y (2-5) respectivamente.

$$\theta = \text{ArcTan}\left(\frac{V_y}{V_x}\right) \tag{2-4}$$

$$M = \sqrt{V_x^2 + V_y^2} \tag{2-5}$$

Los ángulos son cuantificados en 9 valores, es decir 9 direcciones principales, y se almacena el valor de magnitud de cada gradiente que coincide con cada uno de tales direcciones. De esta manera, la imagen original se divide en $V_x U$ particiones y se crean histogramas por cada región, tomando la mitad del área de la región precedente y la mitad de la región posterior. Finalmente, el histograma de cada región R_i es normalizado con la norma L2, según la ecuación (2-6) , y el descriptor final de características de la imagen resulta ser la concatenación de los histogramas de cada región.

$$HOG(R_i) = \frac{H(R_i)}{\sqrt{\sum_{i=1}^n |x_i|^2}}, i = 1, \dots, 9 \tag{2-6}$$

En la Figura 2-4 se puede apreciar la representación de la característica HOG de una imagen y en la Figura 2-5 se observa el histograma obtenida de dicha representación.

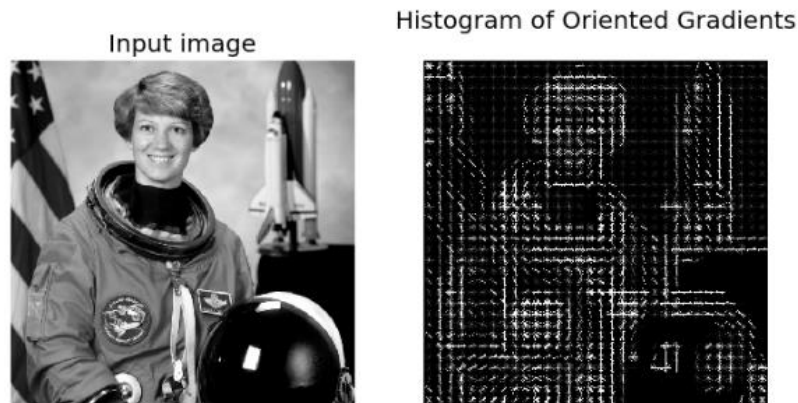


Figura 2-4: Ejemplo de la representación HOG de una imagen.

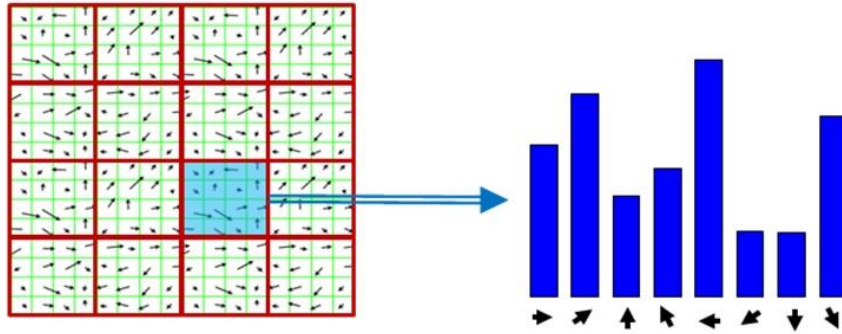


Figura 2-5: Histograma obtenido del vector de características HOG.

2.2.3 Local Derivative Pattern (LDP)

Los patrones derivativos locales (LDP) fueron propuestos por primera vez en [12], donde se postula como un método de patrones binarios de orden superior al utilizar las derivadas direccionales de la imagen original.

El método LDP de orden superior genera micro patrones haciendo uso de la información en base a las diferencias presentes en las derivadas direccionales locales de la región del rostro. El algoritmo LDP, por lo tanto, obtiene información más detallada que LBP debido al uso de derivadas direccionales de alto orden.

LDP funciona con cuatro derivadas direccionales de la imagen original en 0° , 45° , 90° y 135° . Teniendo en cuenta estas direcciones, implícitamente se tienen las derivadas direccionales en 180° (el opuesto a 0°), 225° (el opuesto a 45°), 270° (el opuesto a 90°) y 315° (el opuesto a 135°). El descriptor se denota por LDP_α^n , donde n es el orden del descriptor utilizando la derivada n-1 de I (imagen original) en la dirección de α (0° , 45° , 90° , 135°). Para aplicar el descriptor LDP de orden n, primero se debe derivar la imagen en la dirección α , usando las de ecuaciones (2-7) tomando como referencia la vecindad de un pixel Z_0 del mismo modo que LBP en la Figura 2-1.

$$\begin{aligned}
 I_{0^\circ}^{n-1}(Z_0) &= I_{0^\circ}^{n-2}(Z_0) - I_{0^\circ}^{n-2}(Z_4) \\
 I_{45^\circ}^{n-1}(Z_0) &= I_{45^\circ}^{n-2}(Z_0) - I_{45^\circ}^{n-2}(Z_3) \\
 I_{90^\circ}^{n-1}(Z_0) &= I_{90^\circ}^{n-2}(Z_0) - I_{90^\circ}^{n-2}(Z_2) \\
 I_{135^\circ}^{n-1}(Z_0) &= I_{135^\circ}^{n-2}(Z_0) - I_{135^\circ}^{n-2}(Z_1)
 \end{aligned}
 \tag{2-7}$$

Donde n es el orden de la derivada LDP que se aplica, Z_i son los pixeles vecinos a Z_0 que es el pixel central. Según las ecuaciones (2-8) se desprende que el mínimo orden debe ser 2, donde se debe obtener la primera derivada direccional o derivada cero. Para obtener el LDP de orden 2 se debe obtener la derivada direccional de orden 2 a partir de la derivada de orden 1 y así sucesivamente. Para obtener el descriptor LDP se debe aplicar la función binaria de la ecuación (2-9).

$$f(I_\alpha^{n-1}(Z_0), I_\alpha^{n-1}(Z_i)) = \begin{cases} 0 & \text{si } I_\alpha^{n-1}(Z_0) \cdot I_\alpha^{n-1}(Z_i) > 0 \\ 1 & \text{si } I_\alpha^{n-1}(Z_0) \cdot I_\alpha^{n-1}(Z_i) \leq 0 \end{cases} \text{ con } i = 1, \dots, 8
 \tag{2-8}$$

De esta forma se obtienen 8 valores para cada dirección, los que son ordenados según la ecuación, para obtener así un valor binario de 8 bits, para cada valor de α .

$$LDP_{\alpha}^n = \left\{ f\left(I_{\alpha}^{n-1}(Z_0), I_{\alpha}^{n-1}(Z_1)\right), f\left(I_{\alpha}^{n-1}(Z_0), I_{\alpha}^{n-1}(Z_2)\right), \dots, f\left(I_{\alpha}^{n-1}(Z_0), I_{\alpha}^{n-1}(Z_8)\right) \right\} \quad (2-9)$$

Una alternativa a las ecuaciones (2-2) es la aplicación de los filtros de la Figura 2-6.

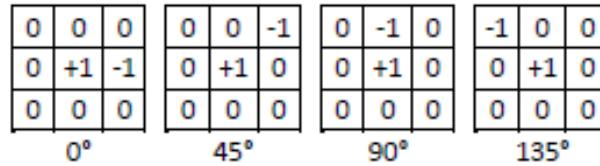


Figura 2-6: Filtros para la obtención de las derivadas direccionales. Para obtener la derivada de orden n, se debe aplicar el filtro m veces.

En la Figura 2-7 se puede apreciar un ejemplo de la aplicación de LDP de orden 2, 3 y 4 en una imagen del rostro en el espectro visible. En la práctica no es factible representar gráficamente la imagen derivada, puesto que posee valores comprendidos entre [-255,255]. Debido a la naturaleza de LDP no es posible aplicarlo a los bordes, por lo que estos no se derivan. Se aprecia además que cuando mayor es el orden, de la derivada, mayores son patrones que va obteniendo la imagen, perdiéndose casi en su totalidad la imagen original. Dado a los buenos resultados obtenidos en reconocimiento de rostros LDP se usará en este trabajo solo con derivada de tercer orden.

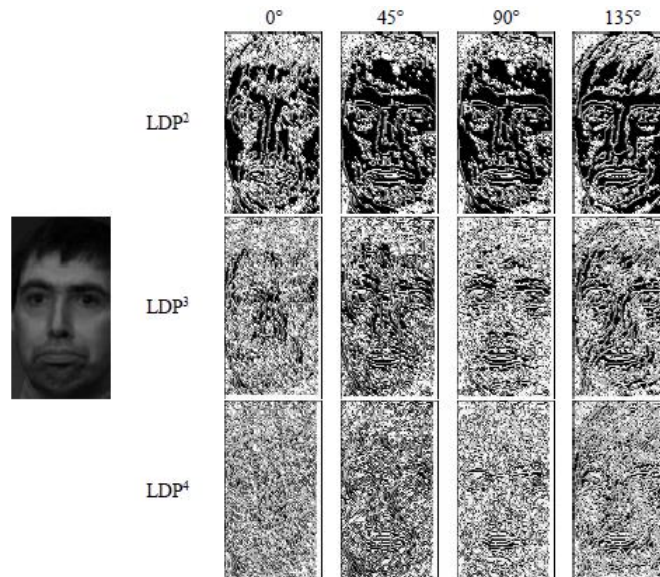


Figura 2-7: A la izquierda se muestra la imagen original den el espectro visible. A la derecha se observa la representación del descriptor de imagen LDP con dirección 0° , 45° , 90° y 135° . Hacia abajo se aprecia el orden LDP de 2, 3 y 4 [26].

Para finalizar el proceso las imágenes obtenidas con sus respectivas derivadas direccionales son divididas en $V \times U$ regiones y al igual que LBP se calcula su histograma por cada región, posteriormente los histogramas son concatenados obteniendo solo un vector de características.

2.2.4 Medidas de similitud

Con los vectores de características representados en forma de histogramas, y siguiendo el esquema de la Figura 1-1, se procede a realizar el reconocimiento, para esto es necesario definir dos conceptos: “galería” y “test”. La galería es un conjunto de imágenes del rostro de un grupo de personas guardadas en el sistema de reconocimiento. El conjunto de test, se compone de varias imágenes de rostros de distintas personas, éstas se encuentran sometidas a variaciones, de iluminación, pose y expresiones. Estos conjuntos son extraídos de alguna base de datos ya definida. La medida de similitud, compara los histogramas de las imágenes del conjunto de “test” con las imágenes del conjunto de “galería” obteniendo así un puntaje que se define como medida de similitud.

Interseccion de histogramas

Es una medida de similitud entre dos histogramas, que corresponden a dos imágenes de rostros que se desean comparar. Si su valor se acerca a 1, indica que los histogramas son parecidos y se podría tratar de la misma persona. En el caso que este valor sea cercano a 0 o simplemente intersección nula, los histogramas no se parecen y por tanto, la imagen corresponde a otro sujeto. En definitiva, esta medida entrega el número de píxeles con mínimo valor, que coinciden en los histogramas de ambas imágenes. Esta medida se muestra en la ecuación (4-5), donde S_i es el vector característico de la imagen de prueba y M_i es el vector característico de la imagen de galería a comparar, i corresponde al índice de la imagen analizada.

$$HI(S, M) = \sum \min(S_i, M_i) \quad (2-10)$$

Distancia Euclidiana

En un espacio euclídeo, es la distancia entre 2 puntos, equivalente al teorema de Pitágoras. En un espacio no euclídeo es la distancia más corta entre dos puntos en un segmento de curva. En esta medida, también se comparan dos vectores característicos uno correspondiente a la imagen de prueba S_i y el otro a la imagen de galería M_i . En este caso cuando los rostros son similares, entrega un valor cercano a cero.

$$EU(S, M) = \sum (S_i - M_i)^2 \quad (2-11)$$

Chi-cuadrado

Es una medida ponderada de las diferencias cuadráticas entre los valores que la componen, estos valores corresponden a la comparación de dos vectores característicos, uno correspondiente a la imagen de prueba S_i y el otro a la imagen de galería M_i . En este caso cuando los rostros son similares, entrega un valor cercano a cero.

$$X^2(S, M) = \frac{\sum (S_i - M_i)^2}{S_i + M_i + \varepsilon} \quad (2-12)$$

Dado a los resultados obtenidos en [7] donde se usa una metodología similar a la que se aplica en este estudio, se selecciona la medida de similitud de intersección de histogramas para el sistema de reconocimiento de rostros desarrollado, ya que este tiene mejores resultados que chi-cuadrado y que la distancia Euclidiana.

2.3 Conclusiones del capítulo

En el capítulo se presentaron los algoritmos y metodologías que permitieron desarrollar el trabajo de tesis. Primeramente se plantean las distintas técnicas de fusión de sistemas biométricos, que al fusionar dos o más de ellos forman un sistema multibiometrico. Estos pueden ser clasificados según su naturaleza de obtención de información, entre esto se destaca la fusión multimodal, ya que el sistema propuesto consta de dos modos a fusionar: el rostro visible y el rostro térmico. Existen tres niveles de fusión multimodal: nivel de características, de puntajes y de decisión. Como se busca fusionar descriptores visibles y térmicos se aplica una fusión a nivel de característica.

Los métodos de reconocimiento de rostros son los que extraerán las características de las imágenes del rostro. Los métodos utilizados de este trabajo trabajan a nivel local, es decir en base a la comparación de sus vecinos, y cada uno de ellos utiliza diferentes metodologías para la obtención del vector de característica, LBP trabaja en base a códigos binarios, mientras que HOG trabaja en base a los gradientes de los píxeles y finalmente LDP, trabaja aplicando derivadas direccionales. A pesar de esto todos logran una representación en forma de histogramas.

Con la representación en histogramas de cada uno de los descriptores, se define posteriormente los conjuntos de “test” y de “galería”, los cuales mediante las medidas de similitud realizan el reconocimiento de rostros.

3 Técnicas de optimización

El presente capítulo de este trabajo se centra en el estudio de algunas de las principales técnicas metaheurísticas de optimización. Primeramente se define que son las metaheurísticas y su desarrollo, posteriormente se clasificarán de acuerdo a su popularidad en la literatura actual. Dicha clasificación la divide en técnicas metaheurísticas basadas en la trayectoria y las técnicas basadas en el comportamiento de la población, las cuales cuentan con distintos algoritmos los cuales serán mencionados.

3.1 Definición de metaheurística

En la inteligencia artificial se emplea el término heurístico en un sentido genérico, donde se relaciona con la tarea de resolver inteligentemente problemas reales usando el conocimiento disponible acerca de un problema, y de las técnicas aplicables, tratando de aportar soluciones (o acercarse a ellas) usando recursos (generalmente tiempo) razonables. Este término se aplica a procedimientos de resolución de problemas de optimización con una concepción diferente, en donde se califica de heurístico al procedimiento que tiene un alto grado de confianza para encontrar soluciones de alta calidad con un coste computacional razonable, aunque no garantice que sea óptimo o factible e incluso, que en algunos casos no se llegue a establecer la cercanía al óptimo [4] [38].

Dado a que los problemas de optimización han tenido una gran relevancia a lo largo de la historia de la informática se han desarrollado una variedad de métodos para resolverlos. Estos métodos se pueden clasificar en técnicas exactas y aproximadas. Las técnicas exactas garantizan que encontrarán la solución óptima para cualquier instancia de cualquier problema en un tiempo acotado. La dificultad de estos métodos es que tardan demasiado tiempo, aunque se dice que es acotado, dicho tiempo crece exponencialmente dependiendo de la complejidad del problema. Esto provoca que en la mayoría de los casos la resolución del problema sea inabordable, ya que podría tardar cientos de años. Por otro lado las técnicas aproximadas sacrifican la garantía de encontrar la solución “óptima” a cambio de encontrar una solución lo suficientemente buena en un tiempo considerable [4].

Dentro de las técnicas aproximadas, se pueden encontrar métodos heurísticos de naturaleza diferente, por lo que resulta complicado clasificarlas. Además, muchos de estos métodos han sido diseñados para la resolución de un problema en específico, con posibilidad de generalización o

aplicación a otros problemas que guarden cierto grado de similitud. Entre ellos podemos encontrar: los heurísticos constructivos, los métodos de búsqueda local y las técnicas metaheurísticas. Esta última será detallada a continuación.



Figura 3-1: Clasificación de las técnicas de optimización

3.2 Técnicas metaheurísticas

Las técnicas metaheurísticas son algoritmos de optimización no exactos, surgen por la idea de combinar distintos métodos heurísticos para formar nuevos algoritmos capaces de conseguir un aumento en el nivel de efectividad en la exploración del espacio solución de un problema [4] [5] [39]. Estas técnicas implementadas por primera vez en [40] fueron denominadas técnicas metaheurísticas. De las diferentes definiciones de metaheurística encontradas en la literatura es posible decir que son estrategias de alto nivel que combina diferentes métodos para explorar el espacio de búsqueda. En otras palabras, una metaheurística es una plantilla general no determinista que debe ser rellenada con datos específicos del problema y que permite abordar problemas con espacios de búsqueda de gran tamaño por lo que es importante encontrar equilibrio apropiado entre diversificación e intensificación [39] [40].

Con diversificación se refiere a la exploración del espacio de búsqueda mientras que con intensificación se refiere a la exploración de alguna área específica de ese espacio. Como se menciona anteriormente, su equilibrio es de gran relevancia ya que por un lado deben identificarse las regiones prometedoras a ser buenas soluciones dentro del espacio, y por otro lado no malgastar tiempo en áreas que no proponen soluciones “buenas”.

Es posible clasificar y describir las técnicas metaheurísticas de distinta forma en este caso particular serán clasificadas en técnicas basadas en la trayectoria y técnicas basadas en la población.

3.2.1 Metaheurísticas basadas en trayectoria

Las técnicas metaheurísticas basadas en la trayectoria tienen como principal característica que de una posición inicial comienzan a explorar la vecindad del espacio de búsqueda actualizando la solución actual, formando así una trayectoria. Estos algoritmos surgen como una extensión de los algoritmos de búsquedas locales simples a lo que se le agrega la característica de no caer en mínimos locales. Ocasionalmente la búsqueda es terminada cuando se alcanza un valor máximo de iteraciones que es predefinido, cuando se encuentra una solución con una calidad aceptable o se detecta un estancamiento en el proceso [4]. Dentro de estos tipos métodos podemos encontrar:

- **Enfriamiento simulado (SA):**

Es la metaheurística posiblemente más antigua, la cual se creó con la estrategia explícita de escapar de los óptimos locales. Esta se inspira en el proceso de recocido del metal y cristal, por lo que para evitar caer en óptimos locales, en cada iteración se selecciona la mejor solución actual y la mejor solución del vecindario, entonces si la solución del vecindario es peor que la actual es aceptada con cierta probabilidad que depende de la temperatura y la variación de la función fitness [4] [6].

- **Búsqueda Tabú (TS):**

Es una técnica bastante exitosa a la hora de resolver problemas de optimización combinatoria. La idea básica de la búsqueda tabú es el uso del historial de búsqueda, tanto para evitar caer en óptimos locales como para idear una estrategia de exploración y evitar buscar en la misma región varias veces. El historial de búsqueda es implementada con una lista “tabú” donde se guardan las soluciones visitadas recientemente y así excluirlas de los futuros movimientos [4] [40] [41].

- **Procedimiento de búsqueda miope aleatorizado y adaptativo (GRASP)**

Este método es una metaheurística simple que combina heurísticos constructivos con técnicas heurísticas de búsqueda local. Este es un procedimiento iterativo que cuenta con dos etapas: construcción de solución y mejora. La solución mejorada es la solución al proceso de búsqueda [4].

- **Búsqueda del vecindario variable (VNS)**

El VNS (por sus siglas en inglés de Variable Neighborhood Search) es una metaheurística que aplica una estrategia para cambiar entre distintas estructuras de vecindario de un conjunto de éstas, que deben ser definidas previamente. Este algoritmo es genérico y consta de muchos grados de libertad a la hora de diseñar variaciones e instancias particulares.

- **Búsqueda local iterada (ILS)**

La búsqueda local iterada o Iterated Local Search (ILS) es una técnica basada en un concepto simple pero que es bastante efectivo. Cada iteración actual es perturbada y a esta nueva solución se le aplica la búsqueda local para mejorarla, Este nuevo óptimo local

obtenido, puede ser aceptado como la nueva solución actual, si es que pasa los estándares de aceptación definidos.

3.2.2 Metaheurísticas basadas en población

Los metaheurísticas basadas en el comportamiento de una población tienen como principal características que trabajan con un conjunto de soluciones (que son llamadas población) que van cambiando en cada iteración. El resultado de este algoritmo va a depender de la forma en la cual se va a manipular la población [4] [40] [41]. Entre estos algoritmos encontramos:

- **Algoritmos Genéticos (GA)**

Los algoritmos genéticos cuentan con la naturaleza para evolucionar seres para adaptarlos a cambios de su entorno. Es un proceso iterativo y estocástico en donde se crea una población de individuos que buscan sobrevivir. Por lo que los mejores se reproducen y/o mutan para adaptarse al ambiente.

- **Búsqueda dispersa (SS)**

La búsqueda dispersa o Scatter Search, cuenta con un algoritmo que se basa en mantener un conjunto preferentemente pequeño de posibles soluciones que tiene la propiedad de contener soluciones lo suficientemente buenas como soluciones diversas. Este conjunto se divide en subconjuntos de soluciones a las cuales se les aplica un proceso de mejora y se re combinan. Para mejorar las soluciones se utilizan mecanismos de búsqueda local [6].

- **Sistemas basados en colonias de hormigas (ACO)**

Las Ant Colony Optimization (ACO) están inspiradas en el comportamiento de las hormigas cuando están buscando comida. Esto se da de la siguiente manera: primeramente las hormigas exploran el área cercana a su nido aleatoriamente, cuando esta encuentra la comida la lleva al nido. Mientras realiza este camino al nido la hormiga bota una feromona que guiará al resto de las hormigas a encontrar comida. Este camino creado por el rastro de la feromona les ayuda a las demás hormigas a encontrar el camino más corto entre el nido y la comida. Este rastro de feromona se crea mediante un modelo probabilístico [40] [42].

- **Algoritmos basados en cúmulos de partículas (PSO)**

Es una técnica metaheurística inspirada en el comportamiento social del vuelo e las bandadas de aves. La toma de decisión tiene características de comportamiento social e individual por cada posible solución creada, con las cuales se determina el movimiento de éstas en el espacio solución. Es aplicado este método para resolver problemas de optimización [15] [8] [43].

En las siguientes secciones se profundizará el estudio del Algoritmo Genético (GA) y el algoritmo PSO que son las dos técnicas presentes, en el método de reconocimientos propuesto.

3.3 Algoritmo basado en cúmulos de partículas (PSO)

El algoritmo basada en cúmulos de partículas o PSO (por su nombre en inglés Particle Swarm Optimization) es una técnica metaheurística basada en poblaciones e inspirada en el comportamiento social de las bandadas de aves o el movimiento de los cardúmenes de peces. Este método fue introducido por primera vez en [15] desarrollado por el psicólogo-sociólogo James Kennedy y por el ingeniero electrónico Russell Eberhart en 1995, basándose en un enfoque de “metáfora social” que describe a este algoritmo, la cual es posible describir como sigue: Los individuos que conviven en una sociedad tienen una opinión que es parte de un conjunto de creencias (espacio de búsqueda) compartido por todos los posibles individuos. Cada individuo puede modificar su propia opinión basándose en tres factores [41]:

- Conocimiento de su entorno.
- Conocimiento histórico (memoria).
- El conocimiento histórico o experiencia de otros individuos situados en su vecindario.

Estos algoritmos son altamente aplicados en diferentes campos de investigación. Como las redes de distribución, fusión de imágenes pixel a pixel, fusión de imágenes mediante la transformada de coseno y wavelet, fusión multimodal a nivel de scores [29], la reconstrucción del diagrama de radiación de antenas [41] [42], etc.

3.3.1 Descripción del algoritmo PSO

El cúmulo de partículas es un sistema multiagente, en otras palabras las partículas generadas son agentes simples que se mueven por el espacio de búsqueda y guardan la mejor posición que han encontrado. Cada partícula cuenta con un “fitness”, una posición y un vector de velocidad que condiciona su movimiento en el espacio. Las partículas son guiadas por las mejores partículas (líder) en el momento (iteración).

- El vector $x_i = \langle x_{i1}, x_{i2}, \dots, x_{in} \rangle$ almacena la posición actual (localización) de la partícula en el espacio de búsqueda.
- El vector $pBest_i = \langle p_{i1}, p_{i2} \dots p_{in} \rangle$ va a almacenar la posición de la mejor solución encontrada por cada partícula hasta el momento.
- El vector de velocidad $v_i = \langle v_{i1}, v_{i2} \dots v_{in} \rangle$ almacena el gradiente, es decir la dirección según la cual se moverá la partícula dentro del espacio solución.
- El valor de $fitness_xi$ almacena el valor de aptitud de a solución actual (vector x_i).
- El valor de $Gbest$ nos da el valor asociado a la mejor solución de entre todas las partículas encontrada hasta el momento.



Figura 3-2: Bandada de ave en el cual se inspira el algoritmo PSO [42].

Las velocidades se generan aleatoriamente, con cada componente en el intervalo $[-v_{max}, v_{max}]$, donde v_{max} será la velocidad máxima que pueda tomar una partícula en cada movimiento. No es conveniente fijarlas en cero ya que los resultados no son lo suficientemente buenos [43].

Una vez creada la nube de partículas, éstas se deben mover dentro del proceso iterativo. Una partícula se mueve desde una posición del espacio de búsqueda hasta otra, simplemente, añadiendo el vector posición x_i el vector velocidad v_i para obtener la nueva posición que se guardará en dicho vector. Para esto se actualiza la posición como se ve en la ecuación (3-1).

$$x_i = x_i + v_i \quad (3-1)$$

Cuando ya se actualice la nueva posición de la partícula se procede a evaluarla en el $fitness_{xi}$. Luego se procede a comparar con los demás fitness y si este es el mejor fitness encontrado hasta el momento, se deben actualizar los valores de mejor posición $pBest_i$ y $fitness_{pBest_i}$. La velocidad de cada partícula es modificada en cada iteración u utilizando como referencia la velocidad de la iteración anterior.

Este algoritmo cuenta con un componente individual y un componente social. El modelo matemático resultante y es representativo del PSO y este viene dado por las siguientes ecuaciones:

$$v_i^{k+1} = w \cdot v_i^k + \varphi_1 \cdot rand_1 \cdot (pBest_i - x_i^k) + \varphi_2 \cdot rand_2 \cdot (Gbest - x_i^k) \quad (3-2)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (3-3)$$

En la ecuación (3-2) se aprecia el vector de velocidad de cada partícula i en una iteración k . El componente individual o cognitivo dado por $\varphi_1 \cdot rand_1 \cdot (pBest_i - x_i^k)$ representa la distancia entre la posición actual y la mejor conocida por esa partícula, es decir, la decisión que tomará la partícula influenciada por su propia experiencia a lo largo de su existencia. La componente social se modela con $\varphi_2 \cdot rand_2 \cdot (Gbest - x_i^k)$ y representa la distancia entre la posición actual y la mejor posición del vecindario, esto implica que es la decisión que tomará la partícula influenciada por el resto de la población de enjambre en la que la partícula está inserta. A continuación se detallan cada uno de los factores utilizados en el algoritmo:

- v_i^k : velocidad de la partícula i en la iteración k
- w : factor de inercia
- φ_1 y φ_2 : Son constantes de aprendizaje (pesos) que controlan el comportamiento cognitivo (individual) y social.
- $rand1, rand2$: números aleatorios entre 0 y 1
- x_i^k : posición actual de la partícula i en la iteración k
- $pBest_i$: mejor posición encontrada en la partícula i hasta el momento
- $Gbest$: representa la posición con el mejor $pBest_fitness$.

Como se pudo apreciar en la ecuación (3-3) las partículas i en cada iteración k generan un movimiento aleatorio, este se puede apreciar en la Figura 3-3. Las flechas con líneas discontinuas representan la dirección de los vectores de velocidad actual, v_i^{kpbest} es la velocidad de la mejor posición tomada por la partícula, v_i^{gbest} es la velocidad de la mejor partícula encontrada en el vecindario y v_i^k es la velocidad actual de la partícula. La flecha de línea continua representa la dirección que toma la partícula para moverse desde la posición x_i^k hasta la posición x_i^{k+1} . El movimiento de esta partícula dirigida por la flecha continua es dependiente de la posición anterior y la velocidad inicial de la partícula.

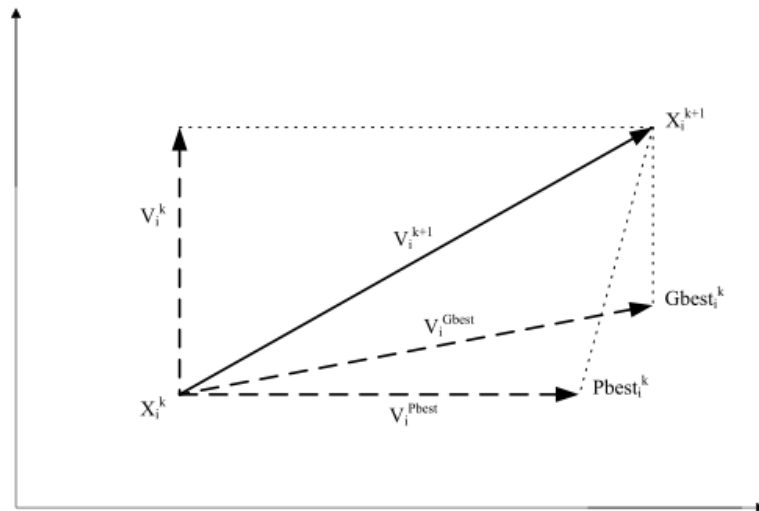


Figura 3-3: Movimiento de las partículas en el espacio solución [42].

3.3.2 Tipos de PSO

Se pueden obtener diferentes tipos de PSO si se realizan distintos cambios en los factores que condicionan su funcionamiento, por ejemplo podemos clasificarlos según la importancia de los pesos cognitivo y social, como también según el tipo de vecindario a utilizar. Primeramente, al clasificarlos dependiendo de la influencia de los factores de cognitivo y social, (con valores $C1$ y $C2$ respectivamente), se obtendrán 4 variantes definidas por Kennedy en [8], ya que estos valores afectan directamente la dirección de la velocidad que toma la partícula en el movimiento. Estos 4 tipos de algoritmos serían:

- Modelo Completo: Donde $\varphi_1, \varphi_2 > 0$ lo que indica que tanto el componente de memoria individual de cada partícula, como el social intervienen en el movimiento. Dentro de este modelo también podemos tener en cuenta los siguiente:
 - Si $\varphi_1 > \varphi_2$ la experiencia propia predominará por sobre el movimiento colectivo del cúmulo.
 - Si $\varphi_1 < \varphi_2$ La experiencia colectiva predominara en el movimiento de las partículas.
 - Si φ_1 y φ_2 son muy grandes el movimiento de las partículas será muy abrupto.
 - Si φ_1 y φ_2 son muy pequeños los movimientos de las partículas son más suaves.
- Modelo sólo Cognitivo: $\varphi_1 > 0$ y $\varphi_2 = 0$. Esto indica que las partículas se moverán de manera independiente buscando su mejor posición.
- Modelo solo Social: $\varphi_1 = 0$ y $\varphi_2 > 0$. El movimiento social será el único existente en el movimiento.
- Modelo sólo Social exclusivo: $\varphi_1 = 0$ y $\varphi_2 > 0$ y $G_{best} \neq x_i$. La posición de la partícula en sí no puede ser la mejor de su entorno.

Las constantes de inteligencia son de una gran importancia en el comportamiento del algoritmo ya que estas determinan como se moverán las partículas en el espacio de búsqueda. Es por esto que en este estudio se realiza un ajuste de parámetros para adaptar el algoritmo al problema.

El sistema de reconocimiento de rostros que se propone aplica fusión de descriptores visibles y térmicos, donde algoritmo PSO cumple un rol fundamental, ya que este es la herramienta que busca los pesos que permitirán destacar, ciertas áreas de las imágenes del rostro, además de indicar que espectro predomina, es las regiones de la imagen fusionada. De esta forma se espera que las tasas de reconocimiento aumenten.

3.4 Algoritmo Genético

Los algoritmos genéticos han sido usados desde mediados de los años 70 [10] [31] intentando emular la teoría de Darwin sobre la evolución por la sobrevivencia del más fuerte. Su uso está asociado inherentemente a la maximización o minimización del resultado de una función, la cual se denomina como “Función de Fitness” (fitness function), por medio de la selección natural y mutación de genes, logrando así un método robusto que encuentra soluciones que no son guiadas matemáticamente. Los algoritmos genéticos son básicamente un método iterativo para encontrar soluciones a un problema planteado. El pseudocódigo de un algoritmo genético (GA, por sus siglas en inglés) es el siguiente:

1. Inicialización: Se crea una “población” inicial de N individuos, llamados códigos genéticos que representan las posibles soluciones al problema definido. Dicha población es generalmente generada en forma aleatoria dentro del dominio de la función.
2. Evaluación: Una vez que la población es inicializada o una descendencia es creada, sus valores de fitness (el resultado de la función de fitness) es calculado.

3. Selección: Dicha población es sujeta a la sobrevivencia del más fuerte, donde se eligen 2 o más individuos que aportan sus genes para la creación de nuevos individuos. La selección se realiza por medio de una “ruleta circular” en la cual a cada individuo se le asigna un espacio dentro de la ruleta proporcional a su valor de fitness; selección estocástica universal; selección por eficiencia; selección por torneo entre otros métodos.
4. Recombinación: Dichos genes se cruzan entre sí de diversas maneras, incluidos el cruce en puntos fijos, cruce en puntos variables, cruce probabilístico, entre otros.
5. Mutación: La descendencia puede mutar en uno, dos o más puntos, con una probabilidad de suceso o bien mutación forzada.
6. Reemplazo: la descendencia tras ser evaluada puede ser incorporada a la población, o bien descartada según los parámetros elegidos, los cuales pueden ser el reemplazo del peor individuo, reemplazo aleatorio, entre otros.
7. Repetir pasos 2 al 6 hasta encontrar la condición de término, la cual puede ser un valor objetivo o una cantidad de iteraciones.

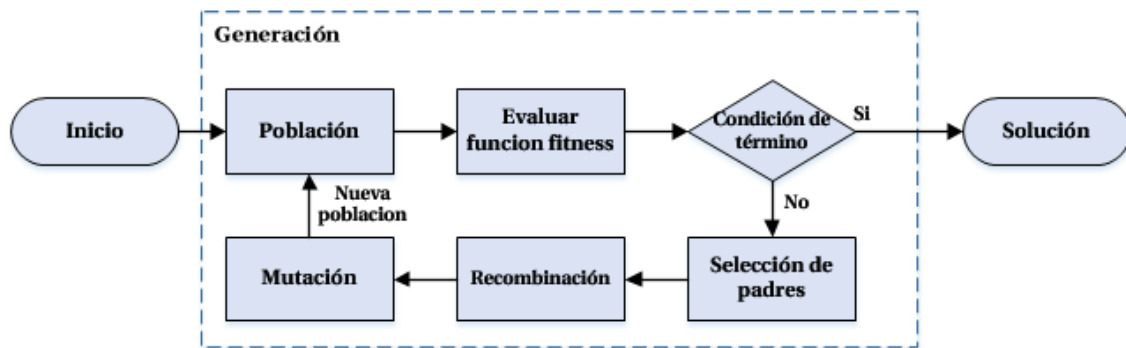


Figura 3-4: Esquema general de un algoritmo genético.

3.5 Conclusiones de capítulo

Las técnicas metaheurísticas han tenido una gran aplicación a lo largo de los años en variadas áreas, estas se pueden dividir en técnicas basadas en trayectoria y en población. Aquí encontramos al algoritmo PSO el cual es el principal foco de este trabajo, de él se comprende la relevancia de sus constantes de aprendizaje, las cuales definen el comportamiento del algoritmo. El algoritmo genético es un proceso estocástico en el cual la población creada ira tomando los mejores resultados, y los irá mutando y cruzando para generar, resultados aún mejores.

Es necesario comprender el comportamiento de ambos algoritmos en este estudio ya que ellos son los que entregan los pesos óptimos del método de fusión utilizado, además de que el motivo de aplicar genético, es para realizar una comparación directa de quien funciona mejor en reconocimiento de rostros.

4 Sistema propuesto

Los diferentes métodos de reconocimiento de rostros se basan en algoritmos de extracción de características en el espectro visible, sin embargo se presentan situaciones indeseables en el reconocimiento cuando existe variación de la iluminación y gestos en las imágenes. Estos métodos se han aplicado en el espectro térmico [2] [3] [20] [18], donde se demuestran ciertas ventajas como robustez a los cambios de iluminación e incluso a la ausencia de la misma. A pesar de esto, el espectro térmico presenta otro tipo de limitaciones, como los cambios de temperatura en el ambiente, alteración de los procesos metabólicos de los sujetos (enfermedad, ejercicio o consumo de alcohol), la susceptibilidad de la cámara a factores externos y la respuesta variable del sensor cuando la cámara lleva trabajando largos periodos de tiempo [2] [3] [33] [20] [44].

Por lo anterior se busca tomar ventaja de ambos espectros aplicando la fusión de ellos. La fusión se puede lograr de formas distintas, tales como la fusión espacial de las imágenes (promedio ponderados de las imágenes, pixel a pixel), fusión aplicando transformadas, por ejemplo la Transformada de Fourier o Discreta de Wavelet [32] [37], y fusión de descriptores de imágenes [7].

De esta manera el método de fusión de imágenes visibles y térmicas propuesto, usará la metodología de fusión de descriptores de imágenes mediante el uso del algoritmo PSO, el cual se encarga de encontrar pesos multiplicativos con tal de combinar la información más relevante de las imágenes y así maximizar las tasas de reconocimiento del sistema.

A modo de comparación se aplica la fusión de descriptores utilizando el Algoritmo Genético, el cual cumple la misma función que el algoritmo PSO, de encontrar los mejores pesos multiplicativos para realizar la fusión, con la finalidad de aumentar las tasas de reconocimiento.

4.1 Fusión de descriptores usando el algoritmo PSO

El sistema de reconocimiento de rostros propuesto usa el algoritmo PSO para encontrar pesos multiplicativos, que luego son aplicados a los descriptores visibles y térmicos, teniendo como función objetivo la tasa de reconocimiento la cual se quiere maximizar. Para esto este sistema se divide en dos etapas. La obtención de los pesos multiplicativos óptimos reconocimiento es llamada en este trabajo como “Etapas de entrenamiento”. Posteriormente estos pesos óptimos son aplicados a una base de datos la cual es llamada “Etapas de validación” del sistema. Esto se explica con mayor detalle en las siguientes secciones.

4.1.1 Etapa de entrenamiento

La etapa de entrenamiento del sistema es aquella que busca los pesos óptimos que maximicen la tasa de reconocimiento de rostros mediante el algoritmo PSO. Esta etapa se compone de tres procesos: Aplicación de métodos de reconocimiento de rostros, “Matching” y fusión de descriptores usando el algoritmo PSO. La etapa de entrenamiento puede apreciarse en la Figura 4-1.

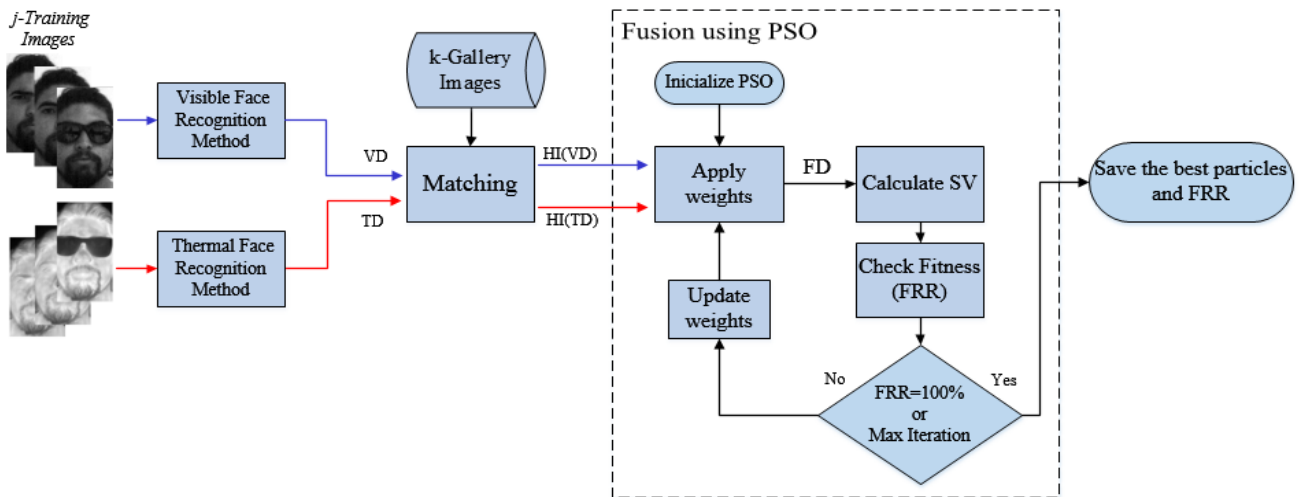


Figura 4-1: Etapa de entrenamiento del sistema.

Aplicación de métodos de reconocimientos de rostros

Un sistema de reconocimiento contiene dos conjuntos de imágenes, uno de “galería y otro de “test”, donde el conjunto de “galería” contiene imágenes que se encuentran almacenadas en el sistema y que tienen solo una imagen de rostro por sujeto, como este sistema usa la fusión de imágenes se tienen pares de imágenes, una visible y otra térmica para cada sujeto. En el caso del conjunto de “test” se encuentran pares de imágenes sometidas a distintas características como: variaciones de expresión, iluminación y/o pose.

Este proceso contiene entonces, un conjunto de pares de imágenes visibles y térmicas, a la cual se denomina “*j-Training images*”, el cual representa el conjunto de “test”. Se deben aplicar dos métodos de reconocimientos de rostros, uno para las imágenes visibles y otro para las imágenes térmicas, estos nos darán como resultado un par imágenes divididas en P regiones, las cuales están representadas en forma de histogramas por región, que luego se concatenan para formar la representación global del rostro.

De los pares de imágenes de “training” se obtienen dos histogramas, donde uno representa el descriptor aplicado en el espectro visible y el otro en el espectro térmico. El descriptor visible será llamado “VD” y el descriptor térmico será llamado “TD”. Este proceso se realiza para cada par de imágenes presentes en el conjunto denominado “*Training*”, donde j representa cada par de imágenes. En la Figura 4-2 se puede apreciar una representación de esto aplicando el método LBP.

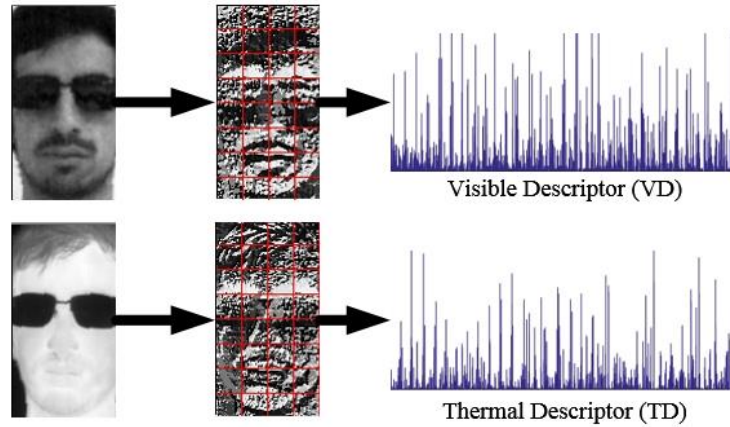


Figura 4-2: Obtención de los descriptores VD y TD.

Matching

Una vez obtenidos los descriptores VD y TD de los pares de imágenes j -*Training* se procede a realizar el “matching”. Este proceso consiste en combinar cada región de los descriptores de cada par de imágenes del conjunto j -*Training* con cada región i de cada par de imágenes del conjunto k -*gallery* utilizando la intersección de histogramas HI (), este proceso combina las imágenes visibles del conjunto k -*gallery* con todas las imágenes visibles del conjunto j -*Training*, generando la intersección de histogramas HI(VD), de igual forma las imágenes térmicas de los conjuntos generan la intersección de histogramas HI(TD). La intersección de histograma está dada por la ecuación (4-1).

$$HI(S^{j,k}) = \sum \min(S_j, S_k) \quad (4-1)$$

Así se obtienen $J \cdot K$ vectores de intersección de histogramas, HI (VD) y HI (TD), donde J corresponde al número de imágenes que contiene el conjunto de *Training* y K es la cantidad de imágenes presentes en el conjunto de galería. Cada vector tendrá un largo que dependerá de las divisiones P en las cuales se divide la imagen, el vector contiene un puntaje asignado al histograma de cada región que genera la intersección de histogramas a cada región. El vector HI(VD) y HI(TD) son enviados al siguiente proceso.

Fusión usando PSO

Para comenzar el proceso los vectores HI (VD) y HI (TD) son concatenados para obtener un único vector de características visibles y térmicas. Este vector final tiene un tamaño de $2P$ correspondiente a la concatenación de las divisiones P del par de imágenes visibles y térmicas. El proceso de fusión asigna un peso multiplicativo aleatorio a cada región del vector de características, modificando la información que este contiene, posteriormente este vector es tomado por el algoritmo PSO quien realiza la fusión de descriptores optimizando los pesos multiplicativos de cada región con el fin de resaltar la información más relevante de ambos espectros. El proceso considera los resultados de reconocimiento de rostros obtenidos para cada imagen de prueba j , imagen de galería k y las particiones p , con $p \in [1, P] \subset \mathbb{N}$ dados por los vectores de características de las imágenes visible y térmica. Matemáticamente el descriptor de

fusión FD de vectores de características viene dado por la ecuación (4-2), para un par de imágenes visible-térmico de test j y un par de imágenes de galería k el cual es multiplicado por el vector de pesos.

$$FD^{j,k} = [w_{visible\ i} \cdot HI(VD_i^{j,k}), w_{térmico\ i} \cdot HI(TD_i^{j,k})] \quad \forall p \in [1, P] \subset \mathbb{N} \quad (4-2)$$

Donde j es la imagen de prueba, k es la imagen de galería, P es la cantidad de particiones de las imágenes, VD es el descriptor de las imágenes visibles y TD es el descriptor de las imágenes térmica. Si el conjunto de prueba contiene J imágenes y el conjunto de galería tiene K imágenes, entonces se obtendrán $J \cdot K$ vectores FD por cada imagen j en el conjunto de test llamado *Training*.

Antes de realizar la fusión es necesario elegir apropiadamente las constantes de inteligencia cognitiva y social del algoritmo, además de su factor de inercia. Una vez escogidos estos valores y habiendo obtenido el vector de característica visible y térmico se procede a realizar la fusión de los descriptores obtenidos, pero primeramente se debe seleccionar la función fitness que en este caso corresponde a la tasa de reconocimiento aplicada a los conjuntos de test y galería. Para esto el algoritmo PSO genera una población x_i de partículas, donde cada partícula será de un largo $2P$ y contiene pesos aleatorios en el intervalo $[0,1]$, estos valores corresponden al peso $w = [w_{visible}, w_{térmico}]$, que es la concatenación de los pesos P visibles y P térmicos. Estos pesos serán multiplicados punto a punto con los vectores de características obtenidos por la intersección de histogramas, donde cada peso modificara la información correspondiente cada una de las regiones en las cuales fueron divididas las imágenes.

Para realizar un análisis más completo no solo se utilizarán pesos aleatorios para realizar la fusión, sino que también se aplicarán pesos visibles y térmicos de manera complementaria durante toda la ejecución del proceso, es decir la suma del peso visible y el peso térmico de la región i debe ser igual a 1. Cada vez que las partículas se actualicen usando la ecuación(3-1), estas pasarán por un proceso de normalización de tal forma que se mantengan de forma complementaria. En la Figura 4-3 se propone un ejemplo del vector de características visible-térmico con una longitud de 10, donde los 5 primeros valores corresponden entonces a los pesos visibles, y los siguientes 5 corresponden a los pesos térmicos.

Estos luego se normalizan aplicando las ecuaciones (4-3) y (4-4).



Figura 4-3: Ejemplo de vector de características visible-térmico a normalizar.

$$VN_n = \frac{V_n}{T_n + V_n} \quad (4-3)$$

$$TN_n = \frac{T_n}{T_n + V_n} \quad (4-4)$$

Donde VN será el nuevo peso visible en la posición n, tomando los pesos V_n y T_n que se encuentran en la misma posición, de igual forma TN será el nuevo peso térmico en la posición n, tomando los pesos visibles y térmicos en la misma posición, dejando los pesos de forma complementaria.

El reconocimiento finalmente es realizado, obteniendo un valor final de intersección de histogramas, el cual se obtiene sumando todos los componentes de FD, este será el valor de similitud de cada sujeto. Este proceso se puede representar matemáticamente como se muestra en la ecuación (4-5):

$$SV^{j,k} = \sum_1^{2P} FD_p^{j,k} \quad (4-5)$$

Este proceso se realiza en cada iteración por lo que significa un gran coste computacional, por lo que cada iteración tarda alrededor de 5 segundos. Cuando ya finaliza o se alcanza el número de iteraciones establecidas el algoritmo entrega los pesos óptimos encontrados.

Pasos del algoritmo para generar la fusion de descriptores

1. Asignamos los valores correspondientes al factor de inercia (w), los coeficientes de inteligencia individual (φ_1) y el de inteligencia social (φ_2).
2. Se genera una población inicial X de N partículas las cuales tendrá 64 posiciones para cada una de ellas, estos pesos deben estar en un intervalo de [0,1].
3. Se define la matriz de velocidad con valores aleatorios limitada por [-vmax, vmax].
4. Asignamos Pbest=X, ya que la mejor posición hasta el momento es X.
5. Obtenemos el valor de fitness de Pbest, para esto cada partícula es aplicada a la base de datos de descriptores FD mediante producto punto de la partícula con cada descriptor FD. El valor final obtenido es el valor de similitud $SV^{jk} = \sum_1^D FD_p^{j,k}$.
6. El reconocimiento es efectuado encontrando el máximo valor de similitud para cada imagen j, de entre los K valores obtenidos debido a la comparación de a imagen de prueba j con cada imagen de galería k. El valor de fitness se obtendrá para cada partícula, la cual será la tasa de reconocimiento obtenida considerando los correctos reconocimientos, el cual será optimizado.
7. La partícula que obtenga el mejor fitness dentro del vector Pbest será asignado como Gbest quien liderará el cúmulo de partículas.
8. La velocidad es modificada considerando los coeficientes de inteligencia individual y social. Luego de que se modifica se actualizan las posiciones.
9. Se obtiene el fitness de X y el fitness de Pbest. Si el fitness de X es mejor Pbest, entonces Pbest=X.
10. Se obtiene el mejor fitness de Pbest y a este lo asignamos como Gbest.
11. Si la iteración es menor a la Iteración máxima se repiten los pasos del 6 al 9.

Como se dice anteriormente al ser un proceso iterativo toma un desgaste computacional considerable sin embargo esta etapa solo se debe realiza una vez, luego solo se deben usar los mejores pesos obtenidos.

4.1.2 Etapa de validación

Esta etapa es mucho más sencilla que la anterior ya que solo toma los pesos óptimos encontrados y lo aplica a una base de datos completa, dándonos el valor real del desempeño del método que se propone. En este trabajo las bases de datos utilizadas están divididas en conjuntos por lo que la tasa de reconocimiento final de la validación es un promedio de cada conjunto de prueba con cada conjunto de galería. En la Figura 4-4 se observa el esquema de la validación de bases de datos.

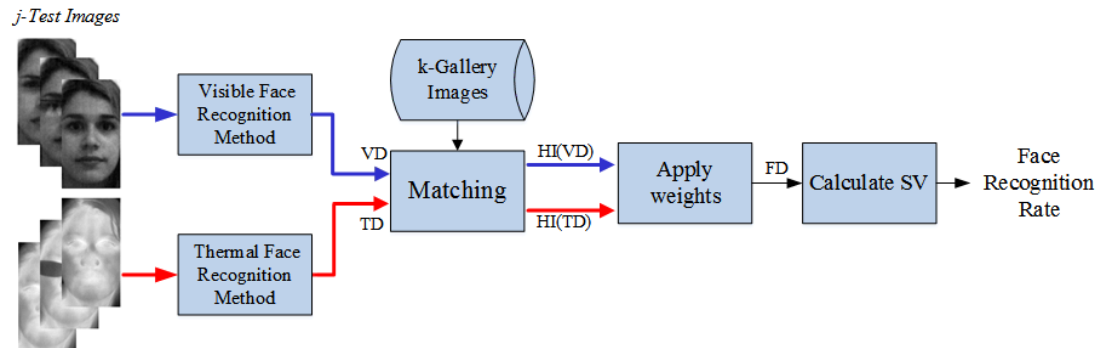


Figura 4-4: Esquema de la etapa de validación propuesto.

El reconocimiento de rostros se realiza aplicando la ecuación (4-5), que nos dará la medida de similitud de cada imagen de galería con todas las imágenes de test. Al entrenar solo se usan pequeños conjuntos de la base de datos, por lo cual al validar se deben considerar todos los conjuntos de esta. Finalmente se debe promediar los resultados de todos los conjuntos y así se obtiene la tasa de reconocimiento final.

4.2 Fusión de descriptores usando algoritmo genético

La fusión de descriptores usando el algoritmo genético se usa como medida de comparación del estudio principal que es la aplicación del algoritmo PSO. El sistema funciona de la misma manera que el explicado en la sección anterior solo que los pesos son optimizados usando el algoritmo genético. Este a diferencia de PSO en cada iteración, va tomando los mejores valores de fitness y les aplica un cruce aleatorio esperando obtener mejores resultados, también se aplica la mutación, esperando exactamente lo mismo. A continuación se explica en detalle la obtención del código genético que nos entrega los pesos óptimos del sistema. Los parámetros para el algoritmo genéticos son tomados de [7]. El esquema de la Figura 4-5, nos muestra el comportamiento de esta metodología.

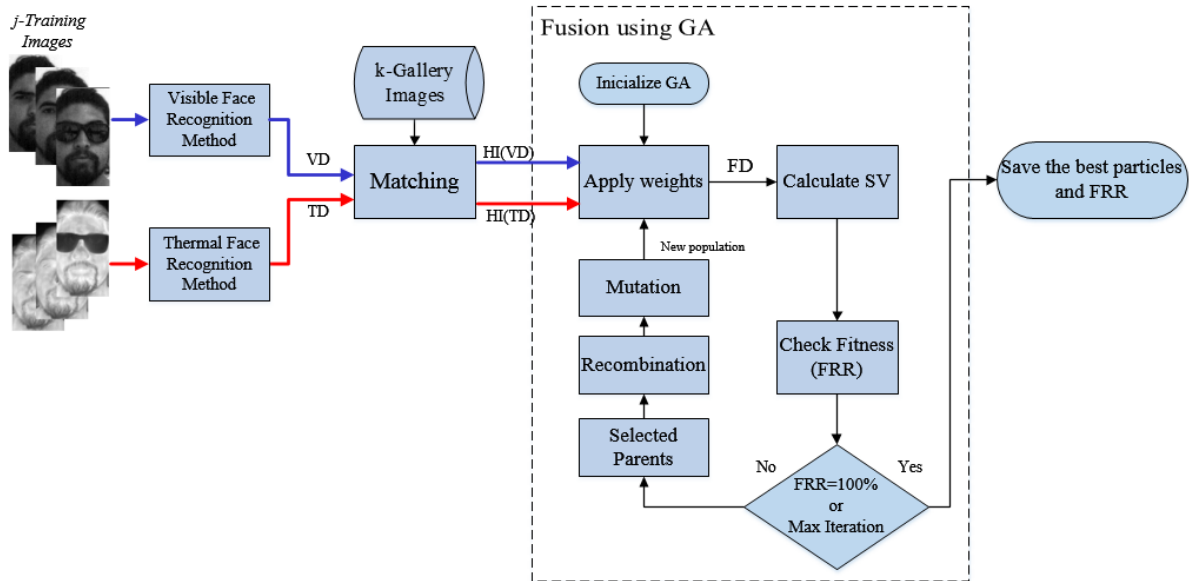


Figura 4-5: Esquema de fusión de descriptores usando algoritmo genético.

Pasos para generar la fusión usando GA

1. Una población inicial de 100 códigos genéticos se crea en forma aleatoria con $2P$ pesos con valores en el intervalo $[0,1]$. La población inicial tiene pesos complementarios entre las regiones visibles y térmicas. Cada código genético es aplicado a la base de datos de descriptores FD mediante producto punto del código con cada descriptor FD . El valor final obtenido corresponde al valor de similitud $SV^{j,k} = \sum_1^{2P} FD_p^{j,k}$.
2. El reconocimiento de rostro se efectúa encontrando el máximo valor de similitud para cada imagen j , de entre los K valores obtenidos debido a la comparación de la imagen de prueba j con cada imagen de galería k .
3. El valor de fitness de cada código genético, es la tasa de reconocimiento obtenida considerando los correctos reconocimientos, el cual será optimizado.
4. Se seleccionan 2 códigos genéticos que actuarán como padres, utilizando el método de la ruleta circular. En la ruleta, cada individuo tiene asignado un espacio proporcional a su valor de fitness.
5. Los padres se cruzan (*crossover*) con una probabilidad de 25,14% en cada punto, obteniendo dos hijos. Los pesos de las regiones visible y térmica ya no son complementarios. Debido a esta probabilidad, en situaciones extremas es posible que los hijos resulten idénticos a los padres, o bien el cruce sea en cada uno de los 64 puntos que componen cada código.
6. Los hijos tienen una probabilidad de 25,14% de mutación en 3 puntos definidos aleatoriamente. Debido a la naturaleza probabilística de mutación y la aleatoriedad de los puntos, como situaciones extremas es posible que ningún punto sufra de mutación, que 3 puntos muten, o que un punto mute 3 veces.
7. El valor de fitness de cada hijo se calcula de la misma manera que los pasos 2 y 3.
8. Si el valor de fitness de un hijo dado es mayor que el valor de fitness más bajo de la población, entonces este hijo se inserta en la población, eliminando un código de manera

aleatoria, de lo contrario, se descarta. Con esto, en caso de ser insertado, es posible que el nuevo hijo reemplace incluso al mejor de los códigos, aunque sí y sólo si es mejor que el peor.

9. Si el número de iteración es menor que 50.000, entonces repetir pasos 5 al 10, de lo contrario, terminar.

De esta forma se obtiene el código genético óptimo que representa los pesos multiplicativos visibles y térmicos que maximizaran la tasa de reconocimiento. Los cuales posteriormente serán validados de igual forma que en el método de fusión usando PSO.

4.3 Conclusiones del capítulo

En este capítulo se presenta un nuevo sistema de reconocimiento de rostros basado en la fusión de descriptores de imágenes visibles y térmicas aplicando el algoritmo PSO, cuyo objetivo es aumentar el desempeño de las tasas de reconocimiento.

El sistema consta de dos etapas fundamentales de las cuales el entrenamiento es la más importante ya que es aquí donde se encuentran los pesos que darán los pesos óptimos que priorizará un espectro por sobre otro, en ciertas áreas de las imágenes, que al fusionarlas maximizara la tasa de reconocimiento.

El algoritmo PSO consta de parámetros fundamentales para su funcionamiento que son los ratios de inteligencia individual y cognitiva, y el factor de inercia. Es necesario encontrar los valores apropiados para que el algoritmo se comporte de la mejor manera posible.

La aplicación del código genético es usada en [44] aplicando la misma metodología de fusión por lo que en este estudio se utiliza solo como medida de comparación con la fusión usando PSO.

5 Experimentos y resultados

En presente capítulo se describen los experimentos y resultados obtenidos utilizando el sistema de fusión propuesto. Primeramente se realiza un ajuste de parámetros del algoritmo PSO para que su funcionamiento sea de la mejor manera posible para el problema. Seguido a esto se realiza el entrenamiento usando del sistema utilizando distintos conjuntos de test y galería de la base de datos Equinox, incluyendo conjuntos que poseen sujetos usando lentes, los cuales como se menciona anteriormente es una limitación en el reconocimiento de rostros térmico. Para finalmente validar los resultados obtenidos en la etapa de entrenamiento aplicando los pesos a bases de datos: Equinox y PUCV-VTF.

5.1 Bases de datos utilizadas

Para poder realizar el reconocimiento de rostros es necesario contar con uno o más conjuntos de imágenes, los cuales están contenidos dentro de una base de datos. Para ello existen distintas bases de datos creadas con la finalidad de evaluar el desempeño de los métodos de reconocimiento facial. Para este estudio se utiliza Equinox [45] y PUCV-VTF [7] debido a que ambas poseen imágenes térmicas y visibles. En este capítulo se describirán las bases de datos utilizadas.

5.1.1 Base de datos Equinox

La base de datos Equinox ha sido usada en una gran variedad de estudios para comparar y evaluar los métodos de reconocimiento de rostros, tanto en el espectro visible como en espectro térmico. Esta base de datos consiste en distintos conjuntos de imágenes frontales tomados en el espectro visible comprendido de longitudes de ondas de 0,38 a 0,75 μm y el espectro infrarrojo con longitudes de onda SWIR de 0,9 a 1,7 μm , MWIR de 3 a 5 μm y LWIR de 8 a 12 μm . Las imágenes fueron capturadas simultáneamente y posee 18629 imágenes en cada espectro, las cuales están en escala de grises y tienen un tamaño de 240x320 píxeles.

La base de datos es dividida en distintos conjuntos, que contienen 81 individuos distintos y donde la captura es realizada con distintos focos de iluminación (Frontal, lateral izquierdo y lateral derecho), con tres distintas expresiones (sonrisa, enojo y sorpresa, imágenes de los sujetos con anteojos e imágenes de los sujetos vocalizando). A partir de estas imágenes se crean 6 sets de imágenes destinadas a ser utilizadas como galerías, las cuales solo tienen una imagen por

individuo, estas son llamadas: “EA”, “VA”, “VF”, “EF”, “VL” y “EL”, donde la primera letra indica el tipo de captura, estas son: “E” cuando posee expresiones y V cuando se está vocalizando. La segunda letra significa el tipo de iluminación, “A” se refiere a que tiene iluminación frontal y lateral, “F” es iluminación frontal y “L” es iluminación lateral. Así mismo se crean 9 conjuntos para utilizar como prueba o bien test, estos son “EA”, “VA”, “VF”, “EF”, “VL”, “EL”, “VG”, “EG” y “RR”, donde “G” indica que los sujetos están utilizando lentes y RR es un conjunto de imágenes aleatorias [14]. El detalle de los conjuntos se puede apreciar en la Tabla 5-1.

Tabla 5-1: Detalle de los conjuntos de la base de datos Equinox.

Nombre	Descripción	Características	Iluminación	imágenes galería	Imágenes prueba
VA	Vocales	Todos los sujetos	Todas las direcciones	81	729
EA	Expresiones	Todos los sujetos	Todas las direcciones	81	729
VF	Vocales	Todos los sujetos	Frontal	81	243
EF	Expresiones	Todos los sujetos	Frontal	81	243
VL	Vocales	Todos los sujetos	Lateral	81	486
EL	Expresiones	Todos los sujetos	Lateral	81	486
VG	Vocales	Usando anteojos	Todas las direcciones	No aplica	324
EG	Expresiones	Usando anteojos	Todas las direcciones	No aplica	324
RR	Vocales y expresiones	Escogidos aleatoriamente	Todas las direcciones	No aplica	500

Todas las imágenes de los distintos conjuntos fueron alineadas con respecto a los ojos, redimensionadas de tal manera que entre ellos hay una distancia de 42 pixeles y cortadas de tal forma que las imágenes tienen un tamaño final de 150 pixeles de alto y 81 pixeles de ancho. Los ojos de los individuos tienen la misma posición espacial y horizontalmente centrados, de manera que solo se muestra la cara de los individuos. En la Figura 5-1 se aprecian imágenes de uno de los sujetos de la base de datos Equinox donde posee distintas expresiones en ambos espectros.



Figura 5-1: Imágenes de la base de datos Equinox: Arriba imágenes del espectro visible y abajo del espectro térmico [14].

5.1.2 Base de datos PUCV-VTF

La base de datos PUCV-VTF consta de 5 set de imágenes frontales con 20 imágenes por tipo de captura de 76 individuos. Estos set de imágenes tienen por nombre “Normal” que consiste en imágenes sin expresión, “Frown” que poseen expresiones de enojo, “Glasses” que poseen imágenes de individuos usando anteojos ópticos u oscuros en forma aleatoria, “Smile” donde las personas están sonriendo y “Vocal” en la cual a las personas se les pide que pronuncien una vocal a su elección. Por cada una de las imágenes visibles, se captura también una imagen térmica y para la elaboración de los conjuntos de imágenes útiles se selecciona una imagen en el espectro visible por individuo y la imagen térmica con el mismo número de captura. Las imágenes son alineadas por la posición de los ojos, de la misma manera en que Equinox, dejando su tamaño de 150 píxeles de alto y 81 píxeles de ancho.

En la Figura 5-2 se aprecian ejemplos de los distintos conjuntos que usa PUCV-VTF y en la Tabla 5-2 se detalla cada uno de los conjuntos.

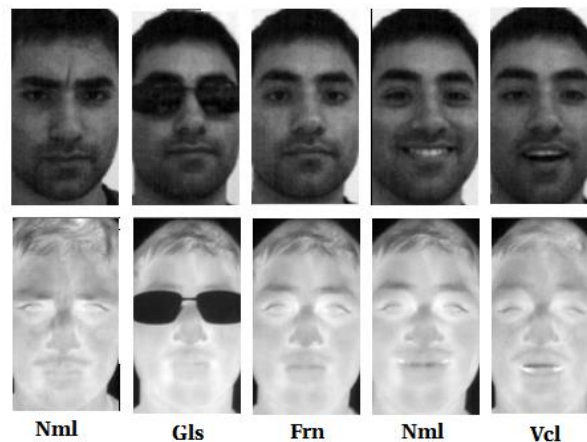


Figura 5-2: Ejemplo de las imágenes de la base de datos PUCV-VTF en sus diferentes conjuntos y en ambos espectros, arriba visible abajo térmico.

Tabla 5-2: Detalle de los 5 conjuntos de imágenes térmicas de la base de datos PUCV-VTF.

Conjunto	Nº de imágenes	Descripción
Nml	76	Imágenes de individuos sin expresiones
Frn	76	Imágenes de individuos con expresiones de enojo
Gls	76	Imágenes de individuos con anteojos
Sml	76	Imágenes de individuos sonriendo
Vcl	76	Imágenes de individuos pronunciando una vocal

5.2 Experimentos para la etapa de entrenamiento

Luego de haber definido las bases de datos a utilizar, se plantean distintos experimentos para la etapa de entrenamiento del sistema. Para esto se debe definir los métodos de reconocimientos de rostros que serán a aplicados en cada espectro, en este estudio se presentan tres métodos: LBP, HOG y LDP. Con ellos se harán distintas combinaciones de descriptores visibles y térmicos para el evaluar el comportamiento y eficiencia del sistema que se propone. Los métodos utilizados fueron seleccionados en base a los experimentos realizados en [7]. Como se menciona en el Capítulo 3, las imágenes son divididas en regiones para obtener el histograma de cada una de esas regiones para luego concatenarlas, y así obtener el descriptor de cada espectro, en este estudio las imágenes serán divididas en 32 (4 regiones de ancho y 8 de alto) y 256 regiones (16 regiones de alto y 16 de ancho).

El descriptor fusionado será denotado de la siguiente manera: $FD - A - VD - TD - P$. Donde **FD** nos dice que es una fusión de descriptores, **A** será la metodología utilizada para fusionar, que en este estudio es PSO, pero de igual forma se aplica GA para comparar los resultados. **VD** es el descriptor que se aplica a las imágenes del espectro visible, **TD** es el descriptor aplicado a las imágenes del espectro térmico, y finalmente **P** es la cantidad de regiones en la cual se dividirá la imagen. En el caso de los pesos normalizados al final de se agregará una **N** indicando que es pesos son complementarios.

Las distintas combinaciones de descriptores visibles-térmicos que se usarán para la fusión de imágenes en este trabajo son:

- LBP-LBP
- LBP-HOG
- HOG-LBP
- HOG-HOG
- LDP-LBP*

Estos están escritos de la forma VD-TD, es decir descriptor visible primero y luego descriptor térmico. La fusión de tipo LDP-LBP es un caso especial en este estudio ya que su intersección de histogramas se hace de forma distinta a los otros métodos. Esto es porque se aplica el descriptor LDP de grado tres a la imagen visible con 256 regiones, mientras que el LBP aplicado a las imágenes térmicas solo tendrá 32. Entonces para poder obtener el tamaño necesario de la intersección de histogramas, la parte visible será agrupada en grupos de ocho regiones para generar histogramas que del mismo tamaño que el de la imagen térmica. Este proceso se emplea para realizar una comparación con los resultados expuestos en [7]. En la Figura 5-3 se puede observar el proceso de agrupación de las regiones para la fusión de tipo LDP-LBP.



Figura 5-3: Ilustración de la conversión de 256 regiones a 32 para realizar la fusión LDP-LBP

La etapa de entrenamiento será evaluada utilizando la base de datos estándar Equinox, la cual es usada en una gran cantidad de estudios de reconocimiento de rostros. Se seleccionará un conjunto de galería de los definidos por la misma base de datos y un conjunto de prueba o de test.

5.2.1 Ajuste de parámetros del algoritmo PSO

El primer experimento realizado en este estudio es el ajuste de parámetros del algoritmo PSO, ya que este depende de varios parámetros que condicionan su rendimiento, entre ellos encontramos las constantes de inteligencia del algoritmo, la velocidad máxima que puede alcanzar cada partícula, el tamaño del cúmulo de partículas y la cantidad de iteraciones [4] [43] [46]. Por lo tanto para definir que este algoritmo se desempeñe de buena manera en el sistema creado, se ajustan estos parámetros antes mencionados. Primeramente se ajustan los parámetros de las constantes de inteligencia que se observan en la ecuación (3-2) las cuales definen el movimiento en el espacio de búsqueda de las partículas, luego se ajustará la velocidad máxima de ellas, seguido del tamaño de la población y finalmente con la cantidad de iteraciones.

Ajuste de las constantes φ_1 y φ_2

Para ajustar los ratios o constantes de inteligencia del algoritmo se selecciona cualquiera de los descriptores que se fusionarán, en este caso se selecciona FD-LBP-LBP-32, ya que es el que tubo mejores resultados en [44]. Posteriormente se seleccionan los conjuntos de entrenamiento y de test a utilizar, que son EA como conjunto de galería y RR como conjunto de test. Se usa EA ya que contiene imágenes de los sujetos con iluminación en todas las direcciones y se usa RR ya que posee imágenes de “random” de todos los conjuntos de la base de datos Equinox por lo que se espera que represente un sistema más real.

Para evaluar los valores de las constantes se decide variar su valor en un rango de [0,2] que es lo que son los valores recomendados en la literatura [15] [43]. El valor de φ_1 partirá en 0 mientras que φ_2 variará en el rango de [0,2] con un delta de 0.2, entonces cuando el valor de φ_2 llegue a 2 el valor de φ_1 aumentará en un delta de 0.2, y así hasta que φ_1 llegue a 2. Esto nos da 121 combinaciones posibles de las constantes. Para evaluar las constantes se hacen 5 pruebas por cada combinación utilizando una inercia típica igual a 0,9, una población de 10 partículas y 100 iteraciones. Luego de las 5 pruebas se calcula el promedio, y la más alta representa la combinación óptima de las constantes.

En la Tabla 5-3 se aprecian los 5 mejores resultados se las 121 pruebas realizadas, donde se observa que el mejor desempeño se tiene cuando la constante de inteligencia individual φ_1 tiene un valor igual a 2, mientras que la constante de inteligencia social φ_2 tiene un valor igual a 1.6, lo que significa que el algoritmo tiene un funcionamiento eficiente cuando se tiene una presencia individual y social igual de alta. Es decir que las partículas se encuentran su equilibrio siguiendo un desempeño individual de acorde a sus mejores posiciones obtenidas y de igual forma se ven influenciada por la mejor partícula del cúmulo.

Tabla 5-3: Resumen de los 5 mejores desempeños en el ajuste de parámetros de las constantes de inteligencia del algoritmo PSO

φ_1	φ_2	Tasa reconocimiento prueba					Promedio (%)
		N°1	N°2	N°3	N°4	N°5	
2	1.6	97.60%	98.00%	97.60%	98.00%	97.60%	97.80%
2	0.6	97.20%	97.80%	97.40%	97.80%	97.60%	97.65%
2	0.2	96.80%	98.20%	97.20%	97.40%	97.60%	97.60%
1.6	1.2	97.00%	97.00%	97.40%	97.60%	97.40%	97.35%
1.2	1.2	97.80%	97.60%	97.80%	96.80%	97.80%	97.56%

Luego de obtener los parámetros de las constantes de inteligencia, se procede a ajustar la velocidad máxima que pueden alcanzar las partículas, lo cual se hará de manera similar a lo hecho anteriormente. Como ya se tienen los parámetros de las constantes de inteligencia cognitivas seleccionados, se procede a variar la velocidad máxima de las partículas, esto se hará variando la velocidad de 0,25 a 6, con un delta de 0,25 donde se aplicarán 5 pruebas y finalmente se calculará el promedio, el cual definirá la velocidad que tiene mejor rendimiento. Cabe destacar que no se recomienda la inicialización de la velocidad en 0, ya que existen mayores posibilidades de caer en óptimos locales [32] [38] [43]. En la Tabla 5-4 se muestran los 5 mejores resultados de esta prueba.

Tabla 5-4: Ajuste de parámetros para la velocidad máxima que pueden tomar las partículas del algoritmo PSO en el sistema propuesto

Vmax	Tasa reconocimiento prueba					Promedio (%)
	N°1	N°2	N°3	N°4	N°5	
0.25	98.00%	97.00%	96.60%	97.40%	97.80%	97.36%
0.50	97.40%	97.20%	97.00%	97.40%	97.20%	97.52%
0.75	96.20%	97.00%	96.80%	97.60%	96.40%	97.28%
1.00	97.80%	97.20%	97.80%	97.20%	97.80%	97.24%
1.75	97.40%	96.80%	97.60%	97.20%	97.40%	97.28%

Como se observa en la tabla anterior el mejor resultado se obtiene cuando la velocidad máxima es igual a 0.5, por lo que en cada una de las fusiones de descriptores desarrollados en este trabajo donde se usa el algoritmo PSO, la velocidad máxima será de 0.5. Nótese que los valores más altos en las tasas de reconocimiento se obtienen cuando el valor de la velocidad está entre el rango de valores que pueden tener las partículas es decir de 0 a 1. De igual forma a medida que la velocidad aumenta las tasas de reconocimiento se ven disminuidas.

Finalmente con los parámetros de los ratios de inteligencia y la velocidad se define la cantidad de partículas y el número de iteraciones necesarias para que el algoritmo PSO converja de forma adecuada. En la Tabla 5-5 se muestran distintas poblaciones de partículas y número de iteraciones para el algoritmo PSO destacado su mejor rendimiento, cuando se tiene un cúmulo de partículas de 30 partículas y un total de 300 iteraciones.

Tabla 5-5: Ajuste del número de partículas y el número de iteraciones que usará el algoritmo PSO-

N° Partículas	Número de Iteraciones				
	100	300	500	700	1000
20	97.60%	97.80%	97.80%	97.60%	97.20 %
30	97.60%	98.20%	97.80%	98.00%	97.80%
50	97.40%	98.00%	97.80%	97.08%	97.80%
100	97.60%	97.80 %	97.60%	97.80%	97.80%

Se observa además en la Tabla 5-5.que los valores son bastantes cercanos, no varían más que en una unidad, pero esto basta para que el algoritmo se vea afectado y baje sus rendimiento. La menor tasa es obtenida cuando se tienen 20 partículas y 1000 iteraciones, esto se debe a que el algoritmo se queda en óptimos locales con el paso de las iteraciones, por lo que no converge a un óptimo. Se observa además que con 300 iteraciones se obtiene valores altos en tasas de reconocimiento que van desde los 97.8% a los 98.2%, por lo que se da a entender que este es el valor apropiado para el método de fusión que se propone en este trabajo. Al usar mayor cantidad de partículas pasa lo mismo que al tener una gran cantidad de iteraciones, las partículas no puede moverse adecuadamente por el espacio solución por lo que caen en óptimos locales y no converge.

En la Figura 5-4 se aprecia la convergencia del algoritmo PSO, con todos los parámetros ajustados, las tasas de reconocimiento suben gradualmente en las primeras 20 iteraciones, luego se mantienen en una tasa de 98% constante hasta la iteración 170 donde, tiene su último aumento de tasa manteniéndose finalmente en 98.2%

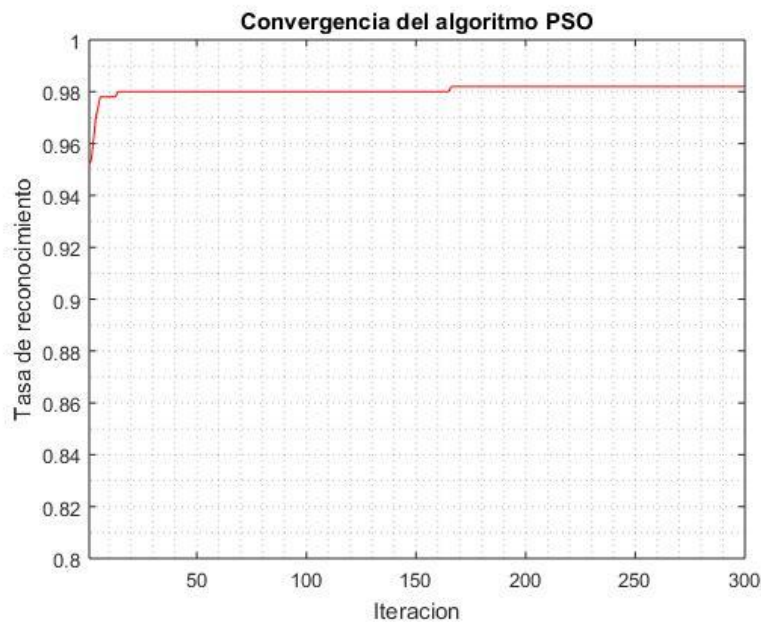


Figura 5-4: Convergencia del algoritmo PSO cuando se tiene un cúmulo de 30 partículas y 300 iteraciones

5.2.2 Entrenamiento utilizando distintos conjuntos

En esta sección se propone la utilización de distintos conjuntos de la base de datos Equinox para realizar la etapa de entrenamiento del sistema. En otros estudios [7] [28] se ha utilizado el conjunto VA y EA como galería, ya que estas son las que poseen mejor iluminación y se espera que con estos conjuntos obtengan mayores detalles del rostro a la hora de realizar el reconocimiento. Por la variabilidad de las imágenes que posee el conjunto RR se ha utilizado en estudios como conjunto de prueba o test, ya que posee imágenes de los sujetos sometidos distintas iluminaciones, distintas expresiones y además posee imágenes de los sujetos utilizando lentes. En la Figura 5-5 se aprecia los resultados obtenidos en [7], utilizando distintos métodos de reconocimiento a la hora de validar la base de datos Equinox. Se puede observar que los conjuntos que poseen el rendimiento más bajo es cuando los sujetos se encuentran utilizando lentes (EG y VG). De esta manera se propone tomar los conjuntos donde se obtienen los promedios más bajos de reconocimiento, para realizar el entrenamiento del sistema, ya que se espera que al usar el algoritmo PSO este busque las principales regiones del rostro, destacando lo más relevante para realizar el reconocimiento, debido a que el uso de anteojos significa pérdida de información en el área de los ojos, el algoritmo PSO debe ser capaz de buscar otras áreas que aporten información de tal manera que aumenten las tasas de reconocimiento del sistema, generando pesos óptimos con distintas características que los obtenidos al usar el conjunto de prueba RR, donde hay imágenes con distintas características como se mencionó anteriormente.

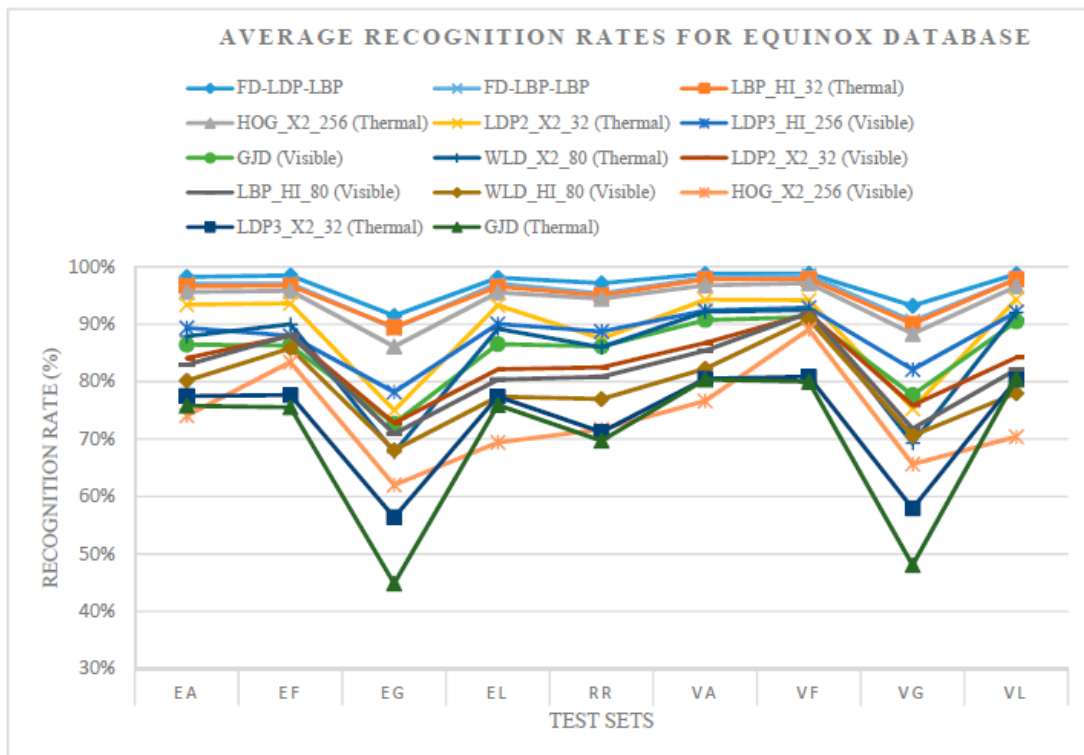


Figura 5-5: Tasas de reconocimiento promedio de la base de datos de Equinox. Diferentes conjuntos de pruebas contra todos los conjuntos de galería [7]

Para este experimento se aplicarán entonces cada una de las cinco metodologías de fusión que fueron propuestas, y cada una de ellas serán entrenadas con 4 diferentes conjuntos de la base de datos Equinox, en la Tabla 5-5 se observan los resultados del entrenamiento de cada conjuntos usando la nomenclatura “galería-test”, es decir que si se tiene por ejemplo el conjunto “EA-RR”, EA es la galería de imágenes del sistema y RR es el conjunto de entrenamiento. Los conjuntos utilizados serán los siguientes:

- EA-RR
- EF-EG
- EL-EG
- EL-VG

Para incluir mayor variabilidad de los pesos multiplicativos, se aplican pesos normalizados para cada una de los métodos y sus combinaciones con los distintos conjuntos de entrenamiento. La normalización de los pesos significa que el peso los pesos visibles-térmicos son complementarios es decir su suma debe ser igual a 1 en cada región.

Tabla 5-6: Resultados obtenidos para cada una de las metodologías de fusión propuestas aplicando diferentes conjuntos de entrenamiento de la base de datos Equinox, utilizando pesos aleatorios y complementarios.

Método de fusión	Tasas de reconocimiento para cada conjunto de entrenamiento (%).			
	EA-RR	EF-EG	EL-EG	EL-VG
FD-PSO-LBP-LBP-32	98.20	94.14	91.13	93.21
FD-PSO-LBP-LBP-32-N	98.00	94.14	91.36	93.82
FD-PSO-LBP-LBP-256	95.20	89.51	85.49	86.73
FD-PSO-LBP-LBP-256-N	98.00	89.81	86.11	87.96
FD-PSO-LBP-HOG-32	91.80	89.51	82.10	82.10
FD-PSO-LBP-HOG-32-N	92.60	90.12	79.63	84.26
FD--PSO-LBP-HOG-256	95.20	87.04	78.70	75.00
FD--PSO-LBP-HOG-256-N	94.80	85.18	76.85	75.00
FD-PSO-HOG-LBP-32	97.80	91.36	93.52	92.90
FD-PSO-HOG-LBP-32-N	97.60	93.59	93.83	93.83
FD-PSO-HOG-LBP-256	95.20	89.81	89.51	92.28
FD-PSO-HOG-LBP-256-N	96.40	89.20	88.27	91.67
FD-PSO-HOG-HOG-32	81.80	79.63	55.86	57.72
FD-PSO-HOG-HOG-32-N	83.40	75.61	51.23	52.16
FD-PSO-HOG-HOG-256	95.20	91.35	83.64	87.07
FD-PSO-HOG-HOG-256-N	95.40	91.05	82.41	84.26
FD-PSO-LDP-LBP	98.2	91.97	91.05	93.21
FD-PSO-LDP-LBP-N	98.60	92.59	91.04	93.21

En la Tabla 5-6 se destacan los métodos que tienen mayor tasa de reconocimiento para cada combinación de conjuntos de entrenamiento en los distintos métodos de fusión. De esta manera es posible descartar algunos métodos de fusión para la etapa de validación.

Como se menciona anteriormente, las tasas de reconocimiento para el entrenamiento donde los sujetos usan lentes posee tasas mucho más bajas, con respecto al conjunto EA-RR. Además se observa que la metodología con las peores tasas es FD-PSO-HOG-HOG-32, esto se debe a que HOG necesita mayor cantidad divisiones para funcionar de manera óptima, ya que así se tiene mayor detalle de las gradientes de los pixeles y de sus magnitudes, generando un histograma más detallado.

La mejor combinación es FD-PSO-LDP-LBP-N usando los conjuntos típicos de entrenamiento,” EA-RR, llegando a una tasa de reconocimiento de 98.6%. Se observa que algunos casos la aplicación de pesos normalizados aumenta la tasa de reconocimiento, el ejemplo más claro es el de FD-PSO-LDP-LBP que su tasa pasa de 98% a 98.6% subiendo un 0.6%. En otros casos las tasas se mantienen o algunas disminuyen, esto se debe a que cada descriptor funciona de forma distinta, y la combinación de ellos al ser con pesos complementarios, determina más aun la relevancia de un espectro por sobre el otro al realizar la fusión.

Tabla 5-7: Resumen de los mejores desempeños de fusión para las distintas combinaciones de fusión con sus respectivos conjuntos de entrenamiento, que serán validados con Equinox.

Método de fusión	Conjunto de entrenamiento	Tasa de reconocimiento en la etapa de entrenamiento (%)
FD-PSO-LBP-LBP-32	EA-RR	98.20
FD-PSO-LBP-LBP-32	EF-EG	94.14
FD-PSO-LBP-LBP-32-N	EL-EG	91.36
FD-PSO-LBP-LBP-32-N	EL-VG	93.82
FD-PSO-LBP-HOG-256	EA-RR	95.20
FD-PSO-LBP-HOG-32-N	EF-EG	90.12
FD-PSO-LBP-HOG-32	EL-EG	82.10
FD-PSO-LBP-HOG-32-N	EL-VG	84.26
FD-PSO-HOG-LBP-32	EA-RR	97.80
FD-PSO-HOG-LBP-32-N	EF-EG	93.59
FD-PSO-HOG-LBP-32	EL-EG	93.52
FD-PSO-HOG-LBP-32-N	EL-VG	93.83
FD-PSO-HOG-HOG-256	EA-RR	95.20
FD-PSO-HOG-HOG-256	EF-EG	91.35
FD-PSO-HOG-HOG-256	EL-EG	83.64
FD-PSO-HOG-HOG-256	EL-VG	87.07
FD-PSO-LDP-LBP-N	EA-RR	98.60
FD-PSO-LDP-LBP-N	EF-EG	92.59
FD-PSO-LDP-LBP	EL-EG	91.05
FD-PSO-LDP-LBP	EL-VG	93.21

En la Tabla 5-7 se muestra de forma resumida los mejores resultados para cada método de fusión propuesto para los distintos conjuntos utilizados, los cuales serán tomados para validarlos en la sección siguiente. Cada metodología de fusión será validada en cada una de sus combinaciones de conjuntos de entrenamiento donde se compara su desempeño con la aplicación de reconocimiento de rostros en cada espectro por separado. Finalmente se selecciona el conjunto de entrenamiento que tenga mejor desempeño para cada metodología y se comparan con la aplicación de fusión usando el algoritmo genético. Cabe destacar que los valores de las tasas de reconocimiento en la etapa de entrenamiento no es el resultado real del método propuesto, ya que los pesos se adaptan a los conjuntos con los que se está entrenando, pero estos pueden aumentar las tasas de igual forma al validar las bases de datos.

5.2.3 Experimentos para la etapa de validación

En la sección anterior se obtuvieron los pesos óptimos para la metodología de fusión propuesta, a pesar de obtener varios resultados con tasas sobre el 90%, esto no quiere decir nada, ya que solo nos dice que para esos conjuntos, los pesos aumentan las tasas de reconocimiento. Es por eso que para comprobar que la metodología es la adecuada se procede a la etapa de validación, donde los pesos óptimos obtenidos son aplicados a las bases de datos completas, que nos da el verdadero desempeño del sistema de reconocimiento de rostros propuesto.

5.2.4 Validación con la base de datos Equinox

La etapa de validación usará los pesos óptimos de cada una de las metodologías de fusión propuestas con sus cuatro conjuntos de entrenamiento y serán aplicados a la base de datos Equinox completa. A diferencia de los conjuntos de entrenamiento donde se tenían 500 imágenes al entrenar con RR y 324 imágenes al entrenar con los conjuntos con anteojos (VG y EG), en la validación se aplican los pesos a 4064 imágenes de entrenamiento con 6 galerías de 81 imágenes cada una. La validación de Equinox consiste en tomar cada conjunto galería establecidos anteriormente y compararlo con cada conjunto de prueba. Donde cada comparación entrega una tasa de reconocimiento que luego es promediada con las demás, dándonos así la tasa de reconocimiento global del sistema.

El análisis de la validación de Equinox se realiza para cada método de fusión propuesto en este trabajo, cada uno de ellos posee cuatro combinaciones de conjuntos de entrenamiento, donde se selecciona el que tenga mejor rendimiento, seguido de esto el que tenga mayor desempeño será comparado con la aplicación de reconocimiento de rostros en el espectro visible y térmico por separado y la con el desempeño de la fusión de descriptores usando el algoritmo genético.

Resulta interesante analizar los resultados de las distintas combinaciones cuando están sometidas a un entrenamiento con sujetos con lentes, ya que, en estas situaciones el promedio de reconocimiento suele ser más bajo, esto sucede debido a que en el área de los ojos se produce una pérdida de información cuando se extrae las características, produciendo que en el histogramas hayan valores más bajos en la representación, por lo que al realizar la intersección de los mismos se obtienen valores más pequeños de medidas de similitud produciendo una baja en las tasas de reconocimiento final.

Fusión de descriptores LBP-LBP

El primer método de fusión propuesto usa el descriptor LBP para el espectro visible y térmico, este algoritmo ha sido utilizado anteriormente en ambos espectros, por lo que su elección se basa en resultados previos encontrados en la literatura [2] [7] [11] [45]. Los resultados de la fusión de descriptores usando el algoritmo PSO pueden observarse en la Tabla 5-8 para cada una de las combinaciones de conjuntos de entrenamiento.

Tabla 5-8: Resultados de la validación de la base de datos Equinox usando el mejor método LBP-LBP para cada conjunto de entrenamiento.

Método	Conjunto de entrenamiento	TR de validación de base de datos Equinox (%)
FD-PSO-LBP-LBP-32	EA-RR	96.60
FD-PSO-LBP-LBP-32	EF-EG	96.25
FD-PSO-LBP-LBP-32-N	EL-EG	96.66
FD-PSO-LBP-LBP-32-N	EL-VG	96.79

El resultado más alto en LBP-LBP se obtiene al entrenar con el conjunto EL-VG, esto se debe a que los conjuntos con lentes, obligan al sistema a buscar otras regiones del rostro para tomar la información, evitando que los pesos más altos se encuentren en las regiones de los ojos. A continuación en la Tabla 5-9 se observa la comparación del método FD-PSO-LBP-LBP-32 con el algoritmo genético, y los espectros por separado.

Tabla 5-9: Comparativa del método de fusión usando PSO con GA y cada espectro por separado.

Método	Conjunto de entrenamiento	TR de entrenamiento (%)	TR de validación de base de datos Equinox (%)
LBP-visible-32	-	-	80.11
LBP-térmico-32	-	-	80.11
FD-PSO-LBP-LBP-32	EA-RR	98.20	96.60
FD-PSO-LBP-LBP-32-N	EL-VG	93.21	96.79
FD-GA-LBP-LBP-32 [7]	EA-RR	98.00	96.99
FD-GA-LBP-LBP-32	EL-VG	97.80	96.08

Se observa que el mejor entrenamiento se obtuvo a usar la metodología propuesta alcanzando una tasa de 98.20%, sin embargo al validar la base de datos Equinox, se observa que la fusión usando el algoritmo PSO con descriptores LBP-LBP, tiene mejor rendimiento al usar los conjuntos con lentes EL-VG, pasando de una tasa de entrenamiento de 93.21% de entrenamiento a 96.79%. El método de fusión propuesto aumenta las tasas con respecto a los espectros por separado, pero no así con respecto al algoritmo genético, donde se tiene una tasa de 96.99%. Las tasas de la fusión

de descriptores para ambas técnicas metaheurísticas son cercanas al 97%, haciendo de esta metodología bastante robusta.

Fusión de descriptores LBP-HOG

La segunda metodología de fusión usa el descriptor LBP para el espectro visible y HOG para el espectro térmico, estos son seleccionados con en base a lo expuesto en [13] [44], donde se aplica reconocimiento de rostros usando estos descriptores.

En la Tabla 5-10 presentan los resultados obtenidos al validar la base de datos Equinox utilizando la fusión de descriptores LBP-HOG donde de los mejores métodos obtenidos para cada conjunto de entrenamiento, se seleccionara el de mayor desempeño en Equinox y será comparado con GA, y los espectros por separado.

Tabla 5-10: Resultados de la validación de la base de datos Equinox usando el mejor método LBP-HOG para cada conjunto de entrenamiento.

Método	Conjunto de entrenamiento	Tasa de reconocimiento al validar Equinox
FD-PSO-LBP-HOG-256	EA-RR	92.52%
FD-PSO-LBP-HOG-32-N	EF-EG	75.88%
FD-PSO-LBP-HOG-32-N	EL-EG	75.88%
FD-PSO-LBP-HOG-32-N	EL-VG	77.53%

En esta metodología la combinación de conjuntos que tiene el desempeño más alto es EA-RR, superando por demasía al entrenamiento usando los conjuntos con lentes. En la siguiente tabla se compara el resultado obtenido con el desempeño de cada espectro por separado y la fusión aplicando el algoritmo genético.

Tabla 5-11: Comparativa del método de fusión usando PSO con GA y cada espectro por separado.

Método	Conjunto de entrenamiento	TR de entrenamiento	TR de validación de Equinox
LBP-visible-256	-	-	95.32%
HOG-térmico-256	-	-	89.52%
FD-PSO-LBP-HOG-256	EA-RR	95.20%	92.52%
FD-GA-LBP-HOG-256	EA-RR	94.00%	93.07%

En la Tabla 5-11 se observa que para el caso de la combinación LBP-HOG PSO tiene mejor rendimiento que GA, obteniendo una tasa de 92.52%. En esta combinación de descriptores sucede que no se supera el rendimiento de los espectros por separado, sobre esto es posible decir que al ser distintos descriptores, donde ambos funcionan de maneras distintas a la hora del cálculo de su histograma, esto podría provocar que la intersección de histogramas en el proceso

de “matching” se vea afectada, produciendo medidas de similitud bastante bajas. En otras palabras los pesos multiplicativos al aplicarlos en el vector de características produce que los espectros se anulen, en lugar de resaltar lo mejor de ellos.

Fusión de descriptores HOG-LBP

La siguiente metodología de fusión usa HOG y LBP, para el espectro visible y térmico respectivamente. Tomando los resultados del de los mejores entrenamientos para cada conjunto se tiene la Tabla 5-12.

Tabla 5-12: Resultados de la validación de la base de datos Equinox usando el mejor método HOG-LBP para cada conjunto de entrenamiento.

Método	Conjunto de entrenamiento	Tasa de reconocimiento al validar Equinox
FD-PSO-HOG-LBP-32	EA-RR	95.80%
FD-PSO- HOG-LBP-32-N	EF-EG	95.71%
FD-PSO- HOG-LBP-32-N	EL-EG	96.40%
FD-PSO- HOG-LBP-32-N	EL-VG	96.09%

Para la fusión usando HOG-LBP el mejor conjunto para entrenar es EL-EG esto se encuentra destacado en la Tabla 5-12. Esto quiere decir que al combinar los descriptores HOG para el espectro visible y LBP para el térmico, es posible tener un correcto entrenamiento de sujetos con lentes. Al igual que en los anteriores combinaciones en la Tabla 5-13 se realiza la comparación con el algoritmo genético y los espectros por separado.

Tabla 5-13: Comparativa del método de fusión usando PSO con GA y cada espectro por separado.

Método	Conjunto de entrenamiento	TR de entrenamiento	TR de validación de Equinox
LBP-visible-256	-	-	80.66%
HOG-térmico-256	-	-	89.52%
FD-PSO-HOG-LBP-32	EL-EG	95.20%	96.40%
FD-GA-HOG-LBP-32	EL-EG	93.83%	95.25%

En la Tabla 5-13, se observa que al realizar la fusión de descriptores HOG-LBP usando el algoritmo PSO aplicando un entrenamiento con conjunto donde los sujetos llevan lentes, se obtiene un mejor rendimiento que el algoritmo genético. Además el rendimiento con respecto al espectro visible aumenta casi un 16%, mientras que en el térmico alrededor de un 6%.

Fusión de descriptores HOG-HOG

Esta metodología usa el mismo descriptor para ambos espectros, los mejores desempeños para las combinaciones de conjuntos se dan cuando las imágenes visibles y térmicas son divididas en 256, esto se debe a que HOG, analiza con mayor detalle cada región al calcular la magnitud y la gradiente. En la Tabla 5-14, se muestran los resultados para la validación de Equinox usando la fusión de descriptores HOG-HOG

Tabla 5-14: Resultados obtenidos al validar la base de datos Equinox usando la fusión de descriptores HOG-HOG con cada uno de los conjuntos de entrenamientos correspondientes

Método	Conjunto de entrenamiento	Tasa de reconocimiento al validar Equinox
FD-PSO-HOG-HOG-256-N	EA-RR	92.51%
FD-PSO-HOG-HOG-256	EF-EG	95.08%
FD-PSO-HOG-HOG-256	EL-EG	93,07%
FD-PSO-HOG-HOG-256	EL-VG	93,14%

El mejor rendimiento de la fusión de descriptores HOG-HOG, donde se utiliza un conjunto de entrenamiento de sujetos usando lentes, al igual que el caso anterior, por lo que HOG, es el mejor descriptor para realizar entrenamientos de sujetos con lentes. Ahora en la siguiente tabla se realiza una comparativa del mejor método empleando HOG-HOG, con su equivalente en GA y el desempeño en el espectro visible y térmico por separado.

Tabla 5-15: Comparativa del método de fusión usando PSO con GA y cada espectro por separado.

Método	Conjunto de entrenamiento	TR de entrenamiento	TR de validación de Equinox
HOG-visible-256	-	-	80.66%
HOG-térmico-256	-	-	89.52%
FD-PSO-HOG-HOG-256	EL-EG	91.35%	95.08%
FD-GA-HOG-HOG-256	EL-EG	92.40%	93.40%

En el caso de la fusión HOG-HOG en el entrenamiento algoritmo genético posee mayor tasa de reconocimiento, pero al validar con los pesos óptimos la fusión usando PSO posee mayor tasa, la validación es el resultado verdaderamente importante ya que aquí se obtiene un promedio que involucra una mayor cantidad de imágenes donde los sujetos tienen distintas expresiones e iluminaciones, además de emular el sistema en su comportamiento en tiempo real.

Fusión de descriptores LDP-LBP

Esta última metodología de fusión de descriptores tiene una particularidad, ya que se usan distintas divisiones en las imágenes visibles y térmicas. Las imágenes visibles son divididas en 256

regiones al aplicar el descriptor LDP, provocando un histograma con mayor detalle, mientras que las imágenes térmicas son divididas en 32 regiones cuando se realiza el cálculo del histograma LBP. Para realizar entonces el proceso de matching, al no tener el mismo tamaño los histogramas se debe realizar un arreglo, donde las regiones del descriptor LDP son agrupadas en 8 regiones, dejando el histograma del mismo tamaño que el térmico. En la Tabla 5-16 se observan los resultados para las distintas combinaciones de conjuntos de entrenamiento de esta metodología de fusión en la tapa de validación usando la base de datos Equinox.

Tabla 5-16: Resultados obtenidos al validar la base de datos Equinox usando la fusión de descriptores LDP-LBP con cada uno de los conjuntos de entrenamientos correspondientes

Método	Conjunto de entrenamiento	Tasa de reconocimiento al validar Equinox
FD-PSO-LDP-LBP-N	EA-RR	98.02
FD-PSO-LDP-LBP-N	EF-EG	96.71
FD-PSO-LDP-LBP	EL-EG	95.98
FD-PSO-LDP-LBP	EL-VG	95.92

Es observa en la Tabla 5-16 se aprecia que se obtiene la mayor tasa de reconocimiento al utilizar el método FD.PSO-LDP-LBP-N, alcanzando una tasa de reconocimiento de 98.02%, siendo esta la tasa más alta de todos los conjuntos que se han estudiado. Este resultado se obtiene al entrenar con dos conjuntos distintos “EA-RR” y “EL-VG” lo cual resulta curioso, ya que son conjuntos con características bastante distintas.

Tabla 5-17: Comparativa del método de fusión usando PSO con GA y cada espectro por separado

Método	Conjunto de entrenamiento	TR de entrenamiento	TR de validación de Equinox
LDP-visible-256	-	-	88.20%
LBP-térmico-32	-	-	95.32%
FD-PSO-LDP-LBP-N	EA-RR	98.60%	98.02%
FD-GA-LDP-LBP	EA-RR	98.00%	96.99

De todas las metodologías estudiadas para la fusión de descriptores visibles y térmicos la que tiene mayor rendimiento es FD-PSO-LDP-LBP-N, este se caracteriza porque usa pesos complementarios, por lo que las influencia de un espectro por sobre otro, o la combinación de ambas será altamente notoria. De los experimentos realizados, se puede decir que PSO tiene una mayor tendencia a tener buenos resultados de validación, sin embargo esto no sucedió en la fusión de tipo LBP-LBP, donde genético tuvo un mejor desempeño, y tampoco el la fusión LBP-HOG, donde el algoritmo de fusión no es capaz de superar el desempeño del espectro térmico por sí solo. La obtención de los pesos óptimos va a depender enormemente del descriptor al cual se le está aplicando la fusión, as regiones en las cuales fueron divididos las imágenes y finalmente los conjuntos de entrenamiento utilizados.

En la Figura 5-6 se observan los pesos resultantes a la fusión FD-PSO-LDP-LBP-N que es la que obtuvo mayor rendimiento en Equinox. Donde el color azul representa los pesos visibles y el rojo el espectro térmico, se observa el carácter de los pesos complementarios, donde el espectro térmico es el “espejo” del espectro visible. Se observa además que en los extremos es decir las primeras 10 regiones tienen pesos que tienden a los extremos, es decir 0 y 1. Mientras que en las regiones más centrales se observan, pesos con medios, donde en esas áreas se aprovechan ambos espectros, para el aumento de las tasas.

Cabe destacar que los pesos no multiplican las imágenes en sí, sino que multiplican la comparación HI del histograma de cada una de las regiones, por lo que a mayor peso en ciertas áreas solo nos dirá que esa región es más abundante en información, por lo que los pesos representan en verdad, es la relevancia de un espectro por sobre otro.

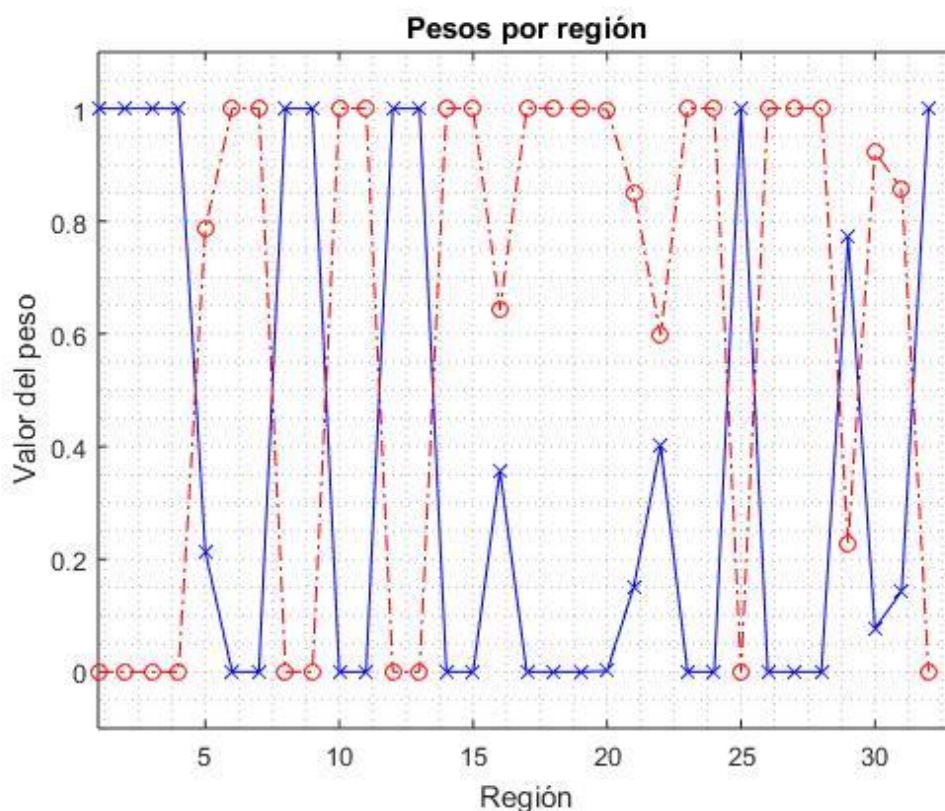


Figura 5-6: Representación de los mejores pesos visibles y térmicos obtenidos al validar la base de datos Equinox, donde se usa la fusión FD-PSO-LDP-LBP-N

Ya que la fusión empleada no entrega una imagen resultante, es posible visualizar la relevancia de los pesos óptimos en cada región, para esto se crea una imagen RGB, donde el canal verde es llevado a 0 quedando solo a disposición el canal azul y el canal rojo. Al canal azul se le asignan los pesos visibles, dando valores proporcionales al rango [0,255] de dicha imagen, el canal rojo es asignado con los pesos térmicos con el mismos valores proporcionales en un rango de [0,255]. Los colores generados como el morado o el café, es una mezcla de visible y térmico. Así la representación de la nos muestra la relevancia de cada espectro en cada una de las regiones del esquema de fusión FD-PSO-LDP-LBP-N, que es la que tiene mejor resultado usando Equinox.

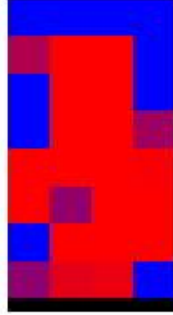


Figura 5-7: Representación de los pesos óptimos, mostrando la relevancia de un espectro por sobre otro usando una imagen RGB. Estos pesos son el promedio de todas las imágenes involucradas en la etapa de entrenamiento del sistema.

Observando la Figura 5-7, se deduce la importancia que tiene PSO, ya que este le asigna relevancia a las distintas áreas de la imagen. Teniendo en cuenta la naturaleza de la intersección de histogramas en la que se suma el mínimo valor del histograma de prueba comparado con el de galería, se puede decir que al disminuir el valor de similitud dado por el $HI()$, significa también disminuir la importancia de tal valor al ser comparado con los demás valores de similitud obtenidos en otras comparaciones. Sin embargo también se puede decir que una región con un peso bajo puede significar una gran similitud entre distintas imágenes, así como una región con peso alto signifique una gran variabilidad entre las distintas imágenes que conforman el conjunto. Por lo cual al analizar la ilustración se observa que el espectro térmico predomina en las zonas centrales del rostro, mientras que el espectro visible funciona más bien como un descriptor de bordes. Esta representación muestra los pesos óptimos obtenidos en el entrenamiento que luego fueron validados, estos representan el promedio de todas las imágenes presentes en la base de datos, no a una sola imagen.

Tabla 5-18: Resumen de los promedios de la tasa de reconocimiento para los mejores métodos de fusión propuestas en comparación con el método de fusión utilizando GA [7] para la base de datos Equinox.

Método de Fusión	Conjuntos de entrenamiento	TR de entrenamiento (%)	TR al validar la base de datos Equinox complete (%)
FD-PSO-LBP-LBP-32-N	EL-VG	98.20	96.79
FD-PSO-LBP-HOG-256	EA-RR	95.20	92.52
FD-PSO-HOG-LBP-32	EL-EG	93.52	96.40
FD-PSO-HOG-HOG-256	EL-VG	87.07	93.14
FD-PSO-LDP-LBP-N	EA-RR	93.21	98.02
FD-GA-LBP-LBP [7]	EA-RR	98.00	95.64
FD-GA-LDP-LBP [7]	EA-RR	97.80	96.99
FD-GA-LBP-LBP	EL-VG	97.80	96.81
FD-GA-LBP-HOG	EA-RR	94.00	92.52
FD-GA-HOG-LBP	EL-EG	93.83	95.25
FD-GA-HOG-HOG	EL-VG	92.40	95.08

En la Tabla 5-18 se muestra de manera resumida los mejores métodos de fusión propuestos en comparativa con la utilización del algoritmo genético para la fusión donde se incluyen los resultados de [7]. Se observa que para el caso de la fusión LDP-LBP se obtiene la tasa de reconocimiento de Equinox contra todos los demás métodos, superando la fusión GA. Sin embargo al comparar directamente la fusión LBP-LBP entre GA y PSO, se observa que GA tiene un desempeño 0.20% más alto, por lo que se puede decir que GA tiene un mejor desempeño en la fusión de tipo LBP-LBP. La tasa más baja se da en LBP-HOG que como se ha mencionado en los análisis anteriores es la fusión con desempeño más bajo.

5.2.5 Validación con PUCV-VTF

El principal objetivo de aplicar los pesos óptimos obtenidos a una base de datos distinta, es comprobar la robustez de las distintas metodologías de fusión, y así validar los resultados obtenidos. De esta forma, se evalúan los distintos métodos de estudio usando PUCV-VTF.

De igual forma que la base de datos Equinox, para validar la base de datos PUCV-VTF, se debe calcular el promedio de tasa de reconocimiento, de cada galería con cada conjunto de test. En este caso se usa solo una galería y 4 conjuntos de test. La galería será el conjunto normal (Nml), mientras que los conjuntos de prueba serán Frown (Frn), Glasses (Gls), Smile (Sml) y Vocal (Vcl). Como ya se acordó en el sistema propuesto la medida de similitud utilizada para este sistema es la intersección de histogramas.

Para realizar un análisis más completo de los resultados se muestran las 4 combinaciones de fusión de descriptores que se proponen en este trabajo, en comparación con el desempeño de la validación usando los espectros por separado. Además se incluye el desempeño de la fusión de descriptores usando el algoritmo genético que se plantea en [7], esperando que el desempeño de la metodología usando PSO, se comporte de mejor manera.

Fusión de descriptores LBP-LBP

Primeramente se analizara el desempeño final de la fusión de descriptores LBP –LBP, donde los resultados puedes observarse en la

Tabla 5-19: Resultados obtenidos de la validación de la base de datos PUCV-VTF aplicando la fusión de descriptores LBP-LBP con distintos conjuntos de entrenamiento.

Método	Conjunto de entrenamiento	TR de validación de base de datos PUCV-VTF (%)
LBP-visible-32	-	96.05
LBP-térmico-32	-	93.75
FD-PSO-LBP-LBP-32	EL-VG	99.34
FD-GA-LBP-LBP-32	EL-VG	97.04

En la Tabla 5-19 se aprecia el mejor resultado al validar la base de datos PUCV-VTF aplicando la fusión de descriptores visibles y térmicos LBP-LBP, donde algoritmo FD-PSO-LBP-LBP-32 tiene el mejor resultado. A diferencia de Equinox, aquí las tasas de las fusiones tienden a ser mucho más altas, donde se encuentran valores de 99%, haciendo de este descriptor bastante óptimo para la base de datos PUCV-VTF. El desempeño de FD-GA-LBP-LBP-32 es 2.3% menor que el obtenido por la fusión usando PSO, por lo que este último está muy cerca del ideal que es el 100% de reconocimiento.

Fusión de descriptores LBP-HOG

A continuación se presentan los resultados obtenidos al validar la base de datos PUCV- VTF utilizando la fusión de descriptores LBP-HOG y se compara su desempeño con el desempeño de los espectros por separado.

Tabla 5-20: Resultados de la validación de la base de datos PUCV-VTF usando la fusión de descriptores LBP-HOG aplicando distintos conjuntos de entrenamiento.

Método	Conjunto de entrenamiento	TR de validación de PUCV-VTF (%)
LBP-visible-256	-	98.36
HOG-térmico-256	-	95.39
FD-PSO-LBP-HOG-256	EA-RR	98.35
FD-GA-LBP-HOG-256	EA-RR	98.03

Al igual que la validación de Equinox esta metodología de fusión no logra superar las tasas de los espectros por separada, siendo la única que falla en el objetivo de este proyecto.

Fusión de descriptores HOG-LBP

Siguiendo con la validación de la base de datos PUCV-VTF, se procede a evaluar el desempeño del esquema de fusión HOG-LBP.

Tabla 5-21: Resultados obtenidos al validar la base de datos PUCV-VTF usando la fusión de descriptores HOG-LBP con cada uno de los conjuntos de entrenamientos correspondientes.

Método	Conjunto de entrenamiento	TR de validación de PUCV-VTF (%)
HOG-visible-32	-	55.21
LBP-térmico-32	-	93.75
FD-PSO-HOG-LBP-32-N	EL-EG	99.34
FD-GA-HOG-LBP-32	EL-EG	98.68

En la Tabla 5-21 se aprecia que en esta metodología de fusión cumple el objetivo de maximizar las tasas e los espectros por separado, donde su máximo rendimiento se obtiene usando FD-PSO-HOG-LBP-32 con el conjunto de entrenamiento El-EG. A diferencia de la propuesta de fusión

LBP-HOG, este diagrama de fusión no se anula disminuyendo sus tasas. Se infiere que se debe a que LBP es un buen descriptor en el espectro térmico por lo que su presencia gracias a los pesos óptimos es mayor.

Fusión de descriptores HOG-HOG

Continuando con las distintas combinaciones de descriptores del método de fusión propuesto, ahora tenemos nuevamente al descriptor HOG, este se desempeña tanto en el espectro visible como en el espectro térmico.

Tabla 5-22: Resultados obtenidos al validar la base de datos PUCV-VTF usando la fusión de descriptores HOG-HOG con cada uno de los conjuntos de entrenamientos correspondientes.

Método	Conjunto de entrenamiento	TR de validación de PUCV-VTF (%)
HOG-visible-256	-	88.81
HOG-térmico-256	-	82.89
FD-PSO-HOG-HOG-256	EL-EG	98.02
FD-GA-HOG-HOG-256	EL-EG	97.04

En la Tabla 5-22 se observa que el descriptor HOG tiene un mejor desempeño en general cuando se aplica la fusión de descriptores con una división de 256 regiones. Nuevamente se supera el rendimiento de los espectros por separado con casi un 10 % para el espectro visible y casi un 15 % por sobre el espectro térmico, además nuevamente la fusión usando PSO es mucho mayor que GA, por aproximadamente 5%.

Fusion de descriptores LDP-LBP

Finalmente se tiene en la fusión de descriptores LDP-LBP, que es el que tuvo mejor resultado usando la base de datos Equinox usando dos conjuntos de entrenamiento distintos.

Tabla 5-23: Resultados obtenidos al validar la base de datos PUCV-VTF usando la fusión de descriptores LDP-LBP con cada uno de los conjuntos de entrenamientos correspondientes.

Método	Conjunto de entrenamiento	TR de entrenamiento	TR de validación de PUCV-VTF (%)
LDP-visible-256	-	-	84.54
LBP-térmico-32	-	-	93.75
FD-PSO-LDP-LBP-N	EA-RR	98.60%	98.03
FD-GA-LDP-LBP	EA-RR	98.00%	98.68

En la Tabla 5-23 la que para el caso de la base de datos PUCV-VTF la fusión LDP-LBP baja su resultado con respecto a las otras técnicas como FD-PSO-LBP-LBP-32, quien obtiene un 99.34% de tasa de reconocimiento. Estos se debe a que la base datos PUCV-VTF sea más susceptible al

ruido, lo que provoca al realizar las derivadas en grado de 3 del descriptor LDP se pierda información dificultando así el reconocimiento. De igual forma se aprecia que la baja más considerable se da al entrenar con sujetos con lentes, ya que al aplicar derivadas direccionales mucha más pérdida de información hay en la parte de los ojos de las imágenes. En este caso no la fusión usando GA obtiene resultados más altos que PSO.

Para la validación de la base de datos PUCV-VTF el mejor resultado fue obtenido al utilizar la fusión FD-PSO-LBP-LBP-32 aplicando los conjuntos de entrenamiento EA-RR. En la Figura 5-9 se aprecia el comportamiento de los pesos óptimos que permiten que la tasa de esta fusión sea la más alta.

Se observa que a diferencia de la validación en Equinox, los pesos en la Figura 5-8 son completamente aleatorios, y no siguen orden alguno. De igual forma se destaca que en los valores de los extremos es decir las primeras y últimas 10 regiones tienen a tener valores extremos es decir que sus valores son 0 o 1, lo que hace que el aporte de la región en cada espectro sea absoluto o nulo. Por otro lado hay varias regiones en las cuales tanto el espectro visible y térmico tienen valor 0, esto representa que esa región no aporta ninguna información importante en el desempeño del sistema.

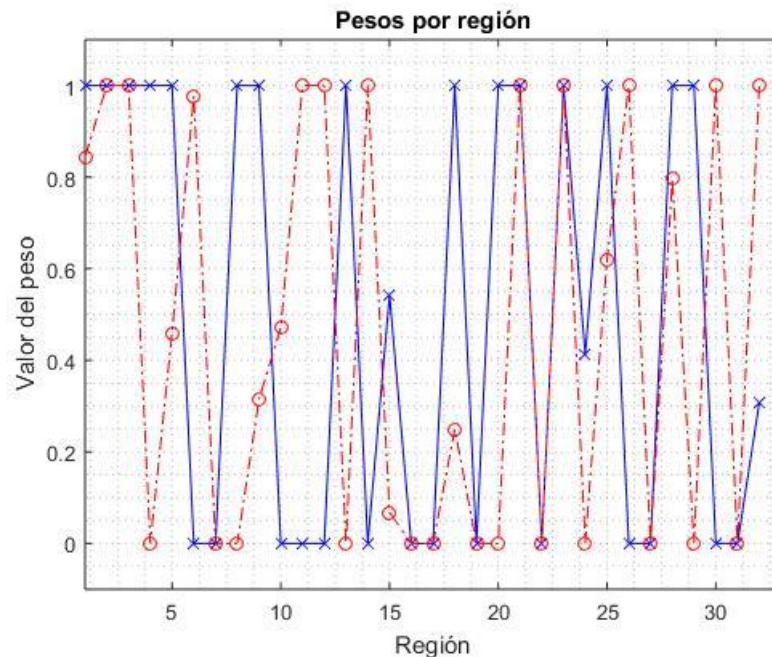


Figura 5-8: Esquema de los pesos óptimos obtenidos usando la fusión de descriptores FD-PSO-LBP-LBP-32.

De igual forma que en Equinox se visualiza la relevancia de cada espectro, usando la imagen RGB, esta es mostrada en la Figura 5-9 donde aprecia una distribución menos uniforme de colores, donde los pesos tienen a tener valores, extremos, combinaciones o valores nulos. En rojo se aprecia la relevancia del espectro térmico, que podría identificarse como el área de la frente y mentón, en azul es espectro visible, donde este destaca los bordes de los rostros. En morado una combinación de ambos espectros, y finalmente en negro se aprecia nula información, es decir

son zonas que el algoritmo PSO identificó como irrelevantes para la maximización de las tasas de reconocimiento.

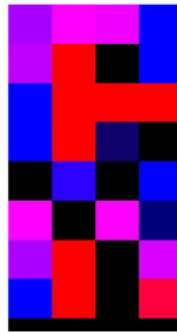


Figura 5-9: Representación de los pesos óptimos, mostrando la relevancia de un espectro por sobre otro usando una imagen RGB, utilizando los pesos con mejores resultados al validar PUCV-VTF. Estos pesos son el promedio de todas las imágenes involucradas en la etapa de entrenamiento del sistema.

Finalmente se plantea la Tabla 5-24 la cual presenta una comparativa de todos los métodos propuestos a lo largo de este trabajo, además comparándose directamente con los resultados que se obtienen en [7].

Tabla 5-24: Resumen de las mejores metodologías al validar la base de datos PUCV-VTF para cada uno de los métodos de fusión propuestos en comparativa con la aplicación de fusión usando GA [7].

Método de fusión	Conjuntos de entrenamiento	TR al validar la base de datos PUCV-VTF (%)
FD-PSO-LBP-LBP-32-N	EL-VG	99.34
FD-PSO-LBP-HOG-256	EA-RR	98.35
FD-PSO-HOG-LBP-32	EL-EG	99.34
FD-PSO-HOG-HOG-256	EL-VG	98.68
FD-PSO-LDP-LBP-N	EA-RR	98.03
FD-GA-LBP-LBP [7]	EA-RR	99.01
FD-GA-LDP-LBP [7]	EA-RR	98.68
FD-GA-LBP-LBP	EL-VG	97.04
FD-GA-LBP-HOG	EA-RR	98.03
FD-PSO-HOG-LBP	EL-EG	98.68
FD-PSO-HOG-HOG	EL-VG	97.04

En la Tabla 5-24 se observan los mejores métodos de fusión para cada combinación de descriptores en comparativa con lo desarrollado en [7], estos se encuentran destacados. Se puede apreciar que en ambas bases de datos PSO tiene un mejor desempeño, pero si analizamos directamente la combinación de descriptores presentes en ambos estudios es decir FD-LDP-LBP y FD-LBP-LBP, se tiene que el primero en el caso de Equinox sigue siendo mejor PSO, sin embargo para PUCV-VTF, es genético quien tiene un mejor desempeño. Para el segundo se sucede lo mismo, que el anterior. Esto puede deberse a que Equinox tiene mayor variabilidad de imágenes

y por ende mayor cantidad de imágenes, por lo que el algoritmo genético, posee se adapta de mejor manera a los conjuntos con los cuales se está entrenando, mientras que en PSO los pesos tienden a tener un carácter más general, debido a que en cada iteración se van modificando todos los conjuntos involucrados, mientras que en genético la mutación y el cruce solo modifican los pesos que tienen una menor tasa de reconocimiento.

5.3 Conclusiones del capítulo

En este capítulo se mostraron los resultados obtenidos al aplicar la fusión de descriptores aplicando PSO y GA, donde para obtener sus resultados se usan 2 bases de datos: Equinox y PUCV-VTF. Estas bases de datos están divididas en conjuntos de pares de imágenes térmicas y visibles, cada uno de estos conjuntos posee distintas características, como lo son la variabilidad de la iluminación y las expresiones de los sujetos.

En este trabajo se utilizan conjuntos de la base de datos Equinox para la etapa de entrenamiento del sistema que se propone, la cual tiene como objetivo encontrar los pesos óptimos para cada región en que las imágenes sean divididas con el fin de maximizar la tasa de reconocimiento del sistema.

Con los pesos óptimos obtenidos se procede a validar los resultados, aplicando los pesos óptimos encontrados a toda la base de datos Equinox y a PUCV-VTF. Esta etapa es llamada etapa de validación. De acuerdo a esto se realizan distintos experimentos de los cuales se concluye lo siguiente:

Respecto al entrenamiento con distintos conjuntos de la base de datos Equinox, se puede decir que se esperaba que el desempeño de los conjuntos con lentes no fuera el máximo, pero sí, que subieran en algún aspecto. El mejor rendimiento en general de todos los conjuntos aplicados y metodologías de fusión usadas fue FD-PSO-LBP-LBP-32 con una tasa de 98.2% superando lo hecho en [7] donde, con el algoritmo genético se obtienen entrenamientos de 98%. Se debe destacar que el entrenamiento solo define los parámetros necesarios para el conjunto de imágenes que se está utilizando, por lo que no significa necesariamente que esto se vea reflejado en la validación, que en fin es el resultado real del sistema.

El entrenamiento con pesos complementarios para los diferentes conjuntos, es un experimento que se realiza para saber si las tasas aumentarían, dejando una norma estándar para los pesos. Para este caso la única mejora significativa la obtuvo la fusión FD-PSO-LDP-LBP-N, la cual alcanza la tasa más alta de todos los entrenamientos con 98.60%.

La validación en las bases de datos, en general tienen bastantes buenos resultados, cumpliendo con el objetivo de que el sistema propuesto tenga tasas más altas que las que se tienen al realizar reconocimiento en los espectros por separado. Curiosamente en el caso de Equinox el mejor rendimiento a la hora de entrenar, también obtuvo la mejor tasa de reconocimiento en la validación dando 98.02%. Este resultado se obtiene usando dos conjuntos de entrenamiento distintos EA-RR y EL-VG, en este punto se cumple la hipótesis que el sistema debiera mejorar, si se aumentan las tasas de los peores rendimientos en validación. Puede ser que varias de las

imágenes dentro de “dando” sean procedentes del conjunto EL-VG, haciendo que el sistema se comporte de una manera similar.

La validación usando la base de datos PUCV-VTF entrega los resultados más altos con una tasa de reconocimiento de 99.67% donde se usa FD-PSO-LBP-LBP-32 con los conjuntos de entrenamiento EA-RR. Este rendimiento es casi el ideal que sería obtener una tasa del 100%.

En este estudio se aplican distintas metodologías de fusión entre ellas la aplicación del algoritmo genético como medio de comparación del método propuesto usando PSO. Analizando los resultados obtenidos en la validación de ambas bases de datos, se puede decir que el rendimiento de PSO es superior al de genético aunque no en todos los casos, en la mayoría de los métodos así lo es. La otra desventaja del algoritmo genético es que para generar el código óptimo de pesos tarda alrededor de 3 horas mientras que PSO solo tarda 10 minutos.

Discusión y conclusiones

En este trabajo de tesis se estudió el funcionamiento de los sistemas de reconocimiento de rostros visibles y térmicos, entendiendo que estos poseen algunas limitaciones. El espectro visible como principal limitación tiene la dependencia a la luz y a las variaciones de las expresiones faciales, mientras que el espectro térmico tiene como limitaciones, la variación de la temperatura corporal, la variación de la captura de la imagen cuando la cámara trabaja largos periodos de tiempo y la oclusión. Con el fin de mejorar los bajos rendimientos dados por estas limitaciones se busca en la literatura, distintas metodologías de fusión de imágenes visibles y térmicas, donde se hay variadas metodologías donde se destaca, la utilización de técnicas metaheurísticas como herramienta de fusión [7], tomando estas referencias se crea un sistema de reconocimiento de rostros donde se utiliza la fusión de imágenes visibles y térmicas usando metaheurísticas, donde se selecciona el algoritmo PSO. Pero antes de desarrollar el sistema es necesario entender los procesos que este conlleva.

Primeramente se realiza un estudio de algunos métodos de reconocimientos de rostros existentes de los cuales se seleccionan 3, tomando como referencia otras investigaciones [7][11][28][33][44], estos son: Local Binary Pattern (LBP), Local Derivative Pattern (LDP) y Histogram of Oriented Gradients (HOG), que también son llamados descriptores de patrones locales. Seguido de esto se estudian las distintas metodologías de fusión, que puede realizarse a distintos niveles: pixel, características, puntaje y decisión, donde se selecciona la fusión a nivel de características. Finalmente es necesario entender el algoritmo de optimización utilizado para el sistema, este es el algoritmo PSO, que forma parte de los algoritmos bioinspirados basados en el comportamiento de las bandadas de aves en búsqueda de comida, de este algoritmo se desprende la importancia de sus constantes de inteligencia que determinan el comportamiento del algoritmo, por lo que la elección correcta de estos valores garantiza una convergencia óptima.

El sistema de reconocimiento de rostros propuesto busca fusionar los descriptores aplicados a las imágenes térmicas y visibles, asignando pesos multiplicativos a cada región en la que los descriptores son divididos que concuerdan con los regiones en las cuales las imágenes son divididas al aplicar el descriptor. Estos pesos son creados de forma aleatoria, pero posteriormente son optimizados por el algoritmo PSO que permite aumentar la tasa de reconocimiento. A modo comparativo con [7] se aplica también el algoritmo genético usando misma lógica.

El sistema consta de dos etapas: etapa de entrenamiento y etapa validación. La primera consiste en utilizar conjuntos de la base de datos Equinox de tal manera que mediante el algoritmo PSO o el GA se encuentre un vector de pesos óptimos que maximicen las tasas de reconocimiento. Para la etapa de validación se toman estos pesos óptimos y son aplicados a la base de datos Equinox completa. De la misma manera, se usan los pesos óptimos en la base datos PUCV-VTF y así realizar una comparativa de los métodos.

Con los datos obtenidos en este estudio mediante los distintos experimentos que involucraron distintos métodos, es posible concluir:

- El algoritmo PSO necesita de un ajuste de parámetros para poder funcionar manera óptima, es necesario definir los valores de los ratios de inteligencia individual y social para que las partículas puedan moverse libremente por el espacio solución sin caer en óptimos locales.
- La fusión de descriptores de imágenes visibles y térmicas de distintos individuos mediante la aplicación de pesos multiplicativos correspondientes a regiones de los descriptores obtenidos de la divisiones de las imágenes , que son optimizados mediante metaheurísticas, permite aumentar las tasas de reconocimiento otorgando grados de relevancia a cada región de las imágenes visibles y térmicas.
- El entrenamiento es el proceso más importante del sistema propuesto ya que son los pesos quienes determinan la relevancia de los espectros para realizar la fusión y aumentar las tasas de reconocimiento, aunque se debe destacar que un entrenamiento con una tasa de reconocimiento alta, no siempre se va a ver reflejado en la validación, esto se debe a que los pesos se adaptan al conjunto con el que se está entrenando, y se espera que estos se ajusten de alguna forma al sistema completo.
- El entrenamiento con distintos conjuntos de sujetos en distintos gestos e iluminación nos ayuda a que el sistema sea más robusto y se acerque al comportamiento de un sistema que funciona en tiempo real. En este estudio se plantea utilizar conjuntos de sujetos con lentes intentando superar la oclusión que estos producen. Esto ayudó al rendimiento del sistema en la validación ya que usando la fusión FD-PSO-HOG-HOG-256 entrenado con EF-EG se obtiene una tasa de reconocimiento de 99.01% en la validación de PUCV-VTF. Del mismo modo usando FD-PSO-HOG-LBP-32 entrenado con EL-VG se obtiene una tasa de 99.34% en PUCV-VTF, siendo una de las tasas más altas.
- En todos los casos el sistema propuesto aumenta el rendimiento de las tasas de reconocimiento con respecto a las tasas obtenidas en la aplicación de los espectros por separado, exceptuando la fusión de tipo LBP-HOG, ya que esta no logra superar el rendimiento del espectro térmico, en ninguna de las dos bases de datos, haciéndolo de la fusión con menor rendimiento y aquella que no cumple el objetivo del proyecto.
- La aplicación de pesos complementarios permite aumentar el rendimiento de la fusión FD-PSO-LDP-LBP-N, esto se ve reflejado en la validación de la base de datos Equinox donde obtiene el mejor resultado con una tasa de 98.02%. Sin embargo esto no se ve reflejado en la base de datos PUCV-VTF, esto se debe principal mente a que LDP, no es susceptible al ruido presentes en PUCV-VTF, lo que hace que su rendimiento baje, por lo cual hay que tener cuidado, en su aplicación en distintas bases de datos.

- En general PSO presenta un mejor rendimiento en la metodología de fusión que el algoritmo genético, además que el proceso de obtención de los pesos óptimos de fusión demoran menos tiempo. Donde genético demora 3 horas 45 minutos en obtener el código, mientras que PSO lo hace en 10 minutos.
- El método más robusto es FD-PSO-LBP-LBP-32 ya que presenta altas tasas en ambas bases de datos, siendo un método que soporta el ruido de las imágenes. Su resultado en Equinox nos entrega una tasa de 96.9% y en PUCV-VTF 99.60% siendo la tasa más alta obtenida por el sistema acercándose al ideal que sería el 100% de tasa de reconocimiento.

Trabajos futuros.

Para posteriores trabajos, se propone utilizar otra técnica metaheurística como el ACO (Ant Colony Optimization y también otros descriptores de patrones locales como WLD (Weber Local Descriptor) y GJD (Gabor Jet Descriptor).

Utilizando las imágenes térmicas es posible obtener la información vascular del rostro, por lo que resulta interesante y desafiante aplicar fusión de imágenes visibles, térmicas y la red vascular, así se podrán obtener descriptores más específicos de cada persona, lo que podría aumentar las tasas de reconocimiento.

Bibliografía

- [1] J. Unar, W. C. Seng y A. Abbasi, «A review of biometric technology,» *Pattern Recogniton*,, 2013.
- [2] M. Akhloufi, A. Bendada y J.-C. Batsale, «State of the art in infrared face recognition,» *Quantitative InfraRed Thermography Journal*, vol. 5, 2008.
- [3] D. Socolinsky y A. Selinger, «Thermal face recognition over time,» *Patter Recognition*, vol. 4, pp. 187-190, 2004.
- [4] B. Melián, J. A. M. Pérez y J. M. M. Vega, «Metaheurísticas: una visión global,» *Revista Iberoamericana de Inteligencia Artificial*, nº 19, pp. 7-28, 2003.
- [5] F. Glover, «Future paths for integer programming,» *Computers and Operations Research*, vol. 5, 1986.
- [6] E. Alba, M. Laguna y R. Martí, «Métodos evolutivos en problemas de optimización,» *REVISTA INGENIERÍA UC*, vol. 10, nº 3, pp. 80-89, 2003.
- [7] G. Hermosilla, Gallardo, F., G. Farias y C. S. Martin, «Fusion of visible and thermal descriptors using genetic algorithms for face recognition systems,» *Sensors*, vol. 5, nº 8, pp. 17944-17962., 2015.
- [8] J. Kennedy, «The Particle Swarm: Social Adaptation of Knowledge,» *IEEE International Conference on Evolutionary Computation*, pp. 303-308, 1997.
- [9] T. Khadhraoui, S. Ktata, F. Benzarti y H. Amiri, «Features Selection based on Modified PSO Algorithm for 2D Face Recognition,» de *13th International Conference Computer Graphics, Imaging and Visualization*, 2016.

-
- [10] S. Wang, S. He, Y. Wu, M. He y Q. Ji, «Fusion of visible and thermal images for facial expression recognition,» *Frontiers of Computer Science*, vol. 8, nº 2, pp. 232-242, 2014.
- [11] T. Ahonen, A. Hadid y M. Pietika, «Face Description with Local Binary Patterns: Application to Face Recognition,» *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 28, nº 12, pp. 2037-2041, 2006.
- [12] B. Zhang, S. Y. Gao, Zhao y J. Liu, «Local Derivative Pattern Versus Local Binary,» vol. 19, nº 2, pp. 533-544, 2010..
- [13] N. Dalal y B. Triggs, «Histograms of Oriented Gradients for Human Detection,» *EEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. pp. 886-893. , 2005.
- [14] Equinox Corporation., «Equinox Database,» [En línea]. Available: <http://www.equinoxsensors.com/products/HID.html>.
- [15] J. Kennedy y R. Eberhart, «Particle Swarm Optimization,» *In Proceedings*, vol. 4, pp. 1942-1948, 1995.
- [16] A. K. Jain y S. Z. li, «Handbook of face recognition.,» *Springer*, 2005.
- [17] T. Kanade, «Picture processing system by computer complex and recognition of human,» 1974.
- [18] G. Hermosilla, J. Ruiz-del-Solar, R. Verschae y M. Correa, «A comparative study of thermal face recognition methods in unconstrained environments,» *Pattern Recognition*, vol. 45, nº 7, pp. 2445-2459, 2011.
- [19] O. A. a. R. Hammoud, «Multi-Sensory Face Biometric Fusion (for Personal Identification),» *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pp. 128-128, 2006.
- [20] G. Hermosilla, «Reconocimiento Robusto de Rostros Usando Imágenes Térmicas,» 2012, Santiago de Chile.
- [21] J. Chen, S. Shan, C. He, M. P. G. Zhao, X. Chen y W. Gao, «WLD: a robust local image descriptor.,» *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, nº 9, 2010.
- [22] N. Krishnan y G. L. Aaron, «Recognition of Non-symmetric Faces Using Principal Component Analysis».

- [23] P. Comon, «Independent component analysis, A new concept?,» *Signal Processing*, vol. 36, nº 3, pp. 287-314, 1994.
- [24] D. G. Lowe, «Distinctive Image Features from Scale-Invariant Keypoints,» *International Journal of Computer Vision*, vol. 60, nº 2, pp. 91-110, 2004.
- [25] H. Bay, A. Ess, T. Tuytelaars y L. V. Gool, «Speeded-Up Robust Features (SURF),» *Computer Vision and Image Understanding*, vol. 110, nº 3, pp. 346-359, 2008.
- [26] B. Zhang, S. Z. Y. Gao y J. Liu, «Local Derivative Pattern Versus Local Binary Pattern: Face Recognition With High-Order Local Pattern Descriptor,» vol. 19, nº 2, pp. 533-544, 2010.
- [27] N. Wang, Q. Li, A. El-latif, J. Peng y X. Niu, «An Enhanced Algorithm for Thermal Face Recognition,» *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 6, nº 3, pp. 67-80, 2013.
- [28] J. Heo, «Fusion of visual and thermal face recognition techniques: A comparative study,» *The University of Tennessee, Knoxville, Tenn, USA*.
- [29] C. Dalila y H. Imane, «Multimodal Score-Level Fusion Using Hybrid GA-PSO for Multibiometric System,» *Informatica*, vol. 39, nº 2, p. 209–216, 2015.
- [30] A. A. Ross, K. Nandakumar y A. K. Jain, *Handbook of multibiometrics*, Springer, 2006.
- [31] C. Liu y H. Wechsler, «Evolutionary pursuit and its application to face recognition,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, nº 6, pp. 570-580, 2000.
- [32] R. Raghavendra, B. Dorizzi, A. Rao y G. H. Kumar, «Particle swarm optimization based fusion of near infrared and visible images for improved face verification,» *Pattern Recognition*, nº 44, p. 401–411, 2011.
- [33] A. B, H. S. y M. Abidi, «Fusion of visual, thermal, and range as a solution to illumination and pose restrictions in face recognition.,» de *In Proceedings of the IEEE International Carnahan Conference on Security Technology, Albuquerque, NM, USA, 11–14 October*, Albuquerque, NM, USA,, 2004.
- [34] Y. Fan, W. Zhu, G. Bai y T. Li, «Face recognition based on wavelet transform and neural network,» de *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, Xi'an, China, 2016, pp. 1569-1572.
- [35] K. Bong, S. Choi, S. K. C. Kim, Y. Kim y H. J. Yoo, «14.6 A 0.62mW ultra-low-power convolutional-neural-network face-recognition processor and a CIS integrated with always-on haar-like face detector," 2017 IEEE International Solid-State Circuits Conference

- (ISSCC),» de *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, USA, 2017, pp. 248-249.
- [36] F. E. Smith, «Evaluación de estrategias de fusión para un sistema de identificación,» 2015.
- [37] J. Váňa, Š. Mráček, M. Drahanský y A. Poursaber, «Applying fusion in thermal face recognition,» pp. 1-7, 2012.
- [38] D. C. C. Coello, de *Introducción a la Computación Evolutiva*, Mexico D.F., CINVESTAV-IPN, 2004, pp. 74-76.
- [39] S. H. Zanakis y J. R. Evans, «Heuristic “Optimization”: Why, When, and How to Use It,» *Interfaces*, vol. 11, p. 15, 1981.
- [40] P. Glover y G. Kochenberger, *Handbook of Metaheuristics*, Norwell, MA: Kluwer, 2002.
- [41] J. M. G. Nieto, «Algoritmos Basados en Cúmulos de Partículas Para la Resolución de Problemas Complejos,» Septiembre 2006.
- [42] J. R. P. López y J. Verdeja, «Optimización con enjambre de partículas,» de *URSI 2005, XX Simposium Nacional de la Unión Científica Internacional de Radio*, Gandía, 2005.
- [43] J. Kennedy, R. Eberhard y Y. Shi, *Swarm Intelligence*, San Francisco: Morgan Kaufmann Publishers, 2001.
- [44] G. Hermosilla, G. Farias, H. Vargas y F. S.-M. C. Gallardo, «Thermal Face Recognition Using Local Patterns,» *Pattern Recognition, Image Analysis, Computer Vision, And Applications; Lecture Notes in Computer Science; Springer: Puerto Vallarta, Jalisco, Mexico*, vol. 8827, pp. 486-497, 2014.
- [45] J. Ruiz-del-Solar, R. Verschae y M. Correa, «Recognition of Faces in Unconstrained Environments: A Comparative Study,» *EURASIP Journal on Advances in Signal Processing, special issue, Recent Advances in Biometric Systems: A Signal Processing Perspective*, 2009.
- [46] R. Eberhart y Y. Shi, «Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization,» *In Proceedings of the International Congress on Evolutionary Computation*, vol. 1, pp. 84-88, 2000.