

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Estado del arte de los métodos de
extracción de rasgos invariantes a la
iluminación en imágenes de rostros**

Dayron Rizo Rodríguez,
Heydi Méndez Vázquez, y
Edel García Reyes

RT_038

enero 2011





CENATAV

Centro de Aplicaciones de
Tecnologías de Avanzada
MINISTERIO DE LA INDUSTRIA BÁSICA

RNPS No. 2142
ISSN 2072-6287
Versión Digital

SERIE AZUL

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Estado del arte de los métodos de
extracción de rasgos invariantes a la
iluminación en imágenes de rostros**

Dayron Rizo Rodríguez,
Heydi Méndez Vázquez, y
Edel García Reyes

RT_038

enero 2011



Estado del arte de los métodos de extracción de rasgos invariantes a la iluminación en imágenes de rostros

Dayron Rizo Rodríguez¹, Heydi Méndez Vázquez², y Edel García Reyes²

¹Dpto. Ingeniería en Sistemas, Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV),
Ciudad de la Habana, Cuba
drizo@cenatav.co.cu

²Dpto. Reconocimiento de Patrones, Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV),
Ciudad de la Habana, Cuba
{hmendez, egarcia}@cenatav.co.cu

RT_038, Serie Azul, CENATAV
Aprobado: 24 de septiembre de 2010

Resumen: En los años recientes ha cobrado un gran auge el uso de las características del rostro en la biometría. Las variaciones en las condiciones de iluminación en las imágenes de rostros, constituyen un factor que afecta el rendimiento de los sistemas biométricos. Varios investigadores han enfocado su trabajo en encontrar técnicas que contribuyan a atenuar las consecuencias de este problema. Los métodos desarrollados con este objetivo se pueden agrupar teniendo en cuenta el momento en que se aplican dentro del proceso de reconocimiento de rostros. Uno de estos grupos está compuesto por los métodos de extracción de rasgos invariantes a la iluminación. En este reporte, se hace un estudio pormenorizado de los diferentes métodos de extracción de rasgos invariantes a la iluminación en imágenes de rostros que han sido desarrollados.

Palabras clave: reconocimiento de rostros, rasgos invariantes a la iluminación.

Abstract: In recent years has become a boom the use of facial features in biometrics. Variations in face image lighting conditions are a factor that affects the performance of biometric systems. Several researchers have focused their work on finding techniques to mitigate the consequences of this problem. Developed methods for this purpose can be grouped taking into account the step in which they are applied within the face recognition process. One of these groups is composed by illumination invariant feature extraction methods. In this report, a detailed survey of illumination invariant feature extraction methods for face images is presented.

Keywords: face recognition, illumination invariant features.

1 Introducción

Los sistemas biométricos basados en el reconocimiento automático de rostros se encuentran entre los más extendidos debido entre otros factores a que es una técnica no invasiva, natural y fácil de usar [1]. La utilización de las características del rostro se ha convertido en un reto para los especialistas en la materia, especialmente en aplicaciones reales donde existen ambientes no controlados en los que se manifiestan factores que inciden negativamente sobre las potencialidades biométricas de los rostros humanos, como son: la pose, la iluminación, la expresión, los accesorios utilizados, entre otros. Debido a esto se ha analizado la influencia de cada uno de ellos en el rendimiento de los sistemas de reconocimiento de rostros, determinándose que las variaciones de pose e iluminación, son las que en mayor medida los afectan [2] [3].

Adini, Moses y Ullman [4] mostraron que las variaciones de una imagen de rostro debidas a los cambios en la iluminación son más significativas que aquellas debidas a diversas identidades personales. Una contribución muy importante que advirtió sobre el peligro de trabajar con imágenes de rostros bajo condiciones de iluminación no adecuadas, puesto que inducen a errores en el reconocimiento de personas. Esto es sencillamente poner en duda la fiabilidad de aquellos sistemas de reconocimiento automático de rostros que no tengan en cuenta esta cuestión.

Para lidiar con los problemas de iluminación durante el proceso de reconocimiento de un rostro se han desarrollado métodos que se aplican en tres estados diferentes: el pre-procesamiento, la extracción de los rasgos y la clasificación [5]. El segundo estado incluye los métodos que se abordarán en este estudio. El objetivo en esta etapa es obtener una representación de la imagen que permita caracterizar el rostro que en ella se presenta, de manera tal que permita establecer comparaciones con otras representaciones de imágenes de rostros. Al afrontar la incidencia negativa del factor iluminación, la descripción de la imagen debe ser capaz de mantenerse inalterada ante imágenes de una misma persona con iluminación diferente.

Chen et al. [6] hacen un estudio profundo sobre la búsqueda de rasgos invariantes a la iluminación. En él afirman que: “no existen rasgos discriminativos invariantes a la iluminación”. Para comprender el alcance de esta expresión es necesario aclarar los conceptos de rasgo invariante a la iluminación y rasgo discriminativo. En [6] se presentan las siguientes definiciones para estos conceptos, las cuales se asumirán a partir de este momento:

Definición 1

Sea O un conjunto de objetos con sus propiedades ópticas; sea S un conjunto de condiciones de iluminación e I el conjunto de todas las posibles imágenes de un objeto $o \in O$ bajo cada una de las condiciones de iluminación $s \in S$. Se define la función $Q: o \times s \rightarrow i$ que devuelve la imagen $i \in I$, del objeto $o \in O$ bajo la iluminación $s \in S$, o sea, $I = Q(o, s)$.

En esta definición se presenta una imagen como el resultado de una función que depende del objeto o los objetos que en ella se representan y la fuente de luz que incide sobre él o ellos. Esto se ajusta al interés descrito anteriormente de enfrentar los problemas que provocan los cambios de iluminación en el reconocimiento de rostros, por tanto el conjunto de objetos estará restringido a rostros humanos en este reporte.

Definición 2

Una función μ en I es invariante a la iluminación sí y sólo sí $\mu(Q(o, s)) = \mu(Q(o, l))$: $\forall s, l \in S; o \in O$.

En esta definición se especifica el concepto de función invariante a la iluminación. Una función, que aplicada a dos imágenes de un mismo objeto con diferentes fuentes de luz incidiendo sobre él, sea capaz de obtener la misma salida. En este reporte se utiliza rasgo

invariante a la iluminación, que puede ser interpretado como la salida de la aplicación de una función invariante a la iluminación sobre una imagen de rostro.

Definición 3

Una función invariante a la iluminación μ es no discriminativa para un conjunto de objetos O , sí y sólo sí, $\mu(I) = \mu(J), \forall I \neq J$, donde I y J están en el dominio de Q . Si no cumple esta condición se dice que es discriminativa. Esta definición implica que μ no depende de o para ser discriminativa o no.

Esta última definición enmarca el concepto de función invariante a la iluminación no discriminativa como aquella función invariante a la iluminación que no es capaz de obtener una salida diferente cuando se la aplica a dos imágenes de dos objetos (rostros) diferentes, bajo determinadas condiciones de iluminación. Es por esto el nombre de no discriminativa, porque la función no detecta la diferencia de los objetos representados en cada una de las imágenes.

De esta definición se deriva el siguiente Lema [6]:

Lema 1

No existen funciones invariantes discriminativas a la iluminación para O si para cualesquiera dos imágenes I y J en el dominio de Q , hay siempre un objeto $o \in O$ el cual, bajo un par de condiciones de iluminación en S , puede generar tanto I como J .

El lema, que se deriva de la definición anterior, muestra las condiciones que impedirían la existencia de este tipo de funciones: la existencia de otro objeto del dominio que bajo dos condiciones de iluminación diferentes genere las dos imágenes analizadas. Esto provocaría que la salida de la función fuera igual para dos imágenes tomadas originalmente de objetos diferentes, es decir, la aparición de este objeto del dominio implica que el resultado de la aplicación de la función no es discriminativo.

A partir de estas definiciones, Chen et al., en [6], enunciaron el siguiente Teorema, que expresa la no existencia de rasgos invariantes a la iluminación discriminativos:

Teorema 1

Sin tener en cuenta la inter-reflexión, dadas dos imágenes de una función en el rango de Q : I y J , y dos fuentes de luz arbitrarias s y l al infinito, y si la proyección en el plano de la imagen de la suma de los vectores s y l no es cero, existe una familia de superficies Lambertianas¹ suaves f tal que I es la imagen de f bajo s y J es la de f bajo l .

Tesis

Sea μ una función invariante a la iluminación. Dadas dos imágenes arbitrarias I, J en el rango de Q , según el Teorema 1, siempre existe una superficie Lambertiana de un objeto o y dos fuentes de luz al infinito s y l , tal que I y J son imágenes de o bajo s y l , respectivamente, es decir, $I = Q(o, s)$ y $J = Q(o, l)$. Según el Lema 1, si esto ocurre, μ es una función invariante no discriminativa [6].

La demostración de este teorema, de Chen et al. [6], asegura el hecho de que no hay manera de afirmar con certeza que dos imágenes fueron generadas por superficies diferentes.

La estrategia básica de la demostración es escribir dos ecuaciones que describan la formación de las imágenes, con la forma del objeto y el albedo como desconocidos. Después se elimina el albedo, llegando a una ecuación diferencial lineal parcial de primer orden con respecto a la forma. Entonces se muestra que esta ecuación siempre tiene una solución única, correspondiente a un conjunto de condiciones iniciales que determinan una familia de soluciones. Usando la forma del objeto, el albedo es resuelto linealmente, correspondiente a un factor de escala. Es importante señalar que las soluciones obtenidas no contienen sombras lo que justifica la utilización de una descripción lineal simple de la formación de la imagen.

¹ Una superficie Lambertiana es un reflector perfectamente difuso, donde el brillo aparente es igual en todas las direcciones de vista.

Primeramente fijan el eje óptico de la cámara como el eje z . Por tanto un objeto o es observado desde la dirección $(0,0,1)$. Luego la superficie visible de los objetos del conjunto O estará definida por un conjunto de pares (f, α) , donde f es la representación $(x, y, z = f(x, y))$ de una función de suavidad $z = f(x, y)$ y $\alpha(x, y)$ es una función no negativa llamada albedo. La fuente de luz al infinito s es representada por un vector de tres dimensiones $s = (s_x, s_y, s_z)$ en la dirección opuesta a los rayos de luz. Se define la región SR como la región de sombra.

Suponiendo que una imagen I es generada por f bajo la fuente de luz $s = (s_x, s_y, s_z)$, la intensidad en la imagen se describe como:

$$I(x, y) = \begin{cases} \alpha(x, y) \vec{s} \cdot \hat{n}(x, y), & \text{Si } (x, y, f(x, y)) \notin SR \\ 0, & \text{e. o. c} \end{cases} \quad (1)$$

donde, $\alpha(x, y)$ es el albedo² del objeto o , $\vec{n} = \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1\right)$ es el vector normal a la superficie f de o , y $\hat{n} = \vec{n}/\|\vec{n}\|$.

Se expresa cada imagen con respecto a la Ecuación 1 suponiendo que se va a trabajar con la superficie de una imagen que no genera sombras:

$$I(x, y) = \alpha(x, y) \vec{s} \cdot \hat{n}(x, y) \quad (2)$$

$$J(x, y) = \alpha(x, y) \vec{l} \cdot \hat{n}(x, y) \quad (3)$$

Multiplicando cruzado las dos partes de las Ecuaciones 2 y 3, y dividiendo el resultado por α , que no va a ser 0, se obtiene la ecuación diferencial parcial de primer orden:

$$(\vec{l} - J\vec{s}) \cdot \vec{n} = 0 \quad (4)$$

Una vez solucionada la Ecuación 4, se sustituye el resultado en la ecuación original para obtener una ecuación lineal en el albedo.

La Ecuación 4 se resuelve utilizando el método de curvas características. En este método, se efectúa un cambio de variable para obtener una ecuación en una variable. Resolviéndola, este resultado nos dice la altura de la superficie a lo largo de una curva en el objeto. La ecuación completa tendrá una solución única, correspondiente a una condición inicial, si cada punto se encuentra en una curva característica única, de modo que la solución de estas curvas por separado proporciona exactamente una altura para cada punto de la superficie.

Generalmente, las curvas características $\vec{r}(t)$ de la Ecuación 4 satisfacen, según [6]:

$$\frac{d\vec{r}}{dt} = I\vec{l} - J\vec{s} \quad (5)$$

La superficie es construida entonces trenzando las curvas características al elegir valores continuos en los bordes. Por tanto la superficie suave es fabricada asignando cuidadosamente condiciones de suavidad a sus bordes.

El Teorema 1 se demostró en [6] generalizado a cualquier tipo de superficie Lambertiana. Ahora bien, es necesario tener en cuenta que en la práctica, en los sistemas de reconocimiento de

² El albedo es la relación, expresada en porcentaje, de la radiación que cualquier superficie refleja sobre la radiación que incide sobre la misma.

rostro, el paso previo de detección del rostro garantiza siempre trabajar con superficies de rostros, es decir, no se trabajará con otro tipo de superficies. Por tanto, el objeto o superficie que se construye para solucionar la ecuación debe ser un rostro. En otras palabras, no se garantiza por la forma de construcción de la superficie solución de la Ecuación 4, que el objeto resultante, que es el que afecta el carácter discriminativo de la función invariante a la iluminación, esté contenido dentro del dominio de superficies con el cual se va a trabajar. Más aún, los rostros por lo general no cumplen la condición de suavidad en los bordes, debido a los cambios provocados en la superficie por rasgos como la nariz, la boca y los ojos.

A pesar de esta dificultad es necesario destacar el aporte positivo de este trabajo. El énfasis en realzar la importancia de las propiedades discriminativas de los rasgos a extraer constituye una visión imprescindible para adentrarse en el dominio del tema que se pretende investigar. No es suficiente encontrar rasgos invariantes a la iluminación, cuya existencia no se pone en duda. Para el reconocimiento de rostros, es fundamental que estos rasgos demuestren buenas propiedades discriminativas porque de lo contrario no serían integrables a sistemas biométricos donde se requiere diferenciar con la mayor exactitud posible la identidad de las personas.

Por tanto el análisis de cualquier técnica, en lo adelante, no se enfocará solamente en comprobar su insensibilidad frente a las variaciones de iluminación sino que será imprescindible tener en cuenta sus propiedades discriminatorias.

La representación que se obtenga de la imagen de rostro una vez aplicada una técnica cualquiera, no tiene por qué ser discriminatoria por naturaleza. Sería una condición sumamente fuerte para pretender que se lograra solamente durante la extracción de los rasgos. Es posible que estos rasgos a *priori* no proporcionen la discriminación deseada y sea necesario desarrollar una medida de comparación entre rasgos que garantice de cierta manera discriminar los rostros representados.

Surge de esta forma un punto importante a tener en consideración para estudiar los métodos desarrollados por los especialistas en el tema. La definición de una medida de similitud que asegure o complemente la factibilidad de los rasgos extraídos cobra una significación meridiana y por tanto será un factor adicional para evaluar cada método que se analizará.

En la Fig. 1 se ilustra un esquema que describe el proceso básico que debe seguir cualquier técnica de extracción de rasgos invariantes a la iluminación.

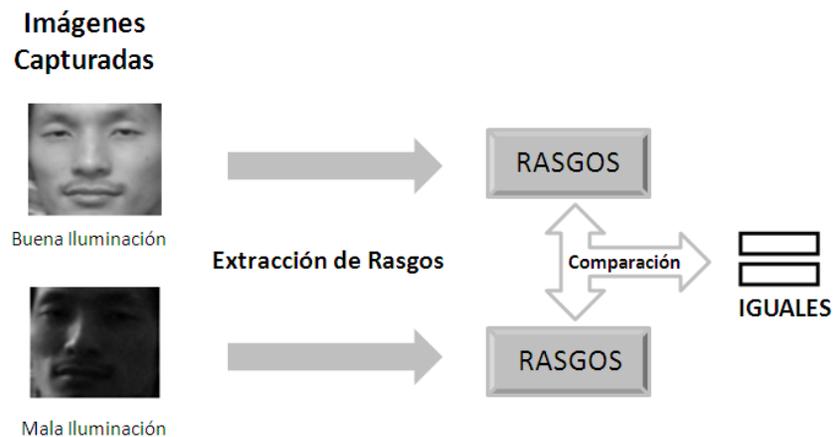


Fig. 1. Esquema representativo de cualquier método ideal para la extracción de rasgos invariantes a la iluminación.

La aplicación de métodos para la extracción de rasgos invariantes a la iluminación posibilita que no sea necesario someter las imágenes de rostros a transformaciones con el objetivo de atenuar la influencia de las variaciones de iluminación en la imagen. En otras palabras, se trabaja siempre sobre la imagen original y se elimina el paso intermedio de tener que aplicar una técnica de pre-procesamiento. Por tanto se podría mejorar el tiempo de ejecución y el consumo de memoria al integrar la extracción de rasgos invariantes a la iluminación a los sistemas de reconocimiento de rostros. Además evitaría la necesidad de poseer una determinada cantidad de imágenes de entrenamiento por cada sujeto bajo diferentes condiciones de iluminación, un factor que atenta contra la puesta en práctica de muchos métodos de los que se estudiarán a lo largo de este reporte. Esto se debe a que, en condiciones prácticas normales, no se cuenta con esa información por cada uno de los sujetos, de los que se almacena, generalmente, una sola imagen.

En este trabajo se persigue como objetivos:

1. Comprobar el impacto y vigencia de la extracción de rasgos invariantes a la iluminación.
2. Hacer un estudio detallado de las técnicas existentes para aumentar el conocimiento alrededor del tema.
3. Comparar las técnicas existentes en cuanto a los resultados reconocidos dentro de las bases de datos estándares para el reconocimiento de rostros.

2 Métodos de Extracción de Rasgos Invariantes a la Iluminación

La extracción de rasgos constituye una de las etapas que conforman el proceso de reconocimiento de rostros. En ella se pretende alcanzar una representación, de la imagen que se analiza, capaz de caracterizar el rostro de manera tal que permita establecer una comparación entre rostros basada en esta representación.

En la Fig. 2 se presenta una taxonomía de los métodos existentes dentro de la literatura que se relacionan con la extracción de rasgos invariantes a la iluminación en imágenes de rostros.

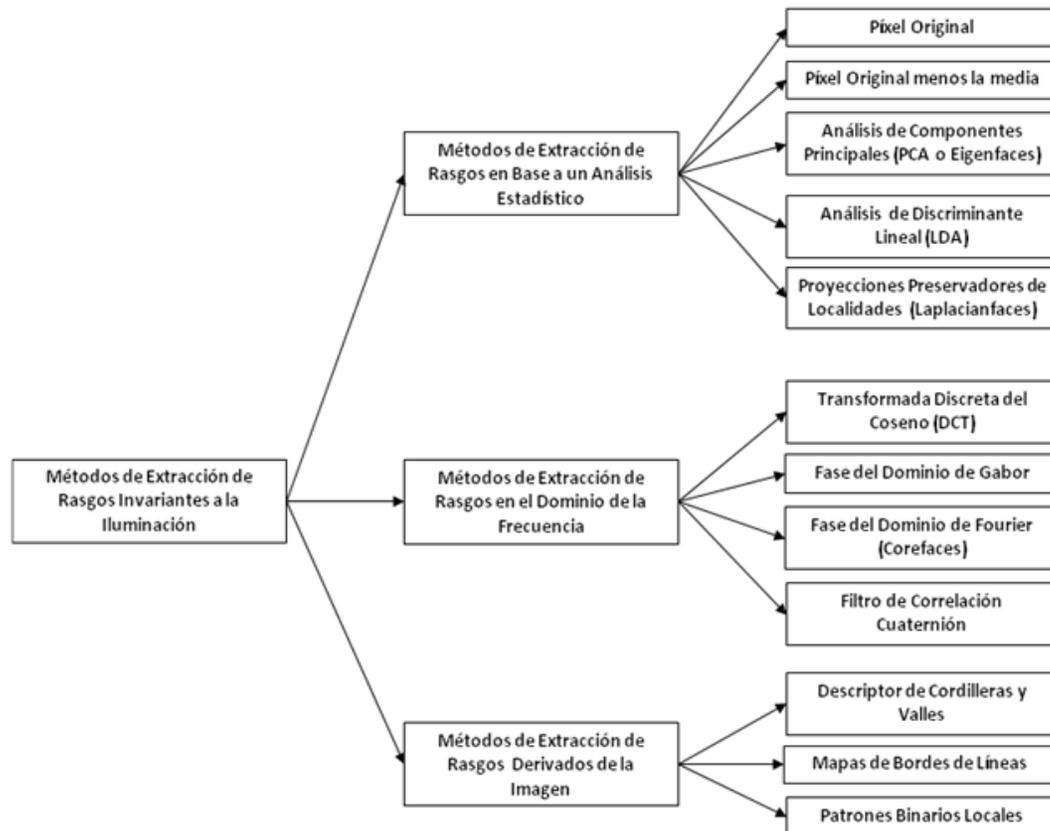


Fig. 2. Taxonomía de los métodos de extracción de rasgos invariantes a la iluminación.

Como puede verse en esta figura, los métodos de extracción de rasgos invariantes a la iluminación pueden dividirse en tres subconjuntos en dependencia de las herramientas utilizadas para el análisis de la imagen. En el primer subconjunto se encuentran agrupados los métodos que utilizan técnicas estadísticas para obtener una representación de la imagen que no dependa de las condiciones de iluminación. El segundo está compuesto por métodos que transforman la imagen al dominio de la frecuencia utilizando diferentes filtros o transformaciones y en este dominio se define la correspondencia entre las imágenes de rostro. El tercer subconjunto está integrado por los métodos que extraen rasgos basados en el procesamiento de la superficie del rostro, o sea, tienen en cuenta las características particulares que se presentan en un rostro.

A partir de esta agrupación se va a orientar el estudio de los métodos existentes. Tratando de abordar por cada técnica las características más importantes que determinan la naturaleza de cada una, enfocándolo siempre a las exigencias de invariabilidad a la iluminación y las propiedades discriminatorias de los rasgos extraídos.

En las próximas secciones se expondrán detalladamente cada uno de estos grupos y métodos.

2.1 Métodos de Extracción de Rasgos en Base a un Análisis Estadístico

Los métodos que se ubican en este conjunto se basan en extraer los rasgos representativos del rostro con herramientas estadísticas a partir de la información de la intensidad de los píxeles de la imagen. Estas herramientas se basan en propiedades que permiten establecer relaciones entre

los píxeles de la imagen, o que por otro lado, pueden utilizarse para reducir la dimensión de la representación de la imagen de rostro.

Las tres primeras técnicas que se describirán: *Píxel original*, *Píxel original menos la media*, y *Análisis de Componentes Principales Local* se basan en un análisis local de la imagen aunque la última fue propuesta inicialmente para aplicarse de manera global [7]. En ellas se utiliza una ventana para el análisis bloque a bloque, siendo cada bloque de $N \times N$ píxeles con $N=8$ a menos que se indique lo contrario. La localización de cada bloque avanza 4 píxeles, provocando un solapamiento de bloques vecinos del 50%. La elección de N y el solapamiento está basada en los resultados obtenidos en [8].

Este grupo también lo integran el algoritmo clásico *Análisis de Discriminante Lineal* y el método de *Proyecciones Preservadoras de Localidad*, los cuales enfrentan, cada uno de manera diferente, el problema de la iluminación en el reconocimiento de rostros.

2.1.1 *Píxel original*

Esta técnica, esencialmente, coloca los valores de intensidad de los píxeles que componen la región en un vector de rasgos. Los píxeles son organizados siguiendo el patrón zig-zag utilizado en [7], el cual se muestra en la Fig. 3.

		$\overbrace{\quad\quad\quad}^u$			
		0	1	2	3
{	0	0	1	5	6
	1	2	4	7	12
	2	3	8	11	13
	3	9	10	14	15

Fig. 3. Ejemplo del orden de los coeficientes siguiendo el patrón de zigzag para $N=4$ (bloque de 4×4).

Sin embargo es importante recalcar que al tratar directamente con las intensidades de los píxeles, cualquier patrón se puede usar. Para un bloque ubicado en (b, a) , el vector de rasgos se formaría de la siguiente manera:

$$\vec{x}^{(b,a)} = \left[p_0^{(b,a)} \ p_1^{(b,a)} \ \dots \ p_{N^2-1}^{(b,a)} \right]^T \quad (6)$$

donde, $p_n^{(b,a)}$ representa el valor del n -ésimo píxel en correspondencia con el patrón zig-zag. Esta técnica se conoce como *píxel original*.

Resaltan rápidamente dos desventajas de la misma: el vector puede estar altamente correlacionado y como efecto colateral de la alta correlación, un cambio de iluminación tiene efecto potencial sobre todos los elementos. Además, en los experimentos realizados en [7], se mostró que posee un pobre poder discriminatorio.

2.1.2 *Píxel original menos la media*

Con el objetivo de atenuar la correlación entre los elementos del vector de rasgos de la técnica *píxel original*, se sustrae la media a cada elemento [7]. Esta sustracción puede ser interpretada como una normalización de la iluminación. Formalmente, el vector restándole la media quedaría:

$$\vec{x}^{(b,a)} = \left[p_0^{(b,a)} - p_\mu^{(b,a)} \ p_1^{(b,a)} - p_\mu^{(b,a)} \ \dots \ p_{N^2-1}^{(b,a)} - p_\mu^{(b,a)} \right]^T \quad (7)$$

donde:

$$p_{\mu}^{(b,a)} = \frac{1}{N^2} \sum_{i=0}^{N^2-1} p_i^{(b,a)} \quad (8)$$

Se conoce este método como *píxel original menos la media*. En [7] se concluye que a pesar de mejorar en cuanto a las características discriminativas, con respecto al *píxel original*, mantiene una alta sensibilidad ante los cambios de iluminación.

2.1.3 Análisis de Componentes Principales Local

En lugar de utilizar Análisis de Componentes Principales (PCA por sus siglas en inglés o Eigenfaces) para la representación global de la imagen como se propone en [10], en [7] se presenta esta técnica aplicada a bloques de la imagen para enfrentar el problema de la iluminación, conociéndose como *PCA Local*. El primer paso es exactamente el mismo que para la extracción con *píxel original*. Denotando el vector de rasgos resultante de un bloque (b, a) como $\vec{r}^{(b,a)}$, un nuevo vector de rasgos, normalmente con menor dimensionalidad, se obtiene de:

$$\vec{x}^{(b,a)} = U^T (\vec{r}^{(b,a)} - \vec{r}_{\mu}) \quad (9)$$

De la Ecuación 9, la matriz de transformación U^T y \vec{r}_{μ} se calculan de la siguiente manera. Definiendo un conjunto de entrenamiento de vectores de rasgos de *píxel original* como:

$$R = \{\vec{r}_i\}_{i=1}^{N_A} \quad (10)$$

A partir de este conjunto se determina su media \vec{r}_{μ} . Luego se calcula una matriz de covarianza:

$$C = \frac{1}{N_A} \sum_{i=1}^{N_A} (\vec{r}_i - \vec{r}_{\mu}) (\vec{r}_i - \vec{r}_{\mu})^T \quad (11)$$

La matriz U se formaría como sigue:

$$U = [\vec{e}_1 \ \vec{e}_2 \ \dots \ \vec{e}_D] \quad (12)$$

donde, \vec{e}_n es el n -ésimo vector propio de C ; los vectores propios son ordenados, de manera descendente, de acuerdo al valor propio que le corresponde a cada uno. Estos representan las direcciones ortogonales que dan cuenta de la mayor magnitud de varianza. D tiene las siguientes restricciones: $D \leq N_A$ y $D \leq N^2$. De aquí se deduce que si $D = N^2$ entonces no hay reducción de dimensionalidad; en este caso, el vector $\vec{x}^{(b,a)}$ es una representación sin correlación del vector $\vec{r}^{(b,a)}$.

En [7] se afirma que una posible desventaja del método PCA Local es que las funciones bases pueden no tener significado en términos de la estructura de la imagen, aunque sí tiene significado estadístico. Además, se alega que las funciones bases varían dependiendo del conjunto de entrenamiento. Por este se propone por sus autores la hipótesis de analizar en esta técnica, el hecho de descartar algunos elementos del vector de rasgos con el objetivo de ganar en robustez frente a los cambios de iluminación.

Para ello tomaron la sección de entrenamiento de la base de datos de imágenes de rostro XM2VTS [11], los primeros vectores se muestran en la Fig. 4. En virtud del análisis de esta

figura, los autores afirman que los tres primeros coeficientes, siguiendo el patrón de zig-zag anteriormente descrito, son los más afectados por las variaciones de iluminación.



Fig. 4. 16 primeros coeficientes obtenidos aplicando el método PCA Local sobre un bloque de tamaño 8 x 8. Los colores claros representan valores más grandes.

A modo de conclusión, en [7] se plantea que la definición de las funciones bases en PCA Local son aprendidas, de esta forma son más representativas del bloque. Además que la reducción de dimensionalidad de PCA Local es óptima en el Error Cuadrático Medio (ECM) [12], es decir, preserva mayor información, por tanto los vectores de rasgos PCA Local pueden tener baja dimensión. Sin embargo, se añade que esto no garantiza que los vectores de rasgos sean óptimos con respecto a sus propiedades discriminativas.

Otra desventaja palpable de este método radica en la necesidad de poseer un conjunto de imágenes de entrenamiento lo suficientemente representativo por cada sujeto, para que se puedan obtener mejores resultados.

2.1.4 Análisis de Discriminante Lineal

El *Análisis de Discriminante Lineal* (LDA o *Fisherfaces*) es una técnica supervisada de reducción lineal de la dimensión que pretende hallar un subespacio discriminante. El objetivo de sus creadores es que los elementos pertenecientes a clases diferentes, en este caso rostros de personas diferentes, se mantengan alejados en el nuevo espacio, mientras que los elementos pertenecientes a la misma clase permanezcan cercanos. Esencialmente se halla la matriz de dispersión intra-clase e inter-clase suponiendo que bajo condiciones ideales las clases son linealmente separables, lo que significa ejecutar la clasificación en el espacio de rasgos de menor dimensión con fronteras de decisión lineales.

Dado un conjunto de imágenes de rostros etiquetadas con la identidad de la persona (conjunto de entrenamiento), estas se usan para hallar una representación en el espacio de menor dimensión. Sea el vector media $u = \frac{1}{N} \sum_{i=1}^N x_i$, la matriz de dispersión inter-clase y la matriz de dispersión intra-clase se definen como:

$$S_B = \sum_{i=1}^N |\chi_i| (u_i - u)(u_i - u)^T \quad (13)$$

$$S_W = \sum_{i=1}^N \sum_{x_k \in \chi_i} (x_k - u_i)(x_k - u_i)^T \quad (14)$$

donde u_i es la media de las imágenes y $|\chi_i|$ es el número de muestras de la clase χ_i .

Luego se resuelve el problema generalizado de valores y vectores propios:

$$S_B v = \lambda S_W v \quad (15)$$

Con los vectores propios se conforma la matriz de cambio de coordenadas y con esta se proyectan las muestras tanto de entrenamiento como otras nuevas en el subespacio lineal hallado.

En [13] se parte de la observación de que las imágenes con diferentes condiciones de iluminación de un rostro (con superficie Lambertiana y sin sombras) yacen en un subespacio lineal 3D [14]. El método no tiene en cuenta las porciones de la imagen que no son representadas por el modelo Lambertiano y desvían a las imágenes del subespacio lineal 3D, por ejemplo sombras moldeadas o manchas especulares, que no son significativas para representar a cada persona ya que tiende a minimizar la varianza intra-clase.

Entonces, para cada rostro se toman 3 o más imágenes bajo diferentes condiciones de iluminación, para construir la base del subespacio para cada clase y se aplica el método. La suposición fundamental del método es que las clases son linealmente separables. En la comparación experimental con el método PCA [13], se muestra que LDA alcanza resultados superiores cuando las imágenes de rostros evaluadas presentan problemas de iluminación, así como problemas de expresión.

Entre las ventajas principales se puede mencionar el bajo costo computacional, ya que una vez que se tienen las direcciones calculadas para cada clase, solo es necesaria una imagen para proyectar y emplear la proyección como modelo de la clase. Entre sus desventajas se encuentra el alto costo del entrenamiento, y que requiere condiciones difíciles de hallar en la práctica, como un conjunto de imágenes por cada clase con diferentes condiciones de iluminación.

2.1.5 Proyecciones Preservadoras de Localidad

El algoritmo *Proyecciones Preservadoras de Localidad* (Laplacianfaces) según [15] encuentra las aproximaciones lineales óptimas a las funciones propias del operador de Laplace-Beltrami en una variedad.

La idea del algoritmo es construir un grafo de vecindades, y a partir de este grafo obtener información que permite hallar la inmersión de los datos en un subespacio de menor dimensión, que preserva características locales que presenta el conjunto de datos en el espacio inicial.

El método consta de los siguientes pasos:

1- Construir el grafo:

Sea $G = (V, E)$ el grafo donde V es el conjunto de nodos representando los datos del conjunto inicial, y E el conjunto de aristas; dos vértices tienen aristas en E si se cumple alguno de los siguientes criterios:

- a) k vecinos más cercanos (del inglés k nearest neighbours): i tiene arista con j si: i es uno de los k vecinos más cercanos de j o j es uno de los k vecinos más cercanos de i .
- b) ε - vecindad: i tiene arista con j si $\|x_i - x_j\|^2 < \varepsilon$.

2- Encontrar los pesos:

A partir del grafo G se construye una matriz W donde en la posición i, j el valor se calcula:

- a) Heat kernel: si los vértices j e i están conectados se hace:

$$W_{i,j} = e^{-\frac{\|x_i - x_j\|^2}{t}} \quad (16)$$

- b) Simple: si los vértices j e i están conectados, entonces $W_{i,j} = 1$.

3- Eigenmaps:

Asumiendo que el grafo es conexo, calcular valores y vectores propios del problema generalizado de vectores propios:

$$XLX^t a = \lambda XD X^t a \quad (17)$$

donde, D es una matriz diagonal cuyas entradas son la suma de las columnas de W , o sea, $D_{ii} = \sum_j W_{ji}$, y $L = D - W$ es la matriz Laplaciana. La i -ésima columna de la matriz X es x_i .

Siendo los vectores columnas a_0, a_1, \dots, a_{l-1} las soluciones de la Ecuación 17, ordenados de acuerdo a sus valores propios, $\lambda_0, \lambda_1, \dots, \lambda_{l-1}$. Entonces, la transformación de los vectores x_i de dimensión n al nuevo espacio, donde el vector y_i de dimensión l es su equivalente en dicho espacio, con $l < n$, sería de la siguiente manera:

$$x_i \rightarrow y_i = A^T x_i; A = (a_0, a_1, \dots, a_{l-1})$$

donde, A es una matriz de dimensión $n \times l$.

En la Fig. 5 [16], se muestran las *Laplacianfaces*, las *Eigenfaces* y las *Fisherfaces* de 10 imágenes de rostro.

En [16] se concluyó que el método *Laplacianfaces* presentó un mejor desempeño que los algoritmos *Eigenfaces* y *Fisherfaces* en bases de datos de imágenes de rostros con problemas de iluminación. En [15] fue demostrado que el algoritmo es capaz de descubrir la estructura no lineal de la variedad de los datos, lo cual se da en el trabajo como justificación al mejor desempeño comparado con las otras técnicas lineales. Esto es muy peculiar ya que con técnicas lineales se está detectando la estructura no lineal de la variedad de rostros y aprovechándola para la clasificación.

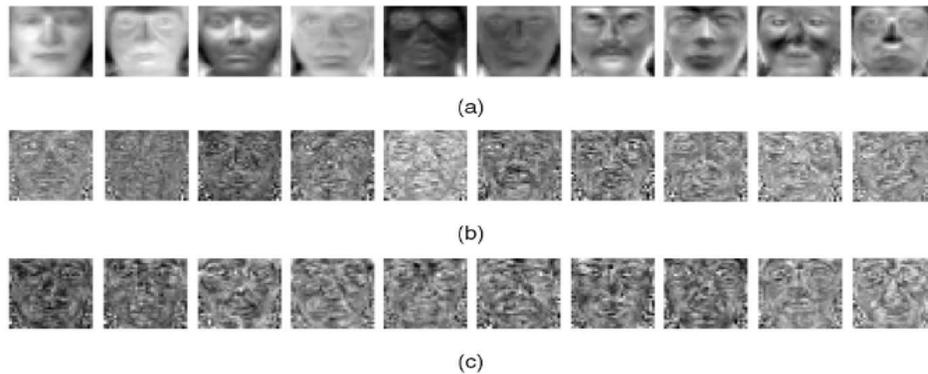


Fig. 5.10 Eigenfaces (a), Fisherfaces (b) y Laplacianfaces (c).

Su principal desventaja es que necesita un conjunto de imágenes de entrenamiento con diferentes condiciones de iluminación por cada sujeto. Difícil de conseguir en las aplicaciones prácticas de los sistemas de reconocimiento de rostro.

2.2 Métodos de Extracción de Rasgos en el Dominio de la Frecuencia

Los métodos que pertenecen a este subconjunto se caracterizan por aplicar transformaciones a la imagen y obtener una representación de la misma en el dominio de la frecuencia. Se verán cuatro técnicas que transforman la imagen al dominio de la frecuencia, dado que es más fácil separar

los efectos de la iluminación en este dominio, estando mayormente presentes las variaciones de iluminación en las bajas frecuencias, siendo poco predominantes en las altas frecuencias [17].

El método *Transformada Discreta del Coseno* transforma la imagen a partir de funciones específicas al dominio de la frecuencia. Por otro lado las técnicas *Fase de la Transformada de Fourier*, *Fase de la Transformada de Gabor* y *Filtro de Correlación Cuaternión* se basan en aplicar filtros de correlación a la imagen para obtener su representación en el dominio de la frecuencia y utilizar la componente que representa la fase del resultado como representación de la información invariante a la iluminación de la imagen.

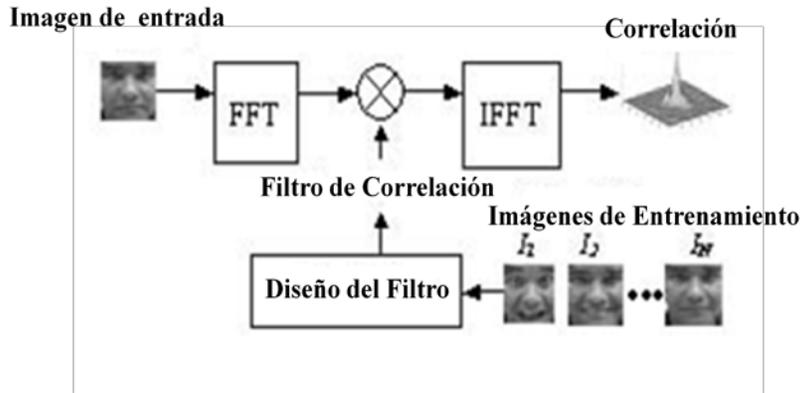


Fig. 6. Esquema de la aplicación de un filtro de correlación.

Los filtros de correlación usados para el reconocimiento de rostros en [18] [19] [20] sintetizan un filtro de correlación simple o una plantilla, de un conjunto de imágenes de entrenamiento. En la Fig. 6 se representa el esquema de la aplicación de un filtro de correlación. Teniendo un conjunto de entrenamiento compuesto por N imágenes de la clase I, se lleva a cabo el proceso de diseño del filtro de correlación para dicha clase. Luego, cuando se desea evaluar una imagen, si esta imagen pertenece a la clase I, entonces la salida devuelve un pico agudo.

2.2.1 Transformada Discreta del Coseno

En la técnica *Transformada Discreta del Coseno* (DCT por sus siglas en inglés), cada bloque $b(y, x)$, donde $y, x = 0, 1, \dots, N - 1$, de dimensiones $N \times N$, con $N = 8$, se descompone a partir de funciones bases ortogonales pre-definidas, como se presentan en la Fig. 7.

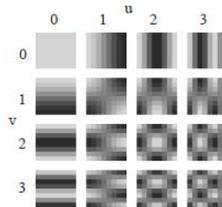


Fig. 7. Funciones bases DCT para $N = 8$; los colores más claros representan valores más grandes.

El resultado de la descomposición de la imagen usando estas funciones es una matriz $C(v, u)$ de $N \times N$ que contiene los coeficientes DCT:

$$C(v, u) = \alpha(v) \alpha(u) \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} b(y, x) \beta(y, x, v, u) \text{ para } v, u = 0, 1, \dots, N - 1 \quad (18)$$

donde:

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}}, & \text{para } v = 0 \\ \sqrt{\frac{2}{N}}, & \text{para } v = 1, \dots, N - 1 \end{cases} \quad (19)$$

y

$$\beta(y, x, v, u) = \cos \left[\frac{(2y + 1)v\pi}{2N} \right] \cos \left[\frac{(2x + 1)u\pi}{2N} \right] \quad (20)$$

El patrón zigzag que se utiliza para formar el vector de rasgos DCT, es el mismo que se presentó en la Fig. 3. Este vector de rasgos determina la cantidad de información almacenada en cada coeficiente [9], siendo considerados los coeficientes de menor orden como los que mayor información contienen. El vector de rasgos DCT se conformaría de la siguiente manera:

$$\vec{x}^{(b,a)} = [c_0^{(b,a)} c_1^{(b,a)} \dots c_{M-1}^{(b,a)}]^T \quad (21)$$

donde $c_n^{(b,a)}$ representa el n-ésimo coeficiente DCT y M es el número de coeficientes retenidos. Para el caso de $N = 8$, M varía de 1 a 64, dependiendo de la reducción de dimensionalidad deseada.

Esta técnica presenta una ventaja obvia: habilidad para reducir la dimensionalidad. Esto se traduce en que menos datos son necesarios para entrenar el clasificador adecuadamente [21] y además, el vector de rasgos contendrá menos ruido, y así será más discriminativo. Otra ventaja de DCT es que interpretando gráficamente las funciones básicas en la Fig. 7, el primer coeficiente será el más afectado por los cambios de iluminación y de esta manera eliminándolo del vector de rasgos se obtiene mayor robustez. Incluso, los dos coeficientes que le siguen podrían estar afectados significativamente por las variaciones de iluminación.

No obstante, la eliminación de estos coeficientes conlleva la pérdida de la información discriminativa de los rasgos del rostro que ellos representan. Una posible mejora sería realizar un estudio más profundo de cuántos y cuáles coeficientes del vector de rasgos DCT se deberían eliminar para lograr la invarianza ante los problemas de iluminación sin que esto implique una disminución del valor discriminativo de los rasgos no eliminados.

En [7] se confirma que a pesar sus ventajas, esta técnica no ofrece buenas propiedades de discriminación cuando se eliminan coeficientes buscando invarianza a la iluminación, avalado por los resultados obtenidos durante la experimentación.

2.2.2 Fase de la Transformada de Gabor

La forma de la onda Gabor 2D puede definirse de la siguiente manera [22]:

$$g(x, y) = e^{-\pi((x-x_0)^2/\alpha^2 + (y-y_0)^2/\beta^2)} e^{-2\pi j(u(x-x_0) + v(y-y_0))} \quad (22)$$

donde (x_0, y_0) especifica la posición de la onda, (α, β) denota la longitud y el ancho efectivo, y (u, v) representa una modulación vector-onda la cual puede ser interpretada en coordenadas

polares como la frecuencia espacial $\bar{\omega} = \sqrt{u^2 + v^2}$ y la orientación $\theta = \arctan(v/u)$. En [22] se utilizan 5 frecuencias espaciales y 8 orientaciones de las ondas de Gabor. La representación Gabor de una imagen de rostro puede obtenerse al convolucionar la imagen con los filtros de Gabor.

Sea $f(x, y)$ la imagen de rostro, su convolución con el filtro Gabor se define como:

$$C(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad (23)$$

Las imágenes de dos personas bajo dos condiciones de iluminación diferentes se presentan en la Fig. 8(a). La fila superior muestra una misma persona bajo dos condiciones de iluminación diferentes, mientras que la inferior muestra otra persona con las mismas condiciones de iluminación que las de la imagen superior. La visualización de la fase de Gabor de estas imágenes correspondiente a la dirección horizontal del filtro de Gabor de mayor escala se muestra en la Fig. 8(b). Las imágenes diferencia de las imágenes de la fase se muestran en la Fig. 8(c), en la primera fila se presentan las de una misma persona y en la segunda fila las que representan diferentes personas con la misma iluminación. Por ejemplo, la imagen superior izquierda de la Fig. 8(c) es la imagen diferencia entre las imágenes superior izquierda y superior derecha de la Fig. 8(b), mientras la imagen inferior izquierda de la Fig. 8(c) es la imagen diferencia entre las imágenes superior izquierda e inferior izquierda de la Fig. 8(b). Finalmente los histogramas de las cuatro imágenes diferencia se muestran en la Figura 8(d).

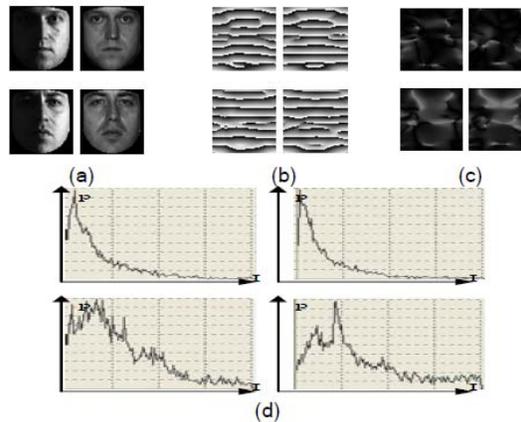


Fig. 8. Imágenes de la fase de 4 imágenes de rostro. (a) Imágenes de 2 personas bajo 2 condiciones de iluminación diferentes. (b) Imágenes de la fase de Gabor de las imágenes de (a). (c) Imágenes diferencia de las imágenes de la fase. (d) Histogramas de las imágenes de (c).

Observando con atención la Fig. 8, se puede percibir que los patrones de la fase de Gabor de cada persona, aún con condiciones de iluminación diferentes, son más distintivos que los de diferentes personas incluso bajo la misma condición de iluminación. Por tanto pudiera decirse que la fase de Gabor es tolerante a las variaciones de iluminación, mientras conserva su habilidad discriminativa. Entonces la información de la fase es usada como una medida insensible a la iluminación para el reconocimiento de rostros [22]. Una posible vía para perfeccionar esto es modelar la diferencia entre las fases de Gabor de dos imágenes utilizando una medida de similitud probabilística parecida a la propuesta que aparece en [23].

Sea Δ la diferencia de la fase de Gabor entre dos imágenes, sea Ω_I las variaciones intra-clase o intra-personales de las fases de Gabor y sea Ω_E las variaciones inter-clases o inter-personales, la similitud se expresa entonces en términos de la probabilidad como:

$$S(I_1, I_2) = P(\Omega_I|\Delta) = \frac{P(\Delta|\Omega_I) P(\Omega_I)}{P(\Delta|\Omega_I) P(\Omega_I) + P(\Delta|\Omega_E) P(\Omega_E)} \quad (24)$$

El próximo paso es el cálculo de las probabilidades $P(\Delta|\Omega_I)$ y $P(\Delta|\Omega_E)$, y las iniciales $P(\Omega_I)$ y $P(\Omega_E)$. En este trabajo se asume que $P(\Omega_E) = P(\Omega_I) = 0.5$, y que los píxeles de las imágenes son independientes, ignorando correlaciones que pueden surgir de la proximidad espacial. Luego la Ecuación 24 se reescribe como:

$$S(I_1, I_2) = \frac{P(\Delta|\Omega_I)}{P(\Delta|\Omega_I) + P(\Delta|\Omega_E)} = \prod_{i \in M} \frac{P(\Delta_i|\Omega_I)}{P(\Delta_i|\Omega_I) + P(\Delta_i|\Omega_E)} \quad (25)$$

Las densidades $P(\Delta_i|\Omega_I)$ y $P(\Delta_i|\Omega_E)$ de la Ecuación 25 se estiman a partir de los histogramas de las diferencias entre las fases de Gabor de cada dos imágenes de la misma persona, y de diferentes personas, respectivamente. Las imágenes de la fase de la misma persona bajo diferente iluminación son casi iguales, mientras que las diferencias entre las imágenes de la fase de dos personas son más aleatorias como se ilustra en la Fig. 8(d).

Este método presenta dos desventajas para su utilización, ambas en el proceso de entrenamiento. Primeramente, es necesario conformar un conjunto de entrenamiento representativo, con imágenes bajo diferentes condiciones de iluminación por cada sujeto para lograr un cálculo efectivo de las densidades $P(\Delta_i|\Omega_I)$ y $P(\Delta_i|\Omega_E)$. En [22] se dice que son necesarias que no menos de 3 imágenes por cada persona. A partir de suponer que se cuenta con esta cantidad de imágenes, algo que en escenarios reales es poco probable, el costo computacional del entrenamiento es muy alto porque, además del número de imágenes, se debe aplicar la convolución de Gabor a cada imagen 40 veces, dado que se utilizarán 5 frecuencias espaciales y 8 orientaciones de las ondas.

2.2.3 Fase de la Transformada de Fourier

Dentro de la gran variedad de filtros de correlación existentes, el filtro conocido como energía de correlación promedio mínima (MACE por sus siglas en inglés) [24] intenta producir picos agudos al minimizar la energía de correlación promedio resultante de las imágenes de entrenamiento y el filtrada, al limitar el pico de correlación en el origen para que sea un valor predeterminado (usualmente este pico es tomado como 1 para las imágenes de entrenamiento). El de mayor relevancia es el filtro MACE sin restricciones, donde, en lugar de restringir el pico en el origen $c_i(0,0)$ a ser un valor predeterminado, simplemente trata de maximizar la magnitud cuadrada del pico. Esta optimización puede ser formulada en el dominio de la frecuencia para obtener una solución más eficiente computacionalmente.

La energía E_i del i -ésimo plano $c_i(x, y)$ puede ser expresada usando la representación del dominio de la frecuencia $C_i(u, v)$ de la siguiente manera:

$$E_i = \sum_{x=0}^{d-1} \sum_{y=0}^{d-1} |c_i(x, y)|^2 = \frac{1}{d^2} \sum_{u=0}^{d-1} \sum_{v=0}^{d-1} |C_i(u, v)|^2 \quad (26)$$

$$= \frac{1}{d^2} \sum_{u=0}^{d-1} \sum_{v=0}^{d-1} |H(u, v)|^2 |X_i(u, v)|^2 = h^+ D_i h$$

donde, se asume que la imagen tendrá dimensión de $d \times d$.

Por tanto, la energía de correlación promedio resultante de las imágenes de entrenamiento se puede escribir como:

$$E_{promedio} = \frac{1}{N} \sum_{i=1}^N E_i = \frac{1}{N} \sum_{i=1}^N h^+ D_i h = h^+ D h \quad (27)$$

donde, $D = \frac{1}{N} \sum_{i=1}^N D_i$ es una matriz diagonal con el espectro de potencia promedio de la imagen de entrenamiento ubicado a lo largo de la diagonal. El vector columna h es de dimensión d^2 .

El pico de correlación en el origen es el producto interno de la función de difusión puntual del filtro $h(x, y)$ y la imagen de entrenamiento $f_i(x, y)$. En el dominio de la frecuencia sería:

$$c_i(x, y) = \sum_{u=0}^{d-1} \sum_{v=0}^{d-1} F_i(u, v) \cdot H(u, v) * e^{\frac{j2\pi ux}{d}} e^{\frac{j2\pi vy}{d}} \quad (28)$$

La Ecuación 28 para el origen $c(0,0)$ quedaría de la siguiente manera:

$$c_i(0,0) = \sum_{u=0}^{d-1} \sum_{v=0}^{d-1} F_i(u, v) \cdot H(u, v) = h^+ x_i \quad (29)$$

donde, x_i es la transformada de Fourier de la i -ésima imagen de entrenamiento denotada como $F_i(u, v)$.

Teniendo N imágenes de entrenamiento $x_{1...N}$, el objetivo es maximizar la magnitud promedio cuadrada de este pico. El cálculo del filtro óptimo h se realizaría a través de la expresión:

$$h = D^{-1} m \quad (30)$$

Este cálculo de h está marcado por el hecho siguiente: D es una matriz diagonal que contiene el espectro de potencia promedio de las imágenes de entrenamiento.

En la Fig. 9 se representa la salida del filtro MACE sin restricciones. La imagen de abajo posee un pico muy agudo resultante de la correlación cruzada del filtro con una imagen de la misma persona cuyas imágenes fueron utilizadas para sintetizar el filtro.

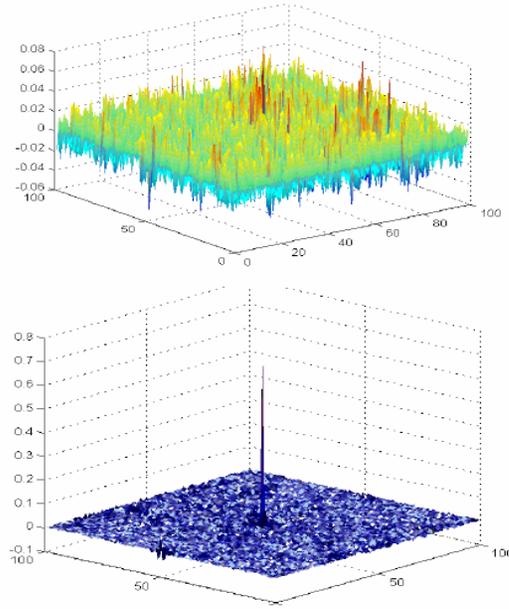


Fig. 9. (Arriba) Salida del filtro de correlación para una imagen de una persona diferente a la que se utilizó para sintetizar el filtro. (Abajo) Salida para una imagen de la misma persona con la que se sintetizó el filtro.

A diferencia de los filtros de correlación utilizados habitualmente, no se tiene en cuenta solamente el valor del pico, ya que es dependiente de las variaciones constantes de iluminación, más bien, se calcula la proporción alrededor del pico, *PSR* (*por sus siglas en inglés*), que es una manera de medir su agudeza:

$$PSR = \frac{peak - mean}{\sigma} \tag{31}$$

En la Fig. 10 se muestra gráficamente el cálculo propuesto en la Ecuación 31.

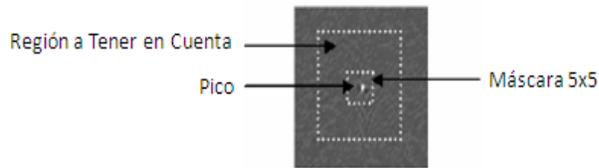


Fig. 10. Región para estimar la proporción alrededor del pico.

Al calcular *PSR*, se examinan varias salidas del filtro de correlación, no solamente el pico, y así se toma una decisión más confiable a la hora de realizar la clasificación. Además, se evita que se afecte la clasificación si la imagen de entrada es sometida a un factor de escala de la iluminación constante.

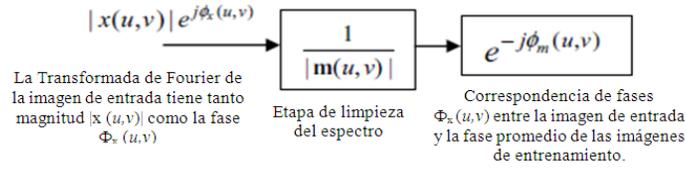


Fig. 11. Diagrama del proceso de un filtro de energía de correlación promedio mínima (MACE).

Los filtros MACE están diseñados para producir picos agudos, que pudieran ser representados como funciones delta en la clase ideal. Una función delta está representada por una transformada de Fourier plana constante, por tanto si se quiere una salida del tipo función delta en la salida del filtro de correlación, este debe de algún modo aplanar el espectro (segunda etapa en la Fig. 11) y cancelar las fases complejas $\phi(u, v)$ (tercera etapa en la Fig. 11) de manera tal, que se termine con un espectro de frecuencia constante.

A partir de lo anterior se puede apreciar que este método es tolerante a la iluminación. Dado que la mayoría de las imágenes tienen la mayor concentración de energía en el espectro de baja frecuencia, el paso $1/|m(u, v)|$ realzará las altas frecuencias en la imagen de entrada, donde los efectos de las variaciones de la iluminación no son predominantes, asumiendo que las variaciones de iluminación se presentan mayormente en el espectro de baja frecuencia. Si la fase de la transformada de Fourier se corresponde con la fase del filtro, entonces se cancelarán, dejando un espectro de magnitud plana constante, suponiendo que el paso de aplanamiento del espectro fue ejecutado exitosamente.

En [25] se expresa que la información más representativa de las imágenes 2D se retiene en la fase y no la magnitud de las transformadas de Fourier. De hecho, la información de la magnitud se puede recuperar usando la información de la fase correspondiente a un factor de escala según [26].

Se propuso, por Savvides et al. [17], usar PCA en el dominio de la frecuencia de la transformada de Fourier para obtener una mejor representación del espectro de la fase de las imágenes de entrenamiento. Con esto se crea un subespacio que modela las variaciones de la fase para una persona en particular. Estos autores en [17] afirman que PCA en el dominio de la frecuencia con técnicas de pre-procesamiento simples evita muchos problemas que se presentan al usarlo en el dominio espacial.

En [17] se determina una relación entre C_s y C_f , las matrices de covarianza de los datos en el dominio del espacio y en el dominio de la frecuencia respectivamente.

$$C_f = \sum_{i=1}^N \{T_{DFT}(x - m)\} \{T_{DFT}(x - m)\}^+ = T_{DFT} X X^+ T_{DFT}^{-1} \quad (32)$$

donde:

$$X X^+ = \sum_{i=1}^N (x - m)(x - m)^+ = C_s \quad (33)$$

PCA diagonaliza la matriz de covarianza C_f usando los vectores propios obtenidos en la Ecuación 32.

$$C_f v_f = \lambda v_f \quad (34)$$

Sustituyendo C_f por la Ecuación 34 se obtiene:

$$T_{DFT} X X^+ T_{DFT}^{-1} v_f = \lambda v_f \quad (35)$$

Pre-multiplicando la Ecuación 35 por T_{DFT}^{-1} se obtiene:

$$X X^+ T_{DFT}^{-1} v_f = \lambda T_{DFT}^{-1} v_f \quad (36)$$

PCA en el dominio del espacio se define como:

$$C_s v_s = \lambda v_s \quad (37)$$

Sustituyendo la Ecuación 33 en la 37:

$$X X^+ v_s = \lambda v_s \quad (38)$$

Comparando las Ecuaciones 34 y 37 se identifica una relación entre los vectores propios de los dominios de espacio y frecuencia, a través de la transformada inversa de Fourier de la siguiente manera:

$$v_s = T_{DFT}^{-1} v_f \quad (39)$$

A las componentes principales del dominio de frecuencia se les llamó *Corefaces* en [17].

Una vez que se encuentra el subespacio lineal óptimo para representar los datos de entrenamiento en el sentido del error cuadrático medio, se proyecta la imagen de entrada en este subespacio, reconstruyéndolo de la forma siguiente:

$$e^{\phi_R(u,v)} = \sum_{i=1}^N v_{f_i}(u,v) e^{\phi_T(u,v)} \quad (40)$$

donde, $e^{\phi_R(u,v)}$ es el espectro de la fase reconstruido, que es resultado del espectro de la fase de la imagen $e^{\phi_T(u,v)}$ proyectado en las “Corefaces” (vectores propios) v_{f_i} .

Por tanto, la representación de una imagen de entrada en el dominio de frecuencia quedaría, cancelando las fases:

$$C(u, v) = e^{-\phi_R(u,v)} e^{\phi_E(u,v)} = e^{\phi_E(u,v) - \phi_R(u,v)} \quad (41)$$

Mientras tanto, para llevarla al dominio espacial se aplicaría la transformada inversa de Fourier, como se muestra en la Ecuación 42, lo que permitirá ver en la salida del filtro de correlación la agudeza del pico y calcular *PSR* como se propuso en la Ecuación 31.

$$c_T(x, y) = \sum_{u=0}^{d-1} \sum_{v=0}^{d-1} e^{\phi_E(u,v) - \phi_R(u,v)} \cdot e^{\frac{j2\pi ux}{d}} e^{\frac{j2\pi vy}{d}} \quad (42)$$

Este método presenta como principal desventaja que para el entrenamiento es necesario contar con imágenes del mismo sujeto bajo diferentes condiciones de iluminación, para poder construir el subespacio con el cual comparar la imagen de entrada.

2.2.4 Filtro de Correlación Cuaternión

Este método consta de 2 estados, obtener una descripción de la imagen utilizando la descomposición wavelet y el diseño de un filtro de correlación cuaternión. La descomposición wavelet suministra un análisis multi-resolución de la imagen, con el objetivo de ofrecer rasgos adicionales que son localizados conjuntamente en la frecuencia y en el espacio, el cual ha tenido éxito al aplicarse en el reconocimiento de rostros [27]. En la propuesta que se presenta en [28] se aplica una descomposición wavelet discreta (DWT por sus siglas en inglés), para transformar la imagen de rostro en una representación multi-bandas, y se aplican filtros de correlación cuaternión para ejecutar un proceso multi-banda de forma paralela.

El filtro de correlación cuaternión está diseñado para reconocimiento multi-canal 2D. Estos filtros de correlación cuaternión se basan en el álgebra del cuaternión, originalmente introducido por Hamilton en 1843 [29]. Un número cuaternión es una generalización del número complejo y se puede considerar como un número con una parte real y una parte imaginaria compuesta de tres componentes ortogonales de la siguiente manera [30]:

$$q = a + bi + cj + dk \quad (43)$$

donde, a , b , c y d son reales; mientras que i , j , k son operadores imaginarios.

Por tanto una representación cuaternión puede estar compuesta por cuatro canales, uno para cada sub-banda. El filtro de correlación cuaternión no solamente modela características de la estructura intra-canal como lo hace el filtro de correlación tradicional, sino que también modela características de la estructura inter-canal; dado que, conjuntamente procesa los canales múltiples, y así proporciona una manera natural de combinar las salidas de la correlación de los 4 canales.

La transformación wavelet descompone la señal original en escalas y resoluciones diferentes, proporcionando mayor comprensión de las características espacio-frecuencia juntas de la señal original. La descomposición wavelet discreta (DWT) puede ser vista como una estructura arbórea, donde la señal original es pasada a través de un filtro paso bajo (B) y un filtro paso alto (A), y luego es reducida para obtener las componentes de las bajas y las altas frecuencias de la señal original respectivamente, como se observa en la Fig. 12. La descomposición puede ser aplicada iterativamente a las bandas de bajas frecuencias y a las de altas frecuencias para generar un árbol de descomposición wavelet.

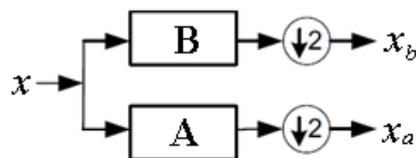


Fig. 12. Árbol de descomposición wavelets.

Este proceso se implementa al proyectar la señal original a las funciones bases wavelet para obtener los coeficientes. Los wavelets 2D son el producto cartesiano de la onda 1D, entonces una imagen 2D será descompuesta en 4 partes correspondientes a las bandas de frecuencia, aplicando una descomposición wavelet de segundo nivel. Las bandas obtenidas serán

representadas a partir de este momento por dos letras que se refieren a la combinación de filtros utilizada para obtener estas descomposiciones. Las combinaciones pueden ser BB , AB , BA o AA . Un ejemplo de la descomposición de ondas de una imagen de rostro se presenta en la Fig. 13. En la parte derecha de dicha figura se muestran las imágenes de cada banda; donde, la imagen de la descomposición BB se encuentra en la posición superior izquierda, la de AB en la superior derecha, la de BA en la inferior izquierda y la de AA en la inferior derecha.



Fig. 12. Izquierda: imagen original; Derecha: descomposición de ondas de primer nivel.

Se introduce entonces, en [28], la combinación de la descomposición wavelet y los filtros de correlación cuaternión para el reconocimiento de rostros. La definición de cómo este método se utiliza incluye tanto la etapa de almacenamiento como la de reconocimiento. Durante la etapa de almacenamiento, se procesan una o múltiples imágenes de cada individuo. La DWT 2D de estas imágenes de entrenamiento es ejecutada para descomponer las imágenes originales en las cuatro sub-bandas. Estas imágenes sub-bandas son llevadas a un arreglo de cuaterniones 2D de la manera siguiente:

$$f(x) = f_{BB}(x) + f_{BA}(x)i + f_{AB}(x)j + f_{AA}(x)k \quad (44)$$

donde, x representa las coordenadas de la imagen, y $f_{BB}(x)$, $f_{BA}(x)$, $f_{AB}(x)$ y $f_{AA}(x)$ representan las cuatro imágenes de sub-banda de onda.

La transformada de Fourier cuaternión (QFT por su siglas en inglés) [30] es ejecutada para transformar la imagen cuaternión al dominio de frecuencia cuaternión. El filtro de correlación cuaternión es diseñado a partir de la QFT de las imágenes de entrenamiento y es almacenado para cada sujeto, como se muestra en la Fig. 14.

En la etapa de reconocimiento, la imagen de entrada es transformada al dominio de frecuencia cuaternión de la misma manera que se hizo para las imágenes almacenadas. Luego a la representación del dominio de frecuencia cuaternión resultante se le aplica la correlación cruzada con cada filtro correlación cuaternión (QCF por su siglas en inglés) en la base de datos, usando la correlación cuaternión 2D especializada propuesta en [30]. Del valor de la magnitud de cada salida de la correlación cuaternión, se calcula un grado de similitud. La imagen de entrada será etiquetada en la clase con mayor valor de similitud. El proceso de reconocimiento se ilustra en la Fig. 15.



Fig. 13. Etapa de almacenamiento del método Filtro de Correlación Cuaternión.

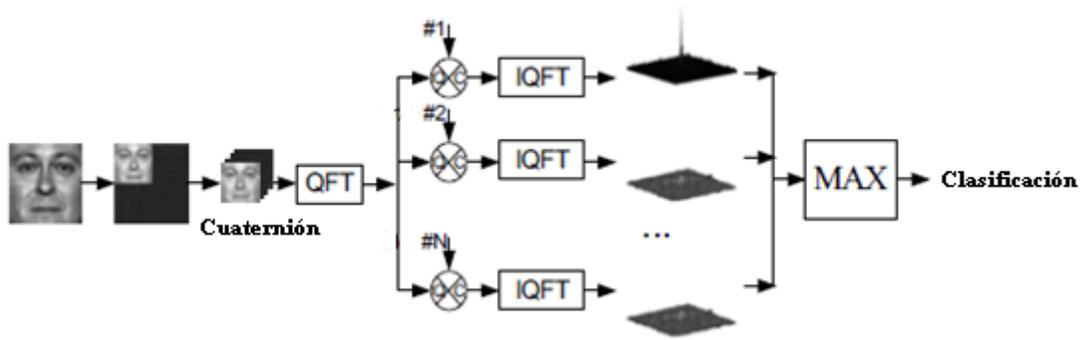


Fig. 14. Etapa de reconocimiento del método Filtro de Correlación Cuaternión.

Un filtro de correlación bien diseñado retornaría valores de pico elevados en el plano de salida de la correlación de una identidad verdadera y no sería capaz de discernir picos para otras personas. Para lograr la agudeza del pico de correlación, el filtro debería ser diseñado para suprimir los valores picos en una vecindad, o sea, los valores de correlación alrededor del pico. A la misma vez, la varianza del ruido de salida de la correlación debería ser minimizada para suprimir el efecto del ruido de entrada. Esto condujo a definir en [28] el filtro cuaternión de equilibrio óptimo sin restricciones (UOTQF por sus siglas en inglés), el cual se define en la Ecuación 45. En [29] se encuentra la definición de este filtro y en [31] su forma original, que es el filtro de equilibrio óptimo sin restricciones (UOTF por sus siglas en inglés). El filtro UOTQF se define de la siguiente manera:

$$h = \gamma(\alpha D + C)^{-1}m \quad (45)$$

donde, h es el filtro diseñado en el dominio de frecuencia que es representado en forma de vector, m representa la media del dominio de frecuencia de las imágenes de entrenamiento en forma de vector. D es una matriz diagonal, donde la diagonal principal es el espectro de potencia promedio de las imágenes de entrenamiento. C es también una matriz diagonal, representando la densidad espectral de la potencia del ruido. Normalmente se asume que el modelo no tiene ruido, por lo que C es una matriz identidad. Finalmente α y γ son parámetros de equilibrio, los cuales pueden ser ajustados para obtener un equilibrio óptimo entre la maximización, la habilidad discriminatoria y la minimización de la varianza del ruido en la salida del filtro h . Para afinar la agudeza del pico se utiliza PSR definido en la Ecuación 31.

En la sección de experimentación de [28] se muestran los excelentes resultados que se alcanzan con esta técnica utilizando una sola imagen con buenas condiciones de iluminación por sujeto para el entrenamiento. Si se utiliza una imagen con variaciones de iluminación esto no sucede. Por tanto, este método posee buenas propiedades discriminatorias si el proceso de entrenamiento es llevado a cabo con una imagen con buenas condiciones de iluminación por cada sujeto.

No obstante, este método propuesto en [28] posee deficiencias que pudieran ser mejoradas, en aras de obtener mejores resultados. Primeramente, no se justifica la utilización de la descomposición wavelet DWT 2D, por tanto, se pudiera realizar un estudio para evaluar otras posibles representaciones que podrían brindar mejores resultados. Además pudieran utilizarse funciones de transformación al dominio de la frecuencia que de manera directa construyan la representación cuaternión a partir de la imagen y no que se forme con coeficientes que se obtienen en un proceso aparte.

2.3 Métodos de Extracción de Rasgos Derivados de la Imagen

Los métodos que se ubican en este grupo se basan en realizar un análisis de rasgos extraídos de la superficie del rostro. Las características del rostro utilizadas para llevar a cabo el análisis pueden ser los bordes, las cordilleras y los valles o patrones derivados de las diferencias de las intensidades de los píxeles. En este epígrafe se presentan los métodos: *Descriptor de Cordilleras y Valles*, *Mapas de Bordes de Líneas* y *Patrones Binarios Locales*.

2.3.1 Descriptor de Cordilleras y Valles.

En la literatura, se puede encontrar una amplia gama de caracterizaciones matemáticas que tratan de formalizar la noción intuitiva de cordillera/valle³. En el método propuesto por Pujol en [34] se estudiaron los operadores *MLSEC*⁴ (por sus siglas en inglés) [32] y el *MLSEC-ST*⁵ (por sus siglas en inglés) [33] debido a su invariancia tanto a movimientos de la imagen rígida, como a los cambios del nivel de gris monótonos y, sobre todo, debido a su elevada continuidad en un rango dinámico significativo, en oposición a otras medidas con la misma invariancia. Para el reconocimiento de rostros, se utilizaron los valles y cordilleras obtenidos al asignar umbrales al operador *MLSEC-ST*.

Una de las propiedades interesantes que ha hecho que se eligiera este descriptor es que se comporta bien frente a los cambios de iluminación. Esta propiedad se debe al hecho que la respuesta del operador depende de las orientaciones de los campos normalizados del gradiente en lugar de sus magnitudes. Además, las relaciones entre estas orientaciones son insensibles al contraste en la vecindad local y a los cambios de intensidad del brillo como los producidos por las variaciones de la iluminación.

Para utilizar los operadores *MLSEC* y *MLSEC-ST* deben seleccionarse varios parámetros. Estos son: i) el valor del umbral y la varianza del kernel gaussiano, en el paso de pre-procesamiento para suavizar al aplicar el operador *MLSEC*; ii) la integral, σ_i , la derivada, σ_d , los parámetros sigma y el umbral final en el caso del operador *MLSEC-ST*. Para ajustar estos parámetros respecto a la información del rostro, en [34] se expresan los factores σ_i y σ_d en proporción a la distancia intra-ocular, que es la distancia entre los puntos medios de los ojos.

En [32] se introduce una medida de confianza para filtrar el ruido, es decir, cordilleras y valles de la imagen que sean irrelevantes. Para ello los autores proponen usar una medida de anisotropía de la estructura de la imagen subyacente. Teniendo en cuenta que los valores propios del tensor estructural miden el gradiente del vector en las direcciones de los vectores propios ellos proponen usar la diferencia entre el mayor y el menor valor propio como una medida de la dirección dominante o la anisotropía. Aplicando una función gaussiana invertida a esta medida se obtiene la medida de confianza $C(x, y; \sigma_d) \in [0, 1]$ de la siguiente manera:

$$C(x, y; \sigma_d) = 1 - e^{-\frac{(\lambda_1(x, y; \sigma_d) - \lambda_2(x, y; \sigma_d))^2}{2c^2}} \quad (46)$$

donde, $\lambda_1(x, y; \sigma_d)$ y $\lambda_2(x, y; \sigma_d)$ son respectivamente el mayor y el menor valor propio del tensor estructural. El coeficiente de confianza $C(x, y; \sigma_d)$ es ajustado usando el valor global c que depende de la característica estructural del rasgo que se desea mejorar. Esta medida de confianza es usada para darle pesos a las respuestas del operador *MLSEC-ST*, cuando se trabaja como un umbral.

³ Las cordilleras y valles son equivalentes en el sentido que las cordilleras de una imagen son los valles de la imagen invertida, y viceversa.

⁴ MLSEC: Multi Local Level Set Extrinsic Curvature.

⁵ MLSEC-ST: Multi Local Level Set Extrinsic Curvature based on Structure Tensor.

A diferencia de la respuesta del operador de pliegues (*MLSEC-ST*), la diferencia entre los valores propios del tensor estructural es altamente dependiente a los cambios del contraste local en la imagen. Este puede provocar la pérdida de las propiedades de invariancia a los cambios de iluminación deseables, que presenta el operador *MLSEC-ST*.

La dependencia del contraste local de la medida de confianza puede ser parcialmente atenuada usando un factor localmente dependiente \tilde{c} ,

$$\tilde{C}(x, y; \sigma_d) = 1 - e^{-\frac{(\lambda_1(x, y; \sigma_d) - \lambda_2(x, y; \sigma_d))^2}{2\tilde{c}(x, y; \sigma_d)}} \quad (47)$$

El factor \tilde{c} es una estimación de la varianza del nivel de gris local, y es usado como un factor de normalización para evitar que el contraste del nivel de gris incida en la diferencia de los valores propios, donde:

$$\tilde{c}(x, y; \sigma_d) = G(x, y; \sigma_d) * (L(x, y) - \mu(x, y; \sigma_d))^2 \quad (48)$$

* denota convolución de la imagen, y:

$$\mu(x, y; \sigma_d) = G(x, y; \sigma_d) * L(x, y) \quad (49)$$

El resultado del operador *MLSEC-ST*, con la función de confianza \tilde{C} se comporta de manera más robusta ante los cambios de iluminación que el criterio global C .

Según [34], una de las ventajas principales del operador *MLSEC-ST*, en comparación con otros operadores de pliegues es su bien definido y significativo rango dinámico. Su respuesta está limitada entre -2 y 2 (en el caso de imágenes bidimensionales). Con el fin de fijar un umbral de cordillera y valle adecuado midieron experimentalmente en datos artificiales la relación entre la estructura y la medida de pliegue, y cómo esta medida varía para estructuras de imágenes diferentes. Al final de este concluyeron que un punto de la imagen se considerará cordillera si su valor, después de aplicar el *MLSEC-ST*, es mayor que 0.75 y será un valle si es menor que -0.75.

Además constataron que aplicando el mismo operador sobre la imagen resultante, se mejoran sucesivamente los resultados obtenidos por lo que construyendo una pirámide de niveles podría obtenerse en el nivel superior una mejor descripción de la imagen.

Las distancias de Hausdorff son técnicas de correspondencia flexibles que permiten medir la distorsión, léase desplazamiento o reorganización de los rasgos de la imagen, en lugar de los cambios en las intensidades de la imagen. En [34] se presentan modificaciones a la distancia de Hausdorff que se deriva de la distorsión observada en un conjunto de entrenamiento lo que conlleva a una medida de similitud más adecuada para el reconocimiento de rostros.

La distancia Hausdorff es una medida de similitud entre dos conjuntos de puntos pertenecientes al mismo espacio métrico. La principal ventaja de esta medida es que no se requiere de una correspondencia explícita entre los puntos. Este requerimiento es superado a través de una correspondencia con el vecino más cercano implícita entre los puntos del conjunto. Formalmente, dado dos conjuntos de puntos A y B , la distancia de Hausdorff $H(A, B)$ entre ellos es definida como:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (50)$$

$$h(A, B) = \max_{a \in A} \left(\min_{b \in B} (d(a, b)) \right) \quad (51)$$

donde, $h(A, B)$ es la distancia de Hausdorff directa del conjunto A al conjunto B , y $d(a, b)$ la distancia, usualmente euclidiana, entre los puntos a y b . Debemos notar que la distancia de Hausdorff no es simétrica.

De manera general la distancia de Hausdorff se puede expresar como:

$$H(A, B) = f^s(h(A, B), h(B, A)) \quad (52)$$

$$h(A, B) = f_{a \in A}^i \left(f^m \left(\min_{b \in B} (d(a, b)) \right) \right) \quad (53)$$

Se llamarán “simetrizador”, “integral” y “mejor-emparejamiento” a las funciones f^s , $f_{a \in A}^i$ y f^m , respectivamente.

Varias modificaciones a la distancia de Hausdorff han sido propuestas combinando distintas funciones como “simetrizador”, “integral” y “mejor-emparejamiento”. La propuesta que se presenta en [34] se llama Distancia de Hausdorff Supervisada (*SHD* en inglés).

Considerando los valles tomados de la respuesta del operador *MLSEC* se define una imagen binaria $I_A(x, y)$. El conjunto $A = \{(x, y) : I_A(x, y) = 1\}$ de las posiciones de la imagen donde se detectaron valles es considerado como el descriptor de la imagen. Utilizando la función media como integral, la distancia de Hausdorff quedaría:

$$h(A, B) = \frac{1}{|A|} \sum_{a \in A} f^m \left(\min_{b \in B} (d(a, b)) \right) \quad (54)$$

$$= \frac{1}{\sum_x \sum_y I_A(x, y)} \sum_x \sum_y I_A(x, y) f^m \left(\min_{\{(i, j) : I_B(i, j) = 1\}} d((x, y), (i, j)) \right) \quad (55)$$

La expresión más a la derecha de la Ecuación 55 es equivalente a la transformación de la distancia, $DT(x, y; I_B)$, de la imagen binaria $I_B(x, y)$, que sería una imagen del mismo tamaño que la imagen binaria, $I_B(x, y)$, donde cada pixel (x, y) es la distancia euclidiana de la posición (x, y) a la posición (i, j) del pixel más cercano distinto de 0 en la imagen binaria $I_B(x, y)$.

Nombrando la imagen binaria (I_A), su $DT(D_A)$ y la función mejor-emparejamiento $f^m(D_A)$ las Ecuaciones 54 y 55 serían:

$$h(A, B) = \left\langle \frac{I_A}{|I_A|}, f^m(D_A) \right\rangle \quad (56)$$

Aunque las tres funciones pudieran ser modificadas, en esta propuesta solamente se modifica la función mejor-emparejamiento y se utiliza como integral y simetrizador la función media. La función mejor-emparejamiento propuesta está diseñada de tal manera que las contribuciones a la distancia de cada punto del rostro estén normalizadas, proporcionalmente al grado de transformación de cada uno (por ejemplo los ojos y la nariz son regiones más estáticas que la barbilla y las cejas). Esta función es aprendida durante una fase de entrenamiento.

Dado un conjunto de clases, o sea, sujetos $C = \{C_1, C_2, \dots, C_n\}$ y sea $C_i = \{D_1^i, D_2^i, \dots, D_m^i\}$ el DT de las imágenes binarias de cada sujeto de la clase C_i . El resultado de aplicar la función

mejor-emparejamiento a un vector de $DT(D_B^i \in C_i)$ es un vector $f^m(D_B^i)$ del mismo tamaño que D_B^i , donde su n -ésima componente, $f_n^m(D_{B,n}^i)$ se define como:

$$f_n^m(D_{B,n}^i) = \frac{D_{B,n}^i}{\sigma_n} \quad (57)$$

donde, σ_n y $D_{B,n}^i$ son las n -ésimas componentes del vector de desviación estándar de la clase-interna y del vector DT , respectivamente.

$$\sigma_n = \sqrt{\frac{1}{|C|} \sum_{C_i \in C} \frac{1}{|C_i|} \sum_{D_{j,n}^i \in C_i} (D_{j,n}^i - \bar{D}_n^i)^2} \quad (58)$$

, y

$$\bar{D}_n^i = \frac{1}{|C_i|} \sum_{D_{j,n}^i \in C_i} D_{j,n}^i \quad (59)$$

donde, $||$ significa la cardinalidad del conjunto.

El vector σ se calcula a partir de la varianza de las imágenes de entrenamiento.

En [34] se manifiesta que la medida de Hausdorff propuesta se ha demostrado que posee mejor rendimiento cuando se tiene una sola imagen por cada sujeto.

Este método posee como principal desventaja que la función utilizada para eliminar las cordilleras y valles irrelevantes no es invariante a la iluminación.

2.3.2 Mapas de Bordes de Líneas

Los estudios de psicología cognitiva [35] [36] han mostrado que los seres humanos reconocen objetos hechos con líneas con la misma exactitud y rapidez que cuando estos objetos están representados en imágenes en tono gris. Takács [37] utilizó mapas de bordes, motivado por el fenómeno anterior, para medir la similitud entre imágenes de rostro. Estas imágenes se convertían en mapas de bordes binarios usando el algoritmo de Sobel para la detección de bordes. La distancia de Hausdorff fue elegida como medida de similaridad entre los conjuntos que representaban a dos imágenes de rostro, debido a que esta métrica se puede calcular sin que exista una paridad explícita entre los puntos de estos conjuntos. La expresión utilizada por el autor en [37] fue $h(A, B) = \frac{1}{N_A} \sum_{a \in A} \min_{b \in B} \|a - b\|$, con el objetivo de que el cálculo fuera menos sensible al ruido, obteniéndose una efectividad de un 92% en los experimentos de reconocimiento de rostros realizados con imágenes que no tenía problemas de iluminación.

La modificación de la distancia de Hausdorff propuesta en [37] tiene en cuenta solamente la información espacial sin considerar características locales de la estructura del mapa de bordes. Esta fue la motivación para que se propusiera en [38] el método *Mapa de Bordes de Líneas* (*LEM* por sus siglas en inglés), el cual se considera una combinación entre la correspondencia por plantilla y la correspondencia por rasgos geométricos. Los autores expresan que *LEM* es invariante a la iluminación y tiene poco requerimiento de memoria.

Según [38] el *LEM* es una representación de los rasgos propuesta para integrar la información espacial y estructural de la imagen de rostro, agrupando píxeles del mapa de bordes en segmentos de líneas. Una vez obtenido el mapa de bordes, se lleva a cabo un proceso de ajuste de líneas para generar el *LEM* del rostro [39]. Un ejemplo del resultado obtenido se puede apreciar en la Fig. 16.



Fig. 15. Ilustración de un mapa de bordes de línea (LEM) aplicado a un rostro.

La representación *LEM* debe ser menos sensible a los cambios de iluminación debido a que es un nivel intermedio entre la imagen y el mapa de bordes según [38]. La unidad base del *LEM* es el segmento de línea agrupado a partir de los píxeles del mapa de borde.

En [38] es presentada una nueva distancia de Hausdorff para comparar representaciones *LEM*, la distancia de Hausdorff entre segmentos de línea (*LHD* siglas en inglés). *LHD* tiene mayor poder distintivo porque puede usar atributos adicionales de la orientación de las líneas, las asociaciones línea-puntos, y la disparidad *LEM*, que se verán más adelante.

Dadas dos representaciones *LEM*: $M^l = \{m_1^l, m_2^l, \dots, m_p^l\}$, del modelo *LEM* de la base de datos y $T^l = \{t_1^l, t_2^l, \dots, t_q^l\}$, de la imagen de entrada; *LHD* se construye con el vector $\vec{d}(m_i^l, t_j^l)$ que representa la distancia entre los segmentos de línea m_i^l y t_j^l . Este vector se define como:

$$\vec{d}(m_i^l, t_j^l) = \begin{bmatrix} d_\theta(m_i^l, t_j^l) \\ d_{//}(m_i^l, t_j^l) \\ d_\perp(m_i^l, t_j^l) \end{bmatrix}$$

donde, $d_\theta(m_i^l, t_j^l)$, $d_{//}(m_i^l, t_j^l)$ y $d_\perp(m_i^l, t_j^l)$ son la distancia de orientación, la distancia paralela y la distancia perpendicular, respectivamente. Estas tres componentes son independientes y se definen como:

$$d_\theta(m_i^l, t_j^l) = f(\theta(m_i^l, t_j^l)) \quad (60)$$

$$d_{//}(m_i^l, t_j^l) = \min(l_{//1}, l_{//2}) \quad (61)$$

$$d_\perp(m_i^l, t_j^l) = l_\perp \quad (62)$$

donde, $\theta(m_i^l, t_j^l)$ determina el menor ángulo de intersección entre las líneas m_i^l y t_j^l , f es una función no lineal de penalización por llevar un ángulo del mapa a un escalar. En el caso de [38]

se utilizó la función cuadrática $f(x) = x^2/W$ donde, W es un peso que se determinó en un proceso de entrenamiento.

La idea del desplazamiento paralelo y el perpendicular se pueden ilustrar con un ejemplo simplificado de dos líneas paralelas como se muestra en la Fig. 17. La componente $d_{//}(m_i^l, t_j^l)$ se define como el mínimo desplazamiento para alinear bien los puntos del extremo izquierdo o los puntos del extremo derecho de las líneas. Mientras que $d_{\perp}(m_i^l, t_j^l)$ es simplemente la distancia vertical entre las dos líneas. En general, m_i^l y t_j^l podrían no ser paralelas, pero se puede rotar la menor línea con su punto medio como centro a la orientación deseada antes de calcular $d_{//}(m_i^l, t_j^l)$ y $d_{\perp}(m_i^l, t_j^l)$.

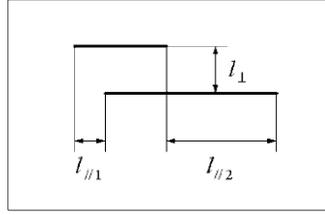


Fig. 16. Posibles desplazamientos para alinear dos líneas.

Los cambios de la paralela $l_{//1}$ y $l_{//2}$ son 0 si las dos líneas son del mismo tamaño, si alguno de los dos desplazamientos es 0 o si ambos desplazamientos son iguales como se muestra en la Fig. 18.

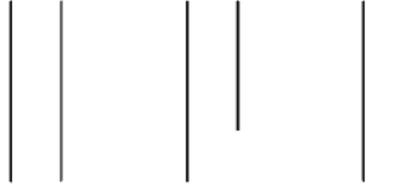


Fig. 17. Casos en los que $d_{//}(m, t) = 0$.

Finalmente, la distancia entre dos segmentos de línea m_i^l y t_j^l se define de la siguiente manera:

$$d(m_i^l, t_j^l) = \sqrt{d_{\theta}^2(m_i^l, t_j^l) + d_{//}^2(m_i^l, t_j^l) + d_{\perp}^2(m_i^l, t_j^l)} \quad (63)$$

La distancia primaria de Hausdorff de segmentos de línea (*pLHD*) en [38] se define como:

$$H_{pLHD}(M^l, T^l) = \max(h(M^l, T^l), h(T^l, M^l)) \quad (64)$$

donde:

$$h(M^l, T^l) = \frac{1}{\sum_{m_i^l \in M^l} l_{m_i^l}} \sum_{m_i^l \in M^l} l_{m_i^l} \cdot \min_{t_j^l \in T^l} d(m_i^l, t_j^l) \quad (65)$$

y l_{m_i} es la longitud del segmento de línea m_i^l . En la Ecuación 65 la contribución de cada línea se asocia a su longitud.

El número de parejas de líneas correspondientes entre la entrada y el modelo es otra medida de similitud. El número de parejas de líneas correspondientes entre dos imágenes idénticas debería ser mayor que entre dos imágenes de objetos diferentes. Por tanto este problema puede ser atenuado al introducir esta información dentro de *pLHD*.

Asumiendo que, para cada línea t_j en el *LEM* de entrada T , su línea correspondiente m_i en el modelo M de un rostro idéntico debería estar ubicada cerca de t_j porque la imagen de entrada y el modelo de entrada han sido alineados y se ha normalizado su escala por un pre-procesamiento antes de la comparación, entonces, una vecindad de la posición V_p y una vecindad del ángulo V_α se introducen. La Vecindad de Similaridad V_s es una combinación de V_p y V_α de la siguiente manera:

$$V_s = V_p \cap V_\alpha$$

Si al menos una línea del modelo M se localiza dentro de la vecindad de similitud de una línea t_j en el *LEM* de entrada T , esto significa que, dicha línea se ubica en la vecindad de posición dada de t_j y la diferencia entre los ángulos está también dentro de la vecindad del ángulo dada. Por tanto, es más probable para t_j encontrar una línea correspondiente correcta entre aquellas líneas dentro de la vecindad de similaridad. Esta línea (t_j) se llama línea de alta confianza. La proporción de líneas de alta confianza (R) es definida como el cociente del número de líneas de alta confianza (V_{hc}) y el número total de líneas en el *LEM* (V_{total}):

$$R = \frac{V_{hc}}{V_{total}} \quad (66)$$

Teniendo en cuenta el efecto de la vecindad de similitud en el cálculo de *LHD*, este quedaría:

$$H_{LHD}(M, T) = \sqrt{H_{pLHD}^2(M, T) + (W_n D_n)^2} \quad (67)$$

donde, $H_{pLHD}(M, T)$ es el definido en la Ecuación 64 y W_n es el peso del número de disparidad D_n . El número de disparidad entre dos representaciones *LEM* es definido como el promedio de los valores R de cada uno:

$$D_n = 1 - \frac{R_M + R_T}{2} = \frac{(1 - R_M) + (1 - R_T)}{2} \quad (68)$$

donde, R_M y R_T son la proporción de líneas de alta confianza de cada representación *LEM*.

Una manera de determinar los parámetros (W, W_n, V_p, V_α) de *LHD* en un sistema de reconocimiento de rostros es tomar los valores con menor tasa de error durante un proceso de entrenamiento sobre una base de datos especializada, usando *LHD* como medida de comparación. Por ejemplo utilizando la base de datos de imágenes de rostros de la Universidad de Berna [40] los valores fueron: $W = 30, W_n = 5, V_p = 6, V_\alpha = 30$.

El principal problema de esta propuesta radica en que los métodos que se utilizan para extraer los bordes no son invariantes a la iluminación. Por tanto, no se garantiza un alto rendimiento a este método para imágenes con problemas de iluminación.

2.3.3 Patrones Binarios Locales

El operador de Patrones Binarios Locales (LBP, por sus siglas en inglés) es uno de los descriptores de textura con mejores resultados y ha sido usado en numerosas aplicaciones. Se ha mostrado que este operador es altamente discriminatorio y tiene varias ventajas como por ejemplo, la invarianza a los cambios monotónicos de los niveles de gris y la eficiencia computacional, lo cual lo hace apropiado para las tareas demandadas en el análisis de imágenes. La idea de usar el LBP para la descripción de rostros está motivada por el hecho de que los rostros pueden ser vistos como una composición de micro-patrones los cuales son bien descritos por él [41].

El operador LBP fue diseñado originalmente para la descripción de textura. Asigna una etiqueta a cada píxel de una imagen usando en una vecindad de 3x3 píxeles el valor del píxel central como umbral y considerando el resultado como un número binario, como se ilustra en la Fig. 19. Luego el histograma de las etiquetas es utilizado como descriptor de la textura.

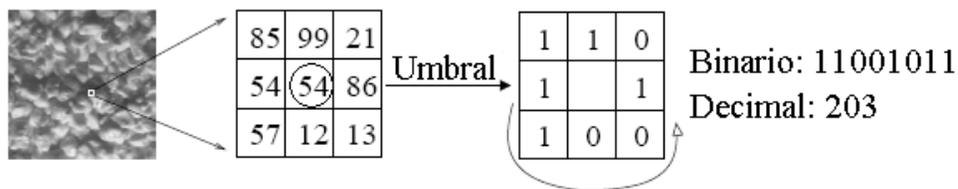


Fig. 18. El operador LBP básico.

El operador fue posteriormente extendido para usar vecindades de diferentes tamaños [42]. Definiendo una vecindad local como un conjunto de puntos uniformemente espaciados en un círculo centrado en el píxel que será etiquetado, permite usar cualquier radio y cualquier número de puntos de ejemplo. Se utiliza interpolación bilineal cuando el punto de ejemplo no cae en el centro de un píxel. Se utiliza la notación (P,R) para vecindades de píxeles, lo cual significa P puntos de ejemplo en un círculo de radio R. La Fig. 20 muestra un ejemplo de vecindades circulares.

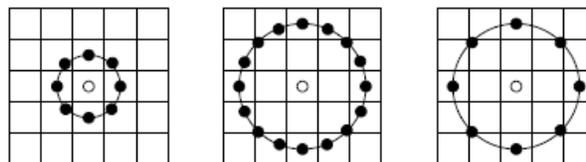


Fig. 9. Vecindades circulares de (8,1), (16,2) y (8,2).

Otra extensión del operador original es la definición de los llamados patrones uniformes [42]. Un patrón binario es llamado uniforme si contiene como máximo dos transiciones de bits de 0 a 1 o viceversa, cuando el patrón es considerado circular. Por ejemplo, los patrones 00000000 (0 transiciones), 01110000 (2 transiciones) y 11001111 (2 transiciones) son uniformes, mientras que los patrones 11001001 (4 transiciones) y 01010011 (5 transiciones) no lo son. En el cálculo del histograma LBP, los patrones uniformes son utilizados de manera que el histograma tiene un depósito (barra) para cada patrón uniforme, mientras que todos los patrones no uniformes son asignados a un mismo depósito.

Posteriormente han aparecido nuevas extensiones del LBP. Por ejemplo, en [43] se advierte que el operador LBP no representa bien la estructura local en determinadas circunstancias y se introduce el LBP mejorado, donde el código binario se establece a partir de la comparación con el promedio de los valores de intensidad de los píxeles en la vecindad, en lugar de con el píxel central.

En cualquiera de sus variantes, los histogramas LBP contienen información acerca de la distribución de los micro-patronos locales, como bordes, manchas y áreas lisas. Para una representación eficiente del rostro, se debe retener también información espacial. Para este propósito la imagen se divide en regiones R_0, R_1, \dots, R_{m-1} y los descriptores son extraídos de cada una de estas regiones independientemente, los descriptores son entonces concatenados para formar el descriptor global de la imagen obteniendo un histograma mejorado espacialmente.

En el histograma mejorado espacialmente, se tiene efectivamente la descripción del rostro en tres niveles diferentes de localización: las etiquetas LBP para los histogramas contienen información acerca de los patrones a niveles de píxeles, las etiquetas se suman en una región pequeña para producir información a un nivel regional y luego los histogramas son concatenados para construir una información global del rostro.

Desde el punto de vista de la clasificación, un problema usual es tener muchas clases y solo muy pocas, probablemente solo una, muestra(s) de ejemplo por clase. Muchas medidas de similitud han sido propuestas para tratar este problema con histogramas: intersección de histogramas, estadística de probabilidad logarítmica, estadística Chi al cuadrado [41].

Cuando la imagen es dividida en regiones, se espera que algunas regiones contengan más información útil que otras, en términos de distinguir un individuo de otros, por ejemplo los ojos y la boca. Para tomar ventaja de esto, un peso puede ser asignado para cada región basándose en la importancia de la información que contiene. Por ejemplo, la distancia pesada Chi al cuadrado puede ser expresada como:

$$\chi_w^2(x, \xi) = \sum_{j,l} w_j \frac{(x_{i,j} - \xi_{i,j})^2}{x_{i,j} + \xi_{i,j}} \quad (69)$$

donde, x y ξ son los histogramas mejorados normalizados para ser comparados, los índices i y j se refieren a la i -ésima barra del histograma correspondiente y a la j -ésima región local y w_j es el peso para la región.

La idea del uso de LBP como rasgos viene dada porque la imagen de rostro puede ser vista como una composición de micro-patronos de bordes, puntos, áreas sobresalientes, entre otros. Ya que el operador se basa en la comparación entre las intensidades de los pixeles vecinos, sin importar en qué magnitud sean mayores o menores. Muchos autores declaran que el operador es invariante a los cambios de iluminación; esto significa que el operador es capaz de describir los rasgos independientemente de las variaciones de iluminación que afecten la imagen.

Si miramos los pixeles bajo el modelo Lambertiano, podemos considerar:

$$I_c = \rho_c \cdot n_c^T \cdot s_c \quad \text{y} \quad I_1 = \rho_1 \cdot n_1^T \cdot s_1 \quad (70)$$

donde ρ es el albedo, n^T la normal a la superficie del objeto y s la luz que incide en el punto, I_c representa la intensidad del pixel central e I_1 la intensidad de un pixel vecino de I_c .

Tanto ρ como n^T dependen de la forma y la textura de la superficie y se espera que en una pequeña vecindad estas sean similares. Luego, lo que realmente representa la diferencia entre dos pixeles vecinos descrita por el operador LBP, es la diferencia entre las iluminaciones incidentes [44]:

$$I_1 - I_c = \rho \cdot n^T \cdot (s_1 - s_c) \quad (71)$$

En ese caso, solamente si las variaciones en la iluminación son monotónicas, es decir, si el signo de la diferencia $(s_1 - s_c)$ se preserva, el operador se comporta invariante. En cualquier otro

caso, más usuales en las aplicaciones de la vida real, la descripción mediante el LBP de una vecindad cambia según cambia la iluminación que incide en ella, esto significa entonces que el operador es sensitivo a este tipo de variaciones de iluminación.

3 Comparación de la efectividad entre los métodos descritos

En esta sección se comparan los métodos analizados. Para ello se presenta una tabla del rendimiento de los distintos métodos que se han presentado aplicados a imágenes de rostros con problemas de iluminación. La comparación se realiza en cuanto a tasa de error igual (EER por sus siglas en inglés), que no es más que el valor en el que la proporción de falsos aceptados es igual a la proporción de falsos rechazados. El valor EER es representativo de la capacidad discriminatoria de cada una de estas técnicas, mientras menor es el valor, mayor es la capacidad discriminativa de los rasgos extraídos por el método.

Tabla 1. Comparación de la eficacia entre los métodos de extracción de rasgos estudiados.

Método	Base de Datos	Rendimiento en imágenes con problemas de iluminación (EER)
Píxel Original	XM2VTS	14.83
Píxel Original Menos La Media	XM2VTS	5.86
PCA Local	YALE	25.3
Fisherfaces	YALE	20.0
Laplacianfaces	YALE	16.0
DCT	XM2VTS	4.83
Gabor	CMU-PIE	0.0
Corefaces	CMU-PIE	0.08
Filtro Correlación Cuaternión	CMU-PIE	0.05
Descriptor Cordillera y Valles	AR	7.0
Mapas de Borde de Líneas	YALE B	14.55
Patrones Binarios Locales	FERET	3.0

Los resultados mostrados en la Tabla 1 son los reportados por los autores en diferentes Bases de Datos, aunque todas son reconocidas por los especialistas, para evaluar los sistemas biométricos de reconocimiento de rostro. Estas Bases de Datos poseen conjuntos de imágenes tomadas con este fin, lo que implica que cada uno de ellos tiene condiciones particulares. Esto puede influir en las tasa de reconocimiento de los sistemas, en otras palabras, que su efectividad se comporte de forma diferente en cada una de ellas. A pesar de esto, las tasas de reconocimiento son bastante representativas de la efectividad de los métodos.

Como puede apreciarse, los métodos de transformación en el dominio de la frecuencia son los que presentan la mayor efectividad de los presentados en la Tabla 1. De hecho, presentan un rendimiento excelente pues sus tasas de error son muy bajas, menores que 0.1. En el caso del método que utiliza la transformada de Fourier y el que usa la transformada de Gabor necesitan imágenes de entrenamiento con diferentes condiciones de iluminación por cada sujeto, por lo

que es de esperar su alto rendimiento. Sin embargo, la presentación de los resultados solamente sobre la Base de Datos CMU-PIE [45] pueden generar dudas, debido al alto grado de efectividad que alcanza y al poco impacto que ha tenido en la comunidad biométrica internacional.

Del resto de los métodos se puede ver que los métodos de extracción de rasgos basados en análisis estadísticos alcanzan un rendimiento inferior a los del grupo de extracción de rasgos derivados de la imagen, en general. En el caso del *Descriptor de Cordilleras y Valles* su rendimiento no es tan bueno porque, a pesar de aplicar el operador MLSEC-ST que es invariante a la iluminación, la selección de los rasgos que caracterizan a la imagen sí es sensible a los cambios de iluminación.

4 Conclusiones

En este reporte de investigación se ha hecho un estudio detallado de los métodos de extracción de rasgos invariantes a la iluminación reportados en la literatura. Para organizar el estudio se subdividieron en tres grupos en dependencia de las herramientas que usan para realizar el análisis y la representación de los rasgos de la imagen.

La mayoría de los métodos abordados necesitan un conjunto de imágenes de entrenamiento con diferentes condiciones de iluminación para modelar las posibles representaciones del rostro bajo dichas condiciones. Por tanto este es un factor muy importante a tener en cuenta dentro de este campo, porque las condiciones teóricas y las prácticas deben ser consideradas por igual, no siendo fácil en sistemas reales contar con varias imágenes en diferentes condiciones de iluminación para cada individuo. Además este factor incide directamente en la tasa de reconocimiento de los métodos pues la cantidad de imágenes de entrenamiento puede verse como una solución ante la poca capacidad discriminativa de la propuesta.

Como pudo verse en cada una de las secciones, los métodos que utilizan una sola imagen de entrenamiento no son invariantes a la iluminación y a la vez discriminatorios, todavía existen problemas en esta área que deben ser investigados.

Otra cuestión importante para evaluar una técnica de reconocimiento de rostros es la elección de la base de datos sobre la cual se va comprobar su rendimiento. Como se vio en la sección anterior, las condiciones de iluminación propias de cada base de datos pueden inducir a que la efectividad del método varíe considerablemente. Ningún de los métodos de extracción de rasgos invariantes a la iluminación ha probado sus capacidades en todas las bases de datos. Luego, no es posible inferir cual es el mejor método ante cualquier base de datos.

En general puede decirse que existe una gran cantidad de métodos desarrollados para la extracción de rasgos invariantes a la iluminación. Esto evidencia la vigencia e importancia del tema de este reporte. De la misma manera se ha podido comprobar que las herramientas utilizadas para resolver esta problemática son diversas, incluyendo matemáticas, físicas y computacionales. Por tanto el estudio de este tema se convierte en una tarea profunda dada su riqueza y amplio espectro. A pesar de ello, los resultados alcanzados son insuficientes para el impacto que tienen los sistemas de reconocimientos de rostros en la sociedad contemporánea.

En los últimos 5 años no han sido publicadas nuevas teorías sobre la extracción de rasgos invariantes a la iluminación para la identificación de rostros. Las teorías existentes han sido combinadas entre ellas buscando métodos más eficaces.

Como conclusión final se puede decir que es imprescindible seguir desarrollando esta temática a nivel mundial debido al gran impacto que tiene los sistemas de reconocimiento de rostros en la sociedad. Existen actualmente un gran número de aplicaciones prácticas para ellos. Por tanto, si la iluminación es uno de los factores que influye negativamente en su rendimiento,

el éxito en el enfrentamiento ante sus efectos se traduciría en una significativa mejora para estos sistemas.

Referencias bibliográficas

1. S. Z. Li and A. K. Jain: "Handbook of face recognition", New York: Springer-Verlag, 2005.
2. W. Y. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips: "Face recognition: A literature survey", *ACM Computing Surveys*, Vol. 35, No. 4, pp. 399–458, December 2003.
3. P. Jonathon Phillips, W. Todd Scruggs, Alice J. O'Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, Matthew Sharpe: "FRVT 2006 and ICE 2006 Large-Scale Results", National Institute of Standards and Technology Gaithersburg, MD 20899, March 2007.
4. Y. Adini, Y. Moses and S. Ullman: "Face Recognition: The Problem of Compensating for Changes in Illumination Direction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 19, Issue 7, 1997.
5. F. Perronnin and J.-L. Dugelay: "A Model of Illumination Variation for Robust Face Recognition", *Multimodal User Authentication Workshop*, pp. 157–164, Santa Bárbara CA, 2003.
6. H. Chen, P. Belhumeur and D. Jacobs: "In search of Illumination Invariants", *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 254–261, Hilton Head, SC, 2000.
7. M. Saban and C. Sanderson: "On Local Features for Face Verification", *IDIAP-RR 04-36*, 2004.
8. S. Eickeler, S. Muller, and G. Rigoll: "Recognition of JPEG Compressed Face Images Based on Statistical Methods", *Image and Vision Computing*, 18(4):279–287, 2000.
9. R. C. Gonzales and R.E. Woods: "Digital Image Processing", Addison-Wesley Reading, MA, 1993.
10. M. Turk and A. Pentland: "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, 3(1):71–96, 1991.
11. K. Messer, J. Matas, J. Kittler, J. Luetin, and G. Maitre: "XM2VTSDB: The Extended M2VTS Database", In *2nd Int. Conf. Audio- and Video-Based Biometric Person Authentication*, pp. 72–77, Washington D.C., 1999.
12. X. Wang: "Feature Extraction and Dimensionality Reduction in Pattern Recognition and Their Application in Speech Recognition", PhD Thesis, Griffith University, Queensland, Australia, 2002.
13. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman: "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 711–720, May 1997.
14. A. Shashua: "Geometry and Photometry in 3D Visual Recognition", PhD thesis, MIT Artificial Intelligence Laboratory, AI-TR-1401, November 1992.
15. X. He and P. Niyogi: "Locality Preserving Projections", *Advances in Neural Information Processing Systems*, Vancouver, Canada, 2003.
16. X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang: "Face recognition using Laplacian faces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
17. M. Savvides, B.V.K. Vijaya Kumar, and P.K. Khosla: "Corefaces-robust shift invariant pca based correlation filter for illumination tolerant face recognition", In *Proc. IEEE CVPR04*, vol. 2, pp. 834–841, Washington, DC, June 2004.
18. M. Savvides and B.V.K. Vijaya Kumar: "Quad-phase minimum average correlation energy filters for reduced-memory illumination-tolerant face authentication", In *Audio and Visual Biometrics based Person Authentication (AVBPA)*, vol. 2688, pp.19-26, Surrey, UK, June 2003.
19. M. Savvides, B.V.K. Vijaya Kumar and P.K. Khosla: "Robust, Shift-Invariant Biometric Identification from Partial Face Images", In *Biometric Technologies for Human Identification (OR51)*, 2004.
20. M. Savvides, B.V.K. Vijaya Kumar and P.K. Khosla: "Cancellable Biometric Filters for Face Recognition", accepted for publication in *International Conference in Pattern Recognition (ICPR)*, 2004.
21. R.O. Duda, P.E. Hart, and D.G. Stork: "Pattern Classification", Wiley, 2001.

22. J.G. Daugman: "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", *J. Opt. Soc. Amer. A*, Vol. 2, No. 7, pp.1160-1169, 1985.
23. B. Moghaddam, T. Jebara, and A. Pentland: "Bayesian Face Recognition", *Pattern Recognition*, Vol.33, No.11, pp. 1771-1782, 2000.
24. A. Mahalanobis, B.V.K. Vijaya Kumar, and D. Casasent: "Minimum average correlation energy filters", *Appl. Opt.* 26, pp. 3633-3630, 1987.
25. A. V. Oppenheim and J. S. Lim: "The importance of phase in signals", *Proc. IEEE*, vol 69, No°5, pp. 529-541, May 1981.
26. M. H. Hayes, J. S. Lim and A. V. Oppenheim: "Signal Reconstruction from Phase or Magnitude", *IEEE Trans. Acoust. Signal Processing*, vol. ASSP-28, pp. 672-680, Dec. 1980.
27. C. Chen, Y. Tseng and C. Chen: "Combination of PCA and Wavelet Transforms for Face Recognition on 2.5D Images", *Proceedings on Image and Vision Computing*, pp 343-347, New Zealand 2003.
28. C. Xie and B.V.K. Vijaya Kumar: "Quaternion Correlation Filters for Color Face Recognition", *Conference on Security, Steganography, and Watermarking of Multimedia Contents VII, SPIE Symposium on Electronic Imaging*, 16-20 Jan. 2005 San Jose, CA.
29. W.R. Hamilton: "Elements of Quaternions", London, U.K., 1866.
30. C. E. Moxey, S. J. Sangwine and T. A. Ell: "Hypercomplex Correlation Techniques for Vector Images", *IEEE Transaction on Signal Processing*, Vol. 51, No. 7, pp. 1941-1953, July 2003.
31. B.V.K. Vijaya Kumar, M. Savvides, K. Venkataramani, and C. Xie: "Spatial Frequency Domain Image Processing For Biometric Recognition", *Proceedings IEEE International Conference on Image Processing*, pp. 53-56, 2002.
32. A.M. Lopez, D. Lloret, J. Serrat, and J.J. Villanueva: "Multilocal creaseness based on the level set extrinsic curvature", *Computer Vision and Image Understanding*, 77:111-144, 2000.
33. A.M. Lopez, F. Lumbreras, J. Serrat, and J.J. Villanueva: "Evaluation of methods for ridge and valley detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4): 327-335, April 1999.
34. A. Pujol: "Contributions to Shape and Texture Face Similarity Measurement", PhD Tesis Universitat Autònoma de Barcelona, ISBN 84-922529-9-5, 2001.
35. I. Biederman and J. Gu: "Surface versus Edge-Based Determinants of Visual Recognition", *Cognitive Psychology*, vol. 20, pp. 38-64, 1988.
36. V. Bruce, E. Hanna, N. Dench, P. Healey and M. Burton: "The Importance of "Mass" in Line Drawings of Faces", *Applied Cognitive Psychology*, vol. 6, pp. 619-628, 1992.
37. B. Takács: "Comparing Face Images Using the Modified Hausdorff Distance", *Pattern Recognition*, vol. 31, pp. 1873-1881, 1998.
38. Y. Gao and M.K. Leung: "Human Face Recognition Using Line Edge Maps", *Proc. IEEE Second Workshop Automatic Identification Advanced Technologies*, pp. 173-176, Oct. 1999.
39. M.K.H. Leung and Y.H. Yang: "Dynamic Two-Strip Algorithm in Curve Fitting", *Pattern Recognition*, vol. 23, pp. 69-79, 1990.
40. Bern Univ. Face Database, Disponible en <ftp://iamftp.unibe.ch/pub/Images/FaceImages/>, Fecha de consulta, 2002.
41. T. Ahonen, A. Hadid and M. Pietikäinen: "Face Recognition with Local Binary Patterns", 8th European Conference on Computer Vision, *Lecture Notes in Computer Science*, Vol. 3021. pp. 469-481, 2004.
42. T. Ahonen, A. Hadid and M. Pietikäinen: "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no7, pp. 971-987, 2002.
43. H. Jin, Q. Liu, H. Lu y X. Tong: "Face detection using improved LBP under bayesian framework", *International Conference on Image and Graphics*, Hong Kong, China. 306-309, 2004.
44. H. Méndez and E. García: "Estado actual de los métodos de reconocimiento automático de rostros basados en la apariencia local", *Reporte Técnico Serie Azul, Reconocimiento de Patrones. RNPS_2142 ISSN 2072-6287, RT_006*, Octubre 2008.

45. T. Sim, S. Baker, and M. Bsat: "The CMU pose, illumination, and expression (PIE) database of human faces", Technical Report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University, Pittsburgh, January 2001.

RT_038, enero 2011

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2011

Editor: Lic. Lucía González Bayona

Diseño de Portada: Di. Alejandro Pérez Abraham

RNPS No. 2142

ISSN 2072-6287

Indicaciones para los Autores:

Seguir la plantilla que aparece en www.cenatav.co.cu

C E N A T A V

7ma. No. 21812 e/218 y 222, Rpto. Siboney, Playa;

La Habana. Cuba. C.P. 12200

Impreso en Cuba

