RNPS No. 2142
ISSN 2072-6287
Versión Digital

SERIE AZUL

REPORTE TÉCNICO
Reconocimiento de Patrones

Estado actual de los métodos de reconocimiento automático de rostros basados en la apariencia local.

Heydi Méndez Vázquez, Edel García Reyes

RT _006 **Octubre 2008**

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

7ma. No. 21812 e/218 y 222,
Rpto. Siboney, Playa;
Ciudad de La Habana.
Cuba. C.P. 12200
www.cenatav.co.cu

RNPS No. 2142
ISSN 2072-6287

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

REPORTE TÉCNICO
Reconocimiento de Patrones

SERIE AZUL

Estado actual de los métodos de reconocimiento automáticos de rostros basados en la apariencia local

Heydi Méndez Vázquez, Edel García Reyes

RT__006 Octubre 2008

7ma. No. 21812 e/218 y 222,
Rpto. Siboney, Playa;
Ciudad de La Habana.
Cuba. C.P. 12200
www.cenatav.co.cu

Estado actual de los métodos de reconocimiento automático de rostros basados en la apariencia local

Heydi Méndez Vázquez, Edel García Reyes

Centro de Aplicaciones de Tecnología de Avanzada. 7a #21812 e/ 218 y 222, Siboney, Playa, Habana, Cuba.
hmendez@cenatav.co.cu, egarcia@cenatav.co.cu

RT__006 CENATAV
21 de Octubre de 2008

Resumen: El reconocimiento automático de rostros es una de las técnicas biométricas más utilizadas en aplicaciones de la vida real, no obstante, aún presenta grandes retos para la comunidad científica. Las variaciones en las condiciones de la iluminación son de las que más afectan el rendimiento de los métodos existentes. Varios métodos han surgido para enfrentar el problema de la iluminación, pero la mayoría de ellos requieren un gran número de imágenes de entrenamiento. Teniendo una sola o muy pocas imágenes de entrenamiento, lo más adecuado es utilizar métodos basados en la apariencia local. En este reporte, se hace un estudio profundo de los diferentes métodos de apariencia local reportados en la literatura, prestando especial atención a la robustez de estos a las variaciones de iluminación.

Palabras clave: Reconocimiento de rostros, normalización fotométrica, métodos basados en la apariencia local.

Summary: Automatic face recognition is one of the most used biometric techniques in real-life applications; however, there are still great challenges for the scientific community in this area. The sensitivity to variations in illumination is one of the major limiting factors for face recognition system performance. Different methods have been proposed in the literature aiming at compensate for illumination variations, but most of them require a large number of training images. Local appearance-based methods are the most suitable for handling the one-sample problem. In this report, the different local appearance-based methods that have been proposed are investigated, with specific reference to their robustness to illumination variations.

Keywords: Face recognition, photometric normalization, local appearance-based methods.

Introducción

El reconocimiento automático de rostros es utilizado en numerosas aplicaciones prácticas, como por ejemplo, sistemas de vigilancia, control de inmigración, control de accesos, autenticación de usuarios para dispositivos electrónicos como teléfonos celulares, cámaras, agendas electrónicas, etc. Es uno de los métodos biométricos más usados debido entre otros factores a que es una técnica no invasiva, natural y fácil de usar [1]. Sin embargo, presenta aún grandes retos para la comunidad científica, especialmente para ambientes no controlados donde la pose de la persona, la iluminación, la expresión, los accesorios utilizados, entre otros, varían considerablemente. En las diferentes pruebas e investigaciones que se han hecho del tema, se identifican las variaciones

de pose e iluminación como los factores que afectan más el rendimiento de los algoritmos de reconocimiento de rostros [2], [3].

Las variaciones en la iluminación afectan el desempeño de los sistemas de reconocimiento de rostros debido a que condiciones de iluminación diferentes pueden producir imágenes muy diferentes del mismo objeto. Adini, Moses y Ullman [4] mostraron que las variaciones de la imagen debidas a los cambios en la iluminación son más significativas que aquellas debidas a diversas identidades personales.

En otras palabras, la diferencia entre dos imágenes del rostro del mismo individuo tomadas bajo condiciones de iluminación diferentes es mayor que la diferencia entre dos imágenes de rostros diferentes tomadas bajo las mismas condiciones de iluminación como puede ser observado en la Figura 1.

a)  b) 

Fig.1. Ejemplo de variaciones de iluminación: a) Mismo sujeto, condiciones de iluminación diferentes; b) Sujetos diferentes, iguales condiciones de iluminación

En un sistema de reconocimiento de rostros, es posible tratar con los problemas de iluminación en tres estados diferentes: durante el pre-procesamiento, a la hora de la extracción de las características y en el momento de la clasificación [5]. El primero de ellos, trata de modificar la imagen de entrada para convertirla en una representación más adecuada para el propósito del reconocimiento [6], [7]; de esta manera se trabaja solamente en el pre-procesamiento sin importar que tipo de clasificador sea utilizado. La segunda vía es derivar de las imágenes de rostros características significativas que sean invariantes a los cambios de iluminación [4], [8]. Por último, están los métodos generativos, los cuales tratan de modelar el objeto de interés bajo todas las posibles condiciones de iluminación y esto no puede ser desasociado del procedimiento de clasificación [5], [9].

Tanto el caso del pre-procesamiento como el de extracción de características invariantes, son independientes de la cantidad de imágenes de una persona que se tengan. Sin embargo, los métodos basados en modelos requieren un conjunto de entrenamiento que contenga muchas imágenes del mismo individuo bajo una gran variedad de condiciones de iluminación. El caso extremo de una sola muestra (imagen) por persona realmente se encuentra con mucha frecuencia en escenarios reales, lo cual es desfavorable para muchas técnicas de reconocimiento de rostros, pero tiene varias ventajas que son deseadas en aplicaciones reales, entre ellas: mayor facilidad en la recolección de las muestras (directa o indirectamente), ahorro en el costo de almacenamiento y ahorro en costo computacional [10]. Por otra parte, intuitivamente es difícil de creer que un ser humano necesite muchas fotos de una persona para desarrollar un buen modelo de su apariencia.

Por tanto, el reconocimiento de rostros a partir de una sola imagen es un problema importante tanto para investigaciones aplicadas como teóricas y presenta nuevos retos para la comunidad de investigadores en reconocimiento de rostros. Uno de estos retos es mejorar la robustez de los algoritmos de reconocimiento de rostros ante diferentes variaciones, entre ellas, de iluminación, cuando los tamaños de muestra son extremadamente pequeños.

Los métodos existentes para el reconocimiento de rostros basados en una sola imagen pueden ser clasificados en dos categorías teniendo en cuenta el tipo de rasgos que utilizan: métodos basados en características geométricas y métodos basados en la apariencia.

Los métodos basados en características geométricas fueron los más populares en los inicios del reconocimiento automático de rostros [11]. Este tipo de métodos hace uso explícito del conocimiento sobre el rostro: los rasgos faciales (nariz, boca, ojos) y las propiedades y relaciones entre estos (áreas, distancias, ángulos). Estos métodos son eficientes y efectivos cuando alcanzan reducción de la información e insensibilidad a las variaciones de iluminación y pose; además requieren detectar los rasgos faciales con una alta probabilidad y contar con imágenes de buena calidad [12]. Por otra parte, estas características geométricas por sí solas son inadecuadas para el reconocimiento de rostros, ya que información importante del rostro contenida en la textura, o en la apariencia facial, es descartada.

Los métodos basados en la apariencia, por su parte, han sido los más dominantes en los últimos años [1]. Éstos operan directamente con las intensidades de los pixeles u otras representaciones basadas en la imagen y han constituido un avance significativo para la eficiencia y efectividad de los sistemas de reconocimiento de rostros [10].

Los métodos basados en apariencia pueden ser utilizados de formas diferentes: de manera holística (global) o de manera local. Los métodos holísticos identifican un rostro utilizando como entrada un vector que representa la imagen completa. Los métodos locales usan vectores de rasgos que representan diferentes regiones de la imagen de rostro.

Para más claridad, en la Figura 2 se puede observar la taxonomía de los métodos de reconocimiento de rostros descritos anteriormente.

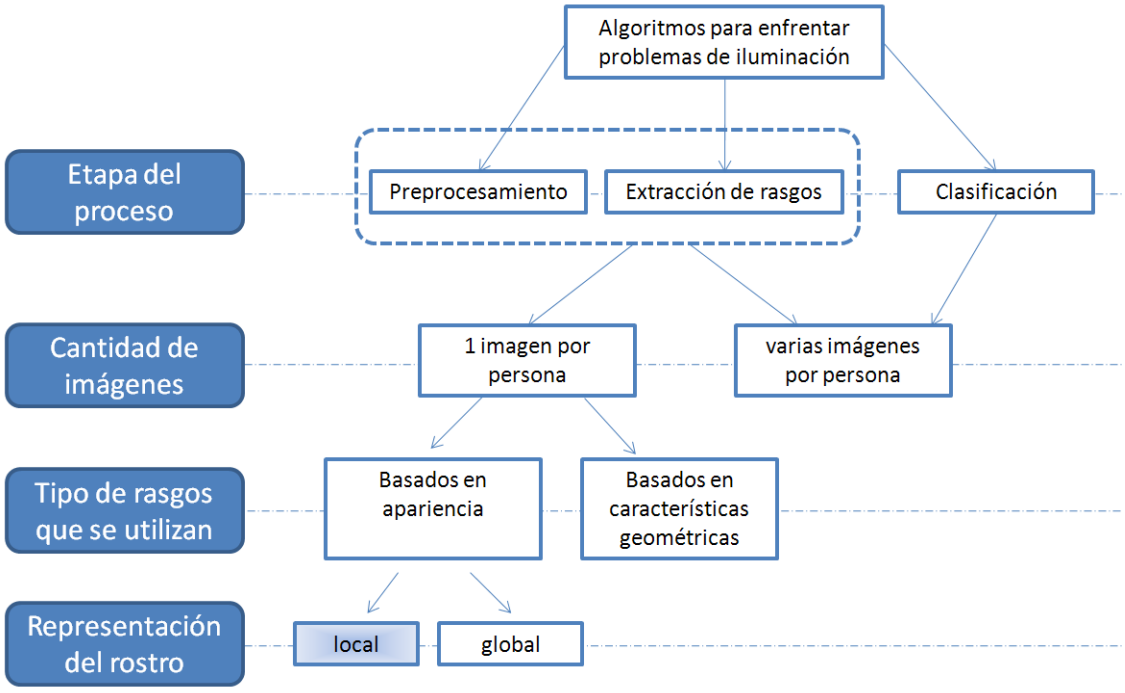

Fig.2. Taxonomía de los tipos de métodos existentes para enfrentar los problemas de iluminación

Los métodos basados en apariencia local, comparados con los globales, son más apropiados para el tratamiento de los problemas con una sola imagen de muestra, entre otras ventajas están:

1. Los datos con menor dimensionalidad son más adecuados para los problemas de clasificación [13], luego, es mejor representar un rostro mediante un conjunto de vectores de características locales de *pocas* dimensiones, en vez de un solo vector de *grandes* dimensiones.
2. Los métodos locales proveen flexibilidad adicional a la hora de reconocer un rostro basado en sus partes, de esa manera, las características comunes y específicas de una clase pueden ser fácilmente identificadas [14].
3. Rasgos faciales diferentes pueden incrementar la diversidad de los clasificadores [15], lo cual es favorable para la identificación de rostros.

Por otra parte, en [16] se muestra que los métodos de normalización locales son más invariantes a los cambios de iluminación que los globales. Si se mira la iluminación de acuerdo al modelo lambertiano, ésta depende de la fuente de iluminación, la reflectancia y la forma del objeto. Fijando los parámetros de la fuente de iluminación se obtiene una imagen con variaciones en los tonos de gris que depende solamente de la reflectancia y la forma local. Excepto para regiones muy particulares del rostro, estos parámetros no deben variar mucho localmente, lo que nos lleva a pensar que las variaciones en la iluminación afectan menos en una pequeña región que en la imagen completa.

Teniendo en cuenta que las afectaciones de iluminación pueden ser enfrentadas de forma más exitosa aplicando técnicas locales en lugar de globales, en este reporte se hará una revisión de los métodos de reconocimiento de rostros basados en la apariencia local existentes, con el objetivo de comparar las ventajas y desventajas de cada uno de ellos, analizando la robustez de estos ante las variaciones de iluminación.

1 Métodos basados en la apariencia local

Los métodos basados en la apariencia local involucran cuatro pasos generales como puede ser observado en la Figura 3:

1. Partición en regiones locales
2. Extracción de rasgos locales
3. Selección de rasgos
4. Clasificación

El primer paso indispensable en los métodos locales, es definir las regiones que van a ser utilizadas. Esto involucra dos aspectos: la forma y el tamaño de las regiones locales. La forma más simple y comúnmente utilizada es dividir la imagen de rostro en ventanas rectangulares [17], [18], [19], [20], las cuales pueden estar solapadas [19], [20] o no [17], [18]. Sin embargo, se han utilizado también otras formas como elipses [21] y franjas [22]. El tamaño de las regiones varía mucho en dependencia del método utilizado, es una cuestión bastante delicada que tiene influencia directa en la robustez de los métodos. Queda aún mucho por investigar en cuanto a qué forma y tamaño son los más indicados a utilizar para el reconocimiento de rostros.

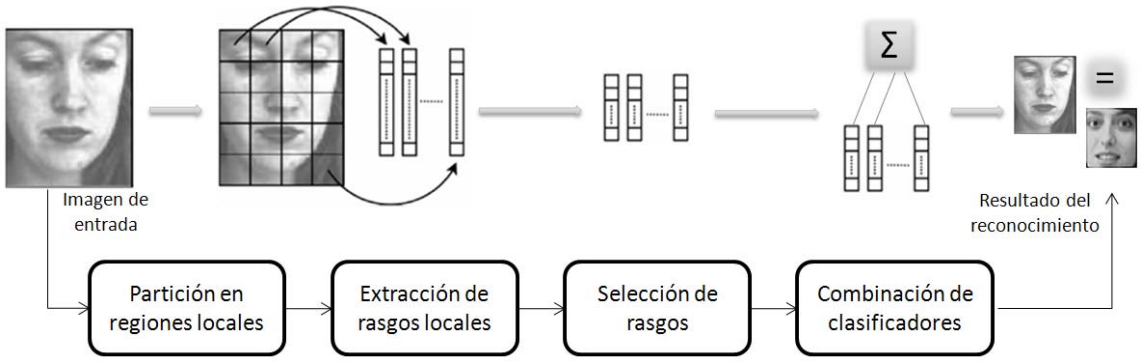

Fig.3. Plataforma general de los métodos basados en la apariencia local

Una vez que las regiones locales han sido definidas, es necesario decidir cómo representar esa información. Éste es un paso crítico para el desempeño de los sistemas de reconocimiento de rostros e igualmente es difícil decidir cuál es el método más adecuado a utilizar. Se utilizan normalmente vectores formados directamente con las intensidades de los pixeles [19], [20], [21] o rasgos derivados de estos como: la transformada discreta del coseno (DCT) [18], *wavelets* de Gabor [22], *wavelets* de Harr [23], rasgos fractales [24], entre otros. En general, utilizar directamente las intensidades de los pixeles es lo más simple sin perder información de la textura, sin embargo otros rasgos derivados pueden ser más robustos a las variaciones de iluminación.

En dependencia del tipo de rasgo utilizado, en algunos casos se hace necesario un paso adicional de selección de rasgos o reducción de dimensionalidad para mejorar la efectividad y eficiencia del método. Este paso consiste en retener solamente un subconjunto del vector de rasgos original, evitando la pérdida de información discriminativa. Entre los algoritmos más utilizados para la reducción de dimensionalidad se encuentran el análisis de componentes principales (PCA) y el análisis discriminante lineal (LDA), que en principio son métodos globales, pero han sido también usados de manera local [8], [26]. Además, algunos métodos de extracción de rasgos incluyen ellos mismos un paso de selección, como es el caso del DCT, en el que sólo una parte de los coeficientes son utilizados para formar el vector de rasgos [18].

Como último paso se encuentra la clasificación, para lo cual se utiliza comúnmente la combinación de clasificadores. Ésta puede ser llevada a cabo de dos formas diferentes: concatenando en un solo vector los rasgos de las diferentes regiones [18] o realizando la clasificación por separado en cada región y luego tomar la decisión final combinando el resultado [15].

Los cuatro pasos anteriores son los más generales, pero no son obligatorios. Cada método tiene sus particularidades y alguno de los pasos pudiera ser cancelado o unido con otro. A continuación se hará una descripción de los métodos de reconocimiento de rostros basados en la apariencia local más representativos.

1.1 Uso de los pixeles directamente

El método más sencillo que puede ser utilizado para describir la apariencia local es el uso directo de los valores de intensidad de los pixeles en un bloque para formar un vector de rasgos que lo

represente. Luego puede usarse, por ejemplo, la correlación cruzada [27] para calcular la similitud entre dos vectores o clasificadores basados en modelos de mezclas gausianas (GMM) [8] o modelos ocultos de Markov (HMM) [23].

Una variante utilizada en [8] fue formar el vector con los pixeles organizados en un patrón de zig-zag (similar al método DCT que se explicará más adelante), de manera que para un bloque localizado en *(b;a)* el vector de rasgos está compuesto por:

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

Donde $p_n^{(b,a)}$ es el valor enésimo de acuerdo al patrón en zig-zag.

En cualquiera de las dos variantes anteriores, esta técnica tiene algunas desventajas, entre ellas, que los elementos que componen el vector pueden estar altamente correlacionados, y como efecto secundario de la alta correlación, un cambio de iluminación afecta a todos los elementos del vector.

Para solucionar esto en [8] propusieron eliminar algo de la correlación entre los elementos de cada vector, sustrayendo el valor medio de las intensidades de los pixeles de cada región a cada elemento. Esto tiene como beneficio adicional que eliminar la media puede ser interpretado como una forma de normalización de la iluminación. Formalmente, el vector de rasgos eliminando la media estaría compuesto por:

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

Donde

$$p_\mu^{(b,a)} = \frac{1}{N^2} \sum_{i=0}^{N^2-1} p_i^{(b,a)} \tag{3}$$

En ambos casos se utilizó un clasificador basado en GMM. El uso de vectores con los valores de intensidad de los pixeles directamente conduce a resultados muy pobres, pero eliminando la media de cada vector se mejora considerablemente el rendimiento. Eliminar la media puede ser entonces considerado como un tipo de normalización fotométrica. Sin embargo, en los experimentos llevados a cabo en ese trabajo puede observarse que, aunque eliminando la media se mejoran los resultados, cualquiera de las dos variantes se ve grandemente afectada por los cambios de iluminación.

1.2 Análisis de componentes principales aplicado de manera local

Uno de los primeros y más utilizados métodos para el reconocimiento de rostros ha sido el PCA. El objetivo que se persigue con este método es extraer la información relevante en una imagen de rostro y codificarla lo más eficientemente posible.

Normalmente ha sido utilizado de manera global. La idea es capturar las variaciones en una colección de imágenes de rostros para luego extraer la información contenida en cada una de las

imágenes individuales. En términos matemáticos, es encontrar los componentes principales de la distribución de rostros, o los valores propios de la matriz de covarianza del conjunto de imágenes de rostros. Estos vectores propios caracterizan las variaciones entre imágenes de rostros y cada imagen contribuye más o menos a cada vector propio, luego el vector propio se puede mostrar como un tipo de rostro fantasmal llamado *eigenface* [28].

Sin embargo, han surgido algunas aplicaciones del PCA de manera local en imágenes de rostros [8], [29] que intentan mejorar los resultados que brinda el método tradicional.

En [29] se presentan dos variantes del uso local de PCA. La primera de ellas es extraer vectores propios independientes de ventanas fijas sobre el rostro que representen determinadas características como por ejemplo, el ojo izquierdo, el ojo derecho y la boca y posteriormente, generar el patrón del rostro proyectando los rasgos faciales particulares en sus vectores propios respectivos. La otra variante es, a partir de los bloques de imagen sobre los rasgos mencionados anteriormente, formar un conjunto de patrones de bloques aleatorios, de manera que en vez de utilizar PCA en cada uno de los bloques individuales o en todos ellos a la vez, se utiliza un PCA más general sobre los bloques aleatorios para generar los vectores propios, de esa manera el sub-espacio es más general. Sin embargo, a pesar de que se han usado en el reconocimiento de rostros, estos métodos están más bien enfocados a la clasificación de emociones y acciones faciales, en donde se quiere buscar generalizaciones y no características de un rostro en particular.

Otra variante, sí dirigida al reconocimiento, se presenta en [8]. La idea es dividir la imagen en bloques regulares y representar los pixeles del bloque directamente en un vector, como en la ecuación (1). Luego, se aplica PCA y se obtiene un nuevo vector con menor dimensionalidad. Con esta representación se aprenden las funciones bases más representativas de los bloques de las imágenes de rostro, por tanto se puede reducir en gran medida la dimensionalidad de los vectores, sin embargo, esto no garantiza que los vectores resultantes sean óptimos discriminativamente. Una desventaja del uso local de PCA es que las funciones bases pueden no tener un significado interpretable en términos de la estructuras de la imagen (opuesto al significado estadístico); además, las funciones bases van a variar mucho en dependencia de la base de datos que se utilice para entrenar. En [8] se utilizó este método como extracción de rasgos para luego utilizar los vectores obtenidos con un clasificador basado en GMM. En imágenes sin problemas de iluminación de la base de datos XM2VTS [30] los resultados obtenidos con este método fueron satisfactorios, pero se vio afectado su rendimiento cuando cambiaron las condiciones en la iluminación. Se hicieron otras pruebas eliminado los coeficientes del vector de rasgos que se estimaron como los más afectados por la iluminación, así como sustituyéndolos por coeficientes delta [8], pero los resultados no cambiaron mucho.

1.3 Método local probabilístico de sub-espacios

En [21] se presenta un método de apariencia local que además enfrenta el problema de contar con una sola imagen por persona.

El primer objetivo que se persigue en ese trabajo es modelar el error de localización, que no es más que el error que se comete en la localización de los rasgos faciales cuando en un método basado en apariencia se le aplica una alineación (*warping*) previamente a la imagen. En la base de datos se tiene una sola imagen por individuo y para estimar el error de localización se ubican manualmente los rasgos faciales para cada persona: posición de los ojos, bocas, etc. Así se calcula la varianza en X y en Y, a partir de lo cual se estima el error.

El error de localización puede ser modelado como una distribución gausiana o una mezcla de gausianas. Una vez que este se conoce, tanto en el eje *X* como en el *Y*, se generan sintéticamente todas las posibles imágenes alineadas (*warped*) para cada clase de rostros: $\hat{s}_i = \{s_{i,1}, \ldots, s_{i,r}\}$, donde *i* se refiere a la clase y *r* al número de posibles localizaciones.

Cuando se tiene este conjunto de imágenes generadas sintéticamente, cada una de estas se divide en 6 áreas elípticas locales como puede verse en la Figura 4. Todas las áreas correspondientes a la misma posición se agrupan en un sub-espacio donde se representa la distribución gausiana asociada al error de localización.

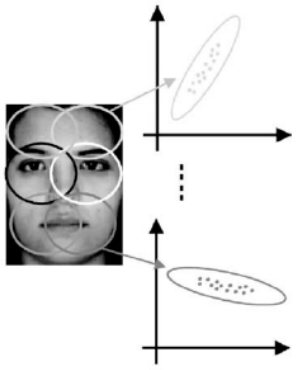

Fig.4. División de la imagen de rostro en 6 elipses y su representación

La dimensionalidad de estos sub-espacios se vuelve muy alta, por lo tanto es conveniente aplicar antes de todo este proceso una reducción de la dimensionalidad del espacio de características original mediante el método PCA. Para lo cual primeramente, para cada una de las 6 áreas en las que se dividen las imágenes, se generan los espacios propios (*eigenspace*) tomando los vectores propios (*eigenvectors*) asociados a los mayores valores propios (*eigenvalues*) de la matriz de covarianza correspondiente a las imágenes de muestra originales, una por cada clase. Entonces, en estos espacios propios creados, es que se proyecta el área correspondiente de cada conjunto $\hat{s}_i$ y se busca el modelo gausiano que representa.

A la hora de identificar una nueva imagen, se divide también en 6 partes, cada una de ellas se proyecta en los *eigenspaces* obtenidos y se calcula la distancia de esa correspondencia local mediante la distancia Mahalanobis. Finalmente, se suman todas las correspondencias locales a cada clase y se escoge aquella de mayor valor.

El método está propuesto para enfrentar más bien los problemas de oclusión y expresión. No se presentan resultados enfrentando las variaciones de iluminación. Aún así, el método presenta varias desventajas, entre ellas:

- Se asume que las muestras generadas con diferentes errores de localización están bien representadas mediante una distribución de gausiana, lo cual puede no ser real.
- La imagen a partir de la cual se generan las imágenes sintéticas y por tanto, la distribución correspondiente a una clase, se fijan con expresiones neutrales y localización correcta, lo cual puede ser difícil de tener en situaciones reales, donde la imagen de muestra que se tenga puede estar en los límites máximos del error permisible para una clase.

- Se generan sintéticamente un gran número de imágenes de muestra, por tanto el costo computacional y de almacenamiento es muy alto.

1.4 Método local utilizando mapas auto-organizados

Este método propuesto en el 2004 [19], es una extensión del método local probabilístico de sub-espacios presentado por Martínez [21], utilizando mapas auto-organizados (SOM) en lugar de modelos gausianos.

Se escoge los SOM, porque incluso cuando el tamaño de la muestra es muy pequeño, para representar fielmente la distribución, este algoritmo puede extraer toda la información significante de los rasgos faciales locales, debido a sus características de ser no supervisado y no paramétrico, mientras que elimina posibles fallas como ruidos, outliers o valores perdidos. Este método se utiliza localmente para evitar el problema de tener un limitado número de imágenes de muestra y estar representadas estas mediante vectores de altas dimensiones.

Se particiona la imagen en M sub-bloques no solapados de igual tamaño y cada uno de estos bloques se representa mediante un vector de rasgos locales (LFV) de bajas dimensiones, que concatena los pixeles del bloque. Una vez que se tienen los vectores locales, se proponen dos estrategias de aprendizaje del espacio topológico de los SOM: 1) entrenar un solo mapa SOM con todas las imágenes de entrenamiento y 2) entrenar un mapa SOM separado para cada clase:

1) Los bloques de todas las imágenes de entrenamiento conforman la entrada del SOM que tiene un conjunto de neuronas $A = \{e_1,e_2,...,e_Q\}$, cada una con un vector de peso w_i. Se agrupan todos los vectores de la entrada pertenecientes a los sub-bloques en conjuntos Voronoi, según los vectores de peso más cercanos a estos. Se calcula el promedio de los vectores en cada región y se afinan con estos valores los vectores de peso de cada neurona. Una vez que el SOM ha sido entrenado, todos los sub-bloques de cada imagen de entrenamiento se mapean en las unidades de mejor correspondencia (BMU*: best matching units*) en el espacio topológico de éste, utilizando una estrategia de vecino más cercano. Los vectores de peso correspondiente a cada BMU serán utilizados como vectores prototipo de cada clase para el propósito de reconocimiento.
2) Con el propósito de evitar el recálculo de los vectores bases cuando se presentan nuevos individuos, se propone crear un subespacio separado para cada rostro. Luego, se sigue el mismo algoritmo descrito anteriormente, pero en el entrenamiento de cada SOM sólo se utilizan los sub-bloques pertenecientes a una clase o rostro.

Una vez entrenado el/los SOM, para identificar un rostro, se propone un esquema de decisión basada en un conjunto (ensemble) de k vecinos más cercanos. Cuando se tiene una nueva imagen a clasificar, esta se divide también en sub-bloques y se proyectan éstos en los SOM entrenados. Luego se calcula la similaridad de cada sub-bloque de la nueva imagen con cada uno de los sub-bloques de las imágenes de entrenamiento. Los valores de similaridad de la j-ésima neurona (que representa a un sub-bloque) del rostro a identificar con sus k vecinos más cercanos (sub-bloques correspondientes en las imágenes de entrenamiento) son ordenados ascendentemente $d_{j1} \leq d_{j2} \leq ... \leq d_{jk}$ y se define el valor de confidencia del vecino k como,

$$c_{jk} = \frac{\log(d_{j1}+1)}{\log(d_{jk}+1)} \qquad (4)$$

donde la clase más similar al sub-bloque tendrá valores cercanos a 1, mientras que las más distantes tendrán valores muy pequeños. Finalmente, la clase a la que pertenece la imagen de entrada se determina mediante un esquema de votación por cada clase *C* de la siguiente manera:

$$Label = \arg\max_{k} \quad (\sum_{j=1}^{M} c_{jk}) \quad , k = 1...C \tag{5}$$

Este método fue probado en la base de datos AR [31], la cual presenta pocas variaciones de iluminación. Enfrentando los problemas de oclusión y expresión brinda mejores resultados que el método local probabilístico de sub-espacios descrito en la sección anterior. Sin embargo, presenta algunas desventajas:

- Si se utiliza un solo SOM, cuando se necesita representar una nueva clase (rostro) es necesario recalcular los vectores bases, es decir, volver a entrenar el SOM, esto puede evitarse cuando se utiliza un solo SOM para cada clase, pero a su vez, puede disminuir la precisión del algoritmo cuando se tienen muchas clases.
- Quedan aún problemas abiertos en cuanto al tamaño de los sub-bloques a utilizar, así como el valor de *k* que debe escogerse.
- Deben estudiarse otras vías de agrupar los vectores representados por cada neurona.
- El uso de un esquema de votación puede traer consigo empates en la decisión final.

1.5 Rasgos fractales

En [24] se presenta un método que utiliza un subconjunto de códigos fractales de la imagen para el reconocimiento de rostros.

La teoría Fractal ha sido muy utilizada en el procesamiento de imágenes y la visión por computadoras [32]. En este método la similaridad entre diferentes partes de una imagen es utilizada para representar una imagen de rostro mediante un conjunto de transformaciones contractivas en el espacio de las imágenes, para el cual el punto fijo está cerca de la imagen original.

Primeramente es necesario un método de particionamiento de la imagen. En este caso se utiliza el particionamiento *quadtree*, que está basado en la división recursiva de la imagen en cuadrantes, permitiendo que la partición resultante pueda ser representada por una estructura de árbol en la cual cada nodo no terminal tiene 4 descendientes. El método de construcción usual *top-down* comienza seleccionando un nivel inicial en el árbol correspondiente a un tamaño de bloque máximo y recursivamente va particionando cada bloque hasta que cumplan un umbral predeterminado. En la Figura 5 se muestra un ejemplo de la división en *quadtree* de una imagen de rostro.

Posteriormente cada bloque que ha sido particionado se representa en un vector de rasgos. Sin embargo, el tamaño de cada vector varía de una imagen a otra y depende del umbral de particionamiento, del tamaño de la imagen, de su complejidad y del tamaño mínimo de los bloques. Con el objetivo de normalizar el tamaño de los vectores se usa la geometría del particionamiento en *quadtree* y se aplica cada valor de rasgo en su posición geométrica.

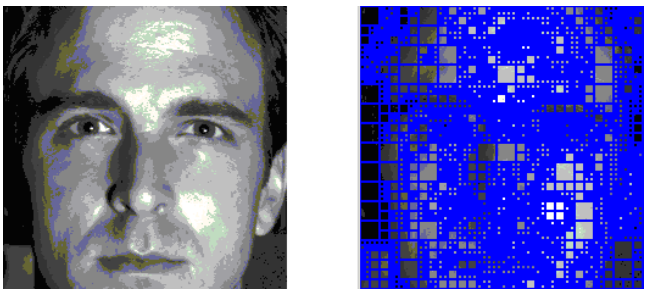

Fig.5. División de la imagen de rostro en *quadtree*

Como el particionamiento en *quadtree* puede ser aplicado a una imagen de tamaño arbitrario, se puede cambiar el tamaño de todos los vectores de rasgos al tamaño de la imagen entrada. Esto hace que el método sea robusto a los cambios en tamaño y escalas.

Estos vectores quedan de un tamaño muy grande, por lo que se reduce su dimensión utilizando PCA. Luego, los vectores de código fractal reducidos son utilizados en la clasificación utilizando el error cuadrático medio para compararlos.

Este método fue probado con imágenes de rostros con problemas de expresión y se obtuvieron buenos resultados. La principal dificultad de este método para enfrentar los problemas de iluminación vendría dada en encontrar un método o umbral de particionamiento que fuera invariante a los cambios de iluminación.

1.6 Modelos ocultos de Markov

Existen varios métodos propuestos para el reconocimiento de rostros utilizando HMM. La diferencia entre ellos radica principalmente en la forma de obtener los vectores de observación.

En [33] por ejemplo, la imagen de rostro es verticalmente escaneada de arriba hacia abajo, formando una secuencia de observación. La secuencia de observación está compuesta por vectores que representan franjas horizontales consecutivas solapadas entre ellas. Cada vector contiene las intensidades del conjunto de pixeles de la franja correspondiente. En este caso, se usa la topología de HMM de una dimensión (1D-HMMs), como puede verse en la Figura 6.

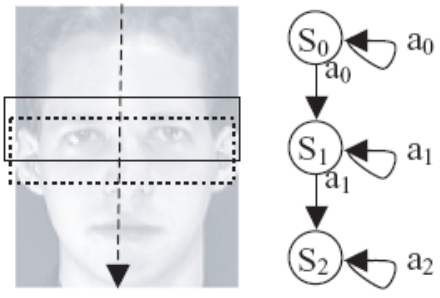

Fig.6. Topología de HMM de una dimensión para el reconocimiento de rostros

Una extensión del método anterior es el pseudo 2D HMM [34], que no es más que un 1D-HMM compuesto de “súper estados” que modelan la secuencia de las columnas en una imagen.

Cada "súper estado" es él mismo un 1D-HMM que modela los bloques dentro de las columnas. Un ejemplo de esto puede verse en la Figura 7.

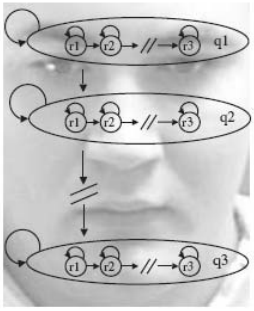

Fig.7. Topología del pseudo 2D HMM para el reconocimiento de rostros

Para manejar la imagen completamente en dos dimensiones sin llevar el problema a una sola dimensión, fue propuesto el método 2D-HMM de baja complejidad (*Low-Complexity 2D-HMM*) [35]. Este método consiste en una constelación rectangular de estados donde tanto las transiciones verticales como las horizontales son tenidas en cuenta. Cada imagen es escaneada en bloques de 8x8 pixeles de izquierda a derecha y de arriba hacia abajo, sin solapamiento. Los bloques en una vecindad diagonal y anti-diagonal se asumen independientes, esto reduce la complejidad de la capa oculta del modelo. La complejidad de este método es considerablemente más baja que la del pseudo 2D-HMM y el rendimiento disminuye en una medida muy pequeña. Sin embargo, en todos los casos se necesitan una cantidad considerable de imágenes para entrenar el modelo.

En [23] se presenta un método usando HMM con una sola imagen de entrenamiento por persona. Este método usa la topología de una dimensión, pero es aplicada tanto en la dirección vertical como en la horizontal. Cada imagen es dividida en franjas verticales solapadas y luego cada una de esas franjas es dividida en bloques solapados. Posteriormente se crean nuevos bloques de rasgos basándose en la diferencia entre dos bloques consecutivos, cada uno de los cuales se normaliza según su media y su varianza. Por último los bloques normalizados son alineados en forma de columna para formar los vectores de observación. Dos factores contribuyen a la viabilidad y la eficacia de ese método. En primer lugar, al generar una gran colección de vectores de observación por cada imagen se logra ampliar el conjunto de entrenamiento. En segundo lugar, se aplica la transformada wavelet Haar a la imagen para disminuir la dimensión de los vectores de observación y mejorar el rendimiento. Los resultados experimentales en la base de datos AR [31] de imágenes de rostros frontales muestran la superioridad de este método respecto a otros como PCA y LDA. Sin embargo, las variaciones de iluminación en las imágenes de esa base de datos son muy pequeñas y en aquellas imágenes que presentan oclusiones junto con problemas de iluminación, los resultados no son muy buenos.

1.7 Jets de Gabor

Los *jets* de Gabor [36] describen la información de la frecuencia local de una región de una imagen. Son una colección de coeficientes de Gabor complejos obtenidos a partir de la misma

porción de imagen. Los coeficientes se generan usando *wavelets* de Gabor de una variedad de tamaños, orientaciones y frecuencias.

Para la configuración estándar, los *jets* de Gabor están basados en 40 *wavelets* (8 orientaciones, 5 frecuencias) complejas, donde cada una de ella tiene una componente real y una imaginaria. Un *jet* para la posición (*x,y*) de una imagen, se produce convolucionando ese punto con cada uno de los *wavelets* obtenidos mediante la ecuación:

$$w(x,y,\theta,\lambda,\varphi,\sigma,\gamma) = e^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} \cos(2\pi\frac{x'}{\lambda}+\varphi) \quad (6)$$

Donde:

$$x' = x\cos\theta + y\sin\theta \quad (7)$$

$$y' = -x\sin\theta + y\cos\theta \quad (8)$$

θ especifica la orientación de la *wavelet* (8 orientaciones: 0,π/8,2π/8,3π/8,4π/8,5π/8,6π/8,7π/8)

λ especifica la frecuencia (5 frecuencias: 4, $4\sqrt{2}$, 8, $8\sqrt{2}$,16)

φ especifica la fase (2 fases: 0, π/2)

σ especifica el radio de la gausiana (σ=λ)

γ especifica la razón del aspecto de la gausiana (λ=1)

Los valores resultantes se convierten en coordenadas polares y se almacenan en un arreglo. Como resultado, los *jets* de Gabor contienen una "descripción" de la información de frecuencia localizada alrededor de un punto en una imagen. Cada coeficiente wavelet captura la información sobre una combinación de fase, orientación y frecuencia, por tanto, en la práctica se obtiene una descripción a través de múltiples frecuencias y orientaciones, de una región de la imagen. Se plantea que estos rasgos de Gabor estimados localmente son robustos a cambios de iluminación, distorsiones y escalamiento [36].

Estos descriptores locales se utilizan en el reconocimiento de rostros con el método conocido cómo *Elastic Bunch Graph Matching* [37]. Este método, es un método basado en características, ya que se conforma un grafo a partir de los puntos característicos obtenidos en la imagen de rostro, pero cada punto característico (nodo del grafo) va acompañado de la información de apariencia local en su vecindad representada por los *jets* de Gabor. En la Figura 7 puede verse un ejemplo de la representación utilizada para este método.

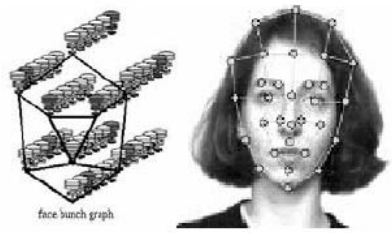

Fig.7. Representación del *Bunch Graph* y los *jets* de Gabor en una imagen de rostro

A pesar de la robustez que muestra este método a los cambios de apariencia en la imagen de rostro, presenta como mayor debilidad la necesidad de localizar certeramente los puntos característicos en la imagen y por otra parte, es muy costoso computacionalmente.

1.8 Patrones binarios locales

El operador de patrones binarios locales (LBP) es uno de los descriptores de textura con mejores resultados y ha sido usado en numerosas aplicaciones. Se ha mostrados que este operador es altamente discriminatorio y tiene varias ventajas como por ejemplo, la invarianza a los cambios de niveles de grises monotónicos 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 este operador [17].

El operador LBP fue diseñado originalmente para la descripción de textura. Asigna una etiqueta a cada pixel de una imagen usando en una vecindad de 3x3 pixeles el valor del pixel central como umbral y considerando el resultado como un número binario, como se observa en la Figura 8. Luego el histograma de las etiquetas es utilizado como descriptor de la textura.

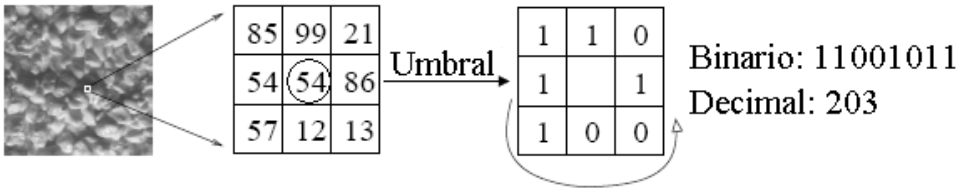

Fig. 8. El operador LBP básico

El operador fue posteriormente extendido para usar vecindades de diferentes tamaños [38]. Al definir una vecindad local como un conjunto de puntos de ejemplo uniformemente espaciados en un círculo centrado en el pixel que será etiquetado, se puede 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 pixel. Se utiliza la notación (*P,R*) para vecindades de pixeles, lo cual significa *P* puntos de ejemplo en un círculo de radio *R*. La figura 9 muestra un ejemplo de vecindades circulares.

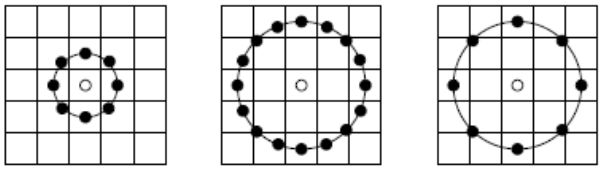

Fig.9. Vecindades circulares de (8,1), (16,2) y (8,2).

Otra extensión del operador original es la definición de los llamados patrones uniformes [38]. 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 (6 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 [39] 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 pixeles en la vecindad, en lugar de con el pixel central.

En cualquiera de sus variantes, los histogramas LBP contienen información acerca de la distribución de los micropatrones 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 $R_0, R_1, \ldots, R_{m-1}$ 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 pixeles, 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 sólo 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 [17].

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:

$$\chi_w^2(x,\xi) = \sum_{j,i} w_j \frac{\left(x_{i,j} - \xi_{i,j}\right)^2}{x_{i,j} + \xi_{i,j}} \qquad (2)$$

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 tasa de reconocimiento utilizando este operador es bastante alta, un 93% si no se utilizan pesos y 97% si se utilizan, mostrando ser superior a otros métodos [17]. Muestra también ser bastante invariante a cambios de expresión facial y de iluminación. Otra ventaja es la eficiencia computacional del operador LBP y no se necesita ninguna normalización previa antes de aplicar el operador a la imagen del rostro.

Desde que se introdujo la aplicación de este operador a imágenes de rostros muchos han sido los estudios realizados. Se deben estudiar métodos más avanzados para dividir la imagen en regiones y encontrar los pesos de estas, el método AdaBoost puede ser una buena base para esta investigación.

1.9 Transformada Census

Aproximadamente en el mismo tiempo que surgió el LBP [17], una estructura local muy similar, denominada transformada Census (CT), fue propuesta como descriptor de textura [40]. La CT, es una transformación local no paramétrica que se define como un conjunto ordenado de comparaciones entre las intensidades de los pixeles en una vecindad local, representando qué pixeles tienen menor valor que el valor central. En general, el tamaño de la vecindad local no está restringido, pero normalmente se utiliza una vecindad de 3x3, motivado por el hecho de que con *kernels* estructurales de tamaño 3x3 se puede resumir la estructura espacial local de una imagen.

Dentro del *kernel* la información de la estructura es codificada en información binaria {0, 1} y los patrones binarios resultantes pueden representar aristas orientadas, segmentos rectos, bifurcaciones, crestas, puntos montados, etc. La Figura 10 muestra algunos ejemplos de estos *kernels* de estructuras.

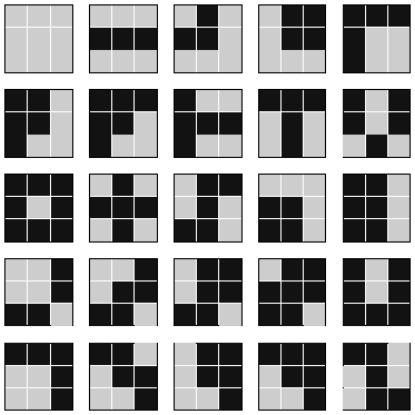

Fig. 10. Un subconjunto de algunos de los posibles *kernels* de estructura local en una vecindad de 3x3

En una cuadrícula de 3x3 existen $2^9 = 512$ de estos tipos de *kernels*. Realmente solo hay $2^9 - 1$ (511) *kernels* razonables dentro de los 2^9 *kernels* posibles porque el que tiene todos los elementos 0 y el que tiene todos los elementos 1 transmiten la misma información (todos los pixeles son iguales) por lo que es redundante y se excluye. Luego, cada punto de la imagen se representa con el kernel que mejor corresponda. El procedimiento completo puede ser pensado como un filtrado no lineal donde a la imagen de salida es asignada el índice del kernel que mejor se ajuste a cada punto.

La CT en su forma original puede ser interpretada como un índice de esas estructuras *kernel* con el centro fijado en 0. Luego, sólo es capaz de representar 2^8 =256 de las 511 estructuras *kernel* definidas en una vecindad de 3x3.

En [41] se cambió la base de la comparación con el objetivo de representar las 511 estructuras posibles. Con esta modificación, se comparan todos los pixeles de una vecindad local de 3x3 con el valor medio de las intensidades en esa vecindad. Se observa también entonces una gran similitud entre la CT Modificada y el LBP Mejorado, publicados ambos en el mismo período.

1.10 Transformada discreta del coseno aplicada de manera local

La DCT ha sido usada como un paso en la extracción de características en diferentes estudios de reconocimiento de rostros, tanto en algoritmos basados en la apariencia global como en otros basados en la apariencia local.

La DCT fue introducida por Ahmed, Natarajan y Rao a principios de los 70 y enseguida empezó a crecer en popularidad y fueron propuestas muchas variantes [42]. Principalmente han sido clasificadas cuatro transformaciones ligeramente diferentes: DCT I, DCT II, DCT III y DCT IV. De ellas, la más utilizada en el reconocimiento de rostros ha sido la DCT II.

Dado una secuencia de entrada *u(n)* de longitud *N*, su DCT, *v(k)*, es obtenido mediante la siguiente ecuación:

$$v(k) = \alpha(k)\sum_{n=0}^{N-1} u(n)\cos\left(\frac{(2n+1)\pi k}{2N}\right) \quad 0 \le k \le N-1 \tag{10}$$

donde:

$$\alpha(0) = \sqrt{\frac{1}{N}}, \alpha(k) = \sqrt{\frac{2}{N}} \quad 1 \le k \le N-1 \tag{11}$$

Alternativamente, se puede pensar en la secuencia *u(n)* como un vector y la DCT como una matriz de transformación aplicada a este vector para obtener la salida *v(k)*. En este caso, la matriz de transformación DCT, $C = \{c(k,n)\}$, se define como:

$$c(k,n) = \begin{cases} \frac{1}{\sqrt{N}} & k = 0, 0 \le n \le N-1 \\ \sqrt{\frac{2}{N}}\cos\left(\frac{(2n+1)\pi k}{2N}\right) & 1 \le k \le N-1, 0 \le n \le N-1 \end{cases} \tag{12}$$

donde *k* y *n* son los índices de las filas y las columnas respectivamente. Usando la Ecuación (12), la DCT de la secuencia *u(n)* (o vector **u**) es simplemente:

$$\mathbf{v} = C\mathbf{u} \tag{13}$$

La inversa de la DCT permite obtener *u(n)* a partir de *v(k)*, la misma está definida por:

$$u(n) = \sum_{n=0}^{N-1} \alpha(k)v(k)\cos\left(\frac{(2n+1)\pi k}{2N}\right) \quad 0 \le n \le N-1 \tag{14}$$

Con *α(k)* como es dado en la Ecuación (11). Utilizando la Ecuación (13), la inversa de la DCT, **u**, de un vector **v** se obtiene aplicando la inversa de la matriz *C* a **v**. Esto es:

$$\mathbf{u} = C^{-1}\mathbf{v} \tag{15}$$

De estas definiciones, se puede observar que aplicando la DCT a una secuencia de entrada, simplemente se descompone en una suma pesada de secuencias básicas de coseno.

La idea del uso de la DCT para el reconocimiento de rostros es calcular la DCT y retener un subconjunto de los coeficientes que conformarán el vector de características que describe el rostro. Este vector de características contiene los coeficientes DCT con las frecuencias de baja a media, que son los que tienen la mayor varianza, estos son extraídos mediante un recorrido en zig-zag. El coeficiente DCT del borde superior izquierdo es eliminado, ya que sólo representa el valor promedio de las intensidades en cada punto.

La diferencia entre los métodos globales y locales radica en que, en el primer caso se halla la DCT para la imagen completa, se conforma el vector con los coeficientes seleccionados y con éste se hace la clasificación, mientras que en el caso de los métodos locales primeramente se divide la imagen en bloques y luego se calcula la DCT y se construye el vector de coeficientes para cada uno de éstos. En éste último caso, una vez obtenidos los vectores por regiones, la fusión de éstos puede hacerse mediante dos vías, una es conformar un vector concatenando los coeficientes DCT obtenidos en cada bloque y con este realizar la clasificación (fusión de características) y la otra es realizar la clasificación por separado en cada bloque y luego combinar los resultados de la clasificación individual (fusión por decisión) [18].

En diferentes experimentos realizados, los métodos locales obtuvieron una tasa mayor de reconocimiento que los globales. Entre ellos, con algunas bases de datos se obtuvo mejor resultados utilizando la fusión por características, mientras que en otras, usando la fusión por decisión [18]. En la Tabla 1 se puede ver un resumen de la comparación de estos métodos en las bases de datos Yale y CMU PIE.

Tabla 1. Comparación del método DCT global y los locales

Método	**Taza de Reconocimiento**	
	Base de Datos Yale	**Base de Datos CMU PIE**
DCT Global	74.4%	44.1%
DCT Local (Fusión por características)	86.7%	70.9%
DCT Local (Fusión por decisión)	98.9%	68.5%

Como puede observarse, aún se hace necesario investigar en la selección de los coeficientes utilizados en la clasificación y en la manera de realizar esta.

2 Comparación entre los métodos existentes basados en la apariencia local

En este epígrafe se hace una comparación entre los resultados reportados en la literatura por los diferentes métodos de apariencia local y se presta especial atención a la robustez de estos a las variaciones de iluminación.

En la Tabla 2 se puede ver un resumen de los resultados reportados en la literatura [10] para algunos de los métodos descritos en el epígrafe anterior:

Tabla 2. Resumen de los resultados experimentales reportados por diferentes algoritmos

Método	Base de Datos de Rostros	Cantidad de Imágenes Probadas	Taza de Reconocimiento	Taza estándar de comparación	Principales variaciones en las imágenes
Método local probabilístico de sub-espacios	AR	600	82.3 %	70.2 %	expresión, tiempo
		400	71.0%	33.0%	oclusiones
Método local utilizando mapas auto-organizados	AR	600	93.7%	70.2 %	expresión, tiempo
		400	76.0%	33.0%	oclusiones
Modelos ocultos de Markov	AR	1440	89.8%	67.2%	expresión, tiempo, iluminación
Jets de Gabor	FERET	1196	95.0%	79.7%	expresión
Patrones binarios locales	FERET	1196	97.0%	67.2%	expresión, tiempo, iluminación

El método que utiliza rasgos fractales no se añadió a la comparación, puesto que no se reportan los porcientos de reconocimiento en la literatura, sin embargo, a pesar de que se muestran resultados favorables ante las variaciones de expresión, su rendimiento disminuye a medida que aumentan los individuos en la base de datos. Por otra parte, la transformada Census, además de que como descriptor, es muy similar al LBP, no se encontraron resultados aplicados directamente al reconocimiento de rostros.

El uso de las intensidades de los pixeles directamente, de PCA local y de DCT local como descriptores de imágenes de rostro fue comparado en [8] utilizando las GMM como clasificador. Para los experimentos se utilizó la configuración I de la base de datos XM2VTS [30], que consta de imágenes de entrenamiento y prueba con condiciones de iluminación controladas, las condiciones de iluminación de las imágenes del conjunto de prueba fueron degradadas artificialmente con transformaciones lineales y no lineales, creando nuevos conjuntos de prueba para evaluar las afectaciones provocadas por estos tipos de variaciones de iluminación. El *Equal Error Rate* (EER*)* se define como el error en el punto en el que se igualan el error de falsos aceptados (FAR) y el error de falsos rechazados (FRR), este punto se determina en el conjunto de evaluación y es utilizado como umbral para determinar los errores cometidos en los conjuntos de prueba, en éstos se calcula el *Half Total Error Rate* (HTER), que no es más que el promedio entre el FAR y el FRR. En la Tabla 3 se muestra un resumen de los resultados obtenidos:

Tabla 3. Comparación del uso de los pixeles directamente, PCA local y de DCT local como descriptores de rostro

Método	Imagenes sin afectaciones de iluminación		Imagenes con afectaciones de iluminación	
			lineales	no lineales
	EER	HTER	HTER	HTER
Pixeles directos	14.83	12.42	45.58	42.90
Pixeles directos eliminando la media	5.86	5.79	9.04	17.87
PCA local	5.68	5.00	6.52	8.53
DCT local	4.83	4.37	4.76	6.29

Una vez analizados cada uno de los métodos basados en la apariencia local existentes, todo parece indicar que los más adecuados para lidiar con el problema de la iluminación son el LBP y el DCT. Por otra parte, en [25] se hace un estudio comparativo de las aplicaciones más representativas que hacen uso del método LBP para el reconocimiento de rostros, prestando especial atención a la robustez frente a los problemas de iluminación. El método con el que se obtuvieron los mejores resultados fue el propuesto por [43] en el que se utiliza LBP como pre-procesamiento y el método DCT para extraer los vectores de rasgos.

Se hace entonces necesario estudiar con más profundidad los métodos LBP y DCT y su robustez a las variaciones de iluminación.

2.1 LBP como rasgo invariante a la iluminación

Como ya se dijo anteriormente, la idea del uso de LBP como rasgos viene dada porque la imagen de rostro puede ser vista como una composición de micro-patrones como 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 faciales independientemente de las variaciones de iluminación que afecten la imagen.

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

$$I_c = \rho_c \cdot {n_c}^T \cdot s_c \quad \text{and} \quad I_1 = \rho_1 \cdot {n_1}^T \cdot s_1 \tag{36}$$

donde ρ es el albedo, n^T las normales 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 éstas sean similares. Luego, lo que la diferencia entre dos pixeles vecinos descrita por el operador LBP realmente representa es la diferencia entre las iluminaciones incidentes:

$$I_1 - I_c = \rho \cdot n^T \cdot (s_1 - s_c) \tag{17}$$

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.

2.2 DCT para compensar las variaciones de iluminación

El método DCT por su parte, normalmente se utiliza en el reconocimiento de rostros, usando los coeficientes asociados a las bajas frecuencias para conformar los vectores de rasgos y con esto se obtienen resultados bastante satisfactorios. Sin embargo, en una imagen de rostro, la iluminación

normalmente cambia suavemente excepto para algunas sombras y especularidades, es decir, que las variaciones en la iluminación normalmente recaen en las bandas de bajas frecuencias. Por tanto, cuáles y cuántos coeficientes escoger para representar eficaz y eficientemente los rasgos del rostro y lograr una mayor robustez ante las variaciones de iluminación que afectan la imagen, representa aún una interrogativa.

No obstante, en [44] se presentó una forma diferente de usar los coeficientes DCT para compensar las variaciones de iluminación. En ese trabajo se mostró que las variaciones de iluminación pueden ser bien compensadas añadiendo o sustrayendo un término de compensación a la imagen dada en el dominio logarítmico. Esto sería fácil si se supiera con exactitud donde están las afectaciones de iluminación y dónde las características propias del rostro, pero en imágenes de rostros, especialmente cuando están afectadas por grandes variaciones de iluminación, la detección de rasgos no es una tarea trivial.

Luego, la DCT es utilizada para transformar la imagen del dominio espacial al dominio de la frecuencia y como las variaciones en la iluminación se espera que caigan sobre las bajas frecuencias, en [44] proponen poner en cero los coeficientes de bajas frecuencias en el dominio logarítmico para compensar las variaciones de iluminación.

Este método fue implementado con la DCT de la imagen completa y fue comparado con un gran número de métodos en la base de datos Yale B, que presenta grandes variaciones de iluminación y los resultados fueron muy superiores.

3 Conclusiones

En este reporte de investigación se ha hecho un estudio detallado de los métodos de apariencia local reportados en la literatura. Una vez hecho esto, se puede comprobar, que a pesar de que el uso de la apariencia local parece más adecuado para enfrentar los problemas de iluminación en las imágenes de rostros cuando contamos con solo una o muy pocas imágenes de entrenamiento, los métodos existentes hasta el momento no son capaces de lidiar eficazmente con este problema.

Los métodos LBP y DCT ofrecen los mejores resultados en los reportes cuando se comparan los distintos métodos en bases de datos con problemas de iluminación. No obstante, aún no son suficientes los resultados alcanzados. En el caso de LBP, vimos como es sensible cuando las variaciones de iluminación en las imágenes no son monotónicas. Mientras que en el caso de DCT aún se hace necesario encontrar cuáles y cuántos coeficientes utilizar y/o desechar para representar correctamente el rostro y alcanzar la invariancia a los cambios de iluminación.

Por otra parte, en los diferentes métodos se observa una gran variedad en las regiones locales que se utilizan. Por tanto se hace necesario determinar cómo dividir la imagen, así como qué forma y tamaño deben tener las regiones locales a utilizar.

Como conclusiones generales se puede decir que para enfrentar el problema de iluminación en el reconocimiento de rostros utilizando métodos de apariencia local, se hace imprescindible el surgimiento de nuevos métodos que logren subdividir la imagen, aplicar un método de normalización y/o extracción de rasgos adecuado a cada una de las partes y luego obtener una medida global de similaridad que permita comparar las diferentes imágenes de rostro.

Referencias bibliográficas

1. Stan Z. Li and Anil K. Jain: “Handbook of face recognition”, *Springer*, 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. Yael Adini, Yael Moses and Shimon 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. Florent Perronnin and Jean-Luc Dugelay: “A Model of Illumination Variation for Robust Face Recognition”, *Multimodal User Authentication Workshop*, 2003.
6. James Short, Josef Kittler and Kieron Messer: “A Comparison of Photometric Normalisation Algorithms for Face Verification”, *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.
7. Zia-ur Rahman, Glenn A. Woodell and Daniel J. Jobson: “A Comparison of the Multiscale Retinex with other Image Enhancement Techniques”, *Proceedings of IS&T 50th Anniversary Conference*, 1997.
8. Marc Saban and Conrad Sanderson: “On Local Features for Face Verification”, *IDIAP-RR 04-36*, 2004.
9. Athinodoros S. Georghiades, Peter N. Belhumeur y David J. Kriegman: “From few to many: illumination cone models for face recognition under variable lighting and pose”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.
10. Xiaoyang Tan , Songcan Chen , Zhi-Hua Zhou , Fuyan Zhang: “Face recognition from a single image per person: A survey”, *Pattern Recognition*, v.39 n.9, p.1725-1745, September, 2006.
11. R. Chellappa, C. Wilson, and S. Sirohey: “Human and machine recognition of faces: a survey”, *Proceedings of the IEEE*, 83-5:704–740, 1995.
12. Sebastien Marcel: “A tutorial on face recognition”, *IDIAP Research Institute*, Martigny, Switzerland, June 2007.
13. Diego A. Socolinsky, Lawrence B. Wolff, Joshua D. Neuheisel and Christopher K. Eveland: “Illumination Invariant Face Recognition Using Thermal Infrared Imagery”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'01) – Vol. 1, p. 527, 2001.
14. Villela P.R. and Juan Humberto Sossa Azuela: “Improving Pattern Recognition Using Several Feature Vectors”, *Lecture Notes in Computer Science*, Springer-Verlag, Heidelberg, 2002.
15. Kuncheva, L.I and Whitaker, C.J.: “Feature subsets for classifier combination: an enumerative experiment”, *Lecture Notes in Computer Science*, Vol. 2096. Springer, Berlin, 2001.
16. M. Villegas and R. Paredes: "Comparison of Illumination Normalization Methods for Face Recognition", *Third COST 275 Workshop*, 2005.

17. Timo Ahonen, Abdenour Hadid y Matti Pietikäinen: “Face Recognition with Local Binary Patterns”, *8th European Conference on Computer Vision*, 2004.

18. Hazim Kemal Ekenel y Rainer Stiefelhagen: “Local Appearance Based Face Recognition using Discrete Cosine Transform”, *13th European Signal Processing Conference (EUSIPCO)*, 2005.

19. Tan X., Chen S.C., Zhou Z.-H., and Zhang F.: “Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble”, *IEEE Transactions on Neural Networks*, v.16 n.4, p. 875-886, 2005.

20. Chen S.C., Liu J., and Zhou Z.-H.: “Making FLDA applicable to face recognition with one sample per person”, *Pattern Recognition*, v.37 n.7, p. 1553-1555, 2004.

21. Martinez, A.M.: “Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class”, IEEE Trans. Pattern Analysis and Machine Intelligence, v.25 n.6, p. 748-763, 2002.

22. Kepenekci B. , Tek F. B., G. Bozdagi Akar: “Occluded Face Recognition based on Gabor Wavelets”, *ICIP 2002*, Sept 2002, Rochester, NY, MP-P3.10.

23. Hung-Son Le y Haibo Li: “Recognizing frontal face images using Hidden Markov models with one training image per person”, *Proceedings of the 17th International Conference on Pattern Recognition(ICPR04)*, . 318 – 321, 2004.

24. Komleh H.E., Chandran V. and Sridharan S.: “Robustness to expression variations in fractal-based face recognition”, *Proc. of ISSPA-01*, Kuala Lumpur, Malaysia, pp. 359-362, 2001.

25. Sebastien Marcel, Yann Rodriguez and Guillaume Heusch: “On the Recent Use of Local Binary Patterns for Face Authentication”, *International Journal on Image and Video Processing Special Issue on Facial Image Processing*, 2007.

26. Loog, M. and de Ridder, D.: “Local Discriminant Analysis”, *18th International Conference on Pattern Recognition ICPR 2006,* Vol. 3, p.328 - 331, 2006.

27. Briechle K. y Hanebeck, U.: “Template Matching using Fast Normalized Cross Correlation”, *Proceedings of SPIE*, Band 4387, AeroSense Symposium, Orlando. Florida, 2001.

28. Matthew Turk y Alex Pentland: "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991.

29. Curtis Padgett y Garrison Cottrell: “Representing Face Images for Emotion Classification” *Cambridge*, MA: MIT Press, 1997.

30. “XM2VTS Face Database”, *http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/home.html*.

31. R. Martinez y R. Benavente: “The AR face database”, *Technical Report 24*, Computer Vision Center (CVC) Technical Report, Barcelona, 1998.

32. Y. Fisher, "Fractal Image Compression: Theory and Application," *Springer- Verlag Inc*., 1995.

33. F. Samaria y A. Harter: “Parameterisation of a Stochastic Model for human Face Identification,” *IEEE Workshop on Applications of Computer Vision*, Sarasota (Florida), December 1994.

34. F. S. Samaria: “Face recognition using hidden markov model,” *Ph.D. dissertation*, University of Cambridge, 1995.

35. H. Othman y T. Aboulnasr: “Low complexity 2-d hidden markov model for face recognition,” *IEEE International Symposium on Circuits and Systems*, vol. 5, Geneva, 2000, pp. 33–36.

36. Lee T. S.: “Image representation using 2-d Gabor wavelets”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 959-971, 18(10), Oct. 1996.

37. David S. Bolme: "Elastic bunch graph matching", *Thesis to obtain the Degree of Master of Science*, Colorado State University, 2003.

38. Timo Ahonen, Abdenour Hadid y Matti 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.

39. 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.

40. R. Zabih y J. Woodfill: "A non-parametric approach to visual correspondence", *IEEE Transactions on Pattern Analysis and Machine intelligence*, 1996.

41. B. Froba and A. Ernst: "Face Detection with the Modified Census Transform", *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

42. Ziad M. Hafed y Martin D. Levine: "Face Recognition Using the Discrete Cosine Transform", *International Journal of Computer Vision*, vol. 43, no. 3, 167–188, 2001.

43. Guillaume Heusch, Yann Rodriguez y Sebastien Marcel: "Local Binary Patterns as an Image Preprocessing for Face Authentication", *IEEE International Conference on Automatic Face and Gesture Recognition*, 2006.

44. Weilong Chen, Meng Joo Er, Member y Shiqian Wu: "Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain", *IEEE Transactions on Systems, Man and Cybernetics-B*, 2006.

RT _006, Octubre 2008
Aprobado por el Consejo Científico CENATAV

Editor: Lic. Miriela Santos Toledo
Diseño de Portada: DCG Matilde Galindo Sánchez
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;
Ciudad de La Habana. Cuba. C.P. 12200
Impreso en Cuba