

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Algoritmo de reconocimiento de
rostros basado en la apariencia local
para aplicaciones reales en
condiciones variables de
iluminación**

Heydi Méndez Vázquez

RT_047

enero 2012





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**

**Algoritmo de reconocimiento de
rostros basado en la apariencia local
para aplicaciones reales en
condiciones variables de
iluminación**

Heydi Méndez Vázquez

RT_047

enero 2012



Centro de Aplicaciones de Tecnologías de Avanzada
Departamento de Reconocimiento de Patrones



Algoritmo de Reconocimiento de Rostros basado en la Apariencia Local para Aplicaciones Reales en Condiciones Variables de Iluminación

Tesis presentada en opción al grado científico de
Doctor en Ciencias Técnicas

por

Ing. HEYDI MÉNDEZ VÁZQUEZ

Supervisada por :

Dr. EDEL GARCÍA REYES (CENATAV)

Dr. JOSEF KITTLER (CVSSP, Universidad de Surrey, UK)



CUJAE
Instituto Superior Politécnico
José Antonio Echeverría

Ciudad de La Habana, 2010

*Para mi madre, este pequeño fruto de mi esfuerzo,
que es poco comparado con su vida entera dedicada a mí.*

Agradecimientos

A todos los que de una forma u otra han contribuido a la realización de este trabajo:

A mi tutor Edel, por la guía y el ejemplo, por todo lo que me ha enseñado y ha contribuido no solo para la realización de esta tesis, sino a mi formación científica y profesional.

A Josef, por sus valiosas ideas y por ser una parte esencial de esta investigación; igualmente a él y a su esposa Antonia, por ofrecerme un poco de calor hogareño al encontrarme sola, lejos de casa.

A Ho, por sus chinicos y enredados, pero importantes consejos, por todos los experimentos que corrió para mí.

A Dayron, por sus resultados tan útiles para esta tesis, por todo lo que me ha ayudado, trabajando bien y con independencia.

A los profesores Juan Lorenzo Ginori y Juan Carlos Sepúlveda por sus observaciones y sugerencias en la predefensa de la tesis, que ayudaron a mejorar la calidad la misma.

A mi mamá, por estar como siempre a mi lado dándome su apoyo y amor incondicional, por las tantas veces que leyó los bocetos de este documento, buscando mejorarlo aún sin comprenderlo; así ha sido siempre en mi vida, gracias a ella he llegado hasta aquí.

A Ricardo, por su amor y comprensión, por ayudarme a encontrar la tranquilidad y a la vez darme fuerzas en los momentos más difíciles, por el gran ser humano que es y lo que aprendo a su lado cada día, por estos cinco años que han sido los mejores de mi vida.

A Noslen, por ser la mejor amiga del mundo y aguantarme tantos años, porque siempre halla un huequito para hacerle un favor a los demás, por su chispa y alegría que contagian, también por el estrés que una vez más hemos compartido y por la portada, claro.

A Dina, por ser más que un amigo, por la sabiduría que esparce a su alrededor, por haber estado siempre pendiente de mí y dispuesto a ayudarme cada vez que lo he necesitado.

A Annette y Candy, por soportarme todas las mañanas y también las tardes, por ser capaces de sacarme una sonrisa hasta en los peores días.

A Iren, Adry, Meyin, Rocío y Emma, por la amistad que perdura, porque sé que están siempre ahí en los buenos y en los malos momentos.

A Chang, por ser casi mi mano derecha; a Yeni, por ser una buena retaguardia; a Yadira y Ricardo Serrano, por soportar mis regaños; gracias a todos el proyecto avanza.

A Diana, porque nunca dice que no, por estar siempre dispuesta a asumir parte de mis tareas, sobre todo en la recta final; también a la One que algo le ha tocado.

A Garea, Calvo, Gil, Mabel, Gabriel y Javier, por ser tan buenos compañeros y soportar mi galillo de vez en cuando, porque RP será siempre mi dpto.

A Shul, por este sueño suyo que se ha convertido en mi realidad, gracias por esta oportunidad y también por confiar en mí.

A Isneris, por su constante preocupación y por permitirnos a veces algunas malcriadeces, sin dejar de velar que no desviemos el camino.

A Palancar y Medina, porque también han contribuido a la preparación que he recibido hasta llegar aquí.

A mis compañeros de CENATAV, porque todos han aportado un granito de arena; muchas gracias en especial a los que me han permitido tomar y usar sus fotos, aún a regañadientas.

A los colegas de Surrey, por los coffee breaks y por hacer más amenas mis estancias.

A mi familia TODA, los que están y los que ya se han ido, a mi papá que siempre estará; también a los que se han unido en el camino; porque toda mi vida se las debo, porque todo lo que soy es un reflejo de la crianza, el cariño y el apoyo que siempre he tenido.

A Tere, Ricardo, Reiner y todo el familión de Morón, por haberme aceptado como a una más de la familia y estar siempre pendientes de mi.

A todos aquellos que puedo estar olvidando ahora, pero que me han guiado y apoyado en el transcurso de este largo camino; a la Revolución cubana, no es una consigna, sin sus conquistas no habría tenido la posibilidad de llegar hasta aquí.

HEYDI MÉNDEZ VÁZQUEZ.
5 de enero de 2012.

Síntesis

El reconocimiento de rostros en condiciones de iluminación variables, es una problemática existente en el desarrollo de sistemas biométricos. Los métodos de pre-procesamiento son los más indicados para aplicar a las imágenes de rostros afectadas por las variaciones de iluminación, pero se ha observado que cuando son aplicados a imágenes de rostros tomadas bajo buenas condiciones de iluminación, afectan la eficacia obtenida por los clasificadores utilizados para discriminar entre un rostro y otro. En esta tesis se propone un nuevo algoritmo de reconocimiento de rostros para ser utilizado en condiciones de iluminación variables. En el algoritmo se evalúa la calidad de las imágenes de rostros respecto a la iluminación y sólo si se determina que están afectadas se pre-procesan antes de extraer los rasgos y realizar la clasificación. Para la evaluación de la calidad de las imágenes de rostros se desarrolla un método novedoso basado en una división triangular de la imagen que permite analizar de manera local las afectaciones provocadas por la iluminación en las distintas regiones del rostro. El pre-procesamiento de las imágenes afectadas se lleva a cabo combinando el mejor método de pre-procesamiento encontrado en la literatura, denominado PS, con un nuevo método basado en la Transformada Discreta del Coseno aplicada de manera local, LDCT. La representación y clasificación de las imágenes se realiza con el método que utiliza los patrones binarios locales multi-resolución (MLBP) y el análisis discriminante lineal (LDA), seleccionado dentro de los métodos basados en la apariencia local existentes en la literatura. Cada uno de los métodos propuestos para utilizar en las diferentes etapas del algoritmo es evaluado en bases de datos de pruebas internacionales, demostrando experimentalmente una mayor eficacia que los reportados en la literatura para esos propósitos. El algoritmo de reconocimiento de rostros que integra los métodos propuestos, se evalúa además con imágenes capturadas en ambientes reales con distintas condiciones de iluminación. Los resultados obtenidos con el algoritmo propuesto se comparan con los obtenidos por un sistema comercial en el reconocimiento de los rostros en estas imágenes, demostrando la factibilidad de su uso en aplicaciones reales en las que no se tenga control sobre la iluminación que incide en los rostros.

Abstract

Face recognition under varying lighting conditions remains an unsolved problem in the development of biometric systems. Preprocessing methods are the most suitable to tackle this problem, but it has been observed that they introduce negative effects or remove valuable discriminatory information when they are applied to face images in good illumination conditions. In this work a new face recognition algorithm for real systems in varying lighting conditions is proposed. The new algorithm evaluates face images illumination quality and only if they are affected they are preprocessed before extract features and make the classification. For evaluating face images illumination quality, a novel method is proposed, based on a new triangular division which allows to analyze illumination affectations in the different face regions. The pre-processing of face images affected by illumination is carried out combining the best state-of-the-art preprocessing method, PS, with a new method based on local Discrete Cosine Transform, LDCT. Face images representation and classification is done with the method that uses Multi-scale Local Binary Patterns (MLBP) with Linear Discriminant Analysis (LDA), which is selected among the local appearance based method in the literature. Each method proposed for the different steps of face recognition algorithm, is evaluated on face images databases for international tests, showing experimentally better results than state-of -the-art methods in all cases. The proposed face recognition algorithm is also evaluated on images captured on real environments with varying lighting conditions. The obtained results are compared with the ones obtained by a comercial system in face recognition of those images, showing that the proposed method can be effectively used in real applications in which the illumination conditions of face images are variable and unknown.

Índice general

Introducción	1
1 El reconocimiento de rostros y las variaciones en las condiciones de iluminación	11
1.1 Algoritmos de reconocimiento de rostros	11
1.2 Métodos basados en la apariencia local	15
1.2.1 Patrones Binarios Locales (LBP)	19
1.2.2 Transformada Discreta del Coseno (DCT)	22
1.3 El problema de la iluminación	25
1.3.1 Métodos de pre-procesamiento	28
1.3.2 Métodos de evaluación de la calidad	34
1.4 Consideraciones finales del capítulo	36
2 Evaluación de la calidad de imágenes de rostros respecto a la iluminación	37
2.1 División en regiones de las imágenes de rostros	37
2.2 Análisis de la iluminación en las diferentes regiones del rostro	40
2.2.1 Análisis de la región de la nariz	45
2.3 Calidad de la imagen respecto a la iluminación	46
2.4 Evaluación experimental	48
2.5 Conclusiones parciales	52
3 Pre-procesamiento de imágenes de rostros basado en el análisis local	53
3.1 Compensación de las variaciones de iluminación en las bajas frecuencias	54
3.2 Método de pre-procesamiento basado en la DCT local	56
3.2.1 Estimación del término de compensación	60
3.3 Evaluación del método de pre-procesamiento propuesto	63
3.3.1 Selección de los parámetros del método	65

3.4	Comparación con otros métodos de pre-procesamiento	68
3.4.1	PS vs. LDCT	70
3.5	Combinación de métodos de pre-procesamiento	72
3.6	Conclusiones parciales	74
4	Algoritmo de reconocimiento de rostros para sistemas reales en condiciones variables de iluminación	75
4.1	Reconocimiento de rostros utilizando métodos locales	75
4.2	Evaluación del Algoritmo Propuesto	78
4.2.1	Verificación	78
4.2.2	Identificación	80
4.3	Pruebas con imágenes en condiciones reales	83
4.4	Algunas consideraciones computacionales	87
4.5	Conclusiones parciales	88
	Conclusiones y Recomendaciones	91
	Referencias Bibliográficas	93
	Producción científica de la autora sobre el tema de la tesis	111
	Glosario de acrónimos	113
	Glosario de términos	115
	Anexos	117
	Anexo 1: Listado de sistemas de reconocimiento de rostros más importantes . . .	117
	Anexo 2: Funcionamiento del operador MLSEC	118

Introducción

La Biometría es una disciplina del Reconocimiento de Patrones que consiste en reconocer automáticamente a los individuos por sus características físicas o su comportamiento. Dentro de las características físicas más utilizadas se encuentran las huellas dactilares, los rasgos faciales, el iris y la geometría de la palma de la mano, mientras que entre las características del comportamiento se incluyen la firma, la forma de caminar y la forma de teclear [Jain 07]. Existen otras características, como la voz, que se consideran una mezcla de características físicas y del comportamiento.

Debido a que es más difícil suplantar la propia identidad, que los documentos de identificación o las contraseñas, el uso de sistemas biométricos ha cobrado una gran importancia, ya que logra sustituir los métodos tradicionales utilizados para reconocer a las personas y hacerlos más seguros. Estos sistemas tienen dos objetivos fundamentales: la *identificación* de un individuo dentro de un conjunto de personas (comparación 1 vs. M) y la *verificación* de la identidad de una persona a partir de dos muestras de alguna de sus características biométricas (comparación 1 vs. 1).

El uso de sistemas biométricos en aplicaciones prácticas ha ido expandiéndose, saliendo del marco de su origen, en la identificación de criminales, para ser utilizados en sistemas de seguridad, gubernamentales y en otros con objetivos comerciales, llegando incluso a ser aplicados como soluciones a problemas en la medicina. No obstante, cada característica biométrica es diferente, y no siempre todas son adecuadas en las distintas aplicaciones. Al diseñar un sistema biométrico se evalúa un conjunto de parámetros, como el poder distintivo de la característica biométrica que se emplee y la facilidad que se tenga para su uso, entre otros, que determinan su utilidad en las aplicaciones reales. En la Tabla 1, cuyos datos fueron extraídos del Manual de Biometría [Jain 07], se muestra el comportamiento de algunos de estos parámetros en sistemas que utilizan las diferentes características biométricas.

En la tabla, se aprecia que los sistemas basados en el rostro o rasgos faciales, presentan la

Tabla 1: Comportamiento de los sistemas que utilizan distintas características biométricas.

	Huella Dactilar	Rostro	Iris	Geometría de la Mano	Voz
Capacidad distintiva	Muy Alto	Alto	Muy Alto	Alto	Medio
Facilidad de uso	Alto	Alto	Bajo	Medio	Alto
Seguridad ante ataques	Media	Media	Muy Alta	Alta	Media
Aceptación	Media	Muy alta	Bajo	Medio	Muy alta
Estabilidad de los datos	Alta	Media	Alta	Alta	Media

mayor aceptación, esto se debe fundamentalmente a que es la característica biométrica más común al ser humano. El reconocimiento automático de rostros se basa en el procesamiento de imágenes, por lo que es considerada una técnica relativamente fácil de usar y no invasiva, que en este contexto significa que no requiere el contacto directo de las personas con los dispositivos de captura, es decir, las cámaras. Esto también facilita que pueda ser utilizado tanto en aplicaciones en las que las personas cooperen y conozcan que están siendo identificadas, como en las que se realice el proceso de manera encubierta. Por las razones expuestas, el reconocimiento automático de rostros es una de las técnicas biométricas más utilizadas en aplicaciones reales [Jain 07].

Los algoritmos de reconocimiento automático de rostros constan de un conjunto de pasos que se ejecutan a partir de que se tiene una imagen digital en la que se encuentra el rostro de la persona que se quiere reconocer. El primero de estos pasos consiste en la detección del rostro dentro de la imagen. El segundo paso es la alineación o normalización, que incluye transformaciones geométricas y/o correcciones de la iluminación, con el objetivo de estandarizar las diferentes imágenes de rostros. Una vez que las imágenes están normalizadas se procede a extraer los rasgos faciales que son usados para el reconocimiento, representados normalmente como un vector de características. Finalmente tiene lugar la clasificación, que no es más que comparar la información extraída en el paso anterior, contra uno o más modelos de rostros, para finalmente obtener una respuesta sobre la imagen de rostro que se está analizando [Li 05].

Las imágenes digitales utilizadas en el reconocimiento automático de rostros pueden ser bidimensionales o tridimensionales [Li 05, Akarun 05, Gokberk 09]. Las imágenes bidimensionales pueden provenir de distintos sensores, como cámaras fotográficas o

de video, que capturan imágenes en el espectro de luz visible, o cámaras infrarrojas [Socolinsky 01, Li 07]. Tanto las cámaras tridimensionales como las infrarrojas tienen elevados precios, por lo que las imágenes bidimensionales son las más utilizadas en el despliegue de aplicaciones reales. Los algoritmos existentes para el reconocimiento de rostros a partir de imágenes bidimensionales pueden ser clasificados en dos categorías teniendo en cuenta los rasgos que se utilizan para representar las imágenes: los basados en características geométricas y los basados en la apariencia.

Los algoritmos basados en características geométricas requieren la localización exacta de los rasgos faciales, lo cual aún no logra hacerse de manera precisa con métodos automáticos [Niu 05, Beumer 06, Shi 06]. Los algoritmos de reconocimiento de rostros basados en la apariencia, han sido los predominantes en los últimos años. Estos operan directamente con las intensidades de los píxeles u otras representaciones basadas en la imagen y han constituido un avance significativo para la eficacia y eficiencia de los sistemas de reconocimiento de rostros [Tan 06]. Los algoritmos basados en la apariencia pueden utilizar la información del rostro de manera global o de manera local. Los que se basan en representaciones globales, identifican un rostro utilizando para ello un vector que representa la imagen completa, mientras que los basados en métodos locales usan vectores de rasgos que representan localidades de la imagen de rostro. Estos últimos han mostrado ser más eficaces y más apropiados para sistemas reales en los que normalmente se cuenta con una sola imagen de muestra de entrenamiento por persona [Saban 04, Tan 06].

En los métodos basados en la apariencia local, la imagen se divide en regiones y cada una de estas se representa y analiza de manera individual, para luego, teniendo en cuenta la información aportada por cada región local, realizar la clasificación. Existen una gran variedad de métodos basados en la apariencia local, entre ellos, los que utilizan los Patrones Binarios Locales (LBP por sus siglas en inglés) [Ahonen 04] y los coeficientes de la Transformada Discreta del Coseno (DCT por sus siglas en inglés) [Ekenel 05] como rasgos extraídos de las regiones de la imagen, han alcanzado los mejores resultados [Tan 06, Marcel 07].

A pesar de los resultados alcanzados, la eficacia de estos algoritmos se ve afectada cuando las condiciones del entorno dónde se capturan las imágenes no son controladas y aparecen distintas variaciones en las imágenes como pueden ser: cambios en la expresión y en la pose, en las condiciones de iluminación, la presencia de accesorios, el envejecimiento, entre otras [Phillips 07]. Cada una de estas variaciones representa un problema en sí. En este trabajo se abordará el problema de las variaciones en las condiciones de iluminación.

Las variaciones en las condiciones de iluminación en el momento que se capturan las imágenes, son de las que más afectan la eficacia de los sistemas de reconocimiento de rostros [Adini 97, Phillips 07]. Muchas han sido las propuestas para enfrentar las afectaciones que provocan estas variaciones de iluminación en las imágenes de rostros [Zou 07]. La mayoría de los métodos que presentan los mejores resultados, requieren un gran conjunto de imágenes de entrenamiento de cada persona en diferentes condiciones de iluminación [Belhumeur 98, Georghiades 01], lo cual es muy difícil de obtener en aplicaciones reales; otros son muy costosos computacionalmente [Zhao 00, Sim 01, Zhang 05b]. Uno de los enfoques de mayor aceptación para enfrentar los problemas de iluminación en aplicaciones reales, es el uso de métodos de pre-procesamiento [Du 05], también conocidos como métodos de normalización fotométrica. Éstos no requieren de un aprendizaje previo de las condiciones de iluminación, por lo que generalmente no necesitan más de una imagen de las personas que serán reconocidas y a la vez son muy eficientes computacionalmente.

Cuando se aplica el pre-procesamiento, este se inserta como el primer paso dentro de la secuencia que normalmente se ejecuta en el proceso de reconocer un rostro luego de detectado y alineado geométricamente, siguiéndole la extracción de rasgos y la clasificación. Los métodos de pre-procesamiento normalizan la imagen de rostro, con el objetivo de obtener una nueva imagen o representación del rostro estable bajo diferentes condiciones de iluminación. Estos métodos pueden basarse en transformar la imagen de rostro de manera global o realizar la transformación por regiones o particiones de la imagen (local). Se ha mostrado en la literatura que los métodos que tienen en cuenta que la afectación provocada por la iluminación es diferente en las distintas regiones de la imagen del rostro y enfrentan el problema con métodos locales, reportan los mejores resultados [Villegas 05, Short 06], aunque estos aún no son los ideales.

Muchos de los métodos existentes [Stockham 72, Rahman 96, Gross 03, Wang 04a, Chen 05, Tan 07] han sido comparados entre ellos, en el pre-procesamiento de imágenes con condiciones variables de iluminación. La principal conclusión a la que se ha llegado es que mientras mejor lidian con imágenes afectadas por variaciones de iluminación, presentan un comportamiento menos eficaz ante imágenes con condiciones de iluminación normales o favorables, principalmente cuando estas imágenes con buenas condiciones de iluminación, presentan otras variaciones de expresión, oclusión o pose [Short 04, Du 05, Tan 07]. En la mayoría de los casos, las transformaciones utilizadas para eliminar las afectaciones de iluminación, eliminan información valiosa para los fines discriminatorios en imágenes iluminadas normalmente [Du 05]. Esto significa, que a pesar de que en los sistemas de reconocimiento de rostros para ambientes con iluminación no controlada, se necesitan los

métodos de pre-procesamiento para enfrentar las afectaciones provocadas por las variaciones de iluminación, si estos se utilizan, se cometen errores al clasificar las imágenes que no estén afectadas.

Ante estas dificultades, se impone el desarrollo de nuevos métodos que presenten un mejor balance entre las ventajas del pre-procesamiento en imágenes afectadas por la iluminación y la disminución de su rendimiento en imágenes con condiciones de iluminación adecuadas. Además de desarrollar nuevos métodos de pre-procesamiento, una alternativa para disminuir las dificultades que se presentan, pudiera ser insertar la evaluación de la calidad de las imágenes respecto a la iluminación al algoritmo de reconocimiento de rostros que se utilice. Es decir, si se detecta que la imagen de rostro está afectada por la iluminación aplicar un método de pre-procesamiento, si no, procesar la imagen normalmente. Sin embargo, en la literatura no se han encontrado trabajos que presenten algoritmos de reconocimiento de rostros en los que se emplee este esquema. Lo más próximo a esto que se ha hecho es detectar regiones de la imagen afectadas por la iluminación y sólo pre-procesar éstas [Alvarez 06, Kryszczuk 05], con la desventaja de enfocarse en un único tipo de afectación por iluminación, ya sea sombras o reflejos especulares.

Por otra parte, existen varios métodos para evaluar la calidad de las imágenes de rostros, pero casi todos miden la distorsión de las imágenes causada por el proceso de adquisición, los métodos de compresión o el desenfoque [Wang 04b, Gao 05, Zamani 08]. Otros trabajos [Ferrara 08] persiguen evaluar algunas de las métricas definidas en el formato para el intercambio de datos de imágenes de rostros [INCITS 06], utilizado por la Organización de la Aviación Civil Internacional (OACI). De estos últimos, sólo se encontró un método que evalúa la calidad de las imágenes respecto a la iluminación [Gao 07] y se basa en la simetría vertical del rostro, lo cual no es del todo eficaz puesto que en ocasiones la afectación de la iluminación en el rostro puede ser simétrica y a su vez, la simetría del rostro puede estar afectada por otras variaciones no necesariamente de iluminación, como son la pose y la oclusión.

Según lo visto, se plantea como **problema de investigación** que los algoritmos de reconocimiento de rostros que utilizan solamente una o muy pocas imágenes de entrenamiento de cada sujeto y son eficientes computacionalmente, son poco eficaces cuando las condiciones ambientales de iluminación no están controladas y esto provoca que algunas imágenes de rostros estén afectadas y otras no.

Para resolver el problema, se propone como **objetivo general** de esta investigación, desarrollar un algoritmo de reconocimiento de rostros, que integre un método para determinar

si una imagen de rostro está afectada por las variaciones en las condiciones de iluminación con un método de pre-procesamiento para enfrentar esas afectaciones, de manera que sea más eficaz que los algoritmos existentes para aplicaciones reales en las que las condiciones de iluminación sean variables y exista solamente una o muy pocas imágenes de entrenamiento por persona.

Para lograr este objetivo se plantean los siguientes **objetivos específicos**:

1. Desarrollar un método que analice la influencia de la iluminación en el rostro y determine si una imagen de rostro se encuentra afectada.
2. Desarrollar un método de pre-procesamiento que sea más eficaz que los métodos existentes en el reconocimiento de imágenes de rostros bajo condiciones de iluminación variables.
3. Seleccionar de la literatura actualizada los métodos de extracción de rasgos y clasificación más adecuados para integrar al método de pre-procesamiento que se proponga.
4. Integrar los métodos propuestos, en un algoritmo que sea eficaz y no implique un alto costo computacional, de manera que pueda ser utilizado en aplicaciones reales de reconocimiento de rostros.

En esta investigación se propone entonces como **hipótesis** principal que si se aplica un método que permita evaluar la calidad de una imagen de rostro respecto a la iluminación, se puede aplicar el pre-procesamiento sólo a las imágenes afectadas, lo que permite elevar la eficacia de los algoritmos de reconocimiento de rostros utilizados en aplicaciones reales en las que las condiciones de iluminación son variables y generalmente no se cuenta con más de una imagen de entrenamiento por persona.

Teniendo en cuenta que el análisis de la literatura sugiere que los métodos de pre-procesamiento y los de reconocimiento basado en la apariencia local son los más eficaces y a la vez eficientes para enfrentar el problema de la iluminación en aplicaciones reales de reconocimiento de rostros [Saban 04, Villegas 05, Short 06, Tan 06], se plantean como hipótesis secundarias que:

H1. Si se modela localmente el comportamiento de la iluminación en los rostros, se puede cuantificar la magnitud de las afectaciones provocadas por sus variaciones en las distintas regiones del mismo y establecer una medida global que permita determinar si una imagen de rostro se encuentra afectada.

H2. Si se utiliza un método de pre-procesamiento basado en transformaciones locales, se puede aumentar la eficacia del algoritmo de reconocimiento de rostros que se proponga para enfrentar condiciones de iluminación variables.

Para darle cumplimiento a los objetivos y demostrar las hipótesis planteadas se trazan las siguientes **tareas**:

1. Estudiar el estado actual de los algoritmos de reconocimiento de rostros adecuados para aplicaciones reales, específicamente los basados en la apariencia local, para identificar las ventajas de cada uno al enfrentar las variaciones de iluminación.
2. Analizar los diferentes métodos propuestos en la literatura para enfrentar el problema de las variaciones en la iluminación en el reconocimiento de rostros con el objetivo detectar sus limitaciones y posibles vías de solución.
3. Desarrollar e implementar un método que permita analizar el comportamiento de la iluminación en las distintas regiones de una imagen de rostro.
4. Desarrollar e implementar un método que a partir del análisis local de las regiones, permita determinar si una imagen de rostro está afectada por la iluminación.
5. Desarrollar e implementar un método de pre-procesamiento local que sea más eficaz que los métodos existentes ante distintas variaciones de iluminación.
6. Diseñar y realizar los experimentos que permitan comparar los resultados que se obtengan con los métodos propuestos, con respecto a los resultados alcanzados por los métodos existentes en la literatura, utilizando para ello bases de datos internacionales de imágenes de rostros.
7. Seleccionar, a partir del estudio del estado actual y los resultados experimentales, el método basado en la apariencia local más adecuado para integrar al método de pre-procesamiento propuesto.
8. Definir los pasos específicos del algoritmo de reconocimiento de rostros que combine los métodos propuestos.
9. Implementar el algoritmo que integra los métodos propuestos y verificar su desempeño con imágenes de bases de datos de pruebas internacionales e imágenes obtenidas en condiciones reales, con iluminación variable.

Entre los **métodos de investigación** utilizados para llevar a cabo la investigación, se destaca el método general *hipotético-deductivo*, guiado por la observación de las problemáticas detectadas y el planteamiento de hipótesis que den respuestas a las preguntas ¿qué hacer? y ¿cómo lograrlo? que luego son corroboradas o validadas.

Para ejecutar el primer paso de este proceso científico se utilizó el método lógico *inductivo-deductivo*, realizándose un estudio del estado actual de los algoritmos de reconocimiento de rostros utilizados en aplicaciones reales, que permitió obtener un conocimiento general de los problemas que afectan la eficacia de estos ante condiciones de iluminación variables y proponer una hipótesis de partida para darle solución a las deficiencias detectadas. Teniendo en cuenta que el problema detectado consta de varios elementos a resolver, se utilizó también el método *analítico-sintético* para descomponerlo en partes y profundizar en el estudio de cada una de ellas por separado, buscando soluciones parciales que luego son nuevamente integradas en el algoritmo general que se propone.

Con el objetivo de representar las imágenes de rostros de manera numérica y poder determinar sus propiedades y relaciones de manera que pueda evaluarse la capacidad del algoritmo que se propone de discriminar entre un individuo y otro, se utilizó el método empírico de la *medición*, apoyado en procedimientos estadísticos y métodos matemáticos *algebraicos y aritméticos*, que permiten a su vez, formalizar los nuevos métodos basados en el análisis local de las imágenes de rostros.

Finalmente, mediante el método empírico *experimental* se llevan a cabo las pruebas necesarias para validar cada uno de los métodos y el algoritmo que se propone, que a su vez se analizan y comparan con los resultados de los algoritmos de reconocimiento de rostros que aparecen en la literatura, apoyado en el método lógico de *comparación-clasificación*. De manera auxiliar se utiliza el método empírico *coloquial* para la presentación y discusión de los resultados obtenidos en sesiones científicas.

La **novedad científica** de este trabajo radica en los aportes que se hacen en el Reconocimiento de Rostros, específicamente en el enfrentamiento de los problemas de iluminación en aplicaciones reales, lo cual es un aspecto no resuelto y de investigación activa en esta área. El principal aporte es el nuevo algoritmo de reconocimiento de rostros, que al utilizar el esquema general representado en la Figura 1, logra ser más eficaz que los algoritmos existentes al reconocer imágenes con condiciones de iluminación variadas. Hay otros aspectos novedosos enmarcados en cada una de las etapas del proceso representado en la figura, que se pueden resumir en:

- Un nuevo método de subdivisión en regiones de las imágenes de rostros, en el que se tiene en cuenta la incidencia de la iluminación en las distintas partes del rostro.
- Un nuevo método de evaluación de la afectación por iluminación de una imagen de rostros basado en el análisis local de la influencia de la iluminación en el rostro, más efectivo que el único método encontrado en la literatura con este fin.

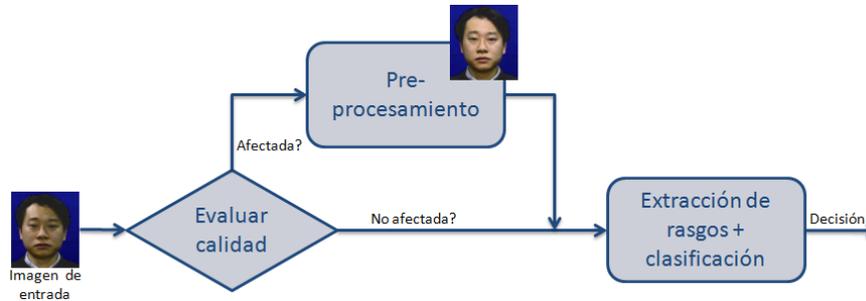


Figura 1: Esquema de algoritmo que se propone.

- Un nuevo método de pre-procesamiento de imágenes de rostros basado en transformaciones locales, en el que se estima y suprime la iluminación incidente sin comprometer la información discriminadora de las imágenes de rostros.
- Una nueva propuesta de combinación de métodos de pre-procesamiento con métodos basados en la apariencia local, que logra una mayor eficacia en el reconocimiento de imágenes de rostros con condiciones de iluminación variadas.

La **significación práctica** de este trabajo viene dada en primer lugar, por el uso del algoritmo propuesto en el desarrollo de sistemas propios de reconocimiento de rostros. Los sistemas de reconocimiento de rostros existentes en el mercado internacional tienen elevados precios (ver Anexo 1), sin contar que en ocasiones existen restricciones para su adquisición y despliegue. El algoritmo que se propone en esta tesis, parte de las premisas de que sea eficiente, de manera que pueda ser utilizado en cualquier aplicación, así como de no estar restringido a la necesidad de varias imágenes de entrenamiento de cada individuo. Es necesario aclarar, que si bien en esta investigación se propone obtener un algoritmo de reconocimiento de rostros más eficaz que los existentes hasta el momento cuando las imágenes son tomadas en condiciones de iluminación variables, en un principio el alcance de la misma no contempla otro tipo de variaciones drásticas en el rostro como la oclusión, el envejecimiento o ángulos de desviación de la pose mayores de 15° . Esto sin embargo, no impide la utilización del algoritmo en aplicaciones reales en las que las condiciones de iluminación son variables y los cambios de apariencia, como los de pose y expresión, no sean drásticos. Ejemplos de estas aplicaciones son los sistemas para chequeos migratorios en los que las condiciones normalmente se controlan pero pueden cambiar de un escenario a otro, sistemas para detectar duplicados o intentos de suplantación de la identidad en registros nacionales, sistemas de control de acceso a lugares restringidos que pueden estar en interiores o al aire libre y se les pide a los usuarios mirar a las cámaras, sistemas de video-vigilancia ubicados en pasillos, entradas de locales o calles estrechas donde se

llame la atención de manera encubierta a la vista de los individuos, entre otras muchas aplicaciones.

Los principales resultados y aportes teóricos que se presentan en esta tesis han sido presentados en eventos y publicaciones tanto nacionales como internacionales, las cuales aparecen listadas al final de este documento. De igual forma como parte de la validación del algoritmo que se propone, se incluyen pruebas realizadas con imágenes capturadas en los puntos de frontera del aeropuerto internacional “José Martí” de Ciudad de la Habana.

Este documento está conformado por cuatro capítulos. En el Capítulo 1 se presenta un estudio de los métodos existentes en el Reconocimiento de Rostros, haciendo énfasis en los que muestran los mejores resultados prácticos, principalmente cuando se enfrentan condiciones variables de iluminación, de manera que se puedan comprender la magnitud de las problemáticas existentes y las soluciones que se proponen. En el Capítulo 2 se propone un método de evaluación de la calidad de una imagen de rostro respecto a la iluminación, que tiene en cuenta la influencia local de la misma en las distintas áreas del rostro. Se comprueba la efectividad del método en la evaluación de imágenes con condiciones de iluminación variadas y se comparan los resultados con los obtenidos por el único método encontrado en la literatura para este propósito. En el Capítulo 3 se presenta un nuevo método de pre-procesamiento basado en la Transformada Discreta del Coseno aplicada de manera local. El método se evalúa y compara con los existentes en la actualidad, mostrando una excelente eficacia, muy similar a la del método que exhibe los mejores resultados en la literatura. Se propone una combinación de clasificadores, que integra y complementa el funcionamiento de ambos métodos utilizando el método de reconocimiento basado en la apariencia local que exhibe los mejores resultados. En el Capítulo 4 se describen los pasos del algoritmo de reconocimiento de rostros que se propone utilizar en sistemas reales en los que las condiciones de iluminación sean variables. El algoritmo propuesto se evalúa con imágenes de bases de datos de pruebas y con imágenes capturadas en ambientes reales y se compara con los algoritmos de mejores resultados reportados en la literatura, mostrando una mayor eficacia. Se presentan además algunas consideraciones computacionales en su implementación para sistemas reales. Finalmente se llega a las conclusiones y recomendaciones de la investigación realizada, se listan las referencias bibliográficas utilizadas, se incluyen los glosarios de acrónimos y términos, y dos anexos que complementan el trabajo presentado.

Capítulo 1

El reconocimiento de rostros y las variaciones en las condiciones de iluminación

En este capítulo se ofrece una panorámica sobre los algoritmos de reconocimiento de rostros existentes. Se provee un marco de referencia para interpretar las soluciones dadas a los problemas planteados, por lo que se presenta un estudio de los métodos basados en la apariencia local y su comportamiento ante las variaciones de iluminación. Se analizan de manera especial los métodos que enfrentan el problema de las variaciones en las condiciones de iluminación, haciendo énfasis en los de pre-procesamiento. Finalmente se referencian los métodos existentes para el análisis de la calidad de imágenes de rostros.

1.1 Algoritmos de reconocimiento de rostros

El reconocimiento automático de rostros comenzó su desarrollo en los años 60 del pasado siglo, cuando un grupo de científicos norteamericanos desarrollaron el primer sistema semiautomático para reconocimiento facial que requería del usuario para localizar características faciales como los ojos, las orejas, la nariz y la boca en las fotografías. A partir de los puntos seleccionados, se calculaban las distancias entre ellos y eran comparadas con datos de referencia almacenados previamente [Zhao 03b]. En la década del 70, Goldstein y su grupo agregaron 21 marcadores subjetivos específicos, entre los que se encontraban el color del cabello y el grosor de los labios, con el objetivo de contar con mayor información para la comparación [Goldstein 71]. Estas primeras técnicas, basadas en la geometría de

las características faciales, no tuvieron mucho éxito en la práctica, por lo engorroso que resulta la anotación manual del gran número de imágenes de rostros que normalmente se procesan.

A principios de los años 90, Brunelli y Poggio presentaron un sistema de reconocimiento de rostros, que extraía automáticamente 35 características faciales correspondientes a puntos de los ojos, la nariz, la boca y la barbilla. Con esto se formaba un vector de 35 dimensiones para representar el rostro, que posteriormente se utilizaba para determinar la correspondencia con otros rostros por medio de un clasificador Bayesiano [Brunelli 93]. Los resultados reportados por este sistema fueron de un 90% de clasificación correcta para una base de datos de 47 individuos, siendo muy bajo el costo computacional y de almacenamiento del sistema. A pesar de los buenos resultados experimentales obtenidos, lo fundamental en este sistema era la localización automática correcta de las características distintivas del rostro, lo cual, incluso en la actualidad es una tarea muy complicada, principalmente cuando hay variaciones de pose, iluminación u oclusión en las imágenes. Con los métodos más modernos [Niu 05, Beumer 06, Shi 06] las tasas de error en la localización de los rasgos del rostro cuando existen distintas variaciones son aún altas. Por otra parte, al representar el rostro solamente por sus características faciales, se pierde información importante sobre la textura y los tonos presentes en la imagen.

Esto dio origen a las llamadas técnicas basadas en la apariencia, que significaron un gran avance para los sistemas de reconocimiento de rostros. En éstas, las intensidades de los píxeles en la imagen de rostro, o transformaciones de estas, son utilizadas como rasgos para la clasificación, empleándose para ello herramientas inteligentes de diversas áreas del conocimiento como la visión por computadoras, el reconocimiento de patrones, las máquinas de aprendizaje y las redes neuronales, entre otros.

El primero y más conocido algoritmo de este tipo, fue propuesto por Turk y Pentland en el año 1991 [Turk 91]. En el algoritmo conocido como *eigenfaces*, primero se construye un subespacio a partir de imágenes de rostros de entrenamiento, utilizando el análisis de componentes principales (PCA). Luego, las imágenes de rostros son representadas en este subespacio, que corresponde a los vectores propios de la matriz de covarianza de las imágenes de rostros. Solamente un subconjunto de vectores propios son utilizados para representar los rostros, con el objetivo de reducir las dimensiones. Para un subespacio de M -dimensiones, los vectores propios seleccionados son los correspondientes a los primeros M valores propios, ordenados de manera descendente de acuerdo a sus magnitudes. De esta forma las imágenes de rostros pueden ser reconstruidas con el mínimo error cuadrado-medio para cualquier dimensión del subespacio. La clasificación se realiza comparando las

imágenes de rostros en este subespacio.

Muchas han sido las extensiones al algoritmo *eigenfaces* con el objetivo de lograr representar mejor las variaciones inter e intra-personales de los distintos rostros. Entre estas extensiones se destacan los algoritmos basados en subespacios con análisis discriminatorio lineal (LDA) [Belhumeur 97, Zhao 99, XiaoJun 04], el basado en probabilidades Bayesianas [Moghaddam 97], el algoritmo basado en el clasificador de máquinas de vectores de soporte (SVM) [Phillips 99], el de clasificación de líneas de rasgos más cercanas [Li 99], el de búsqueda en evolución [Liu 00], y el *Laplacianfaces* [He 05]. Los autores de estos algoritmos los consideran mejores que el *eigenfaces* clásico, sin embargo, todos requieren varias imágenes de entrenamiento por cada persona para poder representar las variaciones intra-personales. De no ser así, estos sólo contarían con la representación de las variaciones inter-personales al igual que el *eigenfaces* y sus resultados pudieran ser incluso peores [Martinez 01]. Además en algunos casos, como el del algoritmo basado en el clasificador SVM y el de líneas de rasgos más cercanas, no funcionan si sólo cuentan con una imagen de entrenamiento de cada individuo [Phillips 99, Li 99].

Para aplicaciones reales, estos sistemas basados en varias imágenes de entrenamiento por persona presentan limitaciones. En primer lugar, no siempre existe la posibilidad de contar con más de una imagen de entrenamiento de las personas que se desean identificar, en aplicaciones de seguridad por ejemplo, no es posible contar con fotos de los individuos circulos con diferentes variaciones, normalmente se cuenta solamente con una foto obtenida de algún organismo oficial. Construir además la base de datos con la colección de imágenes de los individuos, puede ser muy costoso tanto en tiempo y esfuerzo, como en capacidad de almacenamiento. Por tanto, mientras menor sea la cantidad de imágenes a almacenar, mejor. Finalmente el costo computacional de las operaciones involucradas en el sistema de reconocimiento de rostros está directamente relacionado con la cantidad de muestras de entrenamiento, mientras más sean las imágenes en la colección, más costosos se vuelven los procesos de aprendizaje y de clasificación de una nueva imagen.

A partir de la expansión que han tenido los sistemas de reconocimiento de rostros en aplicaciones reales, muchos investigadores han centrado sus esfuerzos en mejorar la robustez de los algoritmos ante diferentes variaciones cuando sólo se cuenta con una imagen de entrenamiento por persona, surgiendo una gran cantidad de algoritmos que pueden ser divididos en dos grandes grupos teniendo en cuenta la forma en que se analiza la imagen: global o local [Tan 06].

El primero de estos grupos lo forman los algoritmos que trabajan con una representación

global de las intensidades de los píxeles en la imagen. Estos tienen la ventaja de preservar la información de la forma y la textura del rostro, lo cual es útil para distinguir entre un rostro y otro. Sin embargo, al utilizar esta representación con sólo una imagen de muestra por persona, aumenta el dilema entre las altas dimensiones del espacio donde se representa la imagen y las pocas muestras, limitándose además el trabajo con muchas de las técnicas de reconocimiento de patrones que necesitan representaciones de las variaciones intra-clases. Hay dos vías fundamentales por las que se ha enfrentado estos problemas. La primera, es tratar de comprimir tanta información como sea posible de una sola imagen de rostro, ya sea en el espacio de altas dimensiones o más comúnmente, en el espacio propio de dimensión reducida, como hacen los métodos $(PC)^2A$ [Wu 02], el $(PC)^2A$ mejorado [Chen 04b] y el de perturbación por descomposición de vectores únicos (SVD) [Zhang 05a], que persiguen enriquecer el espacio propio de rostros perturbando espacialmente la única imagen de muestra que se tiene de cada persona, de manera que sobresalgan las características relevantes y desaparecen las menos importantes. La otra variante consiste en incorporar información previa, ya sea construyendo nuevas vistas o utilizando diferentes representaciones de cada imagen que se tiene. Los métodos que generan nuevas imágenes visuales a partir de la imagen que se tiene de cada persona [Poggio 97, Gutta 04] persiguen aprender información adicional del dominio de la imagen que se tiene como entrenamiento, mientras que los que utilizan diferentes representaciones [Martínez 02, DeLaTorre 05], buscan agrandar de modo efectivo el conjunto de entrenamiento, tratando de explotar más la información que se tiene de la imagen. En cualquier caso, los algoritmos basados en representaciones globales de una sola imagen por individuo, son muy sensibles a los cambios de apariencia que se producen en los rostros debido a las variaciones en la expresión, la iluminación y la pose, entre otros. Estos cambios en la apariencia de un rostro, afectan solamente una región del mismo y a pesar de esto, una pequeña variación afecta la representación global que se tiene de la imagen.

El segundo grupo de algoritmos, corresponde precisamente a los que utilizan métodos basados en la apariencia local para extraer los rasgos y clasificar las imágenes. En éstos, se obtienen vectores de rasgos que representan localidades de la imagen de rostro, que luego son integrados para obtener una clasificación final sobre la imagen. Los algoritmos que utilizan estos métodos basados en la apariencia local presentan ventajas sobre los que utilizan las representaciones globales, dentro de las cuales se destacan:

1. Los datos con menores dimensiones son más adecuados para los problemas de clasificación [Socolinsky 01]. 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 para reconocer un rostro basado en sus partes. Las características comunes y específicas de una clase pueden ser fácilmente identificadas cuando las variaciones, como normalmente ocurre, se producen sólo en una region del rostro [Villela 02].
3. Rasgos faciales diferentes pueden incrementar la diversidad de los clasificadores [Kuncheva 01], lo cual es favorable para la identificación de rostros.

Además de las ventajas anteriores, en diferentes estudios comparativos se ha visto que los algoritmos que utilizan estos métodos basados en la apariencia local obtienen mejores resultados que los basados en representaciones globales de las imágenes, principalmente cuando existen variaciones en las condiciones de captura de las imágenes y no se cuenta con varias imágenes de entrenamiento de cada individuo [Heisele 03, Saban 04, Tan 06]. Con el objetivo de comprender estos métodos y poder seleccionar el más adecuado para utilizar, se revisarán en más detalle y se analizará específicamente su robustez ante las variaciones de iluminación.

1.2 Métodos basados en la apariencia local

Los métodos de extracción de rasgos y clasificación basados en la apariencia local constan de cuatro pasos generales que pueden ser observados en la Figura 1.1. Estos son: 1) Partición en regiones locales, 2) Extracción de rasgos locales, 3) Selección de rasgos y 4) Clasificación basada en la fusión de la información local.

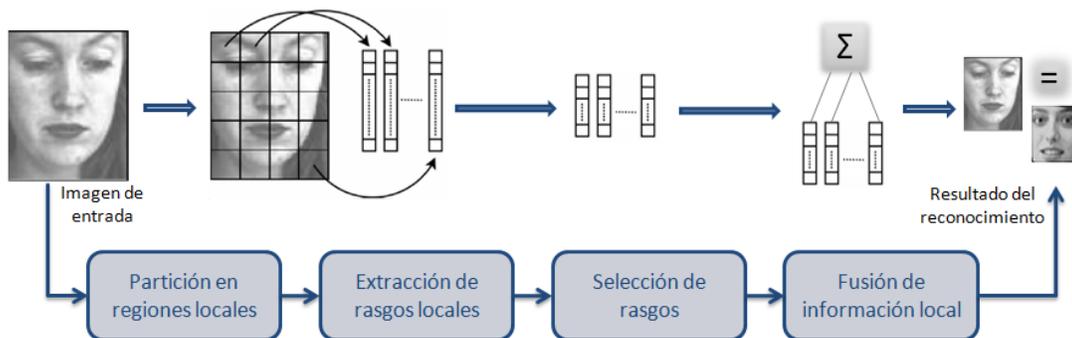


Figura 1.1: Pasos de los métodos basados en la apariencia local.

El primer paso, indispensable en los métodos locales, consiste en definir las regiones que van a ser utilizadas. Para esto se tienen en cuenta dos aspectos fundamentales: 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 [Ahonen 04, Chen 04a, Ekenel 05, Tan 05], las cuales pueden estar solapadas [Chen 04a, Tan 05], o no [Ahonen 04, Ekenel 05]. Sin embargo, se han utilizado también otras formas como elipses [Martínez 02] y franjas [Kepenekci 02, Le 04]. El tamaño de las regiones varía mucho en dependencia del método utilizado, esta es una cuestión bastante delicada que tiene influencia directa en la robustez de los métodos. El principal problema de estas divisiones es que son generales y no toman en consideración la información que brinda la estructura y las formas de los rostros en sí. Queda aún mucho por investigar en cuanto a qué forma y tamaño son los más indicados al utilizar estos métodos en el reconocimiento de rostros.

Una vez que se definen las regiones locales, es necesario representar esa información local. Este segundo paso, denominado extracción de rasgos, es crítico para el desempeño de los sistemas de reconocimiento de rostros e igualmente es difícil decidir cuál es el método más adecuado a utilizar en cada aplicación específica. Lo más simple es utilizar vectores formados directamente con las intensidades de los píxeles [Martínez 02, Chen 04a, Tan 05], sin embargo estos rasgos son demasiado sensibles a cambios de apariencia en la imagen y en particular a los de iluminación, que influyen directamente en los valores de intensidad de los píxeles en las imágenes. Por esta razón, diferentes rasgos derivados de las intensidades de los píxeles originales han sido propuestos, entre estos se destacan los rasgos fractales [Ebrahimpour 01], los *wavelets* de Gabor [Kepenekci 02], los *wavelets* de Haar [Le 04], los patrones binarios locales (LBP) [Ahonen 04] y los coeficientes de la transformada discreta del coseno (DCT) [Ekenel 05].

En dependencia del tipo de rasgo utilizado, en algunos casos se hace necesario un tercer paso adicional de selección de rasgos o reducción de dimensionalidad para mejorar la eficacia y/o eficiencia del método. Este paso consiste en retener solamente un subconjunto del vector de rasgos original o realizar una transformación del mismo que permita seleccionar la información más discriminativa. Entre los algoritmos más utilizados para la reducción de dimensionalidad se encuentran el PCA y el LDA, que en principio son métodos globales, pero han sido también usados de manera local [Saban 04, Loog 06]. 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 [Ekenel 05].

El cuarto y último paso es la clasificación basada en la fusión de la información local. Esta fusión puede ser de dos formas. La primera, consiste en concatenar en un solo vector los rasgos de las diferentes regiones y luego clasificar este único vector. La segunda variante se basa en métodos de combinación de clasificadores [Kuncheva 01], ya que se clasifica

cada región por separado y luego se combinan los resultados para tomar la decisión final. No existe un consenso sobre cual variante ofrece los mejores resultados [Ekenel 05], en cualquier caso la elección depende del método de extracción de rasgos utilizado.

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, en un método específico basado en la apariencia local. En la Tabla 1.1 se resumen los resultados reportados en la literatura por los algoritmos de reconocimiento de rostros más significativos en los que se han utilizado métodos basados en la apariencia local.

Tabla 1.1: Resultados experimentales obtenidos por los algoritmos de reconocimiento de rostros en los que se utilizan métodos basados en la apariencia local.

Método	Base de Datos de Rostros	Cantidad de Imágenes Probadas	Tasa de Reconocimiento	Principales Variaciones en las Imágenes
Método probabilístico local [Martínez 02]	AR	600	76.5 %	expresión, oclusión
SOM [Tan 05]	AR	600	84.85 %	expresión, oclusión
HMM [Le 04]	AR	1140	89.8 %	expresión, iluminación
FLDA Modular [Chen 04a]	FERET	200	86.5 %	expresión
Componentes LDA [Chen 04a]	FERET	350	71.2 %	expresión, iluminación
EBGM [Wiskott 97]	FERET	2345	71.0 %	expresión, iluminación
LBP [Ahonen 04]	FERET	2345	85.0 %	expresión, iluminación
DCT [Ekenel 05]	Yale	150	98.9 %	expresión, iluminación
PCA local [Gottumukkal 04]	Yale	150	72.2 %	expresión, iluminación
GMM [Sanderson 03]	Yale	150	58.9 %	expresión, iluminación

En la Tabla 1.1 puede observarse que no existe uniformidad en las bases de datos utilizadas para realizar las evaluaciones de los diferentes algoritmos. Incluso en una misma base de datos no siempre se utilizan los mismos subconjuntos de imágenes para realizar las evaluaciones. Esto se debe fundamentalmente a que no todas las bases de datos son públicas, la FERET [Phillips 00] por ejemplo, es propiedad del gobierno norteamericano y sólo se vende a determinadas universidades y grupos de investigación. Por otra parte, dentro de las bases de datos públicas, los investigadores seleccionan aquellas con imágenes más representativas de los problemas que enfrentan. A pesar de esto, de la tabla se pueden extraer conclusiones importantes sobre los métodos basados en la apariencia más significativos, que son abordadas de manera más amplia en el reporte de investigación presentado por la autora [Méndez-Vázquez 08a], donde se detallan cada uno de los métodos que aparecen en la tabla y se estudia particularmente cómo afectan las variaciones de la iluminación a cada uno de ellos.

Se puede apreciar que los algoritmos basados en el método probabilístico local [Martínez 02] y en los Mapas Auto-Organizados (SOM) [Tan 05] fueron diseñados para enfrentar el problema del reconocimiento de rostros cuando hay oclusión y variaciones en la expresión. Los basados en componentes LDA [Huang 03] y en el Fisher LDA (FLDA) modular [Chen 04a] están muy relacionados con los dos anteriores y presentan buenos resultados ante imágenes con variaciones en la expresión, pero disminuyen su eficacia cuando se pretenden reconocer imágenes con variaciones en las condiciones de iluminación. El algoritmo que utiliza los Modelos Ocultos de Markov (HMM) [Le 04] aparentemente se encuentra dentro de los de mejores resultados, sin embargo las variaciones de iluminación en las imágenes de la base de datos AR [Martínez 98] no representan un gran desafío, además de que son justamente esas imágenes las que provocan los errores que se cometen en la clasificación al utilizar este método. Por otra parte, la base de datos de Yale [Belhumeur 97], a pesar de tener pocas imágenes, las principales variaciones de estas son precisamente de iluminación. En estas imágenes se ve una superioridad significativa del algoritmo basado en los coeficientes de la Transformada Discreta del Coseno (DCT) [Ekenel 05] sobre los algoritmos que utilizan el PCA local [Gottumukkal 04] y los Modelos de Mezclas Gaussianas (GMM) [Sanderson 03, Zhang 08]. Finalmente el algoritmo conocido como *Elastic Bunch Graph Matching (EBGM)* [Wiskott 97] y el basado en el operador de Patrones Binarios Locales (LBP) [Ahonen 04] son los que con más imágenes aparecen evaluados [Phillips 00]. De estos dos, el basado en los LBP exhibe una mayor eficacia. El EBGM es costoso computacionalmente y presenta como mayor debilidad la necesidad de localizar certeramente los puntos característicos del rostro [Tan 06].

Teniendo en cuenta las consideraciones anteriores, dentro de los algoritmos que utilizan métodos basados en la apariencia local, se seleccionan los basados en el operador LBP y los coeficientes DCT como los más indicados para el reconocimiento de rostros cuando existen variaciones en las condiciones de iluminación.

1.2.1 Patrones Binarios Locales (LBP)

El operador de Patrones Binarios Locales (LBP) es un descriptor de textura que ha sido ampliamente utilizado en diversas aplicaciones de procesamiento digital de imágenes [Ojala 96]. Comparado con otros descriptores de textura, este operador es altamente discriminador y tiene como ventajas principales la invariancia a los cambios de niveles de grises monotónicos que se pueden presentar en las imágenes y una alta eficiencia computacional [Ojala 96].

El operador LBP básico, asigna una etiqueta a cada píxel de una imagen usando, en una vecindad de 3×3 píxeles, el valor del píxel central como umbral y considerando el resultado como un número binario. La forma decimal del código de 8 *bits* resultante se puede expresar como:

$$LBP(x_c, y_c) = \sum_{n=0}^7 s(i_n - i_c) 2^n \quad (1.1)$$

donde i_c corresponde al valor del píxel central (x_c, y_c) , i_n son los valores de los 8 píxeles que rodean al píxel central y la función $s(x)$ se define como:

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad (1.2)$$

Un ejemplo de cómo se obtiene la etiqueta LBP de un píxel se ilustra en la Figura 1.2. Luego de tener las etiquetas LBP de todos los píxeles, se construye un histograma con esos valores, que es utilizado como descriptor de la textura de la imagen.

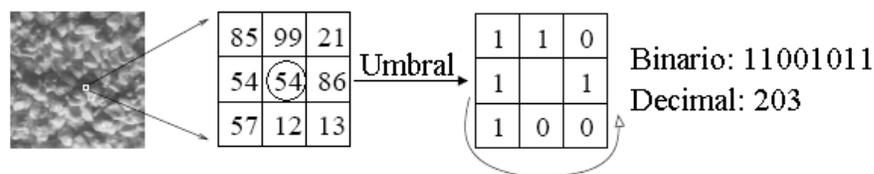


Figura 1.2: El operador LBP básico.

El operador básico fue posteriormente extendido para usar vecindades de diferentes tamaños [Ojala 02]. Al definir una vecindad local como un conjunto de puntos de ejemplo, uniformemente espaciados en un círculo centrado en el píxel que será etiquetado, se puede usar cualquier radio y cualquier número de puntos de ejemplo. Cuando un punto de ejemplo no corresponde con el centro de un píxel, mediante interpolación bilineal se determina su valor. Se utiliza la notación (P, R) para denotar las vecindades de píxeles, lo cual significa P puntos de ejemplo en un círculo de radio R .

Otra extensión del operador original, es la definición de los llamados patrones uniformes [Ojala 02]. Un patrón binario es considerado uniforme si contiene como máximo dos transiciones de bits de 0 a 1 o viceversa, incluyendo la diferencia entre el último y el primer valor del patrón, como si fuera circular. Por ejemplo, los patrones 00000000 (0 transiciones), 11110000 (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 componente (o barra) para cada patrón uniforme, mientras que todos los patrones no uniformes son asignados a un mismo componente del histograma.

Posteriormente han aparecido nuevas extensiones del LBP. Por ejemplo, en [Jin 04] 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.

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 [Ahonen 04]. En cualquiera de sus variantes, los histogramas LBP contienen información acerca de la distribución de los micro-patrones locales como bordes, puntos, áreas sobresalientes y áreas lisas. Para una representación eficiente del rostro se retiene también la información espacial. Con este propósito la imagen de rostro se divide en regiones R_0, R_1, \dots, R_{m-1} y los histogramas de etiquetas LBP se obtienen en cada una de estas regiones de manera independiente. Finalmente son concatenados para formar el descriptor global de la imagen obteniéndose un histograma mejorado espacialmente [Ahonen 04].

En el histograma mejorado espacialmente se tiene de manera efectiva la descripción del rostro en tres niveles diferentes de localización: las etiquetas LBP contienen información acerca de los patrones a niveles de píxeles, estas se suman para formar los histogramas en

una pequeña región lo cual brinda información a nivel regional y luego, los histogramas son concatenados para obtener una información global del rostro.

Los histogramas de cada imagen pueden ser comparados con un clasificador del vecino más cercano y alguna de las medidas de comparación de histogramas propuestas en la literatura, dentro de las que se encuentran: la intersección de histogramas, la estadística de probabilidad logarítmica y estadística Chi al cuadrado (χ^2). Las mejores tasas de reconocimiento reportadas para este operador han sido utilizando Chi al cuadrado [Ahonen 04].

Desde que se introdujo la aplicación de este operador a imágenes de rostros, muchos han sido los estudios realizados y las mejoras propuestas [Marcel 07]. Entre las diferentes extensiones se destacan el *Boosting-LBP* [Zhang 04], que utiliza el método de aprendizaje *Ada-Boost* [Freund 95] para seleccionar las ventanas y los rasgos a utilizar y el LBP Multi-resolución (MLBP) [Raja 06] en el que se tienen en cuenta los rasgos extraídos utilizando diferentes vecindades.

Una de las propuestas más recientes utilizando el LBP, es el uso de la representación MLBP con la clasificación basada en LDA [Chan 07]. Primeramente, el descriptor LBP de cada píxel de la imagen de rostro es calculado para diferentes radios R , obteniéndose R imágenes de etiquetas LBP. Todas las imágenes obtenidas tienen las mismas dimensiones por lo que son particionadas en las mismas regiones rectangulares no solapadas. Se calculan los histogramas para cada región en cada imagen de etiquetas LBP y los histogramas correspondientes a una misma región a diferentes escalas, son concatenados en un solo vector, proveyendo un descriptor regional del rostro de múltiple resolución. Cada descriptor local se proyecta en un subespacio LDA, en el que la correlación normalizada es utilizada como medida de similitud para comparar las proyecciones correspondientes a la misma región de diferentes imágenes. Finalmente, las similitudes de las diferentes regiones son sumadas para obtener una medida global de la similitud entre dos imágenes.

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.

En realidad el operador solamente es invariante a las transformaciones monotónicas en la iluminación incidente sobre el rostro. Una transformación monotónica, T , es una transformación que preserva las desigualdades de sus argumentos. Esto significa que $T(x) > T(y)$ si y solo si $x > y$ y $T(x) < T(y)$ si y solo si $x < y$, $\forall x, y$. En el caso de la iluminación que incide sobre un rostro, se refiere entonces a variaciones en la intensidad de la fuente de luz, que provocan que los valores de los píxeles de la imagen sean mayores o menores

en su totalidad, pero preservando las desigualdades entre ellos.

La etiqueta LBP de cada píxel en una imagen se basa en la comparación entre los valores de intensidad de los píxeles vecinos, sin importar en qué magnitud sean mayores o menores, por tanto, si se preservan las desigualdades entre los píxeles de la imagen, el resultado del operador es exactamente el mismo. Este efecto puede apreciarse en la Figura 1.3 (a), donde se muestran las etiquetas LBP obtenidas para tres imágenes de un mismo individuo con tres condiciones de iluminación diferentes, pero causadas solamente por variaciones monotónicas. Sin embargo, otras variaciones en la iluminación, como por ejemplo en el ángulo de incidencia de la luz, que en el caso de los rostros son las más comunes, afectan de manera significativa la descripción mediante el operador LBP, como puede apreciarse en la Figura 1.3 (b).

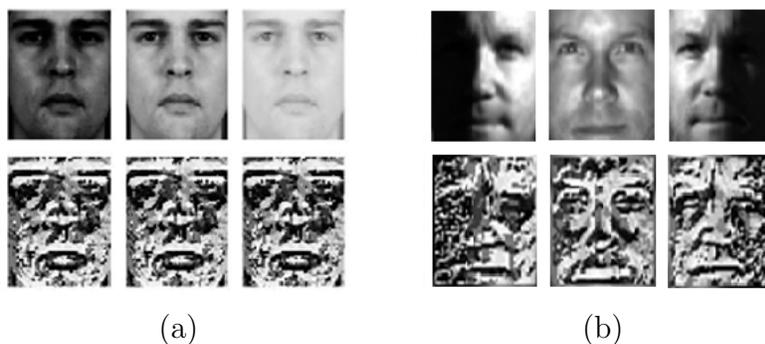


Figura 1.3: Efecto de las variaciones de iluminación (a) monotónicas y (b) no monotónicas en el operador LBP.

1.2.2 Transformada Discreta del Coseno (DCT)

La Transformada Discreta del Coseno (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 [Ahmed 74] y enseguida comenzó a crecer en popularidad proponiéndose diversas variantes. Principalmente han sido clasificadas cuatro transformaciones ligeramente diferentes: DCT I, DCT II, DCT III y DCT IV. De ellas, La DCT II ha sido la más utilizada en el procesamiento de imágenes, especialmente en los estándares de compresión, debido a su poder compacto de representación [Rao 90]. A pesar de que la Transformada Karhunen-Loeve (KLT) es conocida como la óptima en términos de compactación de la información, su naturaleza dependiente de los datos limita su uso práctico en algunas aplicaciones. La DCT

aproxima en gran medida la habilidad de representación compacta de la KLT, lo que la convierte en una herramienta muy útil no sólo en términos de compactación de la información, sino también en términos de complejidad computacional, debido a su naturaleza independiente de los datos [DCT 01].

La DCT de una imagen de $M \times N$ se define como:

$$C(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x, y) \times \cos \left[\frac{\pi(2x+1)u}{2M} \right] \cos \left[\frac{\pi(2y+1)v}{2N} \right] \quad (1.3)$$

donde,

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{M}}, & u = 0 \\ \sqrt{\frac{2}{M}}, & u = 1, \dots, M-1, \end{cases} \quad (1.4)$$

y

$$\alpha(v) = \begin{cases} \frac{1}{\sqrt{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & v = 1, \dots, N-1 \end{cases} \quad (1.5)$$

La idea del uso de la DCT para el reconocimiento de rostros, es calcular la DCT de una imagen y retener un subconjunto de los coeficientes, con los que se conforma el vector de características que describe el rostro [Hafed 01]. Este vector contiene los coeficientes DCT con las frecuencias de baja a media, que son los que tienen la mayor varianza, extraídos mediante un recorrido en zig-zag, como se representa en la Figura 1.4. El coeficiente DCT del borde superior izquierdo es eliminado, ya que sólo representa el valor promedio de las intensidades en cada punto.

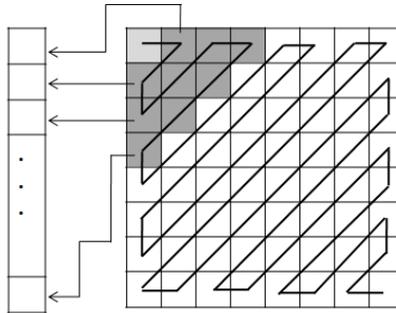


Figura 1.4: Representación de la extracción de rasgos DCT utilizando recorrido en zig-zag.

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 combinando luego los resultados de la clasificación individual (fusión por decisión) [Ekenel 05].

Los coeficientes DCT han sido utilizados con varios clasificadores en el reconocimiento de rostros: redes neuronales, HMM, GMM y LDA, entre otros [Nefian 99, Eickeler 00, Er 05, Scott II 03, Lee 05a]. Sin embargo, con las características de los sistemas de reconocimiento de rostros, principalmente la falta de imágenes de entrenamiento, el clasificador del vecino más cercano combinado con las distancias de la norma L_1 , la norma L_2 y la correlación normalizada, ofrece resultados superiores a los de otras variantes más complejas [Ekenel 09]. Los resultados experimentales muestran que los métodos basados en la DCT local son superiores a los que la utilizan en la imagen global [Ekenel 05], siendo también mejores que otros métodos locales ante imágenes con diferentes variaciones, como se confirma en la Tabla 1.1.

El principal problema de este método radica en cuáles y cuántos coeficientes escoger para representar de manera efectiva los rasgos discriminatorios de un rostro y a la vez lograr robustez ante las distintas variaciones que pueden afectar la imagen. Como se mencionó, normalmente se utilizan los coeficientes DCT asociados a las bajas frecuencias para conformar los vectores de rasgos, estos concentran la mayor información de la imagen y han permitido obtener resultados satisfactorios. Sin embargo, es también en estos coeficientes donde se reflejan las variaciones de iluminación sobre el rostro. En la Figura 1.5 se muestran dos imágenes de una misma persona y sus respectivas DCT. Las imágenes, extraídas de la base de datos Yale B [Lee 05b], fueron capturadas con solo milisegundos de diferencia, siendo las condiciones de iluminación la única variación entre ellas. Como puede observarse en la figura, las variaciones en las condiciones de iluminación inciden directamente sobre los coeficientes DCT de bajas frecuencias, ubicados en la esquina superior izquierda de las imágenes que representan la DCT. A simple vista, las DCT de ambas imágenes parecen muy similares, pero cuando se hace un acercamiento se puede notar con más claridad que los coeficientes DCT de bajas frecuencias contienen información diferente y por tanto, al utilizarlos como representación de las imágenes de ese rostro, no permiten identificar que corresponden a la misma persona. Cuando se utilizan los coefi-

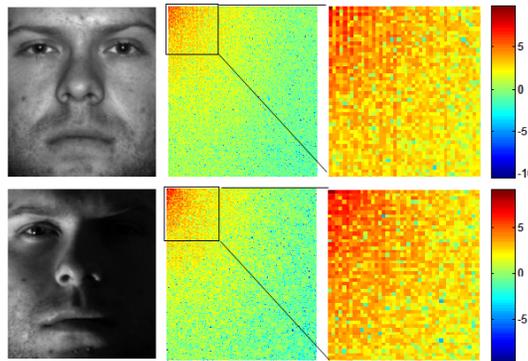


Figura 1.5: Influencia de las variaciones de iluminación en los coeficientes DCT de bajas frecuencias de dos imágenes de un mismo rostro.

cientes DCT de bajas frecuencias obtenidos dividiendo la imagen en bloques, se aminora de cierta forma el problema, puesto que puede haber regiones en las que las diferencias no sean apreciables, pero en aquellos bloques en los que las condiciones de iluminación sean muy diferentes, permanecerá siendo difícil decidir si corresponden o no al mismo individuo.

En [Saban 04] se presenta una modificación del método DCT local, el DCT-mod 2, para lograr mayor robustez a las variaciones de iluminación. Se asume que los tres primeros coeficientes DCT, obtenidos con el recorrido en zig-zag, son los más afectados por las variaciones de iluminación y estos se reemplazan por sus *deltas* horizontales y verticales. Los *deltas* corresponden a la diferencia entre los bloques de coeficientes DCT horizontales y verticales y son interpretados como una representación de la información espacial transitoria. Utilizando esta modificación, se logra aumentar, aunque en no más de 0,50 %, la eficacia del método al reconocer imágenes de rostros bajo diferentes condiciones de iluminación. Sin embargo, el error total en la clasificación permanece alto, lo que demuestra que no son estos primeros tres coeficientes los únicos afectados por las variaciones de iluminación. Otras variantes de modificaciones al método se presentan en ese trabajo [Saban 04], pero ninguna de ellas logra aumentar de manera significativa la eficacia.

1.3 El problema de la iluminación

Las variaciones en la iluminación han sido identificadas como unas de las principales afectaciones en la eficacia de los algoritmos de reconocimiento de rostros [Adini 97, Zhao 03b, Phillips 07]. Se ha visto que condiciones de iluminación diferentes pueden producir imáge-

nes muy distintas del mismo sujeto y que estas diferencias pueden llegar a ser más significativas que las debidas a diversas identidades personales [Adini 97]. En otras palabras, la diferencia entre dos imágenes del rostro del mismo individuo tomadas bajo condiciones de iluminación diferentes, puede ser mayor que la diferencia entre dos imágenes de rostros diferentes tomadas bajo las mismas condiciones de iluminación. Este efecto puede observarse en la Figura 1.6, donde el valor del coeficiente de correlación entre las matrices que representan las imágenes en (a) es de 0,0132, mientras que la correlación entre las imágenes de (b) es 0,7601, mucho mayor.



Figura 1.6: Ejemplo de variaciones de iluminación: (a) Mismo sujeto, condiciones de iluminación diferentes; (b) Sujetos diferentes, iguales condiciones de iluminación.

Desde un punto de vista teórico, asumiendo reflectancia Lambertiana, la intensidad de la luz reflejada en una superficie, capturada por una cámara, es el producto de su reflectancia, R y la iluminación en la escena o luminancia, L [Finlayson 01, Land 71]. De manera que las intensidades de los píxeles en una imagen se pueden representar como:

$$I(x, y) = R(x, y)L(x, y) \quad (1.6)$$

La reflectancia, que representa la forma y textura de la superficie, es única en cada rostro y es lo que permite, de manera ideal, discriminar entre un individuo u otro. Si la intensidad observada en cada píxel y el valor de luminancia son conocidos, la reflectancia puede ser recuperada. Sin embargo, la iluminación que incide en un punto varía según la geometría de la escena, el ángulo de incidencia e intensidad del origen de la luz y el ángulo de observación. Un conocimiento previo de estos factores solamente es posible en un ambiente de laboratorio muy controlado. No es posible contar con esta información en ambientes reales, por lo cual se vuelve difícil la tarea de identificar automáticamente un individuo partiendo solamente de la imagen de su rostro. Se necesitan métodos que estimen la luminancia o compensen de alguna manera las variaciones en las condiciones de iluminación, de manera que se puedan obtener los rasgos discriminatorias relacionados con la reflectancia.

Varios enfoques han sido propuestos para enfrentar las afectaciones que provocan las

variaciones en las condiciones de iluminación a la eficacia de los sistemas de reconocimiento de rostros [Zou 07]. La mayoría de las investigaciones se concentran en el trabajo con imágenes bidimensionales (2D) capturadas normalmente con cámaras fotográficas o de video, pero existen otras técnicas novedosas basadas en sistemas de captura infrarrojos [Socolinsky 01, Li 07, Méndez-Vázquez 09] y de imágenes tridimensionales (3D) [Akarun 05, Gokberk 09], que evitan las afectaciones por iluminación desde el momento de la captura. Las cámaras infrarrojas y 3D tienen elevados costos, por lo que no es fácil su utilización en el despliegue de sistemas reales.

Los métodos que enfrentan el problema de la iluminación a partir de imágenes 2D, pueden ser divididos por su concepción en cinco grupos principales: los de pre-procesamiento, los de extracción de rasgos invariantes, los de modelación de la imagen de rostro, los de aprendizaje de las variaciones de iluminación y los de pos-procesamiento [Zou 07].

Los métodos de pre-procesamiento normalizan la imagen de entrada, con el objetivo de obtener una nueva imagen o representación del rostro estable bajo diferentes condiciones de iluminación [Du 05]. El segundo grupo persigue extraer rasgos de la imagen del rostro invariantes a la iluminación [Chen 00]. Los métodos que pertenecen al tercer grupo, se basan en construir simultáneamente un modelo 3D del rostro y un modelo de la iluminación, que generen la imagen observada [Sim 01]. Una vez que el modelo 3D del rostro es estimado, las condiciones de iluminación son normalizadas re-iluminándolo con una fuente de luz canónica. La cuarta alternativa, implica coleccionar una gran cantidad de imágenes de rostros, tomadas en un amplio rango de condiciones de iluminación, para con ellas entrenar modelos generativos o discriminatorios con la capacidad de representar los rostros bajo todas las posibles condiciones de iluminación [Georghiades 01]. Por último, el efecto de la iluminación puede ser minimizado con métodos de pos-procesamiento [Arandjelovic 06] que normalicen el impacto de los cambios de iluminación sobre el valor de similitud obtenido entre una imagen de consulta y la imagen objetivo de una colección, tomando como referencia los valores de similitud con las imágenes no objetivo.

Cada uno de los cinco grupos de métodos mencionados tiene sus propios méritos, pero puede no ser apropiado para algunos escenarios. Por ejemplo, en aplicaciones reales suele ser muy difícil contar con una colección de imágenes suficientemente representativa de todas las posibles condiciones de iluminación, requeridas en los métodos de aprendizaje [Belhumeur 98, Georghiades 01]. Los métodos que se basan en construir un modelo 3D, por lo general son costosos computacionalmente y requieren en la mayoría de los casos, la localización precisa de los puntos característicos del rostro [Zhao 00, Sim 01], lo cual, como se explicó en el epígrafe 1.1, representa un problema en la práctica. Las técnicas

de pos-procesamiento pueden ser inconvenientes para aplicaciones de verificación en las que normalmente no existe una colección de imágenes [Arandjelovic 06]. En el caso de la extracción de rasgos, existen dudas sobre la posibilidad de desarrollar un método totalmente invariante a la iluminación, debido a la demostración teórica de Moses y Ullman [Moses 92] sobre la inexistencia de una función invariante para ningún objeto, incluso bajo el modelo Lambertiano. En el caso de los métodos de pre-procesamiento, son por lo general muy eficientes computacionalmente y han sido los más extendidos en aplicaciones reales [Du 05, Zou 07], logrando cada vez mejores resultados en el tratamiento de imágenes afectadas por la iluminación.

1.3.1 Métodos de pre-procesamiento

El pre-procesamiento, se inserta como un paso dentro de la secuencia que siguen los algoritmos de reconocimiento de rostros. Una vez que el rostro se detecta y se normaliza geoméricamente, se aplica el pre-procesamiento y luego, la extracción de rasgos y la clasificación.

Los métodos utilizados en el pre-procesamiento reciben una imagen de rostro con el objetivo de normalizar el efecto provocado por las variaciones en las condiciones de iluminación. De ahí que los primeros métodos de pre-procesamiento propuestos para el reconocimiento de rostros incluyen herramientas tradicionales del procesamiento de imágenes que modifican la distribución de los valores de las intensidades de los píxeles en las mismas [Jain 89].

Uno de los más simples es la corrección de ganancia/desbalance, que persigue corregir los efectos de la baja iluminación en una escena, ampliando el rango dinámico de la imagen, de forma tal que se ajuste al rango dinámico del medio de visualización. Si el rango dinámico capturado de una escena es $[r_{max}; r_{min}]$, la operación lineal que se le aplica a cada píxel para recuperar el rango dinámico completo del medio de visualización, con un valor máximo d_{max} , es simplemente:

$$I'(x, y) = \frac{d_{max}}{r_{max} - r_{min}} \cdot (I(x, y) - r_{min}) \quad (1.7)$$

Un método tradicional de procesamiento de imágenes utilizado para rostros, es la ecualización de histogramas (HE) [Petrou 99]. El histograma de una imagen es una distribución que representa el número de píxeles por cada valor de intensidad presente en la imagen. Las imágenes en la que todos los píxeles tienen un valor de intensidad similar muestran

menos detalles que aquellas en las que los valores de las intensidades de los píxeles estén más esparcidos.

La ecualización de histogramas modifica el histograma de una imagen, con el objetivo de obtener una nueva distribución uniforme de las intensidades de los píxeles, en la que la frecuencia de cada valor de intensidad sea igual a $\frac{N}{G}$, donde N es el número total de píxeles en la imagen y G los valores de intensidad. El valor k de cada píxel es sustituido por un nuevo valor de intensidad obtenido a través de la función T , que se basa en integrar y normalizar la distribución de valores de píxeles en la imagen original, p_k , según la ecuación:

$$T(k) = \frac{1}{c} \int_0^G p_k(x) dx \quad (1.8)$$

Para imágenes digitales, este mapeo no produce una distribución perfectamente uniforme, pero mejora el esparcimiento de los valores de intensidad de los píxeles en la imagen y se ha reportado que con este método, se superan los resultados obtenidos utilizando las imágenes de rostros afectadas por la iluminación con sus valores de intensidades originales [Kittler 99].

Una extensión del método anterior es el conocido como especificación o correspondencia de histogramas [Gonzalez 06]. El objetivo en este caso es que la imagen resultante tenga un histograma que se corresponda con el de una imagen con buena iluminación. El problema de este método radica en que para especificar un histograma correspondiente a una iluminación correcta, es necesario contar con una imagen de referencia, lo cual no es siempre posible en las aplicaciones de reconocimiento de rostros.

Existen también los métodos basados en transformaciones no lineales de las imágenes que proveen una compresión del rango dinámico [Gonzalez 06]. Estos consisten en aplicar una función no-lineal a la imagen, como el logaritmo o la potencia, de manera que las regiones oscuras de la escena se realzan. Un efecto no deseado de este proceso es que el contraste se comprime en las regiones de brillo de la imagen, lo que puede provocar pérdida de información de los bordes.

Un método popular es el filtrado homomórfico (HF) [Stockham 72]. Éste opera en el dominio de la frecuencia buscando eficiencia computacional. La imagen es transformada logarítmicamente y luego se le aplica la transformada de Fourier F . En el dominio de la frecuencia la imagen se convoluciona con un filtro de paso alto, H_p , que elimina las bajas frecuencias y retiene las altas, asumiendo que la iluminación es principalmente un fenómeno de bajas frecuencias [Short 04]. La imagen final es obtenida transformando la

imagen filtrada con la inversa de Fourier F^{-1} y aplicando luego el antilogaritmo.

En [Jobson 97] se sugiere el uso de la teoría Retinex [Land 77] para estimar la reflectancia R de una imagen representada según la Ecuación 1.6. El método *retinex* simple (SR) estima la función de iluminación L , suavizando la imagen al convolucionarla con una Gaussiana. La imagen se divide por este estimado para obtener la reflectancia. El algoritmo *retinex* multi-escala (MSR) mejora la forma de estimar la iluminación, al utilizar una suma de muchas imágenes suavizadas con diferentes tamaños de Gaussianas. En [Rahman 97] este último se compara con HE y HF, obteniéndose con él los mejores resultados.

El concepto de suavizar la imagen para estimar la función de iluminación fue continuado por Gross y Brajovic, en el método conocido como suavizado anisotrópico (AS) [Gross 03]. El suavizado anisotrópico varía como función de las medidas de contraste local de la imagen original, de manera que la intensidad del suavizado se reduce en áreas de alto contraste local y se amplifica en áreas de bajo contraste local. Este método tiene en cuenta que la iluminación incide de manera diferente en las distintas regiones de la imagen del rostro, luego el suavizado se reduce en las regiones alrededor de los labios, la nariz y los ojos, donde el contraste es mayor. El método de Gross y Brajovic es uno de los métodos de pre-procesamiento que mejores resultados reporta en la literatura al enfrentar los problemas de iluminación en las imágenes de rostros [Short 06].

El suavizado anisotrópico fue también empleado en [Liu 05] como parte de un proceso iterativo en el que la cantidad de suavizado utilizado es modelado por la distancia de cada píxel al borde más cercano. Se normaliza la imagen dividiéndola por la imagen suavizada y multiplicándola por una imagen suavizada representativa de un rostro bien iluminado. El mapa de bordes es nuevamente determinado en la imagen corregida y utilizado para crear una mejor estimación de la iluminación. El proceso continúa hasta que el cambio en el contenido de la imagen cae debajo de un umbral determinado.

El operador LBP, detallado en el epígrafe 1.2, ha sido utilizado en el pre-procesamiento, como paso previo a la extracción de rasgos [Heusch 06]. Los resultados obtenidos con este método de pre-procesamiento fueron similares a los del AS, pero con un costo computacional menor. No obstante, como se explicó, éste método solamente es invariante a las variaciones monotónicas de la iluminación, en los otros casos no logra suprimir de manera adecuada las afectaciones provocadas por la iluminación.

El método de la imagen cociente (QI), hace uso de la afirmación planteada en [Shashua 97], de que una imagen \hat{I} de un objeto, puede ser representada como una suma lineal de imágenes $I_1 \dots I_j$ de ese objeto, iluminadas por al menos tres fuentes de luz linealmente

independientes:

$$\hat{I} = \sum_j w_j I_j \quad (1.9)$$

donde w_j es el peso correspondiente a la imagen I_j .

La imagen cociente $Q_p(x, y)$ de una imagen $P(x, y)$, se define como:

$$Q_p(x, y) = \frac{R_p(x, y)}{R_b(x, y)} \quad (1.10)$$

donde x y y son las coordenadas del píxel, R_p es la reflectancia de la imagen que se analiza y R_b es la función de reflectancia que representa el promedio de un conjunto de imágenes que son capaces de generar todas las posibles condiciones de iluminación, denominado *bootstrap*.

Estimar la función de reflectancia de la imagen que se analiza es una tarea complicada. El conjunto *bootstrap* de imágenes contiene tres imágenes bajo condiciones de iluminación predefinidas de un grupo de personas. Este se utiliza para determinar los pesos w_1, w_2, w_3 de la Ecuación 1.9 a través de una solución de mínimos cuadrados. Cuando estos pesos son hallados, el conjunto *bootstrap* se utiliza para generar una imagen promedio bajo las mismas condiciones de iluminación que la imagen que se analiza, utilizando la Ecuación 1.9. De esta forma la imagen del cociente se calcula como:

$$Q_p(x, y) = \frac{P(x, y)}{\sum_j w_j \bar{I}_j(x, y)} \quad (1.11)$$

donde \bar{I}_j es el promedio del conjunto *bootstrap* para la condición de iluminación j y w_j es el peso correspondiente.

El inconveniente de este método es que necesita un conjunto de entrenamiento con al menos tres imágenes de cada persona bajo las mismas tres condiciones de iluminación diferentes. Como solución a este problema se propone el método de la imagen del cociente propia (SQI) [Wang 04a], que en lugar de utilizar un conjunto de entrenamiento, el denominador se deriva directamente de la imagen de entrada, a través de la ecuación:

$$Q_p(x, y) = \frac{P(x, y)}{S * P(x, y)} \quad (1.12)$$

donde S es un filtro de suavizado, que no es más que el producto de un *kernel* gaussiano con una ventana del mismo tamaño, en la que cada entrada se pone en cero si el valor de intensidad del píxel correspondiente es menor que el promedio local y en uno, si no.

El problema en esta variante radica en cómo escoger el tamaño del filtro que se utiliza. Si la escala es muy pequeña, la imagen filtrada será muy parecida a la imagen original y por tanto Q_p será cercana a uno, perdiéndose información. Si por el contrario el filtro es muy grande, la imagen filtrada del denominador será más o menos constante, por lo que el resultado es, la normalización de la imagen por su valor promedio. Por esta razón, los autores del método proponen utilizar diferentes escalas. La única diferencia entonces entre este método y el MSR radica en el uso de un *kernel* con pesos, en lugar de una gaussiana pura, que lo hace superior. Muestra también mejores resultados que el QI en imágenes afectadas por la iluminación [Wang 04a, Du 05].

Una de las extensiones más reciente y exitosa del método QI es la basada en el modelo de variación total (TVQI) [Chen 05]. En el modelo de variación total (TV), una imagen f se representa como una suma de la “caricatura” de la imagen u , que contiene los tonos y bordes principales de la imagen y la textura v , que caracteriza los patrones de pequeña escala, como los ojos y la boca en el caso del rostro. Utilizando la descomposición de señales basada en TV con la norma L^1 [Chan 04], se estima u como la representación de la iluminación, por la cual se divide a f para obtener la imagen normalizada.

Uno de los métodos más recientes que ha mostrado ser muy eficiente computacionalmente y superior a los métodos existentes, es la secuencia de pre-procesamiento (PS) presentada por Tan y Triggs [Tan 07]. El método consiste en una serie de pasos, que persiguen reducir los efectos de las sombras locales y las manchas de luz, mientras se mantiene la esencia de la información de la apariencia visual usada en el reconocimiento. El primer paso es aplicar una corrección gamma, que se utiliza como una aproximación a una transformación logarítmica, que transforma los valores de los píxeles en I con I^γ , donde $\gamma \in [0, 1]$. El segundo paso implica un filtrado de pasa-banda con el método de diferencias de Gaussianas (DoG) [Gonzalez 06]. Este filtro de pasa-banda no solo suprime la información de baja frecuencia causada por el gradiente de la iluminación, sino que también reduce el ruido en las altas frecuencias. El último paso es una ecualización de contraste global, donde los mayores valores son truncados para reducir su influencia. Luego de este paso la imagen puede contener todavía valores extremos, así que opcionalmente se utiliza una función no lineal que comprime los valores muy grandes. Este método es de cierta forma similar al HF. Ambos plantean como primer paso, una transformación logarítmica, que realiza el rango dinámico en las regiones oscuras y lo comprime en las brillantes, convirtiendo además, la Ecuación 1.6 en una suma, $\log I(x, y) = \log R(x, y) + \log L(x, y)$, para suprimir L de manera más fácil. Luego de aplicar el logaritmo, en los dos métodos se filtran determinadas bandas de frecuencia, siendo éste su principal problema, ya que es necesario decidir

qué información de frecuencia filtrar para lograr eliminar las afectaciones de iluminación y al mismo tiempo, no comprometer los rasgos discriminatorios de las imágenes.

Un método, similar a estos en su estrategia, en el sentido que se plantea compensar las variaciones de iluminación sustrayendo un término de compensación a la imagen en el dominio logarítmico es el presentado en [Chen 06]. En éste, se suprimen los coeficientes DCT de bajas frecuencias como una aproximación del término de compensación, asumiendo, como en otros trabajos, que las variaciones de iluminación caen en las bajas frecuencias. Los resultados alcanzados por este método fueron incluso superiores a los de métodos de modelación y de aprendizaje [Chen 06], sin embargo, en él se aplica la DCT a la imagen completa, sin tener en cuenta que las afectaciones de iluminación tienen un comportamiento local [Short 06]. De hecho, en [Villegas 05] se comparan varios métodos de pre-procesamiento aplicados de manera global y de manera local, mostrándose que en todos los casos, las variantes locales ofrecen los mejores resultados.

La mayoría de los métodos descritos han sido comparados entre ellos [Short 04, Du 05, Arandjelovic 06, Tan 07]. En la Figura 1.7 se muestra el por ciento de reconocimiento correcto al aplicar algunos de estos métodos a imágenes de rostros en condiciones normales y en presencia de variaciones de iluminación y de otras como la distancia a la que se encuentra la persona de la cámara, la expresión y la presencia de accesorios. Se muestra también en cada caso, el por ciento de reconocimiento correcto cuando no se aplica pre-procesamiento alguno (OI). La principal conclusión a la que se puede llegar al analizar la figura, es que si bien la aplicación de métodos de pre-procesamiento permite obtener

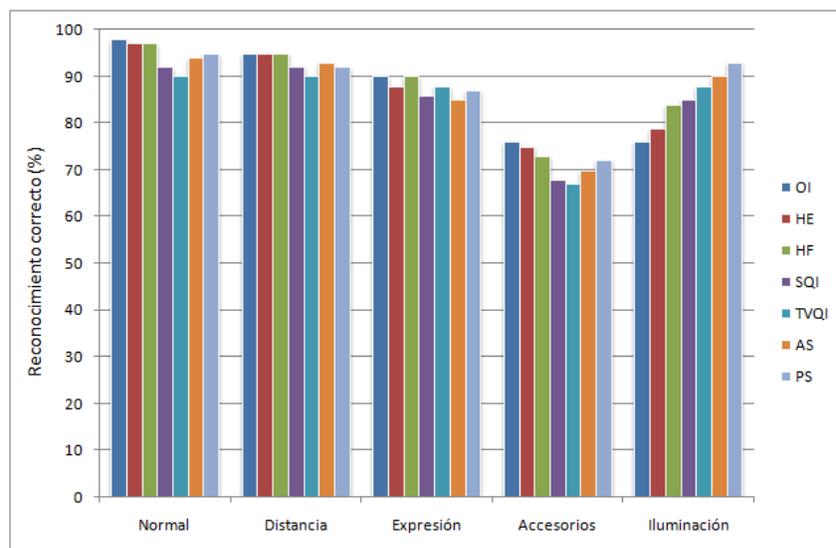


Figura 1.7: Comparación de métodos de pre-procesamiento.

mejores resultados que el uso de las imágenes originales cuando existen afectaciones provocadas por los cambios en las condiciones de iluminación, la eficacia disminuye respecto al uso de las imágenes originales cuando estas están iluminadas normalmente o existen otras afectaciones en la imagen que no están provocadas por la iluminación.

Varios métodos han surgido como solución a este problema. Arandjelovic y Cipolla [Arandjelovic 06] proponen aprender implícitamente cuán similares son las condiciones de iluminación de la imagen de un individuo que llega con el objetivo de ser reconocida, con las condiciones de la imagen que se tiene de ese individuo en la colección, para en base a eso enfatizar el valor de la comparación entre las imágenes originales o el de sus versiones pre-procesadas. Esto se logra calculando la similitud entre las dos imágenes originales y entre las imágenes filtradas y asignándole un peso a cada valor de similitud, que se obtiene a partir de la diferencia entre los dos valores de similitud (de imagen original y de imagen pre-procesada) más altos obtenidos al comparar con todas las imágenes en la colección. Se utiliza finalmente para el reconocimiento el valor de la suma pesada de ambas similitudes. Este proceso es en realidad una combinación de pre-procesamiento con pos-procesamiento y tiene la dificultad de que no puede ser aplicada en sistemas de verificación, en los que no se cuenta con una colección de imágenes.

Otra variante de solución ha sido detectar en las imágenes las zonas afectadas y analizarlas de manera independiente a las no afectadas. Kryszczuk y Drygajlo [Kryszczuk 05] sugieren un método para segmentar las imágenes de rostros que están muy degradadas por la presencia de sombras, separando las áreas de sombras del resto y utilizando para el reconocimiento, solamente las regiones no clasificadas como sombras. De igual forma, pero analizando las regiones sobre-iluminadas, en [Alvarez 06] se presenta un método en el que utilizando *Ada-Boost* [Freund 95] se detectan las regiones del rostro en las que hay manchas de luz y éstas son normalizadas de acuerdo a las regiones no afectadas. La deficiencia de estos métodos radica en que lidian con un solo tipo de afectación, no siendo suficientes para casos generales.

Una posible variante que no ha sido abordada en la literatura consultada, es emplear una medida de calidad para determinar si la imagen de rostro está afectada por la iluminación y sólo aplicarle un método de pre-procesamiento a aquellas que estén afectadas.

1.3.2 Métodos de evaluación de la calidad

A pesar de la popularidad del reconocimiento de rostros, muy poco aparece en la literatura en relación con medidas de calidad para este tipo de imágenes. De manera general, los

métodos que evalúan la calidad de una imagen se dividen en tres grupos [Kusuma 03]: los de referencia completa, los de referencia reducida y los que no utilizan referencia.

Los métodos de referencia completa son los más comunes [Zamani 08, Wang 04b, Gao 05]. Se basan en la comparación de la imagen original de buena calidad, con una imagen obtenida como resultado de la aplicación de algún proceso o transformación a la misma, como pueden ser la compactación, la transmisión o el almacenamiento. Este grupo de métodos no es adecuado para evaluar la afectación de la iluminación en las imágenes de rostros, ya que normalmente en ese caso, no se cuenta con una imagen de referencia.

En los métodos de referencia reducida se conoce una información estructural, representativa de un conjunto de imágenes con características similares, en la cual se apoyan para realizar el proceso de evaluación de la calidad de la imagen. En [Fronthaler 06] se propone un método de referencia reducida de gran eficacia para la calidad de huellas digitales, basado en un tensor de orientación de bordes, descompuesto en componentes determinadas por las estructuras de los puntos característicos de las huellas [Bigun 04]. En ese trabajo, se plantea que el método es también aplicable al caso de las imágenes de rostro, pero sería necesario proponer un tensor de orientación de bordes con estructuras faciales. Es evidente que los rasgos faciales no siguen patrones de orientación como en el caso de las huellas, por lo que podrían concentrarse los esfuerzos en generar un tensor de orientación que al final no sea realmente eficaz.

Los métodos de calidad sin referencia no se apoyan en ninguna información previa para estimar la calidad de una imagen, por lo que son los más factibles para integrar a un sistema de reconocimiento de rostros. A este grupo pertenece el único método encontrado en la literatura con este objetivo, que utiliza la simetría facial para evaluar las degradaciones causadas por una pose no frontal y una iluminación no frontal incidente en el rostro [Gao 07]. El principal problema de este método es que no logra distinguir entre una afectación por la pose y una por la iluminación.

En las imágenes (a) y (b) de la Figura 1.8 se muestra cómo dos imágenes de un individuo,

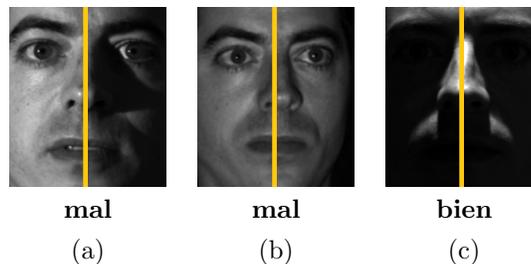


Figura 1.8: Ejemplos de imágenes evaluadas con el método basado en la simetría.

una afectada por la iluminación y la otra por la pose, obtienen la misma evaluación con el método basado en la simetría. Puede observarse también en la imagen (c) de la figura, que en ocasiones las afectaciones que provoca la iluminación en un rostro se comportan de forma simétrica, lo cual no se detecta con este método.

1.4 Consideraciones finales del capítulo

Al estudiar el estado actual de los algoritmos de reconocimiento de rostros, se observó que los que utilizan métodos basados en la apariencia local, son los más indicados para aplicaciones reales en condiciones de iluminación variables, cuando se tiene una sola imagen de entrenamiento por persona. Dentro de éstos, los métodos basados en los rasgos LBP y DCT muestran los mejores resultados. No obstante, cuando se analizó más profundamente el comportamiento de estos métodos, se pudo ver que no son del todo robustos a las variaciones en las condiciones de iluminación y por tanto, se necesitan integrar con otros métodos para enfrentar el problema.

Dentro de los métodos existentes para enfrentar las variaciones en las condiciones de iluminación en el reconocimiento de rostros, los de pre-procesamiento muestran ser los más aplicables a sistemas reales. Sin embargo, estos aún no alcanzan la máxima eficacia, principalmente porque se vuelve difícil suprimir la información relacionada con la iluminación sin afectar la información discriminativa de la imagen de rostro, además de que no todos los métodos han explotado el hecho de que el pre-procesamiento local ofrece mejores resultados que el global. Esto da lugar a la necesidad de desarrollar nuevos métodos de pre-procesamiento, basados en transformaciones locales, que logren suprimir de manera más cuidadosa el efecto de las variaciones de iluminación sin comprometer la información discriminativa de las imágenes de rostros.

Se analizó también que los métodos de pre-procesamiento existentes, disminuyen la eficacia de los algoritmos de reconocimiento de rostros cuando son aplicados a imágenes que no están afectadas por la iluminación. A partir de este análisis, se plantea que se puede desarrollar un algoritmo de reconocimiento de rostro más eficaz, evaluando la calidad de las imágenes de rostro respecto a la iluminación para determinar si se les aplica o no el pre-procesamiento. En la literatura sólo se encontró un método para evaluar las afectaciones provocadas por las variaciones de iluminación en las imágenes de rostro y este no es eficaz para todo tipo de afectaciones, además de que en algunos casos no logra diferenciar estas afectaciones de las provocadas por los cambios de pose.

Capítulo 2

Evaluación de la calidad de imágenes de rostros respecto a la iluminación

En este capítulo se presenta un método para evaluar automáticamente la calidad de una imagen de rostro respecto a la iluminación, que realiza un análisis local de la afectación provocada por las variaciones de iluminación. Para modelar y analizar el comportamiento local de la iluminación en un rostro, se propone un nuevo método de división en regiones de las imágenes de rostros, en el que se tiene en cuenta la forma general de los rostros y a la vez, las características de cada rostro en particular. El método de evaluación de la calidad que se propone, se evalúa en la clasificación de imágenes con afectaciones provocadas por la iluminación e imágenes con una iluminación adecuada y se compara su efectividad con la del único método encontrado para este propósito en la literatura.

2.1 División en regiones de las imágenes de rostros

La representación de los píxeles de una imagen bajo el modelo Lambertiano en la Ecuación (1.6), se puede expresar también como:

$$I(x, y) = \rho(x, y) \cdot \mathbf{n}^T(x, y) \cdot \mathbf{s} \quad (2.1)$$

donde ρ representa el albedo, que es un escalar que define la medida en que una superficie refleja la luz incidente y depende por tanto, de las propiedades físicas de esa superficie, \mathbf{n}^T es el vector normal a la misma y define la forma del objeto, mientras que \mathbf{s} , es un vector que representa la dirección e intensidad de la fuente de luz.

En el rostro, los cambios en las propiedades físicas son normalmente pequeños y locales. Las afectaciones que produce la iluminación en la imagen del rostro están generalmente provocadas por cambios en las normales a la superficie, \mathbf{n}^T . Si se logra dividir una imagen de rostro en regiones, teniendo en cuenta su forma, de manera tal que cada región obtenida pueda ser considerada como un plano, caracterizado por un mismo vector normal para todos los puntos de la imagen comprendidos en dicha región, la combinación de \mathbf{n}^T y \mathbf{s} ($\mathbf{n}^T \cdot \mathbf{s}$), tendría un valor constante, dependiente de la forma del rostro en esa vecindad. Cada región estaría entonces representada por los cambios pequeños en las propiedades físicas, es decir, por el albedo ρ .

Las divisiones en regiones existentes para imágenes de rostros, son las utilizadas por los métodos basados en la apariencia local [Ahonen 04, Tan 05, Martínez 02, Kepenekci 02], descritos en Capítulo 1. Como se explicó en el epígrafe 1.2, todas esas divisiones se basan en particiones regulares de la imagen, en las que no se tienen en cuenta la estructura y la forma de los rostros, lo cual sería necesario para obtener regiones que puedan ser consideradas como un plano, caracterizadas por un único vector normal. Para lograr este objetivo, la autora de esta tesis propone utilizar una malla triangular en la cual los vértices de los triángulos estén ubicados en aquellos puntos del rostro donde la superficie cambia de pendiente [Méndez-Vázquez 08c]. De esta forma se garantiza que cada región triangular sea aproximadamente un plano, como puede apreciarse en la Figura 2.1(a).

Esta malla difiere de la utilizada en los modelos de apariencia activa (AAM) [Cootes 01]. En los modelos AAM, los vértices se sitúan sobre los rasgos del rostro como las cejas, los ojos, la nariz, la boca y la mandíbula, construyéndose la malla para establecer un sistema de coordenadas de referencia respecto a esos puntos faciales. Como se observa en la Figura 2.1(b) los triángulos de la malla del modelo AAM no tienen en todos los casos una superficie plana, con la misma normal \mathbf{n}^T .

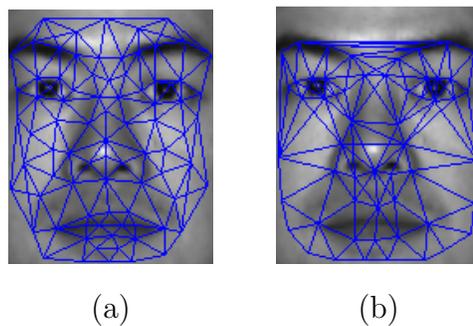


Figura 2.1: Comparación entre a) la malla triangular propuesta y b) la malla del modelo AAM.

Teniendo en cuenta que la forma general de los diferentes rostros humanos es muy similar, se puede construir una malla base y luego adaptarla a cada rostro en particular.

La malla base fue creada marcando manualmente 83 puntos en 10 imágenes de sujetos diferentes. La localización final de los 83 puntos fue determinada por el promedio de las ubicaciones en las 10 imágenes. Los 83 vértices de la malla fueron escogidos cuidadosamente para estar situados en zonas donde la pendiente de la superficie de los diferentes rostros cambia. Se utiliza luego una triangulación Delunay para unir los vértices y obtener las 150 regiones definidas por los puntos escogidos.

Una vez obtenida la malla base, esta es adaptada automáticamente a cada imagen de rostro específica que se quiere dividir, analizando su propia forma. Para extraer la información de la forma de un nuevo rostro se utiliza un descriptor de cordilleras y valles, los cuales representan los máximos positivos y mínimos negativos de la curvatura de relieve de las curvas de nivel de la imagen [López 99], es decir, los puntos de la superficie donde se producen los cambios de pendiente.

El operador de curvatura extrínseca fija de nivel multi-local (MLSEC por sus siglas en inglés) [López 00] ha sido identificado como un descriptor relevante de las cordilleras y valles del rostro, menos sensible a las variaciones de iluminación [Pujol 01], el cual se explica en más detalle en el Anexo 2.

Haciendo uso de este operador, para cada imagen de rostro la malla base es cargada y cada vértice es movido automáticamente al píxel más próximo que corresponda a una cordillera o valle. Estos puntos corresponden a los lugares donde la superficie del rostro cambia de pendiente, por tanto, los triángulos con pendiente constante quedan ajustados a cada imagen en particular. En la Figura 2.2 se muestran algunas imágenes obtenidas después de haber ajustado automáticamente la malla triangular en los respectivos rostros.

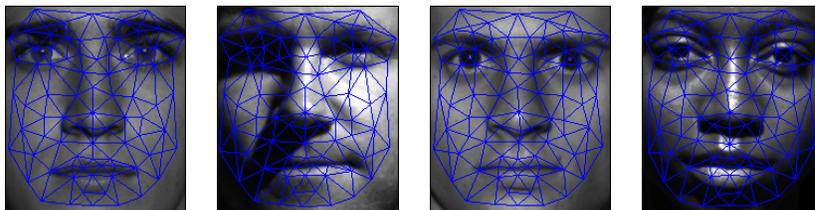


Figura 2.2: Imágenes de rostros de distintas personas en diferentes condiciones de iluminación con la malla triangular ajustada automáticamente.

Se puede observar en la figura que la malla se ajusta a las características individuales de cada rostro, aún cuando esté afectado por la iluminación, pero cada triángulo corresponde

aproximadamente a la misma región de un rostro, es decir, a un área específica de la nariz, la boca o los ojos, por ejemplo.

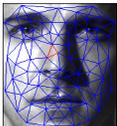
Una vez definidas las regiones triangulares, que tienen en cuenta las características propias de cada rostro, es posible analizar de manera local el comportamiento de la iluminación, que debe ser aproximadamente constante en cada región.

2.2 Análisis de la iluminación en las diferentes regiones del rostro

Para el análisis de la iluminación en las diferentes regiones del rostro, primeramente se convierte la imagen del espacio de colores RGB al espacio HSL. El espacio HSL es un espacio alternativo para la representación de colores, que no es aditivo y con el cual se trata de describir las relaciones perceptuales de los colores de forma más exacta. Este modelo es una forma de reflejar los colores más parecida a como lo capta el ojo humano y está compuesto por tres componentes: tono, saturación y luminancia. Se utiliza el componente de luminancia de cada píxel como una aproximación de la iluminación L .

Para el análisis local de la iluminación en cada región, se propone representar cada triángulo por los valores de luminancia de los píxeles en él, que se distribuyen ordenados ascendentemente en el intervalo $[0;1]$. En la Tabla 2.1 se muestra la distribución de los valores

Tabla 2.1: Distribución de los valores de luminancia de diferentes triángulos de una imagen.

Tipo de región	Imagen con la región seleccionada	Distribución de los valores de luminancia	Media	Mediana	Min.	Max.
Región oscura			0,059	0,06	0,05	0,09
Región brillante			0,98	1,00	0,71	1,00
Región normal			0,78	0,78	0,73	0,87

de luminancia para tres triángulos diferentes de una imagen. Además, se muestra la luminancia media o promedio de cada triángulo, la mediana, que es el valor de luminancia del píxel que marca la mitad de la cantidad de todos los que pertenecen al triángulo y los valores mínimo y máximo de luminancia presentes en la región.

Se puede apreciar en la tabla como los valores de luminancia reflejan la iluminación incidente en una región. En la región oscura la luminancia se concentra en valores bajos, en la región sobre-iluminada se concentra en los valores altos de luminancia, mientras que en la región con una iluminación adecuada, los valores de luminancia son medios. Se ve, además, que para una misma región los valores son aproximadamente constantes.

Después de tener una forma aproximada de describir la iluminación incidente en cada región del rostro, es necesario representar el comportamiento ideal o correcto de la misma, para en función de esto determinar si las regiones correspondientes a los triángulos ajustados a una imagen presentan una adecuada iluminación o no.

Con este objetivo se seleccionó un conjunto de 100 imágenes de diferentes sujetos con una iluminación frontal y homogénea de las bases de datos Yale B [Lee 05b] y XM2VTS [Messer 99]. Para cada uno de los 150 triángulos de la malla, se utilizó el valor promedio de luminancia (la media), obtenido en cada una de las 100 imágenes, con lo que se construyó un histograma representativo de esa región. Estos 150 histogramas, los cuales se nombrarán histogramas medios, describen el comportamiento de la luminancia en cada región cuando estas están iluminadas de manera normal. En la Figura 2.3 se ilustran gráficamente los histogramas obtenidos para algunos de los 150 triángulos de la malla.

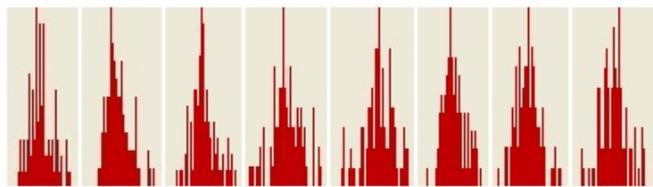


Figura 2.3: Histogramas medios de 8 de los triángulos de la malla.

Al observar estos histogramas se puede apreciar que su comportamiento se aproxima a una distribución normal o Gaussiana. Esto se corrobora utilizando las funciones `histfit` y `kurtosis` de Matlab [MathWorks 09]. La función `histfit`, obtiene la distribución Gaussiana que mejor se ajusta a la distribución que se desea evaluar, lo que permite comprobar cuánto se aproxima la distribución real a una normal. En la Figura 2.4 se muestran ejemplos de las distribuciones Gaussianas, líneas rojas, obtenidas a partir de tres de los histogramas medios de los triángulos, representados por las barras azules. Como puede observarse, en

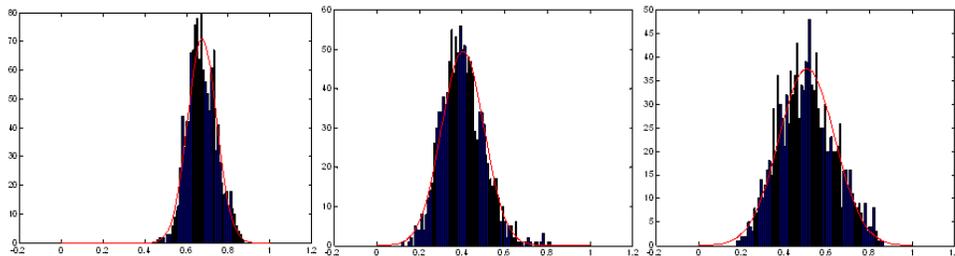


Figura 2.4: Ejemplos de funciones Gaussianas (en rojo) aproximadas a la distribución de histogramas medios (en azul).

los tres casos, no existen grandes diferencias entre las distribuciones Gaussianas generadas y los histogramas medios. Se utilizó, además, la función *kurtosis* de Matlab para llegar a una conclusión más rigurosa. La *kurtosis* de una distribución brinda una medida de cuán propensos son sus valores a ser atípicos. Para distribuciones normales, el valor de la *kurtosis* es 3 y mientras más se aleje de este valor, más valores atípicos presenta la distribución. Utilizando la función de Matlab, se calculó este valor para cada triángulo, obteniéndose un valor promedio de 3,17, lo cual confirma que las distribuciones de los histogramas medios de cada región se aproximan a distribuciones normales.

Teniendo en cuenta el análisis anterior, el comportamiento de la luminancia en una región triangular puede ser modelado como una función de probabilidad con distribución normal, representada por la expresión analítica:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.2)$$

donde μ es la media y σ la desviación estándar de los valores de luminancia media de cada triángulo, calculadas con las ecuaciones 2.3 y 2.4 respectivamente, en las que M es la cantidad de elementos de la muestra.

$$\mu = \sum_i^M X_i \quad (2.3)$$

$$\sigma = \frac{1}{M-1} \sum_i^M (X_i - \mu)^2 \quad (2.4)$$

Se calculan entonces los valores de la media μ y la desviación estándar σ para cada uno de los 150 triángulos, los cuales definen la distribución normal que aproxima a la distribución de valores de luminancia media de cada una de las regiones cuando están

iluminadas correctamente.

Esto permite calcular la probabilidad de que una región k , determinada por uno de los triángulos en una nueva imagen, posea una buena iluminación:

$$P_k = F_k(\mu_k) \quad (2.5)$$

donde k es uno de los triángulos de la malla, μ_k es la media del histograma de luminancia del triángulo para esa imagen y F_k es la función de probabilidad con distribución normal que aproxima al histograma medio del triángulo k , calculada según la ecuación 2.2.

Luego, la probabilidad promedio de que las diferentes regiones de una imagen estén iluminadas correctamente, \bar{P} , se puede calcular como:

$$\bar{P} = \frac{1}{N} \sum_{k=1}^N P_k \quad (2.6)$$

donde N es la cantidad de triángulos de la malla y P_k es la probabilidad de que el triángulo k esté bajo una iluminación adecuada, calculada según la ecuación 2.5.

Finalmente, la calidad de una imagen de rostro respecto a la iluminación, considerando la probabilidad media, puede ser determinada por:

$$MPQ = \begin{cases} 1 & \text{si } \bar{P} > \alpha \\ 0 & \text{e.o.c.} \end{cases} \quad (2.7)$$

donde α es un umbral que se determinará experimentalmente con imágenes bien iluminadas. Si el resultado final de MPQ es 1 significa que la imagen está iluminada correctamente y si es 0 la imagen presenta afectaciones provocadas por la iluminación.

Aunque MPQ se puede utilizar como medida de calidad definitiva, al analizar las diferentes zonas de la superficie del rostro, se puede notar que no todas tienen la misma importancia en el análisis de la iluminación incidente.

En dependencia de las características físicas de las personas, existen zonas del rostro que no serán de utilidad en el momento de realizar la estimación de la calidad de la iluminación de una imagen de rostro. Los ojos y sus regiones aledañas, así como la zona de la boca, no son confiables para determinar un comportamiento medio de la luminancia y pueden inducir a cometer errores en la estimación de la calidad. Los ojos y sus zonas vecinas, pueden tener un comportamiento diferente en personas sometidas a las mismas condiciones de iluminación. Debido a las características físicas específicas de cada individuo, puede ser

que existan personas con los ojos más hundidos o más hacia afuera que otras, generando zonas de sombras o sobre-iluminadas, que no estarán determinadas propiamente por las variaciones de la luz incidente en el rostro. En cuanto a la boca, es una característica que varía mucho en los diferentes individuos, tanto en la forma y el grosor de los labios, como en la textura y el tono, de ahí que el comportamiento del valor de luminancia en esta zona sea inestable aún cuando la iluminación incidente sea la misma. Además, un pequeño cambio en la expresión del individuo, es decir, cerrar los ojos o abrir la boca, puede provocar que el comportamiento de la luminancia en la región sea completamente diferente.

En la Figura 2.5 se muestran las imágenes de dos individuos diferentes sometidos a una luz que incide frontalmente sobre sus rostros, con el mismo ángulo e intensidad. Se puede observar que la iluminación incide de manera diferente en las áreas aledañas a los ojos y la boca. Al tener características totalmente diferentes, las zonas de los ojos de la imagen de la izquierda se perciben más oscuras. En el caso de los labios, los del primer sujeto son más anchos y carnosos lo que provoca un mayor reflejo de la iluminación que en el segundo, en el cuál la zona de la boca parece más oscura porque sus labios son más estrechos.

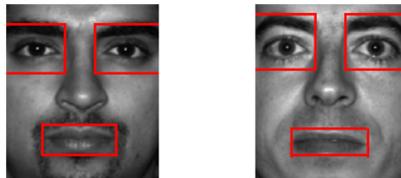


Figura 2.5: Regiones de un rostro en las que las mismas condiciones de iluminación, se comportan diferente para individuos diferentes.

Al analizar por otra parte, el área alrededor de la nariz, se observa que es una zona del rostro que contiene información valiosa sobre la incidencia de la iluminación. Físicamente, la nariz es la región más sobresaliente de la superficie facial, por lo que normalmente genera sombras en la imagen cuando la luz no incide homogénea y frontalmente sobre el rostro. Por otra parte, la nariz, a diferencia de los ojos y la boca, no se mueve de manera independiente, por lo que estando el rostro ubicado en una posición frontal, una sombra en el área de la nariz, no va a estar provocada por movimientos voluntarios del individuo, sino que serán indicios de una iluminación no frontal incidiendo sobre el rostro. En la Figura 2.6 se muestran algunas imágenes del rostro de una persona bajo distintas variaciones de iluminación, en las que se puede observar los efectos de estas en la zona de la nariz y sus alrededores.

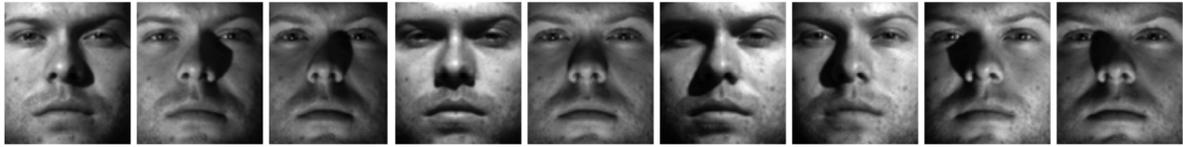


Figura 2.6: Imágenes de un mismo individuo con problemas de iluminación en la nariz y sus alrededores.

Teniendo en cuenta el análisis anterior, los triángulos de la malla que correspondan a las zonas de los ojos y la boca, no serán tenidos en cuenta en el cálculo de \bar{P} en la ecuación 2.6. Mientras que, por la gran influencia de las variaciones de iluminación en el área alrededor de la nariz, se decide utilizar una métrica adicional para analizar los triángulos localizados en esa área.

2.2.1 Análisis de la región de la nariz

Resulta difícil tener una evaluación *a priori* sobre la influencia de cada triángulo del área de la nariz en la calidad de la imagen. Se necesita un método que ajuste dinámicamente el peso de cada triángulo en la decisión final sobre el efecto de la iluminación en la región. Por esta razón se decidió utilizar una red neuronal para clasificar la calidad de la iluminación en el área de la nariz [Bishop 95]. A pesar de las desventajas ya conocidas que se desprenden de su utilización, en este caso no es necesario el conocimiento de la clasificación individual de cada triángulo en particular, sino el comportamiento general de la iluminación en esa zona, teniendo el conocimiento de la probabilidad de estar correctamente iluminado, para cada uno de los triángulo del área.

Con este fin, debido a su sencillez y extendido uso, se seleccionó una red neuronal perceptrón de tres capas: una de entrada, una intermedia y una de salida. La capa de entrada estará constituida por 24 neuronas, una por cada triángulo del área de la nariz. La cantidad de neuronas que integrarán la capa intermedia será determinada experimentalmente, en dependencia de la efectividad de la red. El número de neuronas de esta capa, define si cada neurona se especializará en mayor o menor medida, dependiendo de la proporción existente entre la cantidad de ellas que se incluye en esta capa y la cantidad de neuronas que componen la capa de entrada, que en este caso es 24. Esto influirá en el aprendizaje de la red y en su habilidad para clasificar correctamente una entrada determinada. La capa de salida estará compuesta por dos neuronas, NB y NM , que modelarán las probabilidades de que las regiones de la nariz, tengan buenas y malas condiciones de iluminación, respectivamente.

El proceso de determinar el comportamiento de la iluminación incidente en la nariz y sus alrededores, dada la utilización de la red neuronal, contará primeramente de una fase de entrenamiento con imágenes donde a simple vista se puedan identificar diferentes condiciones de iluminación, tanto buenas como malas, en esta región. En esta fase el valor de activación de NB para las imágenes con buena iluminación será 0,9999 y el de $NM = 0,0001$, mientras que $NB = 0,0001$ y $NM = 0,9999$ para las imágenes con afectaciones de iluminación en la zona de la nariz. Le sigue a la fase de entrenamiento, una de comprobación, para constatar los resultados que se derivan de la aplicación de la red. Una vez seleccionada y entrenada la red a utilizar, la calidad para la región de la nariz, de una imagen que llega para ser clasificada, se determina con:

$$NNQ = \begin{cases} 1 & \text{si } NB > NM \\ 0 & \text{e.o.c.} \end{cases} \quad (2.8)$$

Igual que para MPQ , 1 significará buena iluminación y 0 mala. Sin embargo, los resultados de ambas medidas de calidad, no tienen por que ser iguales siempre. Es necesario definir la forma en que ambos criterios son integrados para obtener una evaluación final de la calidad de la imagen respecto a la iluminación.

2.3 Calidad de la imagen respecto a la iluminación

A partir de los clasificadores MPQ y NNQ , basados en la probabilidad media de que las regiones del rostro estén bien iluminadas y en la red neuronal que analiza la región de la nariz, la autora de esta tesis propone un método para determinar la calidad de una imagen de rostro respecto a la iluminación [Méndez-Vázquez 10c].

Para esto, el primer paso es calcular mediante la ecuación 2.5, la probabilidad que tiene cada región, correspondiente a un triángulo de la malla ajustada a la imagen, de estar correctamente iluminada. A partir de estas probabilidades se evalúa la imagen utilizando los clasificadores MPQ y NNQ , de forma independiente. Si el resultado de ambos clasificadores coincide, ésta será la clasificación final de la imagen, de lo contrario, es necesario determinar cuál de los dos tiene la decisión correcta.

Con este objetivo se define lo que se llamará una *función de confianza* para cada uno de los dos clasificadores, la cual retornará un valor en el intervalo $[0; 1]$ que represente la seguridad de la clasificación, donde 1 significará una alta confianza y 0 muy poca seguridad en el resultado obtenido por el clasificador.

La *función de confianza* para el clasificador de la red neuronal, NNQ , se define de la siguiente forma:

$$NNC = |NB - NM|, \quad NNC \in [0; 1] \quad (2.9)$$

Esta función representa cuán bien separados y por tanto cuán confiables son los valores obtenidos por cada una de las neuronas de salida de la red neuronal.

Para MPQ , la *función de confianza* se define como:

$$MPC = \begin{cases} 0,99 + \frac{|\bar{P} - \alpha|}{100 \cdot \varphi} & \text{si } |\bar{P} - \alpha| > \delta \\ \frac{0,99 \cdot |\bar{P} - \alpha|}{\delta} & \text{si } |\bar{P} - \alpha| \leq \delta \end{cases} \quad (2.10)$$

donde δ representa el valor para el cual el clasificador tiene una confianza de 0,99; o sea, la seguridad de que una imagen con $\bar{P} > (\alpha + \delta)$ está bien iluminada y de que una imagen con $\bar{P} < (\alpha - \delta)$ está mal iluminada. La función está diseñada para asignarle menos confianza a los valores de \bar{P} en el intervalo $[\alpha - \delta; \alpha + \delta]$, siendo menor la confianza mientras más cerca se encuentre \bar{P} de α . El valor φ representa el supremo de las funciones de densidad que aproximan el comportamiento de la buena iluminación de cada región triangular. Si se analiza la ecuación 2.10 se puede apreciar que el primer término de la función toma valores en el intervalo $(0,99 ; 1]$ mientras que la segunda lo hace en el intervalo $[0 ; 0,99]$.

El valor de δ en la ecuación es determinado experimentalmente. Se escoge un valor para el cual más del 90 % de los valores de \bar{P} de las imágenes buenas del conjunto de entrenamiento sean mayores que $(\alpha + \delta)$ y que a la vez más del 90 % de los valores de \bar{P} de las imágenes de entrenamiento seleccionadas como malas sean menores que $(\alpha - \delta)$.

La calidad de una imagen respecto a la iluminación, estará finalmente determinada por la función:

$$CCQ = \begin{cases} MPQ \times NNQ & \text{si } MPQ = NNQ \\ MPQ & \text{si } MPC > NNC \\ NNQ & \text{e.o.c.} \end{cases} \quad (2.11)$$

Como se observa, el criterio de evaluación final se forma de la combinación de los dos clasificadores individuales utilizando sus *funciones de confianza*, obteniéndose como resultado, en el caso que haya discrepancias, la clasificación otorgada por el clasificador que muestre una mayor seguridad en la clasificación.

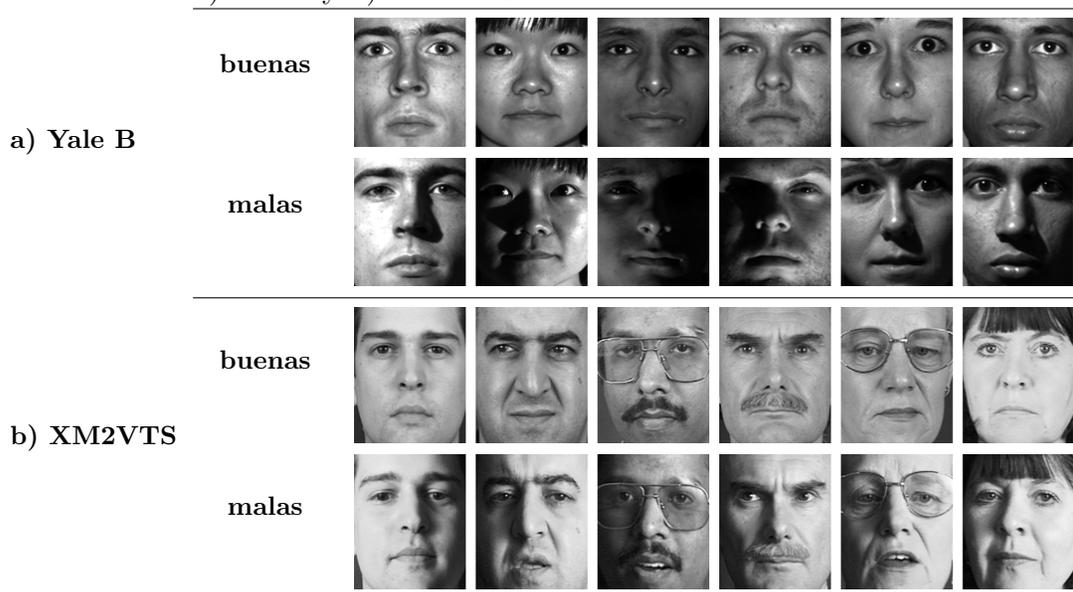
2.4 Evaluación experimental

Imágenes de las bases de datos de pruebas internacionales Yale B Extendida [Lee 05b] y XM2VTS [Messer 99] fueron utilizadas para entrenar y evaluar el funcionamiento del método propuesto en la clasificación de imágenes bien y mal iluminadas.

La base de datos Yale B es muy utilizada en la evaluación de los métodos de reconocimiento de rostros que enfrentan el problema de la iluminación, ya que contiene imágenes de 28 sujetos en 64 condiciones de iluminación diferentes, en las que se controla el ángulo entre la fuente de luz que incide en el rostro y el eje de la cámara fotográfica, de modo que mientras mayor es el ángulo, más afectadas por la iluminación están las imágenes. Para este experimento se seleccionaron como bien iluminadas las imágenes que fueron capturadas con un ángulo menor de 20^0 , de las cuales se utilizaron 50 para entrenamiento y 200 para pruebas. Del resto de las imágenes, con mayores ángulos y por tanto, representativas de distintas afectaciones provocadas por la iluminación, se seleccionaron aleatoriamente 80 para entrenamiento y 320 para pruebas.

Por otra parte, la base de datos XM2VTS contiene 2360 imágenes de 295 individuos tomadas en condiciones de iluminación controladas. Estas fueron divididas aleatoriamente en 160 para entrenamiento y 2200 para pruebas. Esta base de datos contiene un subconjunto de imágenes adicional “Oscuro”, que contiene cuatro imágenes de cada individuo

Tabla 2.2: Ejemplos de imágenes con buena y mala calidad respecto a la iluminación, en las bases de datos a) Yale B y b) XM2VTS.



con condiciones de iluminación desfavorables. Las imágenes de este subconjunto fueron divididas, también de manera aleatoria, en 100 para entrenamiento y 1080 para pruebas. Finalmente, al unir las imágenes seleccionadas de cada base de datos, se cuenta para el entrenamiento con un subconjunto de 210 imágenes con buena iluminación y otro con 180 imágenes afectadas por variaciones en la iluminación. Mientras que para las pruebas, se tiene un subconjunto con 2400 imágenes buenas y otro que contiene 1400 imágenes malas. En la Tabla 2.2 se muestran algunos ejemplos de imágenes seleccionadas en cada una de las bases de datos como representación de una buena y una mala iluminación. Para cada una de las bases de datos, las imágenes en una columna corresponden a un mismo individuo. Puede observarse que cuando la iluminación es adecuada, se distribuye de manera uniforme sobre la superficie de los rostros, mientras que cuando hay afectaciones, se reflejan de manera diferente en las distintas zonas del rostro, especialmente en la nariz, independientemente de cuán severa sea la afectación.

Para evaluar el rendimiento de los clasificadores se utilizó una matriz de confusión:

	Clasificadas como BIEN	Clasificadas como MAL
BIEN	TP	FN
MAL	FP	TN

donde TP representa el número de imágenes bien iluminadas que son clasificadas como buenas, FN el número de imágenes bien iluminadas que son clasificadas como malas, FP el número de imágenes mal iluminadas que son clasificadas como buenas y TN el número de imágenes mal iluminadas que son clasificadas correctamente como mal iluminadas.

A partir de la matriz de confusión se obtienen 3 métricas que describen la efectividad de un clasificador:

$$EB = \frac{TP}{TP+FP} \quad EM = \frac{TN}{TN+FN} \quad ET = \frac{TP+TN}{TP+FP+TN+FN}$$

donde EB representa la efectividad en la clasificación de imágenes bien iluminadas, EM la efectividad de la clasificación de las imágenes mal iluminadas y ET muestra la efectividad total del clasificador evaluado.

Se utilizaron las imágenes del conjunto de entrenamiento para determinar los valores de los parámetros α , δ y φ , así como para entrenar y seleccionar la red neuronal utilizada por NNQ .

El primer valor a determinar es el del umbral α , el cual decide cuándo una imagen está bien iluminada y cuándo no, según el clasificador basado en la probabilidad media, MPQ .

Inicialmente a α se le asigna el valor medio de \bar{P} entre los subconjuntos de imágenes buenas y malas de entrenamiento, el cual es 0,017. Este valor es aumentado y disminuido en 0,005 cada vez, para ir ajustándolo a la mayor efectividad del clasificador sobre los subconjuntos de pruebas. En la Tabla 2.3 se muestra la efectividad total del clasificador MPQ en las imágenes de pruebas para diferentes valores de α , obteniéndose el mejor resultado con el valor inicial seleccionado, por tanto se mantiene $\alpha = 0,017$.

Tabla 2.3: Determinación del umbral α .

α	0,0160	0,0165	0,017	0,0175	0,018
ET(%)	95,2	95,4	95,8	95,6	95,4

Una vez determinado α , se puede escoger el valor de δ que define el intervalo utilizado por la *función de confianza MPC*, fuera del cual se encuentran los valores de \bar{P} para los que se tiene más de un 99% de seguridad en la clasificación. El valor de δ fue seleccionado de forma tal que más del 90% de los valores de \bar{P} de las imágenes buenas de entrenamiento fueran mayores que $(\alpha + \delta)$ y a la misma vez que más del 90% de los valores de \bar{P} de las imágenes de entrenamiento con mala iluminación fueran menores que $(\alpha - \delta)$, obteniéndose bajo estas condiciones, $\delta = 0,004$.

El valor de φ , es el máximo valor que puede tomar la función de densidad que aproxima el comportamiento de la buena iluminación de cada región triangular, ya que tiene como objetivo normalizar el valor del sumando en la ecuación 2.10, para garantizar que el resultado final no sea mayor que 1. El valor supremo de todas las imágenes de entrenamiento, es $\varphi = 0,059$.

Se probó la red neuronal con distintas configuraciones, variando la cantidad de neuronas de la capa intermedia en cada caso. Por cada configuración de la red, se realizaron cinco entrenamientos diferentes. En la Tabla 2.4 se muestran los resultados obtenidos con las dos configuraciones que ofrecieron los mejores resultados.

Tabla 2.4: Determinación de la red neuronal más efectiva.

Configuración Entrenamiento	20 neuronas					30 neuronas				
	1	2	3	4	5	1	2	3	4	5
ET(%)	94,5	92,5	92,7	93,1	93,1	94,0	94,6	96,1	95,8	94,2

Se eligió finalmente para clasificar la calidad de las imágenes respecto a la iluminación, la red obtenida con el tercer entrenamiento de la configuración que utiliza 30 neuronas en la capa intermedia que, como puede observarse en la tabla anterior, es la más efectiva.

En la Tabla 2.5 se muestran los resultados de la evaluación de las imágenes en el conjunto de pruebas, para cada uno de los clasificadores individuales y el clasificador combinado, con los parámetros y entrenamientos escogidos. Los resultados se expresan en términos de las tres métricas definidas, que miden la efectividad clasificando imágenes buenas (*EB*) e imágenes malas (*EM*), así como la efectividad total (*ET*).

Tabla 2.5: Efectividad de los clasificadores individuales y del combinado.

	EB	EM	ET
MPQ	93,29 %	97,06 %	95,80 %
NNQ	97,53 %	94,92 %	96,18 %
CCQ	96,90 %	96,55 %	96,77 %

Se puede observar en la tabla que el clasificador que combina los clasificadores individuales es mejor que estos, no sólo en la efectividad total, sino también obtiene un mejor balance clasificando tanto imágenes buenas como malas. Este no es el caso de los clasificadores individuales: *MPQ* logra clasificar mejor las imágenes con problemas de iluminación, mientras que *NNQ* obtiene una mayor eficacia clasificando las imágenes buenas. Los resultados obtenidos confirman que la clasificación basada en la región de la nariz es más efectiva que utilizar sólo la probabilidad media de los triángulos, así como los beneficios de complementar los clasificadores individuales para tomar la decisión final.

En la Tabla 2.6 se ejemplifica el funcionamiento del método propuesto sobre 4 imágenes de un mismo individuo. Se puede apreciar cómo se resuelve la discrepancia entre los clasificadores individuales utilizando las *funciones de confianza*, garantizándose de esta forma una mayor efectividad en la clasificación.

Tabla 2.6: Ejemplos del funcionamiento del método de evaluación de la calidad.

				
MPQ	0	0	1	1
NNQ	0	1	0	1
MPC	-	0,23241	0,99069	-
NNC	-	0,99949	0,81171	-
CCQ	0	1	1	1

Las imágenes en el conjunto de pruebas, fueron clasificadas además con el único método encontrado en la literatura para evaluar la calidad de una imagen de rostro respecto a la

iluminación, basado en la simetría del rostro [Gao 07] y con el clasificador *MPQ*, basado en la probabilidad media, pero utilizando una división regular en bloques sobre el rostro en lugar de la malla triangular. En la Tabla 2.7 se puede comparar la efectividad total obtenida por estos clasificadores, con la del método *CCQ*.

Tabla 2.7: Comparación de métodos para evaluar la calidad de las imágenes de rostros respecto a la iluminación.

	método basado en la simetría [Gao 07]	MPQ basado en división en bloques	CCQ con división triangular
ET	63,71 %	90,10 %	96,77 %

Se puede apreciar, que es muy significativa la superioridad del método que se propone comparado con el basado en la simetría. Los resultados también demuestran, que la utilización de la malla triangular permite cuantificar adecuadamente las afectaciones provocadas por la iluminación en las distintas regiones del rostro.

2.5 Conclusiones parciales

El método que se propone en este capítulo para evaluar la calidad de una imagen de rostro respecto a la iluminación, basado en el análisis local del comportamiento de las variaciones de iluminación en el rostro, supera en más de un 30 % de efectividad total, al único método para este propósito encontrado en la literatura.

La división triangular que se propone, demuestra experimentalmente, ser efectiva para modelar y analizar el comportamiento local de la iluminación en las imágenes de rostros, ya que al utilizarla se alcanza una eficacia mayor, en más de un 5 %, comparada con la obtenida utilizándose los métodos de división en regiones tradicionales.

Al tener en cuenta, además, que no todas las zonas del rostro tienen la misma relevancia al analizar el comportamiento de la iluminación, se desarrolló un clasificador en el que se combinan el análisis del comportamiento promedio de la iluminación en todas las regiones, con el análisis del comportamiento en las regiones más relevantes. Esto condujo a un clasificador que obtiene una efectividad total de un 96,7 %, superior que la alcanzada por los clasificadores individuales basados en la probabilidad media de que las distintas regiones estén bien iluminadas y en la clasificación de las regiones relevantes, logrando además, un mejor balance al clasificar tanto imágenes buenas como malas.

Capítulo 3

Pre-procesamiento de imágenes de rostros basado en el análisis local

En este capítulo se presenta un nuevo método de pre-procesamiento para enfrentar las afectaciones provocadas por la iluminación en las imágenes de rostros. Para lograr esto, primeramente se analiza el efecto de las variaciones de iluminación en las imágenes de rostros y su influencia en los componentes de bajas frecuencias de la imagen. Se decide utilizar de manera local los coeficientes de la Transformada Discreta del Coseno (DCT) de la imagen en el dominio logarítmico, para estimar las variaciones provocadas por la iluminación en el rostro y este componente estimado, se sustrae de la imagen original en el dominio logarítmico para lograr la compensación deseada. El método propuesto es probado en bases de datos internacionales y comparado con varios algoritmos de pre-procesamiento existentes. Se demuestra experimentalmente la eficacia del método, con un por ciento de error en la clasificación muy similar al método de pre-procesamiento que exhibe los mejores resultados en la literatura consultada. Se realiza un análisis profundo de los dos métodos y se revelan diferencias significativas en su comportamiento. Esta diversidad es explotada en el diseño de un método de fusión de clasificadores que integra y complementa ambos métodos.

3.1 Compensación de las variaciones de iluminación en las bajas frecuencias

La mayoría de los métodos de pre-procesamiento parten de la ecuación (1.6), buscando eliminar el valor de luminancia, L y recuperar la reflectancia, R , para discriminar entre los diferentes rostros. Los métodos más exitosos como el filtrado homomórfico, el suavizado anisotrópico y las diferentes variantes del retinex, suponen que la luminancia cambia suavemente sobre la escena y que por eso es un fenómeno de bajas frecuencias, mientras que la reflectancia, que caracteriza la textura de la piel, contribuye al contenido de altas frecuencias de la imagen.

Debido a la naturaleza multiplicativa del modelo de generación de la imagen, estas dos fuentes de información pueden ser separadas y transformadas en una mezcla aditiva mediante un filtrado en el espacio logarítmico:

$$\log I(x, y) = \log R(x, y) + \log L(x, y) = \log \rho(x, y) + \log(\mathbf{n}^T(x, y) \cdot \mathbf{s}) \quad (3.1)$$

De ahí que varios métodos incluyan un paso de transformación logarítmica para eliminar el efecto de la iluminación.

Volviendo a la ecuación (1.6), el término luminancia es un producto derivado de la luz incidente y la orientación de la superficie. Para una superficie con una lenta variación en la normal a la superficie, la luminancia será una función de baja frecuencia. Ciertamente, para una primera aproximación, muchos investigadores modelan la cabeza humana como un cilindro. Tal perfil 3D conformaría una imagen de luminancia suavizada y de baja frecuencia, que sería fácilmente filtrada. Sin embargo, el rostro contiene características morfológicas como los ojos, la nariz, la boca y las arrugas que inyectan componentes de altas frecuencias a la función de luminancia. Las sombras causadas por estas ondulaciones en la superficie aportan información sobre la estructura 3D de la superficie facial, que debe ser preservada para lograr establecer las diferencias entre las distintas personas.

Similarmente, el término reflectancia contiene información de baja y altas frecuencias. La característica dominante de la piel es de baja frecuencia. La textura de la piel es básicamente homogénea, cambiando muy lentamente sobre la superficie del rostro. Sin embargo, en las vecindades de los rasgos faciales, como las cejas, labios, ojos, defectos de la piel, lunares y bello facial, el albedo cambia rápidamente, introduciendo una señal de alta frecuencia a la función de reflectancia.

El análisis anterior sugiere que ambos componentes, tanto el de luminancia como el de reflectancia de una imagen de rostro, contienen altas y bajas frecuencias. Esto hace difícil el hecho de separar el efecto de la luminancia de la reflectancia meramente sobre la base del contenido de frecuencia. Sin embargo, en ambos casos el contenido de altas frecuencias es el que contiene la información más importante para los fines discriminativos y el de bajas frecuencias el que más varía con los cambios en la iluminación. Por tanto, al eliminar las componentes de bajas frecuencias de la imagen, se compensan las variaciones de iluminación, pero estas no corresponden solamente a la luminancia, sino que contienen también información de la reflectancia, por lo que las bandas de frecuencia a eliminar deben ser cuidadosamente escogidas.

Con el objetivo de identificar un método de compensación adecuado, se analiza en primer lugar el efecto de las condiciones de iluminación variables en una imagen de rostro. En general, la función que representa la luz en la imagen puede ser muy compleja. Para simplificarla, se asume que la escena es iluminada por una sola fuente de luz distante. Las variaciones en la intensidad de la luz, no son un problema, ya que estas pueden ser manejadas sencillamente con un escalamiento. Lo más desafiante es el efecto que provocan los cambios en el ángulo de incidencia de la luz.

En la Figura 3.1 se muestran algunas imágenes correspondientes a un modelo sintético de rostro 3D con albedo constante para distintos ángulos de incidencia de la luz. Por tanto las diferencias que se aprecian en la imagen están dadas por los cambios de $\mathbf{n}^T \cdot \mathbf{s}$, es decir, la luminancia.

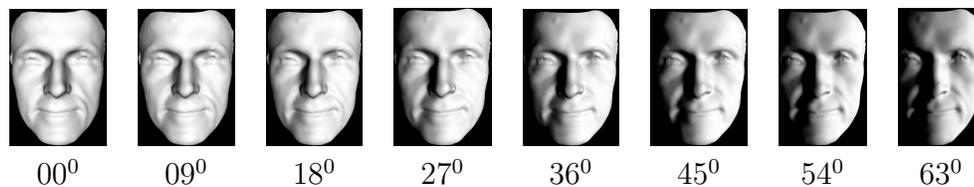


Figura 3.1: Ejemplos de los cambios en la luminancia de imágenes de rostros con albedo constante a diferentes ángulos de luz incidente.

La Figura 3.2 muestra la varianza del logaritmo de las imágenes de luminancia obtenidas para una superficie de rostro 3D, como una función del ángulo de incidencia de la luz. Se puede apreciar que para una iluminación frontal o casi frontal, la varianza en el logaritmo de la luminancia es poca, pero esta aumenta drásticamente para ángulos de iluminación incidente más pronunciados.

Las observaciones anteriores indican que los cambios debidos a la iluminación, a pesar de que varían lentamente, pueden incidir en un amplio espectro de frecuencias. Por tanto,

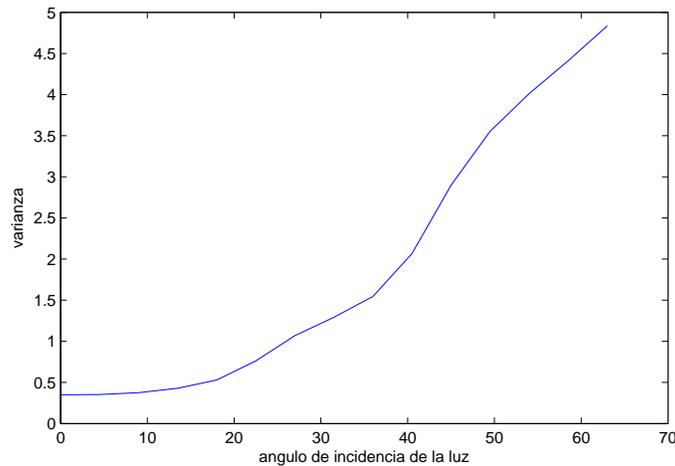


Figura 3.2: Varianza del logaritmo de la luminancia para un rostro 3D con albedo constante, como función del ángulo de la luz incidente.

los métodos que persiguen minimizar el efecto de la iluminación suprimiendo las bajas frecuencias [Short 04, Chen 06, Tan 07] fallan en eliminar completamente la afectación provocada por la iluminación, o peor aún, comprometen el contenido de información útil para el análisis biométrico, en dependencia de la cantidad de componentes de frecuencia que se supriman.

Para maximizar la efectividad del método de normalización fotométrica, es más conveniente estimar y sustraer de manera cuidadosa el contenido de baja frecuencia que está afectado por la iluminación en lugar de suprimirlo radicalmente, para que la información discriminatoria de alta frecuencia no esté comprometida. Más aún, debe tenerse en cuenta que las afectaciones provocadas en la imagen por las variaciones en el ángulo de incidencia de la luz no afectan de la misma forma a todas las regiones del rostro, por tanto no es adecuado utilizar un filtro homogéneo en toda la imagen.

3.2 Método de pre-procesamiento basado en la DCT local

Con el objetivo de optimizar el filtrado del contenido de baja frecuencia de la imagen, esta se puede representar en el dominio de la frecuencia. Varios métodos han sido utilizados para transformar la imagen del dominio espacial al dominio de la frecuencia. El método DCT es comúnmente utilizado en el procesamiento de imágenes y señales por su simpli-

cidad, su bajo costo computacional y por su capacidad de compactación de la energía [Rao 90].

En [Chen 06], se presenta un método que utiliza la DCT para compensar las variaciones de iluminación. Se eliminan los coeficientes DCT de bajas frecuencias de una imagen en el dominio logarítmico como una aproximación del término de compensación que es necesario sustraer de la imagen. Este método muestra resultados superiores a muchos de los métodos existentes que lidian con las variaciones de iluminación al ser comparados en la base de datos Yale B [Chen 06]. Recientemente, en los trabajos [Abbas 09] y [Liau 10] se presentaron dos extensiones de este método. Estas planteaban como objetivo lograr una mayor eficiencia computacional; en ninguno de los dos casos se aumentó la eficacia del método. En la Figura 3.3 (b) puede observarse una imagen de rostro pre-procesada con este método.

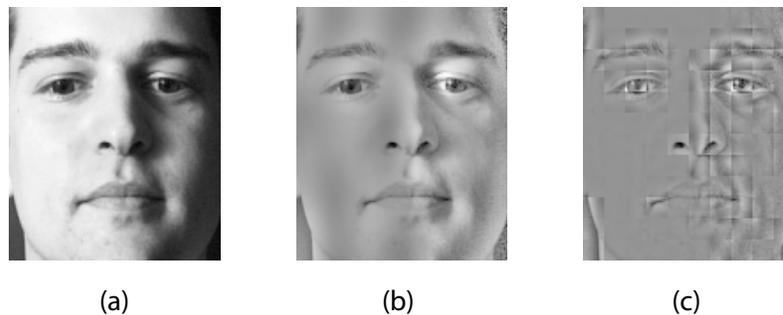


Figura 3.3: Ejemplo de una imagen de rostro (a) original y normalizada fotométricamente utilizando (b) la DCT global y (c) la DCT local.

Poniendo en cero los coeficientes DCT de bajas frecuencias, se crea un filtro ideal con una banda de transición perfecta. Sin embargo, este tipo de filtros crea ondulaciones en el dominio espacial. Este efecto es apreciable en la Figura 3.3 (b), principalmente en los cambios de tonos que aparecen en la frente y el lateral izquierdo de la imagen, que no existen en la imagen original. Para ilustrar mejor el problema, se escoge una imagen, que contiene una función de intensidad de un impulso Dirac mostrada en la Figura 3.4 (a), se transforma utilizando la DCT global y se reemplazan por cero los coeficientes de baja frecuencias. En la Figura 3.4 (b), se muestra la imagen reconstruida en el dominio espacial, en la cual las ondas se extienden sobre toda la imagen.

En el caso de las imágenes de rostros, estas ondulaciones pueden tener un efecto perturbador en el método que se utilice posteriormente para describirlas y clasificarlas. Para

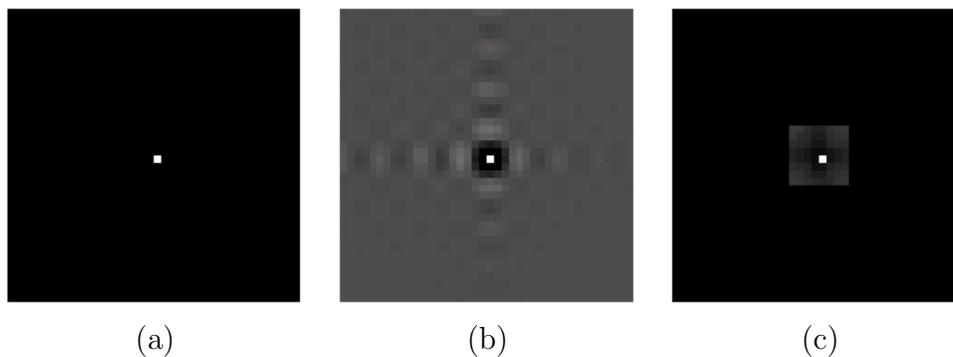


Figura 3.4: Representación del efecto de remover los coeficientes de bajas frecuencias de una imagen de manera global y de manera local. (a) Imagen original, (b) imagen procesada de manera global y (c) imagen procesada de manera local.

minimizar este efecto la autora de esta tesis propuso utilizar la DCT de manera local [Méndez-Vázquez 08b]. El beneficio del procesamiento local, se puede percibir en la Figura 3.4 (c) donde el efecto ondulatorio queda confinado a una pequeña localidad de la imagen.

En el trabajo presentado por la autora de este trabajo en [Méndez-Vázquez 08b], se mostró que al descartar los coeficientes DCT de bajas frecuencias en el dominio logarítmico de manera local, se mejora la eficacia en el reconocimiento de imágenes afectadas por variaciones de iluminación, con respecto al uso de la DCT global. En la propuesta que hace la autora, la imagen de rostro se divide en regiones regulares y los coeficientes DCT de bajas frecuencias son descartados en cada bloque de la imagen. Posteriormente cada región es representada mediante histogramas de Patrones Binarios Locales (LBP) uniformes que son utilizados para la clasificación. En ese caso, el pre-procesamiento realizado con los coeficientes de DCT de manera local, está altamente atado a la división en regiones de la imagen que se utiliza para la extracción de rasgos y la clasificación.

Al utilizarse la misma división de la imagen para el pre-procesamiento que para la extracción de rasgos, se evita el efecto de bloque que provoca el uso del método DCT por regiones, el cual puede llegar a ser altamente pronunciado como se muestra en la Figura 3.3 (c). Si las regiones utilizadas para el pre-procesamiento y para la extracción de rasgos fueran incongruentes, la información de alta frecuencia que se inserta en los bordes de las regiones puede llegar a ser devastadora y llegar a negar incluso, los beneficios del pre-procesamiento, al introducirse información no existente (ruido) en la imagen original.

Desafortunadamente, la cantidad de métodos en los cuales existe una congruencia natural entre la ventana de filtrado utilizada en el pre-procesamiento y la extracción de rasgos,

es muy limitada. Además, en varios trabajos han sido mostrados los beneficios de los métodos de extracción de rasgos basados en el análisis de múltiples resoluciones, es decir, que no se limitan a describir la imagen de rostro procesando bloques de la imagen de tamaño fijo [Zhang 04, Raja 06, Chan 07]. Se necesita entonces desarrollar un método que retenga la sensibilidad local sin introducir el efecto de los bloques, que pueda ser utilizado por cualquier descriptor de rasgos o clasificador sin importar en qué forma se divida la imagen.

La autora de esta tesis propone una modificación del método anterior que mantiene el análisis por regiones y elimina el efecto de los bloques, permitiendo que las imágenes pre-procesadas puedan ser utilizadas por cualquier método de representación de imágenes de rostros [Méndez-Vázquez 10b]. La propuesta tiene como objetivo eliminar el contenido de baja frecuencia de la imagen de rostro que se encuentra afectado por la iluminación.

El primer paso es transformar la imagen al dominio de intensidad logarítmico para obtener una mezcla aditiva, según la Ecuación (3.1). Como se demostró en [Chen 06], la imagen normalizada, I' puede ser obtenida sustrayendo un término de compensación, ε , de la imagen original en el dominio logarítmico, que representa la diferencia entre la iluminación incidente y una iluminación uniforme en el dominio logarítmico, es decir:

$$\log(I') = \log(I) - \varepsilon \quad (3.2)$$

El segundo paso consiste entonces en estimar el término de compensación en el dominio logarítmico, que va a estar basado en los componentes de bajas frecuencias obtenidos con la DCT local. Al sustraer la información de baja frecuencia de la imagen original en el dominio logarítmico, se obtiene la imagen pre-procesada, que contendrá únicamente la información sobre los cambios en la luminancia debido a las propiedades de la superficie de la estructura de los rasgos faciales y sobre los pequeños cambios en el albedo, es decir, la información de alta frecuencia discriminativa. La imagen normalizada en el dominio logarítmico puede ser restaurada a su dominio original, sin embargo, esto es sólo una transformación de escala y se ha visto que puede introducir ajustes incorrectos y ruidos en los valores normalizados [Chen 06]. De hecho, también hay evidencias fisiológicas, de que la respuesta de las células de la retina puede ser aproximada como una función logarítmica de la intensidad [Adini 97]. En consecuencia con esto, en el enfoque propuesto, las imágenes se utilizarán directamente en el dominio logarítmico para el reconocimiento.

En este proceso, estimar el término de compensación basado en los componentes de bajas frecuencias, es fundamental para lograr una normalización fotométrica adecuada.

3.2.1 Estimación del término de compensación

Con el objetivo de utilizar la información local en lugar de la global, la imagen de rostro se divide en bloques rectangulares y la DCT, expresada en la ecuación (1.3) se calcula sobre estos. Utilizando sólo los coeficientes de bajas frecuencias de cada bloque y poniendo el resto en cero, una versión de paso bajo de la imagen logarítmica puede ser reconstruida aplicando la inversa de la DCT.

En un bloque DCT, los coeficientes de la esquina superior izquierda, seleccionados con un recorrido en zig-zag corresponden a la información de baja frecuencia. Sin embargo, el coeficiente $C(0, 0)$, usualmente conocido como coeficiente DC, está relacionado con el valor promedio de intensidades del bloque, representando mucha de la energía de la región. Esto puede ser apreciado en la ecuación (3.3), la cual es una simplificación de la ecuación (1.3), considerando que el coseno de cero es uno.

$$C(0, 0) = \frac{1}{\sqrt{M}\sqrt{N}} \cdot \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x, y) \quad (3.3)$$

El efecto del coeficiente DC puede ser visto en la Figura 3.5. En la primera fila puede observarse la representación de los valores de los coeficientes DC de cada bloque de imágenes de modelos de rostros 3D de diferentes personas transformadas logarítmicamente, mientras que la segunda fila muestra los valores DC para imágenes del mismo modelo bajo diferentes condiciones de iluminación. Todas las imágenes representadas en la figura están re-muestreadas, de manera que cada píxel, representa el valor DC correspondiente a un bloque de la misma. Como puede apreciarse, los coeficientes DC de una imagen están estrechamente vinculados a la iluminación que se percibe en cada bloque y a la vez, en su conjunto, contienen información de alta frecuencia asociada a los cambios en la luminancia provocados por las estructuras de los rasgos faciales como la nariz, los ojos y los labios.

Como se quiere estimar el componente de baja frecuencia asociado a las variaciones de iluminación para sustraerlo de la imagen en el dominio logarítmico, es necesario modificar el coeficiente DC de cada bloque en una forma que refleje los cambios en la iluminación incidente. En la Figura 3.5 se evidencia que en las imágenes con ángulo frontal o cercanos a la frontal, los coeficientes DC de los bloques respectivos muestran una pequeña dispersión en sus valores. Por otra parte, si el ángulo de incidencia de la luz aumenta, la diferencia entre los coeficientes DC comienza a ser mayor. Esto puede ser medido objetivamente en términos de la varianza de los coeficientes DC. En la gráfica de la Figura 3.6 se traza la



Figura 3.5: Representación del coeficiente DC de cada bloque de las imágenes de luminancia de modelos 3D con albedo constante. (a) Diferentes modelos con la misma iluminación incidente. (b) El mismo modelo con diferentes ángulos de incidencia de la luz.

varianza de los coeficientes DC de la luminancia logarítmica de una superficie de rostro 3D con albedo constante como función del ángulo de incidencia de la luz. Puede verse como mientras mayor es la variación en la iluminación, mayor es la dispersión de los valores DC, teniendo un comportamiento creciente de manera similar a la varianza de la luminancia de las imágenes en el dominio logarítmico, representada en la gráfica 3.2.

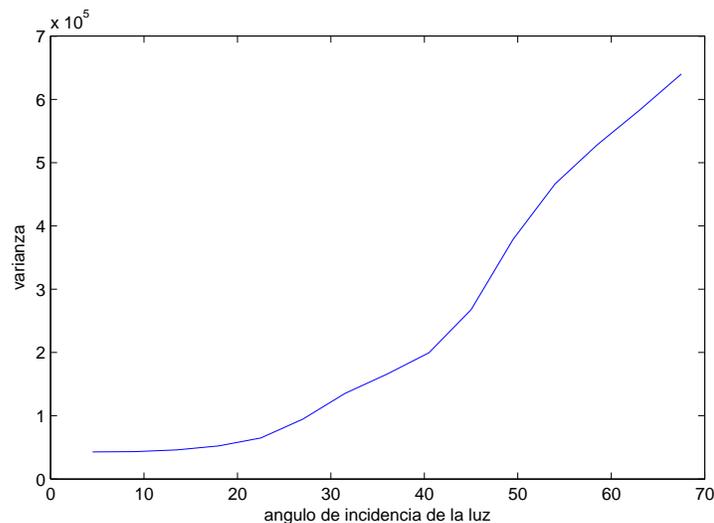


Figura 3.6: Varianza de los coeficientes DC como función del ángulo de incidencia de la luz.

Esto indica que si un valor constante, que represente un valor DC correspondiente a una iluminación uniforme, es sustraído de cada coeficiente DC, el valor obtenido representará la información (positiva o negativa) inyectada a esa región por las variaciones en la iluminación. Para obtener entonces la imagen que representa el componente de baja frecuencia afectado por la iluminación, se utilizarán los coeficientes DCT de bajas frecuencias sobre los bloques de la imagen, restándole a cada coeficiente DC un valor de

referencia constante.

El valor de referencia a utilizar, se determinó calculando el valor promedio de los coeficientes DC en un conjunto de 80 imágenes de entrenamiento de distintos individuos con la misma iluminación frontal incidente. En la Figura 3.7 se muestran los valores promedio obtenidos por cada imagen y el promedio general. Como puede verse, para una misma iluminación el promedio de valores de los coeficientes DCT para imágenes de diferentes individuos es muy similar. Se escoge el valor del promedio general obtenido para esta iluminación frontal, como valor de referencia para normalizar las distintas imágenes de rostros.

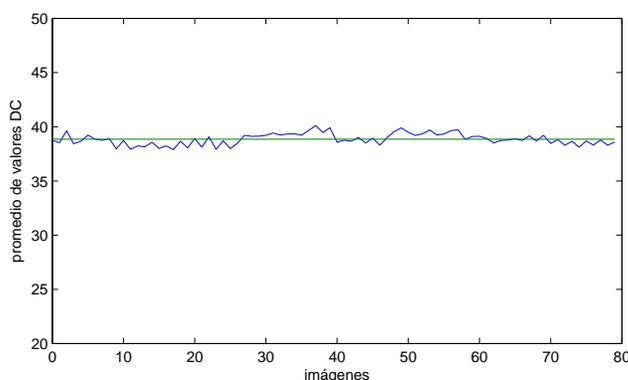


Figura 3.7: Promedio de los coeficientes DC de imágenes de distintos individuos con la misma iluminación frontal.

Una vez reconstruida la imagen de bajas frecuencias, para minimizar el efecto de la división por bloques, se le aplica un filtro de suavizado antes de sustraerla de la imagen original en el dominio logarítmico. Al utilizarse un filtro de suavizado de paso bajo sobre el término de compensación estimado, no interfiere, ni se contradice con el objetivo de eliminar bajas frecuencias, por el contrario, se bloquean las altas frecuencias que pueden haber quedado en la imagen de compensación estimada.

El procedimiento que se propone, para pre-procesar una imagen I , puede ser resumido en los siguientes pasos: 1) aplicar la transformada logarítmica a la imagen original, $\log(I)$, 2) reconstruir la versión de paso bajo de la imagen logarítmica, ε , usando los coeficientes DCT de bajas frecuencias y modificando el valor DC, 3) suavizar la imagen resultante, $\varepsilon' = B(\varepsilon)$, 4) sustraer el término de compensación de la imagen original en el dominio logarítmico, $I' = \log(I) - \varepsilon'$. El efecto de cada uno de estos pasos puede ser observado en la Figura 3.8, mostrando al final la imagen pre-procesada con el método propuesto.

Como puede verse en la imagen obtenida, a pesar de que se pierde cierta información

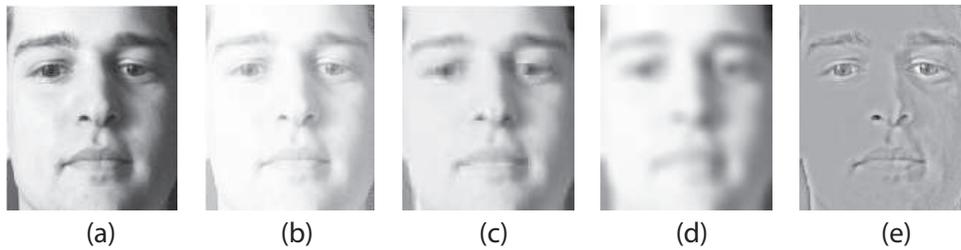


Figura 3.8: Ejemplo del efecto de cada uno de los pasos del método de preprocesamiento propuesto: (a) imagen original, (b) transformación logarítmica, (c) imagen de compensación de iluminación con efecto de bloques, (d) imagen de compensación suavizada, (e) imagen resultante luego de la sustracción.

sobre el tono de la piel original, se suprime toda la información de baja frecuencia que puede verse afectada por las variaciones de iluminación.

3.3 Evaluación del método de pre-procesamiento propuesto

Para evaluar el método propuesto es necesario utilizar un algoritmo de reconocimiento de rostros que reciba como entrada la imagen pre-procesada y la clasifique. Con este objetivo se utilizan los métodos de apariencia local que en el Capítulo 1 se seleccionaron como los de mejores resultados: los basados en los rasgos LBP y DCT. En el caso del basado en DCT, se concatenan los rasgos de cada región en un solo vector que se clasifica con un vecino más cercano ($1nn$) utilizando la distancia euclidiana con la norma L_1 , combinación con la cual se reportan los mejores resultados para este descriptor [Ekenel 05]. Para el LBP se utilizaron tres variantes diferentes: el LBP clásico utilizando el vecino más cercano con la medida de similitud χ^2 , la descripción multi-resolución (MLBP) con el mismo clasificador $1nn$ y medida χ^2 y por último, la extensión más reciente de este método conocida como MLBP+LDA [Chan 07].

Estos algoritmos serán evaluados como sistemas de verificación, siguiendo la Configuración I del protocolo Laussane, diseñado con este objetivo para la base de datos de pruebas internacionales XM2VTS [Messer 99]. La base de datos está compuesta por 2360 imágenes de 295 individuos, capturadas en condiciones de iluminación controladas durante 4 sesiones de 2 fotos por cada individuo. Según este protocolo, las imágenes se dividen en tres subconjuntos: Entrenamiento, Evaluación y Prueba. La distribución de las imágenes en

cada subconjunto, para ser comparadas con imágenes del mismo individuo (genuinos) o con imágenes de individuos diferentes (impostores), se muestra en la Tabla 3.1.

Tabla 3.1: División de las imágenes en la base de datos XM2VTS según la Configuración I del protocolo Laussane.

Sesión	Foto	Genuinos	Impostores	
		200	25	70
1	1	Entrenamiento	Evaluación	Prueba
	2	Evaluación		
2	1	Entrenamiento		
	2	Evaluación		
3	1	Entrenamiento		
	2	Evaluación		
4	1	Prueba		
	2			

Las imágenes en el subconjunto de Entrenamiento se utilizan para crear los modelos o representaciones de los sujetos. Las imágenes del subconjunto de Evaluación se comparan con las imágenes de Entrenamiento para hallar el umbral de decisión al determinar si dos imágenes corresponden a una comparación de genuinos o de impostores. El umbral determinado se utiliza luego para estimar los errores en la clasificación de las imágenes en el subconjunto Prueba al compararlas con las de Entrenamiento. La base de datos cuenta además con otro subconjunto “Oscuro”, que contiene cuatro imágenes de cada uno de los individuos en condiciones desfavorables de iluminación, las cuales son también comparadas con las imágenes del conjunto de Entrenamiento utilizando el umbral determinado en Evaluación. El número total de comparaciones de genuinos y de impostores realizadas por cada subconjunto se resume en la Tabla 3.2.

Tabla 3.2: Cantidad de comparaciones realizadas por cada subconjunto de la base de datos XM2VTS.

	Evaluación	Prueba	Oscuro
Comparaciones de genuinos	600	400	800
Comparaciones de impostores	40000	112000	56000
Total de comparaciones	40600	112400	56800

A partir de estas comparaciones se miden las tasas de errores de falsos aceptados (FAR) y falsos rechazados (FRR) en cada subconjunto, utilizando cada posible valor como umbral:

$$FAR = \frac{\text{falsos aceptados}}{\text{comparaciones de impostores}} \cdot 100 \quad FRR = \frac{\text{falsos rechazados}}{\text{comparaciones de genuinos}} \cdot 100$$

Los falsos aceptados son los casos de comparaciones de impostores que se aceptan como genuinas, mientras que los falsos rechazados son las comparaciones de genuinos que son rechazadas como si fueran impostores. La tasa de error igual (EER o *Equal Error Rate* en inglés) es la tasa de error en la que FAR=FRR. El valor de similitud para el cual se obtiene el EER en el subconjunto de Evaluación, se utiliza como umbral para las comparaciones de las imágenes en los conjuntos Prueba y Oscuro. A partir del FAR y el FRR obtenidos con ese umbral se calcula la tasa de error total (TER) como la suma de ellos, la cual se utiliza como medida para analizar el rendimiento de los métodos.

Después de definido el protocolo de experimentación, antes de integrar el método de pre-procesamiento a los algoritmos de reconocimiento de rostros y evaluarlos, es necesario ajustar los parámetros que se utilizan en su implementación, para con ellos lograr un mejor rendimiento.

3.3.1 Selección de los parámetros del método

El primer parámetro a seleccionar es el tamaño de los bloques en los que será dividida la imagen para aplicar la DCT local. Si el tamaño de los bloques es muy pequeño se necesita mayor esfuerzo computacional. Por otra parte, si el tamaño de los bloques es muy grande, más serán las variaciones en las bajas y altas frecuencias de cada bloque. Se divide la imagen en bloques iguales de $w \times w$ y se prueba el método con diferentes valores de w para seleccionar el mejor.

Otro parámetro crucial en el método propuesto, es el número de coeficientes DCT utilizados en cada bloque para estimar el contenido de baja frecuencia de la imagen. Los coeficientes DCT normalmente son obtenidos con un escaneado en zig-zag, que va desde las bajas frecuencias a las más altas. Tres puntos de parada diferentes en la lista de coeficientes DCT, obtenidos con el escaneado en zig-zag, fueron probados para los diferentes valores de w . En la Figura 3.9 (a) se grafican los por cientos de error total obtenidos en los tres subconjuntos de imágenes de la base de datos XM2VTS, con el método propuesto utilizando diferentes valores de w y diferentes cantidades de coeficientes DCT de bajas frecuencias. En estos experimentos se utilizó el método LBP con la medida de similitud χ^2 , por su simplicidad y eficiencia computacional. No obstante, al probar los otros métodos se percibió un comportamiento similar en las gráficas correspondientes a los diferentes subconjuntos.

Como puede apreciarse, el menor error en todos los casos se obtiene con una división en bloques de 8×8 , la cual es además la división utilizada tradicionalmente por el método



Figura 3.9: Rendimiento (TER) del método propuesto en los conjuntos de Evaluación (*azul*), Prueba (*rojo*) y Oscuro (*verde*) de la base de datos XM2VTS, (a) con diferentes divisiones en bloques $w \times w$ y distintas cantidades de coeficientes DCT y (b) con diferentes filtros de suavizado de distintos tamaños.

DCT cuando se emplea en la compresión [Watson 94]. El número óptimo de coeficientes DCT fue más difícil de seleccionar como era de esperar. Se decidió utilizar 15 coeficientes DCT porque con este valor se muestra el rendimiento más estable para los tres subconjuntos con los diferentes tamaños de bloques probados y el error más bajo en el conjunto de imágenes Oscuro con el tamaño de bloque seleccionado.

Como parámetro final es necesario seleccionar el filtro y el tamaño para la operación de suavizado que se le aplica a la imagen de compensación estimada. Existen una gran

cantidad de filtros de suavizado diseñados para el procesamiento digital de imágenes [Jain 89]. Dentro de estos, por su simplicidad y uso extendido, se seleccionaron para las pruebas un filtro de media cuadrado y uno circular, así como uno Gaussiano. En la Figura 3.9 (b) se reporta el rendimiento del método utilizando el tamaño de bloque y el número de coeficientes DCT escogidos, para los diferentes filtros de suavizado en los tres subconjuntos de la base de datos. El eje x de las gráficas en el inciso (b) representa el tamaño h del núcleo (*kernel*) utilizado por los filtros, a partir del cual, en el caso del filtro Gaussiano, se puede obtener la desviación estándar, σ , que lo caracteriza, con $\sigma = (h - 1)/2$.

En las gráficas se puede observar que los tres filtros exhiben un rendimiento similar. En general, a medida que el tamaño del núcleo aumenta, el error de clasificación disminuye. Sin embargo, mientras más grande es el tamaño del núcleo del filtro, mayor es el costo computacional. El menor error en el conjunto de imágenes Oscuro, fue obtenido utilizando un filtro de media circular con un núcleo de tamaño 11, con el cual se obtiene también un buen rendimiento en los conjunto de imágenes de Evaluación y Prueba. Este tamaño de núcleo representa también un buen valor para el compromiso entre la eficacia y el costo computacional.

Utilizando los parámetros optimizados, se evalúa entonces el método propuesto (LDCT), en el pre-procesamiento de imágenes afectadas por variaciones de iluminación, que son recibidas como entrada por cada uno de los algoritmos de reconocimiento de rostros seleccionados.

Además de los experimentos de verificación en el conjunto Oscuro de la base de datos XM2VTS, se llevan a cabo también experimentos de identificación en la base de datos Yale B Extendida [Lee 05b]. Esta base de datos normalmente se divide en 5 subconjuntos para las pruebas de identificación, teniendo en cuenta el ángulo entre la luz incidente y la cámara que captura la imagen, que mientras mayor es, más afectadas están las imágenes. En este caso, se utiliza una imagen con la luz incidiendo frontalmente para representar a cada uno de los 38 individuos en la base de datos. Se comparan contra estas imágenes, con el objetivo de determinar la identidad de los rostros en las fotos, las 532 imágenes con ángulos entre 35° y 50° que conforman el subconjunto S3, las 456 imágenes del subconjunto S4 que tienen ángulos entre 55° y 70° y las del subconjunto S5, que está compuesto por 570 imágenes con ángulos mayores de 75° llegando hasta 110° . En todos los casos se representaron y clasificaron las imágenes originales (OI) y luego se repitió el proceso con las imágenes pre-procesadas con el método propuesto (LDCT). En la Tabla 3.3 se muestran los por cientos de error obtenidos en cada experimento.

Tabla 3.3: Error (%) en la clasificación de imágenes afectadas por la iluminación.

		XM2VTS	Yale B		
		Oscuro	S3	S4	S5
DCT+L_1	OI	50,2	13,6	24,2	26,8
	LDCT	39,3	9,90	12,7	19,7
LBP+χ^2	OI	95,7	22,0	74,9	88,1
	LDCT	65,2	2,82	24,7	57,0
MLBP+χ^2	OI	89,6	10,5	58,3	85,5
	LDCT	48,4	2,01	16,2	55,3
MLBP+LDA	OI	13,7	10,2	50,6	58,5
	LDCT	4,55	1,65	4,33	18,68

En el caso de la verificación, el error que se reporta es el TER, mientras que en la identificación, es el de las equivocaciones al determinar como imagen más similar en la colección, la de un individuo que no es el mismo que se busca.

Como puede observarse, a pesar de las diferencias entre los resultados de los distintos clasificadores, en todos los casos se alcanza una mejora significativa al utilizar el método de pre-procesamiento propuesto en lugar de clasificar directamente las imágenes originales afectadas por la iluminación.

3.4 Comparación con otros métodos de pre-procesamiento

Con el objetivo de comparar el método de pre-procesamiento propuesto con otros métodos de pre-procesamiento existentes en la literatura, se utilizó la combinación MLBP+LDA para describir y clasificar las imágenes en la base de datos XM2VTS, el cual mostró el menor por ciento de error, al clasificar las imágenes con pre-procesamiento y sin él. En la Tabla 3.4 se muestran los por cientos de error total (TER) para cada subconjunto de imágenes de la base de datos, en la clasificación de las imágenes originales (OI) y con métodos de pre-procesamiento muy conocidos como la Ecualización de Histogramas (HE), el Filtrado Homomórfico (HF) y el Cociente Propio de la Imagen (SQI), así como con otros métodos que reportan muy buenos resultados como el de Suavizado Anisotrópico (AS), el Cociente Total de Variación de la Imagen (TVQI) y la Secuencia de Procesamiento (PS), además del método propuesto (LDCT). En la Figura 3.10 pueden verse, además,

Tabla 3.4: Comparación de diferentes métodos de pre-procesamiento en los tres conjuntos de la base de datos XM2VTS utilizando MLBP+LDA.

	Evaluación	Prueba	Oscuro
OI	1,90	1,16	13,7
HE	2,10	1,17	13,5
HF	2,35	1,35	12,7
SQI	2,31	1,84	11,6
TVQI	2,65	1,98	6,98
AS	2,08	1,50	6,15
PS	2,00	1,56	3,72
LDCT	2,00	1,32	4,55

ejemplos de imágenes pre-procesadas con cada uno de estos métodos. Se puede apreciar que la principal diferencia del método que se propone, LDCT, comparado con el resto, es que logra eliminar toda la información de la imagen relacionada con la iluminación, sin introducir información que no aparece en la imagen original.

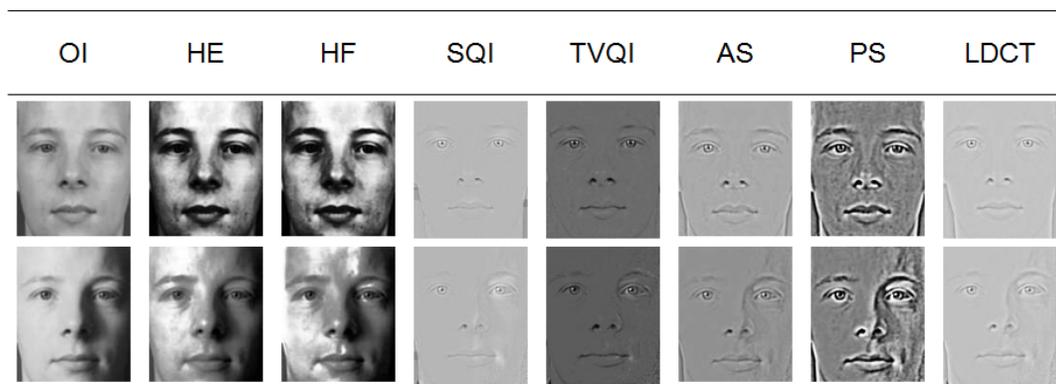


Figura 3.10: Ejemplos de imágenes pre-procesadas con diferentes métodos.

Dentro de todos los métodos, el PS exhibe los mejores resultados al pre-procesar imágenes afectadas por la iluminación (conjunto Oscuro), seguido por el método propuesto en un pequeño rango de por ciento de error. Por otra parte, al pre-procesar las imágenes sin afectaciones de iluminación (en el conjunto Prueba), el PS mostró una eficacia algo menor que el LDCT. De manera general, sus tasas de error total son muy similares. Si se analiza además la complejidad computacional de los dos métodos, puede apreciarse que la cantidad de operaciones y la complejidad de los filtros utilizados por cada uno de ellos es muy parecida, pudiendo implementarse en ambos casos con funciones muy eficientes.

Teniendo en cuenta que tienen puntos en común, pero que a la vez producen imágenes pre-procesadas bien diferentes, es necesario una comparación más profunda entre ellos.

3.4.1 PS vs. LDCT

Al comparar el método PS y el método de pre-procesamiento propuesto (LDCT), la mayor diferencia que resalta entre ellos es en la información de frecuencia que es retenida y suprimida en cada caso. El primer paso de ambos métodos, corrección de intensidades gamma para el método PS y transformación logarítmica para el LDCT, funciona de la misma forma, realizando las regiones oscuras en los valores de intensidad de la imagen y comprimiendo los más brillantes. En los dos métodos el segundo paso es el fundamental. En PS, el filtrado de pasa-banda, DoG, atenúa las más bajas y más altas frecuencias, reteniendo la información en las bandas medias, mientras que el método LDCT propuesto suprime el contenido de baja frecuencia utilizando los coeficientes DCT de baja frecuencia para construir el término de compensación que es sustraído de la imagen logarítmica. La atenuación de algunas altas frecuencias en PS, puede ser la causa de la disminución de su rendimiento en las imágenes bien iluminadas, al no estar eliminando en esos casos el contenido asociado a la iluminación, sino información sobre los rasgos del rostro. El último paso de cada método tiene objetivos diferentes. En el caso de PS, la imagen filtrada es nuevamente procesada para mejorar el contraste global, mientras que en el LDCT, la imagen de compensación obtenida es suavizada para remover el efecto de los bloques. Estas diferencias provocan que las imágenes obtenidas como salida de cada uno de los métodos, sean diferentes.

Se puede percibir que el PS y el LDCT trabajan diferente, pero el error total alcanzado por los dos métodos en la base de datos XM2VTS fue muy similar. Es necesario entonces comprobar si el error específico en la clasificación errónea de las imágenes está correlacionado en ambos casos, es decir, si los métodos fallan clasificando las mismas imágenes o no. Con este objetivo se utiliza la prueba estadística conocida como estadística z [Webb 02], que permite determinar si dos clasificadores obtienen salidas diferentes.

Las diferencias en la clasificación errónea de los métodos PS y DCT, se pueden representar en una matriz de confusión:

		LDCT	
PS	n_{00}	n_{01}	
	n_{10}	n_{11}	

en la cual,

- n_{00} = número de imágenes clasificadas mal por ambos métodos PS y LDCT
- n_{01} = número de imágenes clasificadas mal por PS pero no por LDCT
- n_{10} = número de imágenes clasificadas mal por LDCT pero no por PS
- n_{11} = número de imágenes no clasificadas mal ni por PS ni por LDCT

La estadística z se define a partir de estas variables, como :

$$z = \frac{|n_{01} - n_{10}| - 1}{\sqrt{n_{10} + n_{01}}} \quad (3.4)$$

De manera que si $|z| > 1,96$ se puede decir que ambos métodos no tienen el mismo error en la clasificación (con una probabilidad de 0,05 de decisión incorrecta).

En la Tabla 3.5 se muestra la matriz de confusión obtenida para cada uno de los subconjuntos de imágenes de la base de datos XM2VTS. En la Tabla 3.6 se refleja la estadística z correspondiente a cada caso.

Tabla 3.5: Comparación entre la cantidad de imágenes mal clasificadas utilizando ambos métodos en cada uno de los subconjuntos de imágenes de la base de datos XM2VTS.

		LDCT		
a) Evaluación	PS	581	1042	1623
		699	38278	38977
		1280	39320	40600
		LDCT		
b) Prueba	PS	1724	3028	4752
		1950	105698	107648
		3674	108726	112400
		LDCT		
c) Oscuro	PS	353	640	993
		968	54839	55807
		1321	55479	56800

Tabla 3.6: Estadística z en cada uno de los conjuntos de imágenes de XM2VTS.

	Evaluación	Prueba	Oscuro
$ z $	8,19	15,26	8,15

Se puede observar que en los tres subconjuntos de imágenes el valor de la estadística z es muy superior a 1.96. Esto significa que al utilizar ambos métodos se cometen errores

diferentes en la clasificación individual de las imágenes. Un análisis más profundo de las coincidencias en las imágenes clasificadas erróneamente (n_{00}), reportado en la Tabla 3.7, muestra que menos de la mitad de las imágenes clasificadas incorrectamente, son mal clasificadas por los dos métodos a la misma vez.

Tabla 3.7: Por ciento de coincidencias en la clasificación errónea de las imágenes.

	Evaluación	Prueba	Oscuro
PS	35,80 %	36,28 %	35,55 %
LDCT	45,39 %	46,92 %	26,72 %

El análisis anterior muestra claramente que las diferencias en los procesos de ambos métodos, PS y LDCT, generan diversidad en los resultados obtenidos en el reconocimiento de rostros. Esta diversidad puede ser explotada para mejorar el rendimiento utilizando un método de combinación. En el próximo epígrafe se presenta un esquema de combinación propuesto para integrar y complementar los métodos de pre-procesamiento PS y LDCT.

3.5 Combinación de métodos de pre-procesamiento

Es conocido que la fusión de múltiples clasificadores es un método efectivo para mejorar el rendimiento de los sistemas de reconocimiento de patrones [Kuncheva 01, Kittler 98]. El único requisito necesario es que los componentes individuales que vayan a ser fusionados, provean información complementaria para la clasificación. Existen diferentes vías de lograr diversidad en los clasificadores. Las opciones incluyen diferentes espacios de rasgos, diferentes clasificadores, diferentes medidas de similitud e incluso, diferentes parámetros en el procedimiento de aprendizaje de un clasificador.

En este caso, los métodos de extracción de rasgos y clasificación, son los mismos para los componentes individuales que conformarán el sistema combinado y se propone alcanzar la diversidad utilizando diferentes técnicas de pre-procesamiento, que significan dos representaciones diferentes de la imagen recibidas para la extracción de rasgos y la clasificación [Méndez-Vázquez 10b]. En particular, se combinan las soluciones obtenidas con el método PS y el LDCT, como se ilustra en la Figura 3.11.

Los aspectos esenciales en la fusión de múltiples clasificadores son la arquitectura de fusión, la normalización de los valores de similitud obtenidos como salida de los distintos clasificadores y la regla de fusión que se utilice para integrar los valores de similitud. La normalización de los valores de similitud normalmente se requiere si la regla de fusión

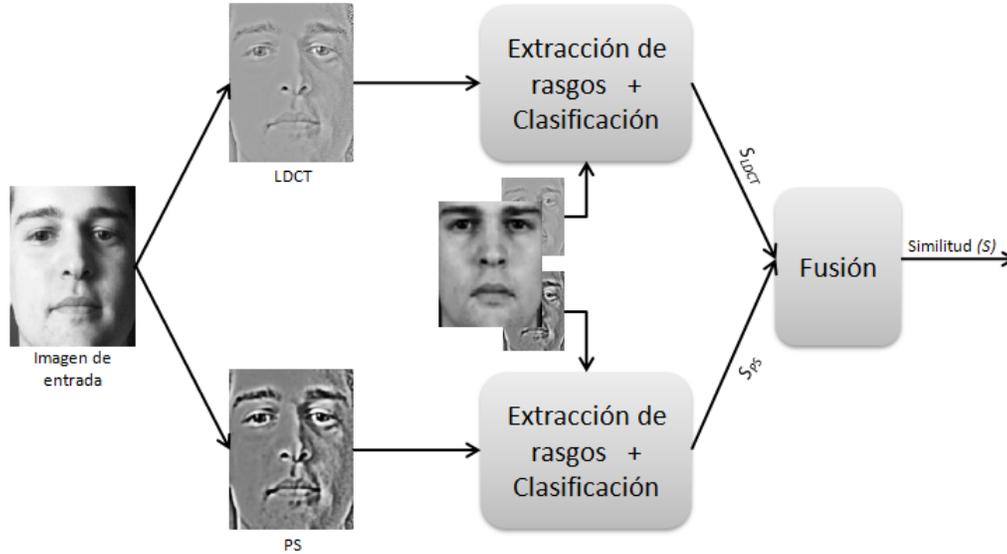


Figura 3.11: Esquema de combinación propuesto.

adoptada es simple, sin ningún entrenamiento. Para reglas de fusión entrenadas, la normalización de los valores no es necesaria, ya que en el proceso de inferencia de la regla de fusión, se aprende un peso adecuado para cada una de las entradas del clasificador múltiple.

Para integrar las salidas de los clasificadores individuales, se decide utilizar una fusión simple con una regla fija, la suma, ya que se quiere una solución que dependa lo menos posible del entrenamiento. Ha sido demostrado que la regla de fusión de la suma es efectiva y además robusta a ruidos en los datos de entrada [Kittler 98]. Además, como en este caso se propone utilizar el mismo clasificador con los dos métodos de pre-procesamiento diferentes, los valores de similitud obtenidos como salidas en los componentes individuales del esquema, que deben ser fusionados, se encuentran en el mismo rango y por tanto no requieren estrictamente una normalización.

Luego, si se denota el valor de similitud obtenido por el sistema al comparar una imagen de rostro de entrada con una imagen en la colección, ambas pre-procesadas con LDCT, como S_{LDCT} y, el obtenido para las mismas imágenes con el pre-procesamiento PS, como S_{PS} , el valor de similitud fusionado estará entonces dado por:

$$S = S_{LDCT} + S_{PS} \quad (3.5)$$

Los méritos de este esquema de fusión simple, pueden ser apreciados en los resultados mostrados en la Tabla 3.8.

Tabla 3.8: Resultados de la fusión utilizando el clasificador MLPB+LDA.

		Evaluación	Prueba	Oscuro
PS	FAR	1,00	1,06	0,47
	FRR	1,00	0,5	3,25
	TER	2,00	1,56	3,72
LDCT	FAR	1,00	1,07	0,42
	FRR	1,00	0,25	4,13
	TER	2,00	1,32	4,55
PS+LDCT	FAR	0,90	0,97	0,37
	FRR	0,89	0,20	2,50
	TER	1,79	1,17	2,87

Utilizando el esquema de fusión se logra una mejora significativa en la clasificación de las imágenes en los tres conjuntos de la base de datos XM2VTS, independientemente de que las imágenes estén afectadas o no por la iluminación.

3.6 Conclusiones parciales

Utilizando los coeficientes DCT de bajas frecuencias, extraídos de manera local de las imágenes de rostros en el dominio logarítmico, para estimar y pre-procesar las afectaciones que provoca la iluminación, se obtienen errores totales en la evaluación de las imágenes de los conjuntos Prueba (que contiene imágenes con buenas condiciones de iluminación) y Oscuro (que contiene imágenes afectadas por las variaciones de iluminación) de la base de datos XM2VTS de 1,32 % y 4,13 % respectivamente, muy similares a los del método que mejores resultados reporta en la literatura.

Utilizando la estadística z se comprobó, que a pesar de que los por cientos de error total son muy similares, ambos métodos no cometen los mismos errores, es decir, no se equivocan al clasificar las mismas imágenes. Esta diversidad fue aprovechada en el diseño de un clasificador múltiple que los integra y logra superar la eficacia obtenida por los métodos individuales, reduciendo el error a 1,17 % para el caso de las imágenes en el conjunto Prueba y 2,87 % en el conjunto Oscuro.

Capítulo 4

Algoritmo de reconocimiento de rostros para sistemas reales en condiciones variables de iluminación

En este capítulo se describe el algoritmo de reconocimiento de rostros que integra los métodos presentados en los dos capítulos anteriores, el cual puede ser utilizado en el reconocimiento de imágenes de rostros en aplicaciones reales en las que las condiciones de iluminación no estén controladas. El esquema propuesto se evalúa tanto en la verificación como en la identificación de imágenes de rostros en bases de datos de pruebas internacionales y los resultados obtenidos son comparados con los de los algoritmos existentes en la literatura actual. Se presentan además los resultados obtenidos utilizando el método propuesto en el reconocimiento de imágenes de rostros en cinco entornos reales. Finalmente se reflejan algunos aspectos relacionados con la implementación y uso en la práctica del algoritmo propuesto.

4.1 Reconocimiento de rostros utilizando métodos locales

Los sistemas de reconocimiento de rostros constan de cuatro pasos fundamentales. El primero de ellos es la detección del rostro dentro de la imagen recibida como entrada, es decir, separar la parte de la imagen que contiene exclusivamente el rostro. El segundo paso tiene como objetivo normalizar las distintas imágenes de rostros para que puedan

ser comparadas entre ellas. Luego de normalizada la imagen, se procede a la extracción de los rasgos que serán utilizados en el reconocimiento, obteniéndose una representación del rostro adecuada al método de comparación automática que se vaya a utilizar. Finalmente tiene lugar la clasificación, que consiste en comparar los rasgos extraídos de la imagen de entrada con los rasgos de una o varias imágenes de una colección, según sea el objetivo verificación o identificación respectivamente, para ofrecer una respuesta sobre la imagen analizada.

Los métodos de evaluación de la calidad y de pre-procesamiento presentados en los capítulos 2 y 3 de esta tesis, se insertan a este proceso como parte del paso de normalización. Esta normalización no se refiere sólo a una alineación geométrica de los rostros sino que incluye una estandarización de las condiciones de iluminación en las distintas imágenes. No obstante, el esquema de fusión propuesto que integra los métodos de pre-procesamiento, influye en los pasos de extracción de rasgos y clasificación, al utilizarse dos representaciones de la imagen para realizar la comparación. Por otra parte, el objetivo de introducir el método de evaluación de la calidad es para no aplicarle el pre-procesamiento a todas las imágenes de entrada, solo a aquellas que lo necesiten por estar afectadas por las variaciones de iluminación. Los pasos específicos que se definen para conformar el algoritmo de reconocimiento de rostros que integra los métodos propuestos [Méndez-Vázquez 10a], se muestran en la Figura 4.1.

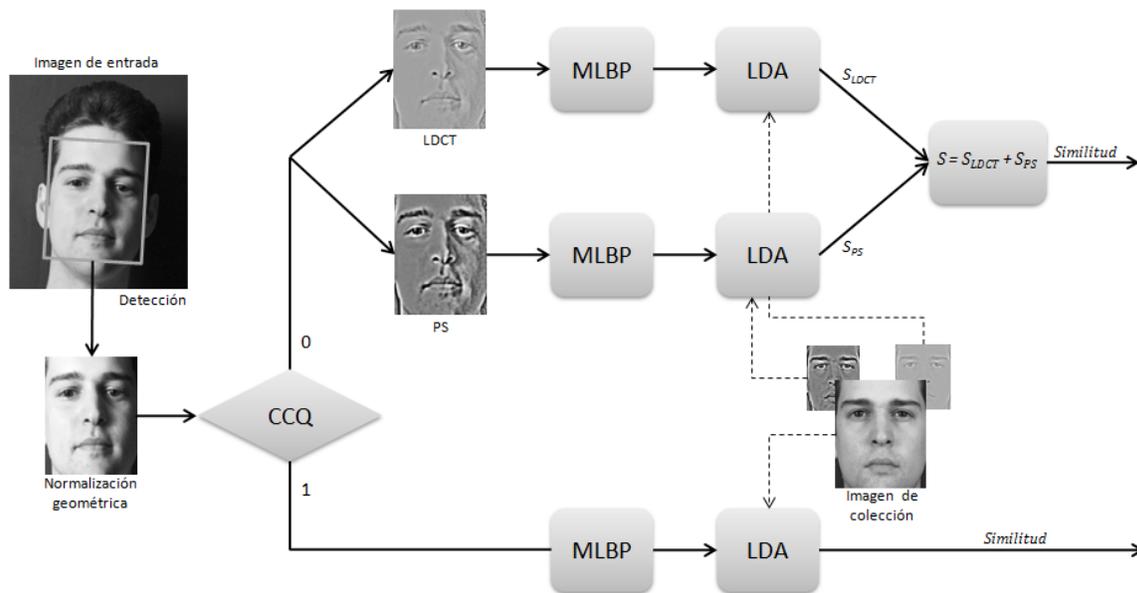


Figura 4.1: Algoritmo de reconocimiento de rostros.

Como puede observarse, se seleccionó el método de extracción de rasgos MLBP con la cla-

sificación basada en LDA, ya que en las evaluaciones experimentales mostró los mejores resultados en el reconocimiento de las imágenes con y sin pre-procesamiento. Es necesario aclarar que esta selección no contradice la suposición inicial de que en aplicaciones reales se cuenta con sólo una o muy pocas imágenes de entrenamiento. Si bien es cierto que el método LDA requiere varias imágenes de entrenamiento de cada sujeto para representar mejor las variaciones intra-personales, en este caso al ser un método basado en la apariencia local que obtiene una representación de cada imagen a múltiples resoluciones, se crea un subespacio por cada región con los histogramas de las distintas resoluciones, lográndose aumentar de esta forma los datos representativos de cada clase. Teniendo además la flexibilidad de enriquecerse mientras más imágenes de muestra existan para cada sujeto.

En el algoritmo representado en el esquema de la Figura 4.1, la detección del rostro y los ojos, se realizará con el algoritmo de Viola y Jones [Viola 01] que es el detector más utilizado en los sistemas de reconocimiento de rostros, por los buenos resultados que ofrece y su bajo costo computacional. Los ojos se detectan automáticamente para a partir de estos realizar un conjunto de transformaciones afines de rotación, traslación y escala que conllevan a la normalización geométrica de las imágenes [Garea 06], de manera que todas las imágenes de rostros son representadas en una plantilla de 72×80 píxeles.

Después de que el rostro se detecta y se normaliza geoméricamente, se evalúa la calidad respecto a la iluminación. En dependencia de la respuesta del método de calidad, CCQ , se decide qué representación de la imagen se utiliza en la clasificación, si la original o las imágenes pre-procesadas con los métodos LDCT y PS. Cuando a cada representación de la imagen se le aplica el MLBP y los histogramas locales de múltiples resoluciones se proyectan en los subespacios regionales LDA, la medida de similitud global de la imagen se obtiene integrando las medidas obtenidas en cada región con una correlación normalizada, la cual ofrece siempre una respuesta en el intervalo $[-1; 1]$, donde 1 significa la mayor similitud entre las imágenes comparadas. Si se detecta que la imagen de entrada está afectada por la iluminación y se decide aplicar el clasificador múltiple que combina los dos métodos de pre-procesamiento, la similitud obtenida al comparar dos imágenes se extiende al intervalo $[-2; 2]$. Por esta razón, para homologar los valores de similitud obtenidos cuando las imágenes son pre-procesadas y cuando no lo son, se multiplica por dos el valor de similitud que se obtenga al utilizar el clasificador con las imágenes que no requieran pre-procesamiento.

Los pasos anteriores se insertan tanto en un proceso de verificación como en uno de identificación. En la verificación, el valor de similitud que se obtenga como salida se compara con un umbral determinado en una fase previa de entrenamiento para determinar si las

imágenes comparadas pertenecen o no a la misma persona. En el caso de la identificación, se realizan tantas comparaciones como imágenes de individuos existan en la colección y se determina que la identidad de la persona en la imagen de entrada es aquella para la cual se obtiene la mayor similitud en la comparación.

4.2 Evaluación del Algoritmo Propuesto

En este epígrafe se evalúa experimentalmente el algoritmo de reconocimiento de rostros propuesto en bases de datos de pruebas internacionales. Para la evaluación del algoritmo en la verificación se utiliza la base de datos XM2VTS con el protocolo Laussane [Messer 99] y para la evaluación del algoritmo en la identificación se utiliza la Yale B Extendida [Lee 05b]. Estas bases de datos han sido ampliamente utilizadas en la evaluación de algoritmos de reconocimiento de rostros para verificación e identificación, respectivamente, lo cual permitirá comparar el esquema propuesto para las dos posibles variantes, con los métodos reportados en la literatura científica actualizada. En ambos casos, primeramente se analizan las ventajas de utilizar el esquema propuesto en lugar de clasificar las imágenes con el proceso tradicional, es decir, pre-procesando o no todas las imágenes. Posteriormente, se comparan los resultados obtenidos con los resultados reportados para esas bases de datos por los algoritmos existentes en la literatura.

4.2.1 Verificación

Los experimentos de verificación se llevan a cabo en la base de datos XM2VTS con el protocolo Laussane que contempla los tres subconjuntos de imágenes: Evaluación, Prueba y Oscuro. En la Tabla 4.1 se muestran los por cientos de error total obtenidos en cada subconjunto. La primera fila de la tabla corresponde a la aplicación del algoritmo MLBP+LDA con las imágenes originales normalizadas sólo geoméricamente y la segunda a la combinación de clasificadores propuesta en el Capítulo 3 que utiliza los métodos de pre-procesamiento LDCT y PS. La última fila exhibe el por ciento de error total en cada uno de los subconjuntos de imágenes al utilizar el algoritmo descrito en el epígrafe anterior, donde se tiene en cuenta la calidad de las imágenes respecto a la iluminación para pre-procesarlas o no.

Como se observa, al tenerse en cuenta la evaluación de la calidad de las imágenes y pre-procesar solamente las que estén realmente afectadas por la iluminación, se obtiene el

Tabla 4.1: TER (%) obtenido en la verificación de las imágenes en la base de datos XM2VTS.

	Evaluación	Prueba	Oscuro
Imágenes Originales	1,90	1,16	13,7
Pre-procesamiento	1,79	1,17	2,87
Algoritmo Propuesto	1,68	1,15	2,84

menor error en la clasificación para los tres subconjuntos de imágenes de la base de datos XM2VTS, en los cuales las imágenes están sujetas a condiciones de iluminación diferentes.

En la Tabla 4.2 se compara el algoritmo propuesto con los resultados reportados en la literatura científica por los algoritmos que exhiben los mejores resultados en esta base de datos.

Tabla 4.2: Comparación con el TER de otros algoritmos en la base de datos XM2VTS.

	Evaluación	Prueba	Oscuro
LBP_ MAP [Marcel 07]	-	2,84	25,8
LBP_ AdaBoost [Marcel 07]	-	7,80	71,20
LBP_ LDA [Heusch 06]	-	9,12	18,22
LBP_ HMM [Heusch 06]	-	2,74	19,22
AS_ LDA [Messer 06]	6,50	9,76	25,24
AS_ HMM [Messer 06]	10,50	8,38	24,00
ICB06-Best¹ [Messer 06]	1,63	0,96	-
ICB06-Best² [Messer 06]	2,35	-	2,02
Algoritmo Propuesto	1,68	1,15	2,84

El error total de 2,84%, obtenido por el algoritmo propuesto en el conjunto Oscuro, que contiene las imágenes afectadas por las variaciones de iluminación, es muy cercano al menor error que se ha reportado en las pruebas de verificación utilizando este subconjunto de imágenes de la base de datos XM2VTS. Sin embargo, el algoritmo que muestra el mejor resultado, que fue el ganador en la última competencia realizada en esa base de datos en la Conferencia Internacional de Biometría ICB 2006, logra el error de 2,02% utilizando un entrenamiento específico para imágenes afectadas por la iluminación. Como puede observarse en la Tabla 4.2, el mismo algoritmo ganador de la competencia, utiliza entrenamientos diferentes para clasificar las imágenes afectadas y las que no presentan afectaciones de iluminación, sin controlar de manera automática este proceso, es decir,

¹Entrenado y probado con imágenes sin afectaciones de iluminación.

²Entrenado y probado con imágenes afectadas por la iluminación.

que necesita conocer de antemano las condiciones a las que va a enfrentarse el sistema. Al utilizar el modelo de aprendizaje de las condiciones desfavorables de iluminación con imágenes no afectadas disminuye su eficacia. El algoritmo que se propone en esta tesis logra obtener un mejor balance clasificando imágenes bajo diferentes condiciones de iluminación.

4.2.2 Identificación

La evaluación del algoritmo propuesto en la identificación se lleva a cabo en los subconjuntos de la base de datos Yale B Extendida. La cantidad de imágenes y las variaciones en los ángulos de incidencia de la luz en cada subconjunto, así como algunos ejemplos de imágenes pertenecientes a estos, se muestran en la Figura 4.2.

	Ángulos	Ejemplos de imágenes						
S1 (228)	5° - 10°							
S2 (456)	15° - 30°							
S3 (532)	35° - 50°							
S4 (456)	55° - 70°							
S5 (570)	75° - 110°							

Figura 4.2: Imágenes en base de datos Yale B Extendida.

La imagen de cada persona de la base de datos, que fue tomada con un ángulo de 0° de variación entre la luz incidente y la cámara, se utiliza para representar a los 38 individuos en la colección. Los por cientos de reconocimiento correcto obtenidos en la identificación de las imágenes en cada uno de los subconjuntos se pueden observar en la Tabla 4.3.

Tabla 4.3: Reconocimiento correcto (%) obtenido en la identificación de las imágenes en la base de datos Yale B Extendida.

	S1	S2	S3	S4	S5
Imágenes Originales	100	99,56	90,97	49,78	42,10
Pre-procesamiento	100	100	99,06	99,34	96,53
Algoritmo Propuesto	100	100	99,06	99,34	96,53

Al igual que en la verificación, la primera fila corresponde a la aplicación del algoritmo MLBP+LDA con las imágenes originales, la segunda a la combinación de los métodos de pre-procesamiento presentada en el capítulo anterior y la tercera a la clasificación con el algoritmo propuesto.

En este caso no es posible apreciar una mejora en los resultados de la clasificación con el algoritmo propuesto respecto a los del clasificador combinado de pre-procesamiento. Esto se debe a que la mayoría de las imágenes en la base de datos están afectadas por la iluminación y por lo tanto, necesitan ser pre-procesadas. Las imágenes del subconjunto S1 y algunas del subconjunto S2, sobre las que incide la luz con ángulos menores de 30° son las únicas que son clasificadas como buenas por el método de evaluación de la calidad y para estos dos subconjuntos, el método combinado de pre-procesamiento obtiene un 100% de clasificación correcta. La ventaja de utilizar el algoritmo propuesto en este caso, radica en mantener la tasa de reconocimiento del 100% en esos dos subconjuntos con una mayor eficiencia computacional, puesto que no se pre-procesan esas imágenes clasificadas como buenas. Para los otros tres subconjuntos el proceso y por tanto los resultados, son similares.

La base de datos Yale B Extendida es una extensión más reciente de la base de datos Yale B original [Georghiades 01], que contiene solo imágenes de 10 sujetos. Varios algoritmos novedosos han sido evaluados en la base de datos extendida. Entre ellos se destacan el método basado en subespacios creados a partir de la imagen del cociente que compensa la iluminación en el espacio de colores HSV (QI₋ HSV), el método del cociente Stiefel canónico (CSQ) y la clasificación basada en representaciones esparcidas (SRC). Algunos han expuesto sus resultados con la misma división en subconjuntos de la base de datos presentada, otros reportan solamente el valor promedio en la comparación de todas las imágenes con el conjunto de entrenamiento. En la Tabla 4.4 se comparan estos algoritmos con el propuesto.

Para una comparación más completa se seleccionaron las imágenes de los 10 individuos en la base de datos original y se determinó el error de clasificación correcta para estas

Tabla 4.4: Comparación con las tasas de reconocimiento correcto (%) de otros algoritmos en la base de datos Yale B Extendida.

	S1	S2	S3	S4	S5	Promedio
CGHP [Lui 09]	-	-	54,28	32,63	15,65	-
Geodesic [Lui 09]	-	-	78,60	63,71	29,30	-
CSQ [Lui 09]	-	-	99,78	97,88	51,78	-
QL HSV [Wang 08]	100	100	93,75	90,63	84,37	93,75
LTP [Tan 07]	100	100	98,0	99,2	94,1	-
NN [Wright 09]	-	-	-	-	-	90,7
NS [Wright 09]	-	-	-	-	-	94,1
SVM [Wright 09]	-	-	-	-	-	97,7
SRC [Wright 09]	-	-	-	-	-	98,1
Algoritmo Propuesto	100	100	99,06	99,34	96,53	98,98

Tabla 4.5: Comparación con el error de clasificación correcta (%) obtenido por otros algoritmos en la base de datos Yale B original.

	S3	S4	S5
Cociente de iluminación [Zhao 03a]	3,3	18,6	-
Subespacios lineales [Belhumeur 98]	0	15,0	-
Conos de iluminación [Georghiadis 01]	0	8,6	-
Armónicos esféricos [Zhang 06]	0,3	3,1	-
9 Puntos de Luz [Lee 05b]	0	2,8	-
Ángulo del gradiente [Chen 00]	0	1,4	-
Fisher en relieve [Kouzani 00]	0,9	12,1	27,3
Re-iluminación del cociente [Shan 03]	0	9,4	17,5
QL HSV [Wang 08]	0	8,3	15,7
LogDCT PCA [Chen 06]	0	0,18	1,71
Algoritmo Propuesto	0	0	0,51

imágenes utilizando el algoritmo propuesto. Esta es la forma en la que han presentado los resultados la mayoría de los algoritmos existentes en la literatura propuestos para enfrentar el problema de la iluminación en el reconocimiento de rostros. En la Tabla 4.5 se muestra la comparación de los errores obtenidos por cada algoritmo con las imágenes de estos 10 individuos en los subconjuntos S3, S4 y S5. En los subconjuntos S1 y S2 todos obtienen 0% de error y para algunos de los algoritmos, como el del ángulo del gradiente [Chen 00], a pesar de sus buenos resultados en los otros subconjuntos no se reportan los errores obtenidos en el subconjunto S5.

De las comparaciones mostradas en las Tablas 4.4 y 4.5 se puede concluir que entre los algoritmos hasta ahora existentes, el que se propone en esta tesis exhibe la mayor eficacia en la clasificación de las imágenes de la base de datos Yale B.

4.3 Pruebas con imágenes en condiciones reales

En este epígrafe se muestran los resultados obtenidos utilizando el algoritmo propuesto en el reconocimiento de imágenes de rostros en condiciones reales. Con este objetivo se obtuvieron imágenes de 80 individuos en 6 condiciones ambientales diferentes.

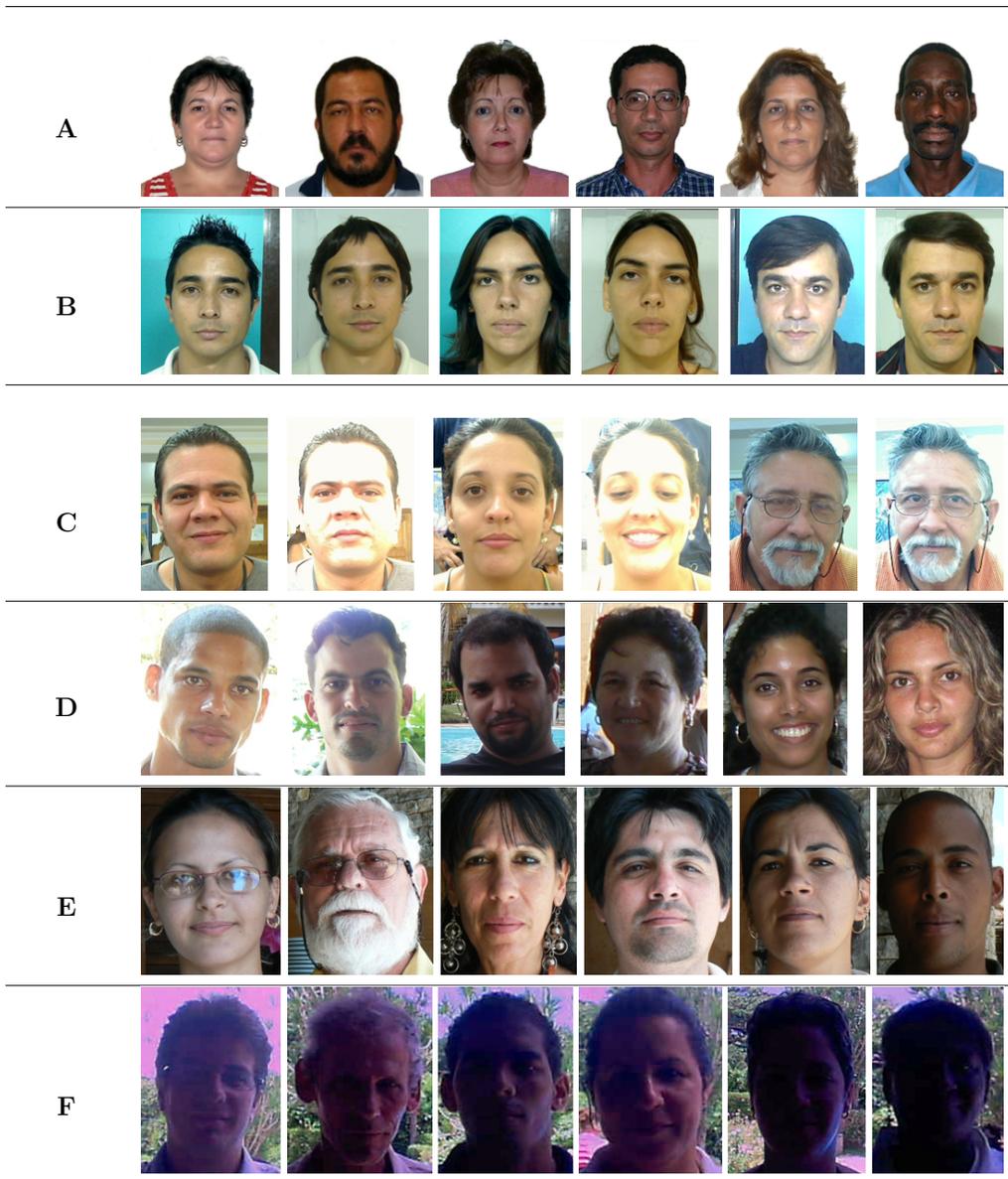
El primer grupo de imágenes (A) contiene una foto de cada uno de los 80 individuos tomadas en condiciones controladas de iluminación, posición y expresión del rostro con una cámara fotográfica de alta resolución. Las fotos en este grupo cumplen los requisitos establecidos para la captura de imágenes de rostros para documentos de identificación [INCITS 06] y se utilizarán para formar la colección o lista de individuos que van a ser reconocidos.

El segundo grupo (B) contiene imágenes de 10 de los individuos, igualmente en condiciones controladas, aunque no iguales entre ellas, ya que fueron tomadas en 3 sesiones de días diferentes con un intervalo de más de tres meses entre ellas. Estas fotos fueron capturadas con la cámara web Orbit AF de Logitech, que captura a una resolución máxima en video de 960x720 y ajusta automáticamente la saturación, la exposición y el balance de color blanco en la escena. Este subconjunto de imágenes fue capturado en las cabinas de inmigración del aeropuerto internacional "José Martí" de Ciudad de la Habana y tiene como objetivo comprobar el funcionamiento del algoritmo en las aplicaciones de cheques migratorios, en las que aunque se tratan de garantizar los requisitos mínimos en las condiciones ambientales de iluminación, así como en la posición y expresión de las personas que son fotografiadas, no obstante, éstas varían entre una foto y otra y, además, la posición de las lámparas, las luces del día y las luces situadas en los salones de espera, inciden de manera variada en los rostros a los que se les toma la fotografía.

Las imágenes del tercer subconjunto (C) también se capturaron con la cámara web Orbit AF en ambientes interiores. En este caso no se controlaron las condiciones de iluminación, unas imágenes están sobre-iluminadas y otras fueron tomadas con una iluminación no frontal. Se capturaron de esta forma dos fotos de 25 de los individuos en la colección. Los cuales además presentaron pequeñas variaciones en la expresión y la pose del rostro.

El cuarto subconjunto (D) se formó con imágenes digitales tomadas a 25 de los individuos

Tabla 4.6: Ejemplos de imágenes reales en cada uno de los subconjuntos con distintas condiciones ambientales.



en distintos momentos, ambientes y condiciones de iluminación. Se adquirieron dos fotos de cada una de estas 25 personas tomadas en situaciones de la vida real, expresiones y posiciones diferentes, aunque la mayoría son frontales o casi frontales. Algunas fotos son en interiores, pero la mayoría son en exteriores y otras incluso en la oscuridad de la noche.

Para formar el quinto subconjunto (E) se tomaron fotografías de 65 de los individuos en ambientes exteriores. Estas imágenes se tomaron con una cámara digital de alta resolución y con la luz del día incidiendo indistintamente en cada uno de los rostros, algunas de forma homogénea, pero otras no. También en este caso las personas muestran pequeñas variaciones en la expresión y la pose.

Para el último grupo (F) se capturaron con una cámara web también de Logitech, pero de una resolución máxima de 640x480 y sin ningún balance automático, 65 fotos de los mismos individuos del subconjunto anterior. En este caso las condiciones de iluminación son totalmente adversas, al igual que la calidad de las fotografías tomadas. Las imágenes en este subconjunto simulan una aplicación de video-vigilancia, en las que normalmente la resolución de los cuadros del video es poca y el ambiente puede ser totalmente adverso.

En la Tabla 4.6 se muestran ejemplos de las imágenes pertenecientes a cada uno de los subconjuntos descritos anteriormente. El objetivo con estas imágenes es evaluar el algoritmo propuesto en condiciones reales en las que las variaciones en las condiciones de iluminación pueden ser muy diversas. Para esto se llevan a cabo 5 experimentos de identificación, en los que cada una de las imágenes de los subconjuntos B, C, D, E y F se comparan con las imágenes en la colección A y se determina el por ciento de reconocimiento correcto.

En la Tabla 4.7 se muestran los resultados de clasificación correcta en la primera posición obtenidos en cada caso por el algoritmo propuesto, comparados con los resultados obtenidos clasificando las imágenes sin ningún pre-procesamiento y pre-procesándolas todas con el método de combinación. Se muestran además en la Figura 4.3, las curvas de reconocimiento correcto acumulado en función de la posición en que se encuentra la respuesta correcta en la lista ordenada de similares.

Tabla 4.7: Reconocimiento correcto (%) obtenido en la identificación de imágenes reales.

	B	C	D	E	F
Imágenes Originales	84,84	16,00	54,00	28,35	02,94
Pre-procesamiento	90,90	42,00	52,00	55,22	08,82
Algoritmo Propuesto	93,93	48,00	58,00	62,68	10,29

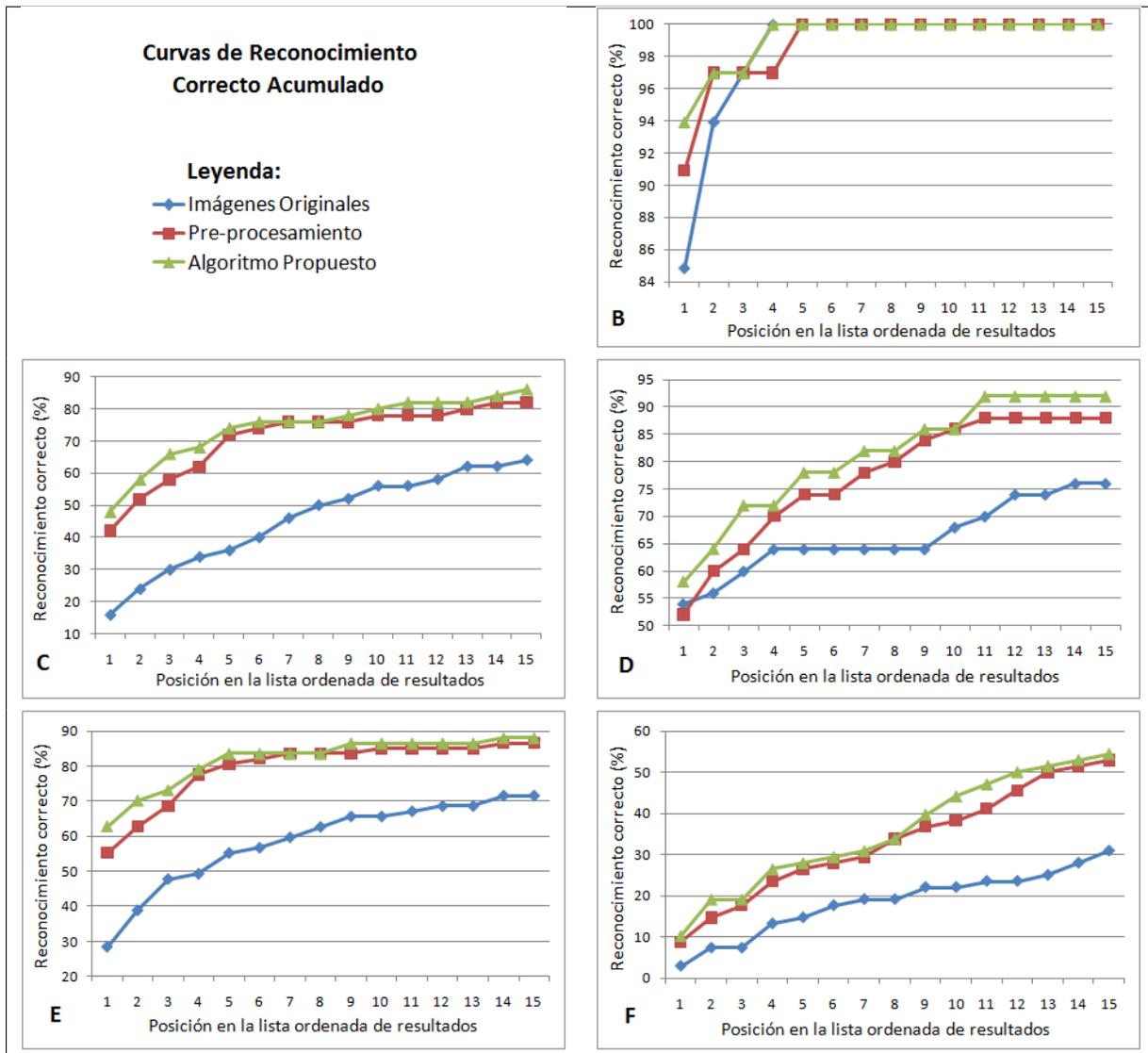


Figura 4.3: Reconocimiento correcto (%) obtenido en la identificación de imágenes reales.

En todos los grupos de imágenes, en las distintas condiciones de captura, se observa la superioridad en la eficacia del algoritmo propuesto sobre la utilización del método de pre-procesamiento en todas las imágenes, apreciándose además una diferencia significativa respecto al uso del clasificador sobre las imágenes originales sin ningún pre-procesamiento.

Con el objetivo de analizar si el algoritmo propuesto para ser utilizado en las soluciones propias, está acorde y compite con las soluciones existentes en el mercado, se compara con la versión más reciente del sistema VeriLook de la empresa *Neurotechnology* [Neurotechnology 10], una de las de mayor prestigio a nivel internacional en el área de la Biometría, al evaluarse con este último los mismos subconjuntos de imágenes de rostros

en condiciones reales.

En la Tabla 4.8 se muestra el rendimiento del sistema VeriLook y el del algoritmo propuesto en esta tesis. Como puede observarse, en las mismas condiciones ambientales reales, el algoritmo propuesto mejora de manera significativa los resultados obtenidos en la identificación por el sistema VeriLook, uno de los que mejores resultados obtiene en las competencias de sistemas comerciales [Phillips 07]. Esto permite asegurar que un sistema de reconocimiento de rostros en el que se utilice el algoritmo propuesto sería competitivo en el mercado internacional.

Tabla 4.8: Reconocimiento correcto (%) obtenido en la identificación de imágenes reales.

	B	C	D	E	F
VeriLook [Neurotechnology 10]	63,64	30,00	32,00	22,06	0,0
Algoritmo Propuesto	93,93	48,00	58,00	62,68	10,29

4.4 Algunas consideraciones computacionales

Los objetivos de esta tesis han estado encaminados a lograr una mayor eficacia que la de los algoritmos existentes en la literatura para el reconocimiento de rostros en condiciones variables de iluminación. No obstante la meta final es la aplicación del algoritmo que se propone en sistemas reales.

Las evaluaciones experimentales llevadas a cabo en la investigación, fueron realizadas en Matlab 2008b. Sin embargo, al concluir este trabajo, el algoritmo descrito se encuentra implementado en C# y se ha comenzado su implementación en C++. A pesar de que el esquema propuesto pueda parecer a simple vista costoso computacionalmente, los métodos que se utilizan en cada una de las fases son muy eficientes. Para comprobar esto se realizaron algunas pruebas utilizando la implementación en C# del algoritmo en una computadora personal de 2.00 GB de memoria RAM y un procesador Intel Core 2 Duo a 1.86 GHz.

El paso de detección del rostro y los ojos, seguido de la normalización geométrica en una imagen de 640×480 píxeles tarda menos de 15 milisegundos, utilizando la implementación de la OpenCV [Bradski 00] de los detectores. El método de evaluación de la calidad implementado en C#, una vez que la red neuronal ha sido entrenada y los parámetros necesarios estimados, tarda aproximadamente 55 milisegundos en determinar si la imagen está afectada o no por la iluminación. Los métodos de pre-procesamiento LDCT y PS son

ambos muy eficientes, ya que utilizan filtros que han sido bien estudiados en el procesamiento digital de imágenes, existiendo implementaciones muy eficientes de ellos. En 100 imágenes de rostros normalizadas geométricamente a un tamaño de 72×80 toman ambos un promedio de 12 milisegundos en ejecutarse.

Para comprobar la eficiencia computacional en la clasificación utilizando la implementación en C# del algoritmo propuesto se cargaron en memoria los rasgos de 1000 imágenes representativas de distintos sujetos en una colección. Los rasgos de cada imagen, conformados por la representación en los subespacios regionales LDA de las imágenes originales y de las imágenes pre-procesadas con LDCT y PS, necesitan 16Kb de memoria. Al tener la información de 1000 imágenes cargadas en memoria, el reconocimiento de una imagen de rostro ya normalizada geométricamente comparada contra la colección completa tarda unos 810 milisegundos.

Es evidente que es necesario establecer un compromiso entre la capacidad de memoria RAM de la que se dispone y la velocidad de procesamiento, en dependencia de la aplicación específica en la que se vaya a utilizar el algoritmo, teniendo en cuenta la cantidad de imágenes en la colección y el tiempo de respuesta mínimo que se requiera. No obstante la eficiencia computacional del método puede todavía mejorarse con la implementación en C++ y versiones paralelas del algoritmo que aprovechen las capacidades de las máquinas modernas de múltiples procesadores.

4.5 Conclusiones parciales

El algoritmo propuesto, que aplica el método de evaluación de la calidad propuesto en el Capítulo 2 para determinar si una imagen de rostro está afectada por la iluminación y si lo está, utiliza entonces la combinación de métodos de pre-procesamiento presentada en el Capítulo 3 a partir del método LDCT propuesto, demuestra experimentalmente ser más eficaz que los algoritmos existentes en la verificación e identificación de imágenes de rostros.

En los experimentos de verificación, llevados a cabo en la base de datos XM2VTS, el algoritmo que se propone, con por cientos de errores de 1,15 % y 2,84 % en los subconjuntos Prueba y Oscuro respectivamente, muy similares a los del algoritmo que mejores resultados ha obtenido en esa base de datos, muestra ser superior a este, el cual necesita conocer de antemano las afectaciones de iluminación a las que va a estar sometido el algoritmo para, en base a eso, utilizar un modelo de aprendizaje específico, basado en un conjunto

de entrenamiento de imágenes en esas condiciones.

En los experimentos de identificación, realizados en la base de datos Yale B, el algoritmo exhibe los mejores resultados que se han reportado en el reconocimiento correcto de los individuos en la primera posición de la lista de candidatos de esta base de datos, ya sea en su versión original o en la extendida.

Los resultados obtenidos en los experimentos realizados con imágenes en cinco entornos reales diferentes, demuestran que en estos entornos con condiciones de iluminación variable, es más eficaz utilizar el algoritmo que se propone en lugar de pre-procesar todas las imágenes o no pre-procesar ninguna.

Se demuestra además el buen funcionamiento del algoritmo propuesto al ser comparado con uno de los sistemas comerciales de mayor prestigio a nivel internacional en el reconocimiento de las imágenes de rostros en condiciones reales y mostrar un rendimiento significativamente superior.

Finalmente, al tener en cuenta algunas consideraciones computacionales en la implementación del algoritmo para su uso en sistemas reales, se muestra que la propuesta además de ser eficaz es factible de utilizar en cuanto al costo computacional que implica para aplicaciones reales.

Conclusiones y Recomendaciones

Conclusiones

Se desarrolló un nuevo algoritmo de reconocimiento de rostros, el cual determina en una primera etapa, si la imagen está afectada por las variaciones de iluminación y sólo en este caso, se aplica el pre-procesamiento propuesto para enfrentar estas afectaciones. El algoritmo desarrollado demostró experimentalmente ser más eficaz que los algoritmos existentes para la verificación e identificación de imágenes de rostros en condiciones variables de iluminación, utilizando solamente una o muy pocas imágenes de entrenamiento por persona.

En particular, se desarrolló un método para evaluar la calidad de las imágenes de rostros respecto a la iluminación, que ajusta una malla triangular diseñada para analizar el comportamiento local de la iluminación en las distintas regiones del rostro. De esta forma se logra determinar si las imágenes se encuentran afectadas o no por la iluminación, con una eficacia de un 96.77 %.

Se desarrolló un método de pre-procesamiento: LDCT, que utilizando la Transformada Discreta del Coseno de manera local, estima y sustrae de la imagen de rostro el contenido de baja frecuencia afectado por la iluminación, sin comprometer la información discriminativa, lo cual se demuestra experimentalmente al obtener resultados similares a los de los mejores métodos de la literatura.

Se seleccionó la representación MLBP con la clasificación basada en LDA para reconocer los rostros en las imágenes pre-procesadas con el método propuesto (LDCT), y con el método PS de la literatura. Se demostró que ambos métodos se complementan y aportan diversidad, a pesar de obtener por cientos de errores muy similares, por lo que se propuso un esquema de combinación de clasificadores, con el cual se obtiene mayor eficacia que la de los métodos de pre-procesamiento existentes para imágenes con condiciones variadas de iluminación.

La implementación del algoritmo de reconocimiento de rostros desarrollado, que se basa

solamente en una imagen de entrenamiento por persona y pre-procesa únicamente las imágenes afectadas por la iluminación, demostró que puede ser ejecutado en un tiempo computacional permisible para ser aplicado en sistemas reales de reconocimiento de rostros.

Recomendaciones

Con los resultados obtenidos no se concluye el trabajo en esta temática. Del estudio realizado se derivan algunos trabajos futuros a modo de recomendaciones:

1. Analizar con más profundidad los resultados del algoritmo propuesto cuando en las imágenes de rostros hay presentes otras afectaciones además de las de iluminación y desarrollar nuevas variantes del algoritmo en función de esto.
2. Investigar otros métodos de clasificación que puedan ser utilizados en las diferentes etapas del algoritmo que se propone.
3. Realizar un análisis del algoritmo propuesto para obtener la implementación más eficiente que se pueda alcanzar y una posible versión paralela.

Referencias Bibliográficas

- [Abbas 09] A. Abbas, M.I. Khalil, S.AbdelHay, and H.M.A. Fahmy. “Illumination invariant face recognition in logarithm discrete cosine transform domain”. In *Internacional Confernece on Image Processing*, 4157–4160, 2009.
- [Adini 97] Y. Adini, Y. Moses, and Sh. Ullman. “Face recognition: the problem of compensating for changes in illumination direction”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:721–732, 1997.
- [Ahmed 74] N. Ahmed, T.Natarajan, and K.R. Rao. “On image processing and a discrete cosine transform”. In *IEEE Transactions on Computers* (C-23), 90–93, 1974.
- [Ahonen 04] T. Ahonen, A. Hadid, and M. Pietikäinen. “Face recognition with Local Binary Patterns”. *Lecture Notes in Computer Science: Computer Vision - ECCV*, 469–481, 2004.
- [Akarun 05] L. Akarun, B. Gökberk, and A.A. Salah. “3D face recognition for biometric applications”. In *European Signal Processing Conference*, 2005.
- [Alvarez 06] E.M. Alvarez, F. Silva, E. Garea, H. Méndez-Vázquez, and M. Herrera. “A fast method for localization of local illumination variations and photometric normalization in face images”. *Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications*, 4225:48–57, 2006.
- [Arandjelovic 06] O. Arandjelovic and R. Cipolla. “A new look at filtering techniques for illumination invariance in automatic face recognition”. In *FGR*

- '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 449–454. IEEE Computer Society, 2006.
- [Belhumeur 97] P.N. Belhumeur, Jo.P. Hespanha, and D.J. Kriegman. “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:711–720, 1997.
- [Belhumeur 98] P.N. Belhumeur and D.J. Kriegman. “What is the set of images of an object under all possible illumination conditions?”. *International Journal on Computer Vision*, 28(3):245–260, 1998.
- [Beumer 06] G. M. Beumer, Q. Tao, A. M. Bazen, and R.N.J. Veldhuis. “A landmark paper in face recognition”. In *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 73–78. IEEE Computer Society, 2006.
- [Bigun 04] J. Bigun, T. Bigun, and K.Nilsson. “Recognition by symmetry derivatives and the generalized structure tensor”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1590–1605, 2004.
- [Bishop 95] Christopher M. Bishop. “Neural networks for pattern recognition”. Oxford University Press, Inc., New York, NY, USA, 1995.
- [Bradski 00] G. Bradski. “The OpenCV Library”. *Dr. Dobb's Journal of Software Tools*, 2000.
- [Brunelli 93] R. Brunelli and T. Poggio. “Face recognition: features versus templates”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.
- [Chan 04] T.F. Chan and S. Esedoglu. “Aspects of total variation regularized L1 function approximation”. Tech. Report 04-07, CLA CAM, 2004.
- [Chan 07] C.H. Chan, J. Kittler, and K. Messer. “Multi-scale local binary pattern histograms for face recognition”. *Lecture Notes in Computer Science: Advances in Biometrics*, 4642/2007:809 – 818, 2007.

- [Chen 00] H.F. Chen, P.N. Belhumeur, and D.W. Jacobs. “In search of illumination invariants”. In *IEEE Conference on Computer Vision and Pattern Recognition*, 254–261, 2000.
- [Chen 04a] S. Chen, J. Liu, and DeLaTorre05 Z.H. Zhou. “Making FLDA applicable to face recognition with one sample per person”. *Pattern Recognition*, 37(7):1553–1555, 2004.
- [Chen 04b] S. Chen, D. Zhang, and Z.H. Zhou. “Enhanced (PC)² A for face recognition with one training image per person”. *Pattern Recognition Letters*, 25(10):1173–1181, 2004.
- [Chen 05] T. Chen, W. Yin, X.S. Zhou, D. Comaniciu, and T.S. Huang. “Illumination normalization for face recognition and uneven background correction using total variation based image models”. In *CVPR’05: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 532–539, 2005.
- [Chen 06] W. Chen, Meng J. Er, and S. Wu. “Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain”. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 36(2):458–466, 2006.
- [Cootes 01] T.F. Cootes, G.J. Edwards, and C.J. Taylor. “Active appearance models”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:681–685, 2001.
- [DCT 01] “The transform and data compression handbook”. CRC Press LLC, Boca Raton, 2001.
- [DeLaTorre 05] F. DeLaTorre, R. Gross, S. Baker, and B.V. Kumar. “Representational oriented component analysis (ROCA) for face recognition with one sample image per training class”. In *CVPR ’05: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2)*, 266–273. IEEE Computer Society, 2005.
- [Du 05] B. Du, S. Shan, L. Qing, and W. Gao. “Empirical comparisons of several preprocessing methods for illumination insensitive face

- recognition”. In *ICASSP '05: IEEE International Conference on Acoustics, Speech, and Signal Processing* (2), 981–984, 2005.
- [Ebrahimpour 01] H. Ebrahimpour, V. Chandran, and S. Sridharan. “Robustness to expression variations in fractal-based face recognition”. In *Sixth International Symposium on Signal Processing and its Applications*, 359–362. IEEE Computer Society, 2001.
- [Eickeler 00] S. Eickeler, S. Müller, and G. Rigoll. “Recognition of JPEG compressed face images based on statistical methods”. *Image and Vision Computing*, 18(4):279 – 287, 2000.
- [Ekenel 05] H. K. Ekenel and R. Stiefelhagen. “Local appearance based face recognition using discrete cosine transform”. In *13th European Signal Processing Conference (EUSIPCO)*, 27–30, 2005.
- [Ekenel 09] H. K. Ekenel. *A Robust Face Recognition Algorithm for Real-World Applications*. PhD thesis, University of Karlsruhe, Germany, 2009. Director - Prof. Dr. A. Waibel.
- [Er 05] M.J. Er, W. Chen, and Sh. Wu. “High-speed face recognition based on discrete cosine transform and rbf neural networks”. *IEEE Transactions on Neural Network*, 16(3):679–91, 2005.
- [Ferrara 08] M. Ferrara, A. Franco, and D. Maltoni. “Evaluating systems assessing face-image compliance with icao/iso standards”. *Biometrics and Identity Management: First European Workshop, BIOID*, 191–199, 2008.
- [Finlayson 01] G.D. Finlayson and S.D. Hordley. “Color constancy at a pixel”. *Journal of the Optical Society of America A*, 18(2):253–264, 2001.
- [Freund 95] Y. Freund and R. E. Schapire. “A decision-theoretic generalization of on-line learning and an application to boosting”. In *EuroCOLT '95: Second European Conference on Computational Learning Theory*, 23–37. Springer-Verlag, 1995.
- [Fronthaler 06] H. Fronthaler, K. Kollreider, and J. Bigun. “Automatic image quality assessment with application in biometrics”. In *CVPRW '06*:

- Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, 30. IEEE Computer Society, 2006.
- [Gao 05] X. Gao, T. Wang, and J. Li. “A content-based image quality metric”. *Lecture Notes in Computer Science: Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 3642:231–240, 2005.
- [Gao 07] X. Gao, S.Z. Li, R. Liu, and P. Zhang. “Standardization of face image sample quality”. In *International Conference on Biometrics*, 242–251. Springer, 2007.
- [Garea 06] E. Garea, J. Kittler, K. Messer, and H. Méndez-Vázquez. “A comparative study of face representations in the frequency domain”. *Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications*, 4225:99–108, 2006.
- [Georghiades 01] A.S. Georghiades, P.N. Belhumeur, and D. 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*, 23(6):643–660, 2001.
- [Gokberk 09] B. Gokberk, A.A. Salah, N. Alyüz, and L. Akarun. “3D face recognition: Technology and applications”. *Handbook of Remote Biometrics for Surveillance and Security*, 10.1007:217–246, 2009.
- [Goldstein 71] A. J. Goldstein, L. D. Harmon, and A. B. Lesk. “Identification of human faces”. *Proceedings of the IEEE*, 59(5):748–760, 1971.
- [Gonzalez 06] R.C. Gonzalez and R.E. Woods. “Digital image processing (3rd edition)”. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
- [Gottumukkal 04] R. Gottumukkal and V.K. Asari. “An improved face recognition technique based on modular pca approach”. *Pattern Recognition Letters*, 25(4):429–436, 2004.
- [Gross 03] R. Gross and V. Brajovic. “An image preprocessing algorithm for illumination invariant face recognition”. *Lecture Notes in Computer Science: Audio- and Video-Based Biometric Person Authentication*, 2688:10–18, 2003.

- [Gutta 04] S. Gutta and H. Wechsler. “Face recognition using asymmetric faces”. In *First International Conference on Biometric Authentication*, 162–168. Springer-Verlag, 2004.
- [Hafed 01] Z.M. Hafed and M.D. Levine. “Face recognition using the discrete cosine transform”. *International Journal of Computer Vision*, 43(3):167–188, 2001.
- [He 05] X. He, S. Yan, Y.o Hu, P.Niyogi, and H.J. Zhang. “Face recognition using laplacianfaces”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:328–340, 2005.
- [Heisele 03] B. Heisele, P. Ho, J. Wu, and T. Poggio. “Face recognition: component-based versus global approaches”. *Computer Vision and Image Understanding: Special Issue on Face Recognition*, 91(1-2):6–21, 2003.
- [Heusch 06] G. Heusch, Y. Rodriguez, and S. Marcel. “Local binary patterns as an image preprocessing for face authentication”. In *FGR’06: 7th International Conference on Automatic Face and Gesture Recognition*, 9–14, 2006.
- [Huang 03] J. Huang, P.C. Yuen, W.Sh. Chen, and J.H. Lai. “Component-based LDA method for face recognition with one training sample”. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 0:120, 2003.
- [INCITS 06] INCITS. “Face recognition format data interchange, version 2.0”. InterNational Committee for Information Technology Standards (INCITS) Secretariat. Information Technology Industry Council, 2006.
- [Jain 89] A.K. Jain. “Fundamentals of digital image processing”. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.
- [Jain 07] A.K. Jain, P. Flynn, and A.A. Ross. “Handbook of biometrics”. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

- [Jin 04] H. Jin, Q. Liu, H. Lu, and X. Tong. “Face detection using improved LBP under bayesian framework”. In *International Conference on Image and Graphics*, 306–309. IEEE Computer Society, 2004.
- [Jobson 97] D. J. Jobson, Z. Rahman, and G. A. Woodell. “A multiscale retinex for bridging the gap between color images and the human observation of scenes”. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997.
- [Kepenekci 02] B. Kepenekci, F.B. Tek, and G.B. Akar. “Occluded face recognition based on gabor wavelets”. In *ICIP02: International Conference on Image Processing (I)*, 293–296, 2002.
- [Kittler 98] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. “On combining classifiers”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.
- [Kittler 99] J. Kittler. “Face authentication using client specific fisherfaces”. In *The Statistics of Directions, Shapes and Images*, 63–66, 1999.
- [Kouzani 00] A.Z. Kouzani, F. He, and K. Sammut. “Towards invariant face recognition”. *Information Sciences*, 123(1-2):75–101, 2000.
- [Kryszczuk 05] K. Kryszczuk and A. Drygajlo. “Gradient-based image segmentation for face recognition robust to directional illumination”. In *International Conference on Visual Communications and Image Processing*, 803–813, 2005.
- [Kuncheva 01] L. Kuncheva and Ch.J. Whitaker. “Feature subsets for classifier combination: An enumerative experiment”. In *MCS '01: Second International Workshop on Multiple Classifier Systems*, 228–237. Springer-Verlag, 2001.
- [Kusuma 03] T.M. Kusuma and H.-J. Zepernick. “A reduced-reference perceptual quality metric for in-service image quality assessment”. In *Joint First Workshop on Mobile Future and Symposium on Trends in Communications*, 71–74. IEEE Computer Society, 2003.
- [Land 71] E.H. Land and J. Mccann. “Lightness and retinex theory”. *Journal of the Optical Society of America*, 1–11, 1971.

- [Land 77] E.H. Land. “The retinex theory of color vision”. *Scientific American Magazine*, 237:108–128, 1977.
- [Le 04] H.S. Le and H. Li. “Recognizing frontal face images using hidden markov models with one training image per person”. In *ICPR '04: 17th International Conference on Pattern Recognition* (1), 318–321. IEEE Computer Society, 2004.
- [Lee 05a] H.J. Lee, H.J. Kim, and W.Y. Kim. “Face recognition using component-based DCT/LDA”. In *Proceeding of IWAIT*, 2005.
- [Lee 05b] K.Ch. Lee, J. Ho, and D.J. Kriegman. “Acquiring linear subspaces for face recognition under variable lighting”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.
- [Li 99] S.Z. Li and J. Lu. “Face recognition using the nearest feature line method”. *IEEE Transactions on Neural Networks*, 10:439–443, 1999.
- [Li 05] S.Z. Li and A.K. Jain. “Handbook of face recognition”. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [Li 07] S.Z. Li, R. Chu, Sh. Liao, and L. Zhang. “Illumination invariant face recognition using near-infrared images”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):627–639, 2007.
- [Liau 10] H.F. Liau and D. Isa. “New illumination compensation method for face recognition”. *International Journal of Computer and Network Security*, 2(3):5–12, 2010.
- [Liu 00] Ch. Liu and H. Wechsler. “Evolutionary pursuit and its application to face recognition”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:570–582, 2000.
- [Liu 05] D.H. Liu, K.M. Lam, and L.S. Shen. “Illumination invariant face recognition”. *Pattern Recognition*, 38(10):1705 – 1716, 2005.
- [Loog 06] M. Loog and D. de Ridder. “Local discriminant analysis”. *Pattern Recognition, International Conference on*, 3:328–331, 2006.

- [López 99] A.M. López, 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, 1999.
- [López 00] A.M. López, D. Lloret, J. Serrat, and J.J. Villanueva. “Multilocal creaseness based on the level-set extrinsic curvature”. *Computer Vision and Image Understanding*, 77(9):111–144, 2000.
- [Lui 09] Y.M. Lui, J.R. Beveridge, and M. Kirby. “Canonical stiefel quotient and its application to generic face recognition in illumination spaces”. In *BTAS09: Third IEEE International Conference on Biometrics Theory, Applications and Systems*, 1–8, 2009.
- [Marcel 07] S. Marcel, Y. Rodriguez, and G. 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. IDIAP-RR 06-34.
- [Martínez 98] A. Martínez and R. Benavente. “The ar face database”, Jun 1998. Cites in Scholar Google: <http://scholar.google.com/scholar?hl=en&lr=&client=firefox-a&cites=1504264687621469812>.
- [Martinez 01] A.M. Martinez and A.C. Kak. “PCA versus LDA”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:228–233, 2001.
- [Martínez 02] A.M. Martínez. “Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.
- [MathWorks 09] MathWorks. “MATLAB: The language of technical computing. Version 7.8.0.347 (R2009a)”. The MathWorks (www.mathworks.com). 3 Apple Hill Drive Natick, MA 01760-2098, Massachusetts, U.S.A, 2009.
- [Messer 99] K. Messer, J. Matas, J. Kittler, and K. Jonsson. “XM2VTSDB: The extended M2VTS database”. In *Second International Conference*

- on Audio and Video-based Biometric Person Authentication*, 72–77, 1999.
- [Messer 06] K. Messer, J. Kittler, J. Short, G. Heusch, F. Cardinaux, S. Marcel, Y. Rodriguez, Sh. Shan, Y. Su, W. Gaod, and X. Chen. “Performance characterisation of face recognition algorithms and their sensitivity to severe illumination changes”. In *International Conference on Biometrics (ICB)*, 1–11, 2006.
- [Méndez-Vázquez 08a] H. Méndez-Vázquez and E. García. “Estado actual de los métodos de reconocimiento automático de rostros basados en la apariencia local”. Tech. Report RT_006, Serie Azul, CENATAV, 2008.
- [Méndez-Vázquez 08b] H. Méndez-Vázquez, E. Garcia, and Y. Condes. “A new combination of local appearance based methods for face recognition under varying lighting conditions”. *Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications*, 5197:535–542, 2008.
- [Méndez-Vázquez 08c] H. Méndez-Vázquez, E.B. García, and Y. Condes. “A new image division for LBP method to improve face recognition under varying lighting conditions”. In *19th International Conference on Pattern Recognition (ICPR)*, 1–4. IEEE Computer Society, 2008.
- [Méndez-Vázquez 09] H. Méndez-Vázquez, C.S. Martín, J. Kittler, Y. Plasencia, and E. García. “Face recognition with LWIR Imagery using Local Binary Patterns”. *Lecture Notes in Computer Science: Advances in Biometrics*, 5558:327–336, 2009.
- [Méndez-Vázquez 10a] H. Méndez-Vázquez and E. García. “A new methodology for face recognition under varying illumination conditions”. Tech. Report RT_026, Serie Azul, CENATAV, 2010.
- [Méndez-Vázquez 10b] H. Méndez-Vázquez, J. Kittler, C.H. Chan, and E. García. “On combining local DCT with preprocessing sequence for face recognition under varying lighting conditions”. *Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications*, 6419:410–417, 2010.

- [Méndez-Vázquez 10c] H. Méndez-Vázquez, D. Rizo, and E.B. García. “An illumination quality measure for face recognition”. In *20th International Conference on Pattern Recognition (ICPR)*, 1–4. IEEE Computer Society, 2010.
- [Moghaddam 97] B. Moghaddam and A. Pentland. “Probabilistic visual learning for object representation”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:696–710, 1997.
- [Moses 92] Y. Moses and S. Ullman. “Limitations of non model-based recognition schemes”. In *2nd European Conference on Computer Vision*, 820–828. Springer Verlag, 1992.
- [Nefian 99] A.V. Nefian. *A hidden markov model-based approach for face detection and recognition*. PhD thesis, Georgia Institute of Technology, Atlanta, USA, 1999. Director - Dr. Hayes Monson.
- [Neurotechnology 10] Neurotechnology. “Verilook SDK: Face identification for pc or web solutions”. NEUROTECHNOLOGY (www.neurotechnology.com). Zamenhofo str. 5, Vilnius 06332, Lithuania, 2010.
- [Niu 05] Z.H. Niu, S. Shan, X. Chen, B. Ma, and W. Gao. “Enhance ASMs based on adaboost-based salient landmarks localization and confidence-constraint shape modeling”. In *IWBRS*, 9–14, 2005.
- [Ojala 96] T. Ojala, M. Pietikäinen, and D. Harwood. “A comparative study of texture measures with classification based on feature distributions”. *Pattern recognition*, 29(1):51 – 59, 1996.
- [Ojala 02] T. Ojala, M. Pietikäinen, and T. Mäenpää. “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [Petrou 99] M. Petrou and P. Bosdogianni. “Image processing: the fundamentals”. John Wiley & Sons Ltd, West Sussex, PO19 IUD, England, 1999.

- [Phillips 99] P.J. Phillips. “Support vector machines applied to face recognition”. In *Advances in Neural Information Processing Systems 11*, 803–809. MIT Press, 1999.
- [Phillips 00] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss. “The feret evaluation methodology for face-recognition algorithms”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1090–1104, 2000.
- [Phillips 07] J.P. Phillips, T.W. Scruggs, A.J. O’toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, and M. Sharpe. “FRVT 2006 and ICE 2006 large-scale results”. tech. report, National Institute of Standards and Technology, March 2007.
- [Poggio 97] T. Poggio. “Image representations for visual learning (invited paper)”. In *AVBPA ’97: First International Conference on Audio- and Video-Based Biometric Person Authentication*, 143. Springer-Verlag, 1997.
- [Pujol 01] A. Pujol, A.M. López, J.L. Alba, and J.J. Villanueva. “Ridges, valleys and hausdorff based similarity measures for face description and matching”. In *International Workshop on Pattern Recognition and Information Systems (PRIS)*, 80–90, 2001.
- [Rahman 96] Z.U. Rahman, D.J. Jobson, and G.A. Woodell. “Multi-scale retinex for color image enhancement”. In *ICIP96: International Conference on Image Processing (III)*, 1003–1006. IEEE Computer Society, 1996.
- [Rahman 97] Z.U. Rahman, G.A. Woodell, and D.J. Jobson. “A comparison of the multiscale retinex with other image enhancement techniques”. In *IS&T’s 50th Annual Conference: A Celebration of All Imaging*, 426–431, 1997.
- [Raja 06] Y. Raja and Sh. Gong. “Sparse multiscale local binary patterns”. In *17th BMVC: British Machinery and Vision Conference*, 2006.
- [Rao 90] K. R. Rao and P. Yip. “Discrete cosine transform: algorithms, advantages, applications”. Academic Press Professional, Inc., San Diego, CA, USA, 1990.

- [Saban 04] M. Saban and C. Sanderson. “On local features for face verification”. Tech. Report RR04-36, IDIAP, 2004.
- [Sanderson 03] C. Sanderson and K. K. Paliwal. “Fast features for face authentication under illumination direction changes”. *Pattern Recognition Letters*, 24(14):2409–2419, 2003.
- [Scott II 03] W.L. Scott II. *Block-level discrete cosine transform coefficients for autonomic face recognition*. PhD thesis, Louisiana State University and Agricultural & Mechanical College, USA, 2003. Director - Dr. Subhash Kak.
- [Shan 03] S.G. Shan, W. Gao, B. Cao, and D.B. Zhao. “Illumination normalization for robust face recognition against varying lighting conditions”. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures: AMFG03*, 157–164, 2003.
- [Shashua 97] A. Shashua. “On photometric issues in 3D visual recognition from a single 2D image”. *International Journal of Computer Vision*, 21:99–122, 1997.
- [Shi 06] J. Shi, A. Samal, and D. Marx. “How effective are landmarks and their geometry for face recognition?”. *Computer Vision and Image Understanding*, 102(2):117–133, 2006.
- [Short 04] J. Short, J. Kittler, and K. Messer. “A comparison of photometric normalisation algorithms for face verification”. In *AFGR04: Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 254–259. IEEE Computer Society, 2004.
- [Short 06] J. Short, J. Kittler, and K. Messer. “Photometric normalisation for component-based face verification”. In *FGR '06: 7th International Conference on Automatic Face and Gesture Recognition*, 114–119. IEEE Computer Society, 2006.
- [Sim 01] T. Sim and T. Kanade. “Combining models and exemplars for face recognition: An illuminating example”. In *CVPR 2001: Workshop on Models versus Exemplars in Computer Vision*, December 2001.

- [Socolinsky 01] D.A. Socolinsky, L.B. Wolff, J.D. Neuheisel, and C.K. Eveland. “Illumination invariant face recognition using thermal infrared imagery”. In *IEEE CS Conference on Computer Vision and Pattern Recognition* (1), 527–534, 2001.
- [Stockham 72] T.G. Stockham. “Image processing in the context of a visual model”. *PIEEE*, 60(7):828–842, July 1972.
- [Tan 05] X. Tan, S. Chen, Z.H. Zhou, and F. Zhang. “Recognizing partially occluded, expression variant faces from single training image per person with som and soft knn ensemble”. *IEEE Transactions on Neural Networks*, 16:2005, 2005.
- [Tan 06] X. Tan, S. Chen, Z.H. Zhou, and F. Zhang. “Face recognition from a single image per person: A survey”. *Pattern Recognition*, 39(9):1725–1745, 2006.
- [Tan 07] X. Tan and B. Triggs. “Enhanced local texture feature sets for face recognition under difficult lighting conditions”. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 168–182, 2007.
- [Turk 91] M. Turk and A. Pentland. “Eigenfaces for recognition”. *J. Cognitive Neuroscience*, 3(1):71–86, 1991.
- [Villegas 05] M. Villegas and R. Paredes. “Comparison of illumination normalization methods for face recognition”. In *Third COST 275 Workshop- Biometric on the Internet*, 27–30, 2005.
- [Villela 02] P.R. Villela and J.H. Sossa. “Improving pattern recognition using several feature vectors”. In *MICAI '02: Second Mexican International Conference on Artificial Intelligence*, 282–291. Springer-Verlag, 2002.
- [Viola 01] P. Viola and M. Jones. “Rapid object detection using a boosted cascade of simple features”. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:511, 2001.

- [Wang 04a] H. Wang, S.Z. Li, and Y. Wang. “Face recognition under varying lighting conditions using self quotient image”. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 819–824, 2004.
- [Wang 04b] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. “Image quality assessment: From error visibility to structural similarity”. *IEEE Transactions on Image Processing*, 13:600–612, 2004.
- [Wang 08] Y.H. Wang, X.J. Ning, C.X. Yang, and Q.F. Wang. “A method of illumination compensation for human face image based on quotient image”. *Information Sciences*, 178(12):2705 – 2721, 2008.
- [Watson 94] A.B. Watson. “Image compression using the discrete cosine transform”. *Mathematica Journal*, 4(1):81–88, 1994.
- [Webb 02] A. R. Webb. “Statistical pattern recognition”, ch. 8.3 8.3, 266 – 271. John Wiley and Sons Ltd, 2dn edition, 2002.
- [Wiskott 97] L. Wiskott, J.M. Fellous, N. Kruger, and C. VonDerMalsburg. “Face recognition by elastic bunch graph matching”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.
- [Wright 09] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma. “Robust face recognition via sparse representation”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, February 2009.
- [Wu 02] J. Wu and Z.H. Zhou. “Face recognition with one training image per person”. *Pattern Recognition Letters*, 23(14):1711–1719, 2002.
- [XiaoJun 04] W. XiaoJun, J. Kittler, Y. JingYu, K. Messer, and W. ShiTong. “A new kernel direct discriminant analysis (KDDA) algorithm for face recognition”. In *British Machinery and Vision Conference*, 2004.
- [Zamani 08] A.N. Zamani, M.K. Awang, N. Omar, and Sh.A. Nazeer. “Image quality assessments and restoration for face detection and recognition system images”. In *AMS '08: Second Asia International*

- Conference on Modelling & Simulation*, 505–510. IEEE Computer Society, 2008.
- [Zhang 04] G. Zhang, X. Huang, S.Z. Li, Y. Wang, and X. Wu. “Boosting local binary pattern (LBP)-based face recognition”. In *Advances in Biometric Person Authentication*, 179–186, 2004.
- [Zhang 05a] D. Zhang, S. Chen, and Z.H. Zhou. “A new face recognition method based on svd perturbation for single example image per person”. *Applied Mathematics and Computation*, 163:895–907, 2005.
- [Zhang 05b] W.C. Zhang, S.G. Shan, H.M. Zhang, W. Gao, and X.L. Chen. “Multi-resolution histograms of local variation patterns (MHLVP) for robust face recognition”. In *AVBPA05*, 937–944, 2005.
- [Zhang 06] L. Zhang and D. Samaras. “Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):351–363, 2006.
- [Zhang 08] Z. Zhang, B. Tian Dai, and A.K.H. Tung. “Estimating local optimums in em algorithm over gaussian mixture model”. In *ICML '08: Proceedings of the 25th international conference on Machine learning*, 1240–1247. ACM, 2008.
- [Zhao 99] W.Y. Zhao, R. Chellappa, and P.J. Phillips. “Subspace linear discriminant analysis for face recognition”. In *UMD*, 1999.
- [Zhao 00] W.Y. Zhao and R. Chellappa. “Illumination-insensitive face recognition using symmetric shape-from-shading”. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:1286, 2000.
- [Zhao 03a] J. Zhao, Y. Su, D.J. Wang, and S.W. Luo. “Illumination ratio image: Synthesizing and recognition with varying illuminations”. *Pattern Recognition Letters*, 24(15):2703–2710, November 2003.
- [Zhao 03b] W.Y. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld. “Face recognition: A literature survey”. *ACM Computing Surveys*, 399 –

458, 2003. Also appeared as UMD Technical Report, CS-TR4167, 2000. Revised 2002, CS-TR4167R.

[Zou 07]

X. Zou, J. Kittler, and K. Messer. “Illumination invariant face recognition: A survey”. In *First IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, Sept. 2007.

Producción científica de la autora sobre el tema de la tesis

1. E. Garea, H. Méndez-Vázquez y otros: “An Illumination Insensitive Representation for Face Verification in the Frequency Domain”. 18th International Conference on Pattern Recognition (ICPR’06). Publicación en: *IEEE Computer Society*, ISBN 0-7695-2521-0, pp.215-218, 2006 ★.
2. E.M. Alvarez, H. Méndez-Vázquez y otros: “A Fast Method for Localization of Local Illumination Variations and Photometric Normalization in Face Images”. XI Congreso Iberoamericano de Reconocimiento de Patrones (CIARP 2006). Publicación en: *Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications*, ISSN 0302-9743, Vol. 4225 pp. 48-57, 2006 ★.
3. E. Garea, H. Méndez-Vázquez y otros: “A Comparative Study of Face Representations in the Frequency Domain”. XI Congreso Iberoamericano de Reconocimiento de Patrones (CIARP 2006). Publicación en: *Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications*, ISSN 0302-9743, Vol. 4225 pp. 48-57, 2006 ★.
4. E. Garea y H. Méndez-Vázquez: “Una Representación de Imágenes de Rostros Insensible a los Cambios de Iluminación en el Dominio de la Frecuencia”. II Taller de Procesamiento Digital de Imágenes y Señales Aplicadas de la UCI. Publicación en *Memorias de UCIENCIA 2006*, ISBN 959-16-0463-7, 2006.
5. E.M. Alvarez, H. Méndez-Vázquez y otros: “Application of Fast Object Detection Algorithms in Biometrics Images”. IV Congreso Nacional de Reconocimiento de Patrones. Publicación en *Memorias de RECPAT 2006*, 2006.
6. L.Chang, H. Méndez-Vázquez y otros: “Obtención de la mejor imagen para el Reconocimiento de Rostros”. X Congreso de la Sociedad Cubana de Matemática y Computación. Publicación en *Memorias de COMPUMAT 2007*, ISSN 1728-6042, Vol.5, Número Especial, 2007.

7. H. Méndez-Vázquez y otros: “A Rank based classifier to improve face recognition”. V Congreso Nacional de Reconocimiento de Patrones. Publicación en *Memorias de RECPAT 2007*, ISBN 978-959-286-006, 2007.
8. H. Méndez-Vázquez y E. García-Reyes: “Estado actual de los métodos de reconocimiento automático de rostros basados en la apariencia local”. Reporte Técnico *RT_006, Serie Azul*, Reconocimiento de Patrones, CENATAV, RNPS_ 2142 ISSN 2072-6287, Octubre 2008.
9. H. Méndez-Vázquez: “Face Recognition using Local Appearance Based Methods”. 5th IAPR International Summer School for Advanced Studies on Biometrics for Secure Authentication. BEST STUDENT PAPER AWARD. 2008.
10. H. Méndez-Vázquez, E. García-Reyes y otros: “A new combination of local appearance based methods for face recognition under varying lighting conditions”. XIII Congreso Iberoamericano de Reconocimiento de Patrones (CIARP 2008). Publicación en: *Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications*, ISSN 0302-9743, Vol. 5197 pp. 535-542, 2008 ★.
11. H. Méndez-Vázquez, E. García-Reyes y otros: “A New Image Division for LBP Method to Improve Face Recognition under Varying Lighting Conditions”. 19th International Conference on Pattern Recognition (ICPR’08). Publicación en: *IEEE Computer Society*, ISBN 978-1-4244-2175-6, 2008 ★.
12. H. Méndez-Vázquez y otros: “Face recognition with LWIR Imagery using Local Binary Patterns”. 3rd International Conference on Biometrics (ICB 2009). Publicación en: *Lecture Notes in Computer Science: Advances in Biometrics*, ISBN 978-3-642-01792-6, Vol. 5558 pp. 327-336, 2009 ★.
13. H. Méndez-Vázquez y otros: “An Illumination Quality Measure for Face Recognition”. 20th International Conference on Pattern Recognition (ICPR’10). Publicación en: *IEEE Computer Society*, ISBN 978-1-4244-2175-6, 2010 ★.
14. H. Méndez-Vázquez, J. Kittler y otros: “On combining local DCT with preprocessing sequence for face recognition under varying lighting conditions”. XV Congreso Iberoamericano de Reconocimiento de Patrones (CIARP 2010). Publicación en: *Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis and Applications*, Vol. 6419 pp. 410-417, 2010 ★.
15. H. Méndez-Vázquez y E. García-Reyes: “A New Methodology for Face Recognition under Varying Illumination Conditions”. Reporte Técnico *RT_026, Serie Azul*, Reconocimiento de Patrones, CENATAV, RNPS_ 2142 ISSN 2072-6287, Marzo 2010.

★ Referenciada en SCOPUS.

Glosario de acrónimos

AAM	Modelo de Apariencia Activa, del inglés <i>Active Appearance Model</i>
AS	Suavizado Anisotrópico, del inglés <i>Anisotropic Smoothing</i>
CCQ	Clasificador combinado que se propone para evaluar la calidad de una imagen de rostro respecto a la iluminación
DCT	Transformada Discreta del Coseno, del inglés <i>Discrete Cosine Transform</i>
DoG	Filtrado de pasa-banda con Diferencia de Gaussianas, del inglés <i>Difference of Gaussians</i>
EBGM	Algoritmo conocido como <i>Elastic Bunch Graph Matching</i>
EB	Efectividad en la clasificación de imágenes buenas
EM	Efectividad en la clasificación de imágenes malas
ET	Efectividad total de un clasificador
FAR	Tasa de error de falsos aceptados, del inglés <i>False Acceptance Rate</i>
FN	Muestras buenas clasificadas como malas, del inglés <i>False Negatives</i>
FP	Muestras malas clasificadas como buenas, del inglés <i>False Positives</i>
FRR	Tasa de error de falsos rechazados, del inglés <i>False Rejection Rate</i>
GMM	Modelos de Mezclas Gaussianas, del inglés <i>Gaussian Mixture Models</i>
HE	Ecuilización de Histogramas, del inglés <i>Histogram Equalization</i>
HF	Filtrado Homomórfico, del inglés <i>Homomorphic Filtering</i>
HMM	Modelos Ocultos de Markov, del inglés <i>Hidden Markov Models</i>
LBP	Patrones Binarios Locales, del inglés <i>Local Binary Patterns</i>
LDA	Análisis Discriminante Lineal, del inglés <i>Linear Discriminant Analysis</i>
LDCT	Método que se propone basado en la Transformada Discreta del Coseno aplicada de manera local
MLBP	Patrones Binarios Locales Multi-resolución, del inglés <i>Multi-scale Local Binary Patterns</i>
MSR	<i>Retinex</i> Multi-resolución, del inglés <i>Multi-scale Retinex</i>

MPC	Función de confianza del clasificador basado en la probabilidad media
MPQ	Clasificador basado en la probabilidad media de las regiones
NB	Neurona de salida de la red neuronal, que modela la probabilidad de que una imagen tenga buena iluminación
NM	Neurona de salida de la red neuronal, que modela la probabilidad de que una imagen tenga mala iluminación
NNC	Función de confianza del clasificador basado en la red neuronal
NNQ	Clasificador basado en la red neuronal
OACI	Organización de la Aviación Civil Internacional
OI	Imágenes originales sin ningún pre-procesamiento
PCA	Análisis de Componentes Principales, del inglés <i>Principal Component Analysis</i>
PS	Secuencia de preprocesamiento propuesta por [Tan 07]
QI	Método de la Imagen Cociente, del inglés <i>Quotient Image</i>
SOM	Mapas Auto-Organizados, del inglés <i>Self Organizing Maps</i>
SR	<i>Retinex</i> Simple, del inglés <i>Simple Retinex</i>
SVD	Descomposición en Vectores Únicos, del inglés <i>Singular Vector Decomposition</i>
SVM	Máquinas de Vectores de Soporte, del inglés <i>Support Vector Machines</i>
SQI	Imagen del Cociente Propia, del inglés <i>Self Quotient Image</i>
TER	Tasa de error total, del inglés <i>Total Error Rate</i>
TN	Muestras malas clasificadas como malas, del inglés <i>True Negatives</i>
TP	Muestras buenas clasificadas como buenas, del inglés <i>True Positives</i>
TV	Modelo de Variación Total, del inglés <i>Total Variation</i>
TVQI	Extensión del método QI basado en el modelo TV, del inglés <i>Total Variation Quotient Image</i>

Glosario de términos

algoritmo	Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema. En este trabajo se refiere al conjunto de procedimientos que se ejecutan de manera ordenada para lograr el reconocimiento automático de los individuos a partir de imágenes del rostro.
conjunto <i>bootstrap</i>	En estadística, el <i>bootstrapping</i> es un método para estimar la distribución muestral de un estimador por remuestreo. El conjunto <i>bootstrap</i> es la muestra que se utiliza como remplazo de los datos originales.
convolución	Se denomina convolución a una función, que de forma lineal y continua, transforma una señal de entrada en una nueva señal de salida.
eficacia	Es la capacidad de alcanzar el efecto que espera o se desea tras la realización de una acción.
eficiencia	Es el uso racional de los medios para alcanzar un objetivo predeterminado.
<i>eigenfaces</i>	Término en inglés por el cual se conoce al algoritmo de reconocimiento de rostros basado en el análisis de componentes principales (PCA).
<i>kernel</i>	En procesamiento de imágenes, normalmente se le denomina <i>kernel</i> a la función de impulso utilizada en una convolución.
histograma	El histograma de una imagen es una distribución que representa el número de píxeles por cada valor de intensidad presente en la imagen.
método	Modo de obrar o proceder. En el contexto del trabajo se refiere a la forma de ejecutar alguno de los pasos necesarios en el algoritmo de reconocimiento de rostros.
<i>wavelets</i>	Una <i>wavelet</i> es una oscilación de onda, con una amplitud que comienza en cero, aumenta y luego disminuye de nuevo a cero. Una serie de <i>wavelets</i> es una representación de una función cuadrática integrable por una cierta serie ortonormal generada por una <i>wavelet</i> .

Anexos

Anexo 1: Listado de sistemas de reconocimiento de rostros más importantes

Sistema	Compañía	País	Precio (USD)	Descripción
VeriLook Extended SDK	Neurotechnology	Lituania	\$1349.00	Herramientas para el desarrollo de software de reconocimiento de rostros en máquinas personales y a través de servicios web, en los lenguajes C, C++, C#, Java 2, VB 6, VB .NET y Delphi 7.
Morpheus SDK	KeeSquare	Italia	\$1875.00	Biblioteca de funciones para el desarrollo de aplicaciones de reconocimiento de rostros en Visual Studio con los lenguajes C, C++, C# y VB.
FaceVACS SDK	Cognitec	Alemania	\$3800.00	Herramientas para el desarrollo de software, con funcionalidades para el registro de imágenes de rostros, la verificación y la identificación.
FaceSDK 3.0	Luxand	USA	\$3950.00	Biblioteca de funciones para el desarrollo de aplicaciones Web, Windows, Linux y Macintosh en Visual Studio C++, C#, VB y Borland Delphi. Con posibilidades para el reconocimiento de rostros a partir de imágenes y de videos en tiempo real.
FRS SDK	Acscys	Canada	\$8000.00	Herramientas para el desarrollo de software, con funcionalidades para el seguimiento y registro de imágenes de rostros, la verificación e identificación, la creación de un servidor de base de datos y la comunicación con éste. Provee la opción de seguimiento en tiempo real y la posibilidad para el desarrollo de aplicaciones autónomas o basadas en arquitecturas cliente-servidor.
Facelt	Identix	USA	\$10000.0	Herramientas para el desarrollo de software, con funcionalidades para el seguimiento y registro de imágenes de rostros, la evaluación de la calidad, la verificación y el reconocimiento.
FaceSnap	C-Vis	Alemania	\$19250.0	Sistema de reconocimiento de rostros en imágenes digitales y de video en tiempo real. Incluye el hardware y software necesarios para crear y mantener el servidor de imágenes de rostros.

Nota: Los precios en la tabla se refieren a la licencia para el uso de los sistemas en una sola computadora. En la mayoría de los casos hay que pagar el mismo precio por cada computadora donde se utilice las herramientas o librerías descritas.

Anexo 2: Funcionamiento del operador MLSEC

Los pasos seguidos para obtener los valles y cordilleras de una imagen de rostro, I , utilizando el operador MLSEC, son los siguientes [Pujol 01]:

- 1) Se aplica un filtro de suavizado, S , a la imagen: $I_s = S(I)$.
- 2) Se calcula el vector gradiente normalizado: $\bar{\mathbf{w}} = (L_{x1}, L_{x2})^t \left(\sum_{j=1}^2 L_{xj}^2 \right)^{-1/2}$,
donde L_{x1} y L_{x2} son las derivadas de la imagen suavizada I_s , sobre los ejes de coordenadas x y y respectivamente.
- 3) Se aplica el operador MLSEC definido para el dominio discreto a partir de la divergencia del vector gradiente [López 00]:

$$\bar{\kappa}_d = -\text{div}(\bar{\mathbf{w}}) = -\frac{d}{r} \sum_{k=1}^r \bar{\mathbf{w}}_k^t \cdot \mathbf{n}_k,$$

donde d son las dimensiones del vector gradiente y r se define como la proximidad a la cual se limita la vecindad del operador.

En este caso, el operador MLSEC 2D para un píxel $P(i, j)$, se calcula como:

$$\bar{\kappa}_2[i, j] = -\frac{1}{2}(\bar{w}^1[i+1, j] - \bar{w}^1[i-1, j] + \bar{w}^2[i, j+1] - \bar{w}^2[i, j-1]),$$

donde \bar{w}^1 y \bar{w}^2 son los componentes de $\bar{\mathbf{w}}$ en 2D.

- 4) La respuesta del operador, que intuitivamente da una medida de las regiones de valles (valores positivos entre 0 y 2) y de cordilleras (valores negativos entre -2 y 0), se umbraliza, de manera que los píxeles para los que la respuesta del operador MLSEC es menor que $-\tau_1$ se consideran cordilleras y para los cuales es mayor que τ_2 se consideran valles. En este caso se utilizan $\tau_1 = \tau_2 = 1$.

En la figura se muestran ejemplos de los valles y crestas obtenidos para cuatro imágenes diferentes correspondientes a un mismo rostro:



RT_047, enero 2012

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2012

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

