



INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE INGENIERÍA MECÁNICA Y ELÉCTRICA
“UNIDAD CULHUACÁN”
SECCIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

**“RECONOCIMIENTO DE ROSTROS Y EXPRESIONES
FACIALES BASADO EN EL ESQUEMA DE FASES PROPIAS
CON SUB-BLOQUES DE TAMAÑO ÓPTIMO”**

TESIS

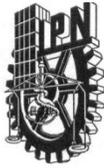
QUE PARA OBTENER EL GRADO DE:
**MAESTRO EN CIENCIAS DE INGENIERÍA
EN MICROELECTRÓNICA**

PRESENTA:
ING. GIBRÁN DE JESÚS BENÍTEZ GARCÍA

ASESORES:
Dr. Héctor Manuel Pérez Meana
Dr. Gabriel Sánchez Pérez



MÉXICO, D.F., NOVIEMBRE DE 2013



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México D. F., siendo las 16:00 horas del día 21 del mes de Noviembre del 2013 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de SEPI-ESIME-CULHUACAN para examinar la tesis titulada:
"Reconocimiento de Rostros y Expresiones Faciales Basado en el Esquema de Fases Propias con Sub-Bloques de Tamaño Óptimo"

Presentada por el alumno:

Benítez García Gibrán de Jesús
Apellido paterno Apellido materno Nombre(s)
Con registro:

B	1	1	0	2	4	0
---	---	---	---	---	---	---

aspirante de:

MAESTRÍA EN CIENCIAS DE INGENIERÍA EN MICROELECTRÓNICA

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de tesis

Dr. Héctor Manuel Pérez Meana Dr. Gabriel Sánchez Pérez
Dr. Volodymyr Ponomaryov Dra. Mariko Nakano Miyatake
Dr. Luis Martín Rojas Cárdenas



PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Gonzalo Isaac Duchén Sánchez



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México, D.F. el día 22 del mes de Noviembre del año 2013, el(la) que suscribe Ing. Gibrán de Jesús Benítez García alumno(a) del Programa de Maestría en Ciencias de Ingeniería en Microelectrónica, con número de registro B110240, adscrito(a) a SEPI-ESIME-CULHUACÁN, manifiesto(a) que es el(la) autor(a) intelectual del presente trabajo de Tesis bajo la dirección del Dr. Héctor Manuel Pérez Meana y el Dr. Gabriel Sánchez Pérez y cede los derechos del trabajo titulado “Reconocimiento de Rostros y Expresiones Faciales Basado en el Esquema de Fases Propias con Sub-bloques de Tamaño Óptimo”, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor(a) y/o director(es) del trabajo. Este puede ser obtenido escribiendo a las siguientes direcciones gibenitezg@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.



Ing. Gibrán de Jesús Benítez García

A mi madre

✓6080ନୁ ନୁ5 ଶକ୍ତ10065ନୁ10 ନୁ3 ଓ469 1017ନୁ9ଓ 3ଓ ର
ନୁ3ଓଓଓଓ 3x5ନୁଓ ନୁ509ନୁ 3ଓ 20ଓଓ 5 3ଓ 41ନୁ90ନୁ
୧୨ଓନୁ0ଓ10 769 10ନୁ୧100 ନୁ10ଓଓ5ଓ6 ରନୁ 4x

AGRADECIMIENTOS

Este trabajo de tesis representa la culminación de una etapa más de mi vida, un periodo de esfuerzo, dedicación, pasión, sacrificio y muchas emociones. Todo gracias a muchas personas que estuvieron conmigo en este tiempo, sin el valioso apoyo de ustedes no hubiese logrado la culminación de este proyecto. GRACIAS.

*En primer lugar a mi padre **Efrén**, porque sin ti simplemente no sería nadie. Por creer en mí, por apoyarme, por hacer las veces de padre y madre, pero sobre todo gracias por tu ejemplo. Realmente no hay logro que pudiera alcanzar en esta vida que se compare con el apoyo que he recibido de tu parte. Aun así, continuaré trabajando arduamente para que te sientas más orgulloso de mí.*

*A mi casa de estudios el **Instituto Politécnico Nacional** y a la **Sección de Estudios de Posgrado e Investigación de la ESIME Culhuacán**, por mi formación y por darme la oportunidad de poner "La Técnica al Servicio de la Patria".*

*Al **CONACYT** y al Programa Institucional de Formación de Investigadores (**PIFI**) por facilitarme el soporte económico para la realización de este trabajo de investigación.*

*A mis asesores el **Dr. Héctor Manuel Pérez Meana** y el **Dr. Gabriel Sánchez Pérez** por su confianza y apoyo, además de sus consejos y enseñanzas las cuales no fueron únicamente académicas, pero especialmente por sembrar en mi ese amor por la investigación. Siempre serán mi ejemplo a seguir, fue un honor haber sido su asesorado.*

*A la **Dra. Mariko Nakano Miyatake** y la **Dra. Linda Karina Toscano Medina** por sus enseñanzas y ayuda en la realización de este proyecto, además de sus consejos fuera del ámbito académico.*

*Al Dr. **Jesús Olivares Mercado** por impulsarme a ser siempre el mejor, por su honestidad y habilidad para hacerme ver mis errores, pero sobre todo gracias por brindarme su amistad al grado de considerarlo como un hermano mayor.*

*A toda **mi familia** por su apoyo y amor incondicional, por ustedes trato de hacer las cosas de la mejor manera posible, aspirando a ser un orgullo para los más grandes y un ejemplo a seguir para los más pequeños.*

*A mi madrina **Lidia**, a **Rodrigo, Hugo, Jimena**, y **Dulce** por apoyarme en todo momento y estar siempre ahí, en las buenas y en las malas. Sobre todo gracias por creer en mí, ya que en muchas ocasiones ustedes confiaban más en mí que yo mismo.*

*A **mis compañeros** y **amigos** de la SEPI, por soportarme, trabajar conmigo y divertirse junto a mí todo este tiempo, aprendí mucho de ustedes y espero que los lazos que hemos creado perduren durante mucho tiempo.*

*Finalmente y no menos importante, a los profesores **Masahide Kaneko** y **Keita Takahashi** así como a la Universidad de Electro-Comunicaciones de Tokio y al programa JUSST, por aceptarme y brindarme todas las facilidades necesarias durante mi estancia en Japón.*

RESUMEN

Este trabajo de investigación presenta tres sistemas referentes a reconocimiento de rostros y reconocimiento de expresiones faciales. Estos sistemas se proponen con el fin de solucionar los problemas relacionados con variaciones de iluminación, tamaño de las imágenes de entrada y la oclusión parcial.

El algoritmo de Fases Propias para el reconocimiento de rostros presenta un rendimiento bastante bueno, bajo condiciones controladas. Sin embargo, su desempeño disminuye cuando las imágenes bajo análisis presentan variaciones en las condiciones de iluminación, así como oclusión parcial. Para superar estos problemas, este trabajo propone una modificación al método de Fases Propias basado en subbloques. Los resultados tanto teóricos como experimentales muestran que el tamaño de bloque óptimo es de 2×2 píxeles, mejorando el porcentaje de identificación alrededor del 3% en comparación con el método convencional, incluso cuando se presenta la oclusión parcial. Los resultados muestran que el sistema propuesto, cuando requiere llevar a cabo la tarea de verificación, es capaz de proporcionar, de forma simultánea, una falsa aceptación y un falso rechazo menores al 0.5%, situación que no es posible obtener utilizando el método convencional.

Referente al método de patrones binarios locales (LBP), este trabajo presenta dos modificaciones. En la primera, llamada patrones binarios por ventanas (WBP), primeramente la imagen del rostro se segmenta en ventanas de $N \times N$ píxeles, las cuales se caracterizan solamente por el coeficiente de LBP correspondiente al píxel central. Enseguida, se aplica el método de análisis de componentes principales (PCA) para la estimación de los vectores característicos. En la segunda modificación, denominada DI-WBP, la imagen del rostro se redimensiona antes de que se aplique el WBP, excluyendo la utilización del PCA. Los resultados experimentales muestran que los métodos propuestos logran resultados superiores al 98% para identificación y mayores al 99% para la tarea de verificación.

En cuanto a los sistemas de reconocimiento de expresiones faciales, en este trabajo se presenta un método basado en la segmentación de la imagen del rostro en cuatro regiones faciales (ojos-cejas, frente, boca y nariz). El método propuesto emplea el algoritmo de Fases Propias con sub-bloques de tamaño óptimo para la conformación del vector característico. Con la finalidad de unificar los diferentes resultados obtenidos a partir de las combinaciones de las regiones faciales, se propone el Enfoque de Valor Modal el cual emplea la decisión más frecuente de los clasificadores. La robustez del algoritmo también se evalúa bajo oclusión parcial, usando cuatro tipos diferentes de oclusión (oclusión de la mitad del rostro izquierdo/derecho, ojos y boca). Los resultados experimentales muestran que el uso de Enfoque del Valor Modal mejora el reconocimiento, alcanzando resultados mayores al 90%, además este porcentaje de reconocimiento puede permanecer alto incluso en presencia de oclusión parcial, mediante la exclusión de las partes ocluidas del rostro en el proceso de extracción de características.

ABSTRACT

This research work presents three systems regarding to face recognition and facial expression recognition. These systems are proposed in order to overcome the problems related to variable illumination, size of the input images and partial occlusion.

The Eigenphases algorithm for face recognition performs fairly good, achieving high face recognition rate, under controlled conditions. However, its performance degrades when the face images under analysis present variations in the illumination conditions as well as partial occlusions. To overcome these problems, this work proposes a modified Eigenphases approach using sub-block processing. Theoretical and computer evaluation results show that a block size of 2×2 improves the identification rate about 3%, in comparison with the conventional one, even when it presents partial occlusions. Evaluation results show that, when the proposed approach is required to perform identity verification tasks, it is able to provide, simultaneously, false acceptance and false rejection rates lower than 0.5%, which is not possible to obtain using the conventional approach.

About the Local Binary Pattern (LBP) based features extraction method, this work presents two modifications. In the first one, called Window Binary Pattern (WBP), firstly the face image is segmented into sub-blocks of $N \times N$ pixels which are characterized only by the LBP coefficient corresponding to the central pixel of each sub-block. After that, the principal component analysis (PCA) is applied for feature vector estimation. In the second modification, called DI-WBP, the face image is firstly decimated before the WBP is applied, allowing the exclusion of PCA process. Evaluation results show that proposed schemes achieves recognition results that exceed 98% and 99%, in identification and verification tasks, respectively.

Regarding facial expression recognition systems, this work presents an approach based on segmentation of a face image into four facial regions (eyes-eyebrows, forehead, mouth and nose). The proposed method employs sub-block Eigenphases with optimum sub-block size algorithm for feature vector estimation. In order to unify the different results obtained from facial region combinations, a modal value approach that employs the most frequent decision of the classifiers is proposed. The robustness of the algorithm is also evaluated under partial occlusion, using four different types of occlusion (half left/right, eyes and mouth occlusion). Experimental results show that using modal value approach improves the average recognition rate achieving more than 90% and the performance can be kept high even in the case of partial occlusion by excluding occluded parts in the feature extraction process.

CONTENIDO

Sesión de Derechos	iii
Dedicatoria	v
Agradecimientos	vii
Resumen.....	ix
Abstract	xi
Contenido.....	xiii
Lista de Figuras.....	xvii
Lista de Tablas	xxiii
1. INTRODUCCIÓN.....	1
1.1 Motivación	2
1.2 Justificación	3
1.3 Objetivo General.....	4
1.4 Objetivos Específicos	4
1.5 Aportaciones Principales	5
1.5.1 Publicaciones Realizadas	6
1.6 Organización de la Tesis.....	7
2. ESTADO DEL ARTE.....	9
2.1 Sistemas Biométricos.....	10
2.1.1 Reconocimiento de Huella Dactilar	14
2.1.2 Reconocimiento de Firmas	15
2.1.3 Reconocimiento de Voz.....	15
2.1.4 Reconocimiento de Iris	16
2.2 Reconocimiento de Rostros	17
2.2.1 Esquema General del Reconocimiento de Rostros	20
2.3 Reconocimiento de Expresiones Faciales (FER).....	22
2.4 Conclusiones.....	24

3. MARCO TEÓRICO	25
3.1 Algoritmos de Pre-procesamiento.....	26
3.1.1 Normalización del Valor del Pixel.....	26
3.1.2 Ecuación del Histograma	28
3.1.3 CLAHE	33
3.2 Métodos de Extracción de Características	38
3.2.1 Caras Propias (Eigenfaces)	40
3.2.2 Fases Propias (Eigenphases).....	43
3.2.2.1 Extracción del Espectro de Fase.....	44
3.2.2.2 Análisis de Componentes Principales (PCA).....	46
3.2.3 Patrones Binarios Locales (LBP).....	48
3.3 Clasificador	51
3.3.1 Máquinas de Soporte Vectorial (SVM)	53
3.3.1.1 Esquema de uno contra todos (OVA).....	57
3.4 Conclusiones	59
4. SISTEMA PROPUESTO USANDO FASES PROPIAS	61
4.1 Diagrama del Sistema	62
4.2 Etapa de Pre-procesamiento.....	64
4.3 Segmentación en Sub-bloques	65
4.3.1 Tamaño de Bloque Óptimo.....	66
4.4 Extracción del Espectro de Fase	72
4.5 Conformación del Vector Característico usando PCA	73
4.6 Clasificación usando SVM	74
4.6.1 Etapa de Entrenamiento	75
4.6.2 Etapa de Identificación	76
4.6.3 Etapa de Verificación.....	77
4.7 Conclusiones	78
5. SISTEMA PROPUESTO USANDO LBP	79
5.1 Diagrama del Sistema	80
5.2 Extracción de características usando LBP	81
5.2.1 LBP usando PCA (LBPC)	82
5.3 Extracción de características usando WBP.....	83
5.3.1 WBP con redimensión de la imagen (DI-WBP).....	86
5.4 Clasificación usando SVM	87

5.5	Conclusiones.....	88
6.	SISTEMA PROPUESTO PARA RECONOCIMIENTO DE EXPRESIONES FACIALES	89
6.1	Diagrama del Sistema.....	90
6.2	Segmentación del Rostro.....	91
6.3	Extracción de Características usando Fases Propias.....	93
6.4	Conformación del Vector Característico.....	94
6.5	Clasificación usando SVM.....	96
6.6	Enfoque del Valor Modal.....	97
6.7	Conclusiones.....	100
7.	RESULTADOS EXPERIMENTALES.....	101
7.1	Bases de Datos.....	102
7.1.1	Base de Datos AR.....	102
7.1.1.1	Transformación de Iluminación.....	103
7.1.1.2	Imágenes Utilizadas para las Pruebas.....	105
7.1.2	Base de Datos Cohn-Kanade.....	106
7.1.2.1	Simulación de la Oclusión.....	108
7.1.2.2	Resolviendo la Oclusión de la Mitad del Rostro.....	109
7.1.2.3	Imágenes Utilizadas para las Pruebas.....	110
7.2	Numero Óptimo de Imágenes de Entrenamiento.....	112
7.3	Resultados de Identificación.....	113
7.3.1	Sistema Propuesto usando Fases Propias.....	115
7.3.2	Sistema Propuesto usando LBP y WBP.....	121
7.3.3	Comparación de los 5 Mejores Resultados.....	125
7.4	Resultados de Verificación.....	127
7.4.1	Sistema Propuesto usando Fases Propias.....	129
7.4.2	Sistema Propuesto usando LBP y WBP.....	140
7.4.3	Comparación de Resultados.....	146
7.5	Resultados de Reconocimiento de Expresiones Faciales.....	146
7.5.1	Resultados Sin Oclusión.....	148
7.5.2	Resultados Con Oclusión.....	153
7.6	Conclusiones.....	158

8. CONCLUSIONES GENERALES Y TRABAJO A FUTURO	159
8.1 Conclusiones Generales	160
8.2 Trabajo a Futuro.....	163
REFERENCIAS	165
PUBLICACIONES	173

LISTA DE FIGURAS

2.1. Clasificación de los sistemas biométricos.	11
2.2. Ejemplo de problemas que afectan al reconocimiento de rostros.....	18
2.3. Diagrama a bloques general de un sistema de reconocimiento de rostros.	21
2.4. Ejemplo de las 6 expresiones básicas	22
3.1. Ejemplo de la aplicación del método de normalización del valor del pixel a una imagen de rostro	27
3.2. Ejemplo del histograma de una imagen.....	29
3.3. Ejemplo de la ecualización del histograma en tiempo continuo.....	31
3.4. Resultado de la ecualización del histograma aplicado a la imagen de la Figura 3.2(a).....	31
3.5. Ejemplo de la aplicación de la ecualización del histograma a una imagen de rostro	32
3.6. Procedimiento para la aplicación del método CLAHE.....	34
3.7. Resultado de aplicar CLAHE con $\beta=0.3$ y $(F,C)=2,2$ en la imagen de la Figura 3.2(a).....	36
3.8. Diferencias entre la aplicación de HE y CLAHE a una imagen de rostro.....	37
3.9. Experimento de Oppenheim	44
3.10. Proceso del método original de LBP	48
3.11. Ejemplos del método LBP con diferentes valores de P y R	49
3.12. Ejemplo de la aplicación del SVM en un problema linealmente separable.....	53
3.13. Ejemplo del Hiperplano de Separación Óptima (OSH).....	55
3.14. Ejemplo de la aplicación del SVM en un problema no separable linealmente..	57
3.15. Diferencia entre clasificación binomial y clasificación multi-clase	58
3.16. Ejemplo de la aplicación del algoritmo de SVM para un caso de multi-clasificación (3 clases) utilizando el método “uno contra todos” (OVA).	59
4.1. Diagrama a bloques del sistema propuesto de fases propias basado en sub-bloques.....	62
4.2. Resultado de la aplicación de las 5 primeras etapas del sistema propuesto	63

4.3. Resultado de la aplicación de los 3 algoritmos de pre-procesamiento a una imagen de rostro.....	64
4.4. Ejemplo de los diferentes tamaños de sub-bloque.....	65
4.5. Representación por fragmentos de un sub-bloque de la imagen del rostro	66
4.6. Ejemplo de la diferencia de los espectros de fase utilizando la segmentación en sub-boques de dos imágenes con diferentes condiciones de iluminación	70
4.7. Ejemplo de la diferencia de los espectros de fase de dos sujetos distintos usando un tamaño de sub-bloque de 2x2	71
4.8. Ejemplo de la extracción del espectro de fase de la misma imagen usando diferente tamaño de sub-bloque.....	73
4.9. Conformación de vectores característicos usando PCA.	74
4.10. Esquema de la etapa de entrenamiento usando SVM.	76
4.11. Esquema de la etapa de identificación usando SVM.	77
4.12. Esquema de la etapa de verificación usando SVM.	78
5.1. Diagrama a bloques del sistema propuesto basado en LBP.....	80
5.2. Procedimiento de la extracción de características llevada a cabo por el método LBPh.	81
5.3. Comparación entre imágenes LBP obtenidas con diferente radio	82
5.4. Procedimiento de la extracción de características llevada a cabo por el método LBPh.	83
5.5. Conformación de vectores característicos usando PCA.	83
5.6. Ejemplo de la implementación de WBP	84
5.7. Diferencia entre los resultados de los métodos LBP(8,1) y WBP	85
5.8. Procedimiento de la extracción de características llevada a cabo por el método WBP.....	85
5.9. Procedimiento para la etapa de extracción de características llevada a cabo por el método DI-WBP.	87
6.1. Diagrama a bloques del sistema propuesto para el reconocimiento de expresiones faciales.	90
6.2. Localización de las 4 regiones fundamentales del rostro basada en la distancia entre irises (DI) y la posición de los ojos.....	91
6.3. Ejemplo de la segmentación del rostro y demarcaciones de las 4 regiones fundamentales: boca, nariz, ojos-cejas y frente.	92
6.4. Procedimiento del algoritmo de fases propias con tamaño de bloque óptimo para el sistema de reconocimiento de expresiones faciales.	93
6.5. Proceso de aplicación del PCA para FER.....	93

6.6. Proceso de conformación del vector característico final utilizando los vectores característicos de las 4 regiones fundamentales	95
6.7. Ejemplo de diferentes vectores característicos finales obtenidos de una misma imagen de entrada.	95
6.8. Esquema de la etapa de reconocimiento para FER usando SVM.....	97
6.9. Diagrama del método de decisión: Enfoque del Valor Modal.....	98
7.1. Ejemplo de las imágenes de dos sujetos de la base de datos AR.....	103
7.2. Ejemplo de las 5 variaciones de iluminación aplicadas a la base de datos AR.	104
7.3. Ejemplo de las imágenes del set A.	105
7.4. Ejemplo de las imágenes del set B.	106
7.5. Ejemplo de imágenes pico de la base de datos Cohn-Kanade.	107
7.6. Ejemplo de la sección del rostro recortada de la base de datos Cohn-Kanade..	107
7.7. Ejemplo de los cuatro tipos de oclusión aplicados a la base de datos.	108
7.8. Proceso para la obtención de las imágenes espejo.....	110
7.9. Ejemplo de los 5 sets de imágenes utilizados para las pruebas.	111
7.10. Resultados usando diferente número de imágenes de entrenamiento.....	112
7.11. Ejemplo de las imágenes de entrenamiento utilizadas para una persona.	113
7.12. Ejemplo de una gráfica de ranking.	114
7.13. Gráfica de ranking del sistema propuesto sin pre-procesamiento usando las imágenes del set A (sin oclusión).....	116
7.14. Gráfica de ranking del sistema propuesto usando normalización del valor del pixel y las imágenes del set A (sin oclusión).	117
7.15. Gráfica de ranking del sistema propuesto usando ecualización del histograma (HE) y las imágenes del set A (sin oclusión).	117
7.16. Gráfica de ranking del sistema propuesto usando CLAHE y las imágenes del set A (sin oclusión).....	118
7.17. Gráfica de ranking del sistema propuesto sin pre-procesamiento usando las imágenes del set B (con oclusión).....	119
7.18. Gráfica de ranking del sistema propuesto usando normalización del valor del pixel y las imágenes del set B (con oclusión).	120
7.19. Gráfica de ranking del sistema propuesto usando ecualización del histograma (HE) y las imágenes del set B (con oclusión).	120
7.20. Gráfica de ranking del sistema propuesto usando CLAHE y las imágenes del set B (con oclusión).....	121
7.21. Gráfica de ranking con las variaciones del método LBPh usando las imágenes del set A (sin oclusión).....	123

7.22. Gráfica de ranking con las variaciones de los métodos LBPC, WBP y DI-WBP usando las imágenes del set A (sin oclusión).	123
7.23. Gráfica de ranking con las variaciones del método LBPh usando las imágenes del set B (con oclusión).	124
7.24. Gráfica de ranking con las variaciones de los métodos LBPC, WBP y DI-WBP usando las imágenes del set B (con oclusión).	125
7.25. Gráfica de ranking con los 5 mejores resultados de identificación usando las imágenes del set A (sin oclusión).	126
7.26. Gráfica de ranking con los 5 mejores resultados de identificación usando las imágenes del set B (con oclusión).	127
7.27. Ejemplo de una gráfica de verificación.	129
7.28. Gráfica de verificación para el método sin pre-procesamiento usando como tamaño de sub-bloque la imagen completa (convencional).	130
7.29. Gráfica de verificación para el método sin pre-procesamiento usando un tamaño de sub-bloque de 12x12 píxeles.	131
7.30. Gráfica de verificación para el método sin pre-procesamiento usando un tamaño de sub-bloque de 6x6 píxeles.	131
7.31. Gráfica de verificación para el método sin pre-procesamiento usando un tamaño de sub-bloque de 4x4 píxeles.	132
7.32. Gráfica de verificación para el método sin pre-procesamiento usando un tamaño de sub-bloque de 2x2 píxeles (óptimo).	132
7.33. Gráfica de verificación para el método de normalización del valor del pixel usando como tamaño de sub-bloque la imagen completa (convencional).	133
7.34. Gráfica de verificación para el método de normalización del valor del pixel usando un tamaño de sub-bloque de 12x12 píxeles.	133
7.35. Gráfica de verificación para el método de normalización del valor del pixel usando un tamaño de sub-bloque de 6x6 píxeles.	134
7.36. Gráfica de verificación para el método de normalización del valor del pixel usando un tamaño de sub-bloque de 4x4 píxeles.	134
7.37. Gráfica de verificación para el método de normalización del valor del pixel usando un tamaño de sub-bloque de 2x2 píxeles (óptimo).	135
7.38. Gráfica de verificación para el método de ecualización del histograma (HE) usando como tamaño de sub-bloque la imagen completa (convencional).	135
7.39. Gráfica de verificación para el método de ecualización del histograma (HE) usando un tamaño de sub-bloque de 12x12 píxeles.	136

7.40. Gráfica de verificación para el método de ecualización del histograma (HE) usando un tamaño de sub-bloque de 6x6 píxeles.	136
7.41. Gráfica de verificación para el método de ecualización del histograma (HE) usando un tamaño de sub-bloque de 4x4 píxeles.	137
7.42. Gráfica de verificación para el método de ecualización del histograma (HE) usando un tamaño de sub-bloque de 2x2 píxeles (óptimo).	137
7.43. Gráfica de verificación para el método CLAHE usando como tamaño de sub-bloque la imagen completa (convencional).....	138
7.44. Gráfica de verificación para el método CLAHE usando un tamaño de sub-bloque de 12x12 píxeles.....	138
7.45. Gráfica de verificación para el método CLAHE usando un tamaño de sub-bloque de 6x6 píxeles.....	139
7.46. Gráfica de verificación para el método CLAHE usando un tamaño de sub-bloque de 4x4 píxeles.....	139
7.47. Gráfica de verificación para el método CLAHE usando un tamaño de sub-bloque de 2x2 píxeles (óptimo).....	140
7.48. Gráfica de verificación para el método LBPh(8,1).....	142
7.49. Gráfica de verificación para el método LBPh(8,2).....	142
7.50. Gráfica de verificación para el método LBPh(8,3).....	143
7.51. Gráfica de verificación para el método LBPh(8,1).....	143
7.52. Gráfica de verificación para el método LBPh(8,2).....	144
7.53. Gráfica de verificación para el método LBPh(8,3).....	144
7.54. Gráfica de verificación para el método WBP.....	145
7.55. Gráfica de verificación para el método DI-WBP.....	145
7.56. Resultados de todas las posibles combinaciones utilizando las 4 regiones del rostro.....	151
7.57. Comparación de los resultados del método propuesto con los de FPBS.....	157
8.1. Ejemplo de la importancia de la boca y frente para poder reconocer una expresión facial.	162

LISTA DE TABLAS

2.1. Aplicaciones de los sistemas biométricos divididas en 3 categorías	14
6.1. Ejemplo de la aplicación del enfoque del valor modal.....	98
6.2. Caso especial para la aplicación del enfoque del valor modal, cuando un cierto número de clasificadores proveen la misma decisión.	99
7.1. Número de imágenes por cada expresión facial.	111
7.2. Resultados de las 4 variaciones y los 5 tamaños de sub-bloques de las pruebas usando Fases Propias realizadas con el set A de la base de datos (sin oclu)....	116
7.3. Resultados de las 4 variaciones y los 5 tamaños de sub-bloques de las pruebas usando Fases Propias realizadas con el set B de la base de datos (con oclu)....	118
7.4. Resultados de las pruebas de identificación usando las variantes del sistema basado en LBP realizadas con ambos sets de la base de datos.....	122
7.5. Mejores 5 resultados de identificación utilizando ambos sets de la base	125
7.6. Punto de equilibrio del ejemplo de la Figura 7.27.....	128
7.7. Resultados de verificación para todas las variaciones del sistema usando Fases Propias, porcentaje de error (F. aceptación = F. rechazo) y umbral.....	130
7.8. Resultados de verificación para todas las variaciones del sistema basado en LBP, porcentaje de error (F. aceptación = F. rechazo) y umbral.....	141
7.9. Mejores resultados de verificación por parte de los sistemas propuestos.....	146
7.10. Ejemplo de una matriz de confusión.	147
7.11. Matriz de confusión del método FPBS.....	148
7.12. Resultados de reconocimiento del sistema utilizando solo una región facial..	148
7.13. Matriz de confusión del sistema propuesto utilizando solo la región de la boca (B).	149
7.14. Resultados de reconocimiento del sistema utilizando dos regiones del rostro.	149
7.15. Matriz de confusión del sistema propuesto utilizando la combinación de Ojos-Boca (OB).....	149
7.16. Resultados de reconocimiento del sistema utilizando tres regiones del rostro.	150

7.17. Matriz de confusión del sistema propuesto utilizando la combinación de Ojos-Boca-Nariz (OBN).....	150
7.18. Matriz de confusión del sistema propuesto utilizando la combinación de las 4 regiones del rostro (OFBN).....	150
7.19. Resultados de reconocimiento del sistema utilizando dos regiones del rostro.	152
7.20. Matriz de confusión del mejor resultado obtenido para FER, usando el método de enfoque del valor modal con la combinación OB–FB–OFBN–FPBS.....	152
7.21. Resultados de reconocimiento del sistema utilizando las regiones anexadas..	153
7.22. Mejores resultados utilizando para cada tipo de oclusión utilizando el método del enfoque del valor modal.	154
7.23. Matriz de confusión para la combinación OBN–OFBN– Ri–Rd con imágenes en presencia de oclusión de la mitad izquierda del rostro.	154
7.24. Matriz de confusión para la combinación OBN–OFBN– Ri–Rd con imágenes en presencia de oclusión de la mitad derecha del rostro.	155
7.25. Matriz de confusión para la combinación B–FB–BN–FBN con imágenes en presencia de oclusión de los ojos-cejas.	155
7.26. Matriz de confusión para la combinación N–ON–FN–OFN con imágenes en presencia de oclusión de la boca.	155
7.27. Resultados del método FPBS para cada uno de los tipos de oclusión.	156
7.28. Matriz de confusión del método FPBS con imágenes en presencia de oclusión de la mitad izquierda del rostro.	156
7.29. Matriz de confusión del método FPBS con imágenes en presencia de oclusión de la mitad derecha del rostro.....	156
7.30. Matriz de confusión del método FPBS con imágenes en presencia de oclusión de los ojos-cejas.....	157
7.31. Matriz de confusión del método FPBS con imágenes en presencia de oclusión de la boca.	157

CAPÍTULO I

1. INTRODUCCIÓN

En este capítulo se expone la motivación por la cual se eligió el tema principal de la presente tesis, respondiendo a la pregunta: ¿Por qué realizar un trabajo de investigación acerca de sistemas de reconocimiento de rostros y expresiones faciales?, una vez resuelta esta incógnita, en la justificación se presentan los problemas y los medios de cómo serán atacados, demostrando que aún hay mucho por hacer en este campo de investigación.

Asimismo, es en este capítulo donde se definen los objetivos y alcances que tendrá este trabajo de investigación, además de citar las principales aportaciones que con ellos se generan. Finalmente la organización de la tesis se presenta en la última sección del capítulo.

1.1 Motivación

El reconocimiento de personas por medio del rostro, es una tarea que los humanos realizamos a diario, y en muchas ocasiones sin necesidad de reparar en ello. Esta acción es llevada a cabo por nuestro cerebro inmediatamente después de observar el rostro de cualquier persona. De esta forma, la manera natural en la que un ser humano reconoce a otro es por medio de su rostro.

Por otro lado el reconocimiento de personas entre seres humanos es un proceso muy complejo, debido a que en ciertas ocasiones podemos reconocer a una persona sin siquiera ver su rostro. En situaciones donde no es posible ver el rostro de una persona, podemos reconocerla por medio de su voz, su forma de caminar, su vestimenta, su olor, etc. Motivados por esta asombrosa habilidad humana, desde el surgimiento de las computadoras, se ha tratado de desarrollar un método automatizado para el reconocimiento de personas igual de eficiente que el realizado por los seres humanos. Los resultados de investigaciones científicas y avances tecnológicos han permitido el surgimiento de los sistemas biométricos, los cuales realizan la tarea de reconocer personas por medio de características humanas. Siendo el reconocimiento de rostro uno de los más populares y ampliamente aceptados.

El reconocimiento de rostros como sistema biométrico presenta grandes ventajas. Una de las más relevantes se basa en la simplicidad para la obtención de la muestra, ya que el sistema solo necesita la imagen del rostro de una persona para poder realizar el reconocimiento. Permitiendo así, el uso de este sistema a distancia e incluso de manera encubierta. Sin embargo, aún existen problemas que aquejan a este tipo de sistemas, los cuales no permiten que se logre un desempeño perfecto.

Asimismo, relacionado al proceso de entendimiento humano, se han desarrollado investigaciones referentes al reconocimiento de expresiones faciales. Tema de investigación enfocado a la interacción entre humano-maquina, propuesto con el objetivo de entender los sentimientos y emociones de los seres humanos. A su vez, este tipo de investigaciones han logrado mejoras significativas y el desempeño de sistemas de reconocimiento de expresiones faciales es muy aceptable, sin embargo aún nos encontramos lejos de alcanzar la perfección en este tipo de sistemas.

Es por eso que este trabajo presenta una investigación y desarrollo de métodos para el reconocimiento de rostros y expresiones faciales, tales que cumplan con los requisitos establecidos y además presenten una mejora, solucionando un problema específico de los relacionados a este tipo de sistemas.

1.2 Justificación

A pesar de que el reconocimiento de rostros es una investigación activa desde hace poco más de tres décadas y que muchos trabajos se han propuesto para resolver los problemas que aquejan a este tipo de sistemas, aún hay temas que resolver. Tal es el caso de los problemas en las variaciones de iluminación, pues siguen siendo un asunto que aún no se ha podido resolver por completo. Referente a este tema, el presente trabajo analiza dos métodos de solución, uno basado en la etapa de pre-procesamiento y otro en la etapa de extracción de características. Proponiendo la utilización del algoritmo de CLAHE para la etapa de pre-procesamiento en conjunto con el método de Fases Propias basado en sub-bloques, y proponiendo el uso de sub-bloques de tamaño óptimo para el algoritmo de Fases Propias, donde se demuestra que al utilizar este método es irrelevante una etapa de pre-procesamiento.

Adicionalmente, en base a uno de los métodos más recientemente propuestos para sistemas de reconocimiento de rostros (LBP, el cual debido a sus propiedades no necesita una etapa de pre-procesamiento), se propuso una modificación que reduce su complejidad (WBP). Siendo la complejidad computación un factor importante que debe tomarse en cuenta cuando se requiere aplicar este tipo de sistemas en dispositivos con poca capacidad de procesamiento.

Por otro lado, los problemas de los sistemas de reconocimiento de expresiones faciales son muy similares a los de reconocimiento de rostro, de manera que problemas como oclusión parcial siguen disminuyendo el rendimiento de estos sistemas hoy en día. Es por eso que como aportación a los sistemas de reconocimiento de expresiones faciales, se plantea la utilización del método de Fases Propias con sub-bloques de tamaño óptimo, además de la propuesta de un método de decisión basado en el valor modal de los resultados obtenidos por varios clasificadores trabajando en paralelo, denominado “Enfoque del valor modal”.

1.3 Objetivo General

Desarrollo de algoritmos para reconocimiento de rostros y expresiones faciales robustos a cambios de iluminación, tamaño de las imágenes y oclusión parcial.

1.4 Objetivos Específicos

- Análisis de algoritmos recientemente propuestos para el reconocimiento de rostros operados en ambientes de iluminación variable, tales como Fases Propias y Patrones Binarios Locales (LBP).
- Optimización de una base de datos, a partir de bases estándar, que permita la adecuada evaluación de los algoritmos de reconocimiento de rostros bajo condiciones variantes de iluminación.
- Análisis de esquemas de pre-procesamiento para mejorar el funcionamiento del algoritmo de Fases Propias cuando opera en ambientes de iluminación variable.
- Desarrollo de un algoritmo de Fases Propias con sub-bloques de tamaño óptimo.
- Desarrollo de un algoritmo basado en LBP con baja complejidad computacional.
- Desarrollo de un algoritmo para el reconocimiento de expresiones faciales robusto a diferentes tipos de oclusión parcial.
- Optimización de una base de datos, a partir de bases estándar, que permita la adecuada evaluación del algoritmo de reconocimiento de expresiones faciales bajo diferentes tipos de oclusión parcial.
- Evaluación experimental de los algoritmos desarrollados.

1.5 Aportaciones Principales

La presente tesis expone una investigación detallada acerca de métodos de pre-procesamiento y extracción de características para el desarrollo de sistemas de reconocimiento de rostros y expresiones faciales, de manera que las aportaciones principales de este trabajo se resumen a continuación:

- Análisis detallado de la aplicación de CLAHE para mejorar las imágenes en esquemas de reconocimiento de rostros.
- Aplicación del método de Fases Propias con sub-bloques de tamaño sub-óptimo en conjunto con el algoritmo de CLAHE (para la etapa de pre-procesamiento) en sistema de reconocimiento de rostros.
- Análisis de varios tamaños de sub-bloque para el algoritmo de Fases Propias, y propuesta del tamaño de sub-bloque óptimo, el cual no requiere de una etapa de pre-procesamiento para mejorar el rendimiento de sistemas de reconocimiento de rostros.
- Propuesta de un método de extracción de características para sistemas de reconocimiento de rostros, basado en las propiedades del algoritmo LBP. Denominado WBP (patrones binarios por ventanas).
- Reducción de la complejidad del método WBP y en general del algoritmo LBP, concluyendo en un nuevo método de extracción de características denominado DI-WBP (WBP con redimensión de imagen).
- Aplicación del método de Fases Propias con sub-bloques de tamaño óptimo, en conjunto con una etapa de segmentación de rostro, para sistemas de reconocimiento de expresiones faciales.
- Propuesta de un nuevo método de decisión para sistemas de reconocimiento de expresiones faciales que utilicen la segmentación del rostro, basado en resultados obtenidos por clasificadores en paralelo. Denominado “Enfoque del valor modal”.

6 | Capítulo I: Introducción

1.5.1 Publicaciones Realizadas

Durante el proceso de realización de la presente tesis, los resultados obtenidos fueron sometidos a la evaluación correspondiente en diferentes artículos científicos. Tal es el caso de revistas indexadas, congresos internacionales y congresos institucionales, las cuales se citan a continuación.

Revistas indexadas JCR:

- “A sub-block-based eigenphases algorithm with optimum sub-block size,” G. Benitez-Garcia, J. Olivares-Mercado, G. Sanchez-Perez, M. Nakano-Miyatake, and H. Perez-Meana. *Knowledge-Based Systems*, vol. 37, pp. 415-426, 2012.
- “New Window Binary Pattern Approach Based on LBP for Face Recognition,” G. Benitez-Garcia, J. Olivares-Mercado, G. Sanchez-Perez, M. Nakano-Miyatake, and H. Perez-Meana. *Sensor*. (aceptado con revisiones menores).
- “Facial Expression Recognition Based on Facial Region Segmentation and Modal Value Approach,” G. Benitez-Garcia, G. Sanchez-Perez, H. Perez-Meana, K. Takahashi and M. Kaneko. *IEICE Transactions on Fundamentals*. (aceptado para su publicación).

Revistas indexadas:

- “HALFTONING: REVIEW AND ANALYSIS,” F. Pelcastre, L. Peregrina, M. García, A. Ochoa, U. Juárez, E. Ríos, G. Benítez, and M. Nakano. *Revista Ingenierías Universidad de Medellín*, vol. 11, pp. 161-174, 2012.

Congresos internacionales:

- "Face Identification Based on Contrast Limited Adaptive Histogram Equalization (CLAHE)," G. Benitez-Garcia, J. Olivares-Mercado, G. Aguilar-Torres, G. Sanchez-Perez, and H. Perez-Meana. in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV'11), WORLDCOMP*, vol. 11, pp. 363-369, 2011.

- “Comparison of Facial Expression Recognition Rates Depending on Combination of Different Facial Regions,” G. Benitez-Garcia, G. Sanchez-Perez, H. Perez-Meana, K. Takahashi, and M. Kaneko. *2012年映像情報メディア学会冬季大会*, December 18-19, 2012.
- “Facial Expression Recognition Under Partial Occlusion Based on Facial Region Segmentation,” G. Benitez-Garcia, G. Sanchez-Perez, H. Perez-Meana, K. Takahashi, and M. Kaneko. *映像情報メディア学会技術報告*, vol. 37, pp. 95-98, 2013.

Congresos institucionales:

- “Facial Expression Recognition under Partial Occlusion using Sub-Block Eigenphases,” G. Benitez-Garcia and M. Kaneko. *Proceedings of the UEC International Mini-Conference for Exchange Students on Electro-Communications and Information Science No.28*, Abstract p.90, Poster p.91, August 9-10, 2012.
- “Facial Expression Recognition under Partial Occlusion Based on Modal Value Approach,” G. Benitez-Garcia, G. Sanchez-Perez, H. Perez-Meana, K. Takahashi, and M. Kaneko. *Proceedings of the UEC International Mini-Conference for Exchange Students on Informatics & Engineering and Information Systems No.29*, pp. 55-62, March 5-6, 2013.

1.6 Organización de la Tesis

- **Capítulo 1.**
Presenta un panorama general acerca del desarrollo de la presente tesis, además de la motivación, justificación, objetivos y principales aportaciones de este trabajo.
- **Capítulo 2.**
Presenta los conceptos básicos referentes al tema principal de esta tesis, concerniente a los sistemas biométricos y ahondando específicamente en el reconocimiento de rostros y el reconocimiento de expresiones faciales.

8 | Capítulo I: Introducción

- **Capítulo 3.**

Presenta los fundamentos teóricos referentes a los métodos utilizados para el desarrollo del proyecto. Los cuales pueden ser utilizados tanto para el reconocimiento de rostros como para el de expresiones faciales, y se dividen en algoritmos para las etapas pre-procesamiento, extracción de características y clasificación.

- **Capítulo 4.**

Presenta la metodología del sistema propuesto basado en el método de Fases Propias con sub-bloques. Además de presentar un análisis detallado acerca del tamaño de sub-bloque óptimo para este método de extracción de características.

- **Capítulo 5.**

Presenta la metodología del sistema propuesto basado en LBP. Donde se detallan los métodos propuestos de WBP y DI-WBP para la etapa de extracción de características de un sistema de reconocimiento de rostros.

- **Capítulo 6.**

Presenta la metodología del sistema propuesto para el reconocimiento de expresiones faciales, especificando la forma de utilización del método de Fases Propias con sub-bloques de tamaño óptimo en este tipo de sistemas. Además de presentar una explicación detallada acerca del método de decisión propuesto, denominado enfoque del valor modal.

- **Capítulo 7.**

Presenta una evaluación extensa y detallada de los resultados conseguidos por los 3 sistemas propuestos. Además de explicar las modificaciones realizadas a las bases de datos con las que se evaluaron dichos sistemas.

- **Capítulo 8.**

Presenta las conclusiones generales obtenidas en este trabajo, además de presentar las posibles líneas de continuación referentes a esta tesis.

CAPÍTULO II

2. ESTADO DEL ARTE

En este capítulo se presenta un panorama conceptual acerca de los temas relaciones con el presente trabajo. Para poder hablar del reconocimiento de rostros es necesario definir primero el tema de “sistemas biométricos”. Asimismo, es de gran ayuda entender el funcionamiento de algunos sistemas biométricos que han presentado grandes aportaciones en el tema, como es el caso de los basados en huella dactilar, firmas, voz e iris.

Una vez introducido el tema de los sistemas biométricos, es pertinente hablar acerca de los sistemas de reconocimiento de rostro. En este capítulo se citan algunas de las aplicación más significativas que ha presentado este tipo de tecnología, como es el caso de los juegos olímpicos de Beijing 2008, donde se utilizó por primera vez este sistema biométrico para permitir el acceso a la inauguración y clausura de este magno evento. De igual forma, algunos de los productos comerciales que brindan este tipo de tecnología también son citados. Finalmente se define el concepto de reconocimiento de expresiones faciales y de cómo este tema de investigación ha ido ganado importancia a pesar de los problemas que a él aquejan.

2.1 Sistemas Biométricos

Los sistemas biométricos son un conjunto de métodos automatizados para el reconocimiento de personas usando características fisiológicas o de comportamiento personal [1, 2].

Para que una característica, ya sea fisiológica o de comportamiento personal, sea apta y funcional para un sistema biométrico, debe de poseer las siguientes propiedades:

- **Universalidad.**
Cualquier persona debe de poseer la característica biométrica a medir. Con ciertas excepciones, como el caso de personas con capacidades diferentes y personas disminuidas físicamente por causa de un accidente.
- **Unicidad.**
La característica biométrica debe de mostrar claras distinciones entre dos personas de la población en general. La única excepción a esta propiedad puede ser el caso de dos gemelos idénticos.
- **Permanencia.**
La característica biométrica debe ser invariante, al menos en el periodo de tiempo en el cual trabajará el sistema biométrico. Una característica biométrica ideal debe ser resistente a cambios de edad, peso, estatura, lesiones, enfermedades, etc.
- **Elusión.**
Las personas que se vean beneficiadas con asumir otra identidad o con ocultar la propia, tratarán de eludir a los sistemas biométricos. Por lo tanto, la característica biométrica debe de ser suficientemente ineludible.

Para elegir adecuadamente el tipo de sistema biométrico, siempre debe tomarse en cuenta el campo de aplicación del mismo, sin descuidar el grado de eficiencia que presente la característica biométrica a las propiedades antes mencionadas. Como se mencionó previamente, las características biométricas pueden dividirse en dos tipos: fisiológicas y de comportamiento personal [1-3]. Sobre las cuales se clasifican en general los sistemas biométricos, la Figura 2.1 muestra un ejemplo de esta clasificación.

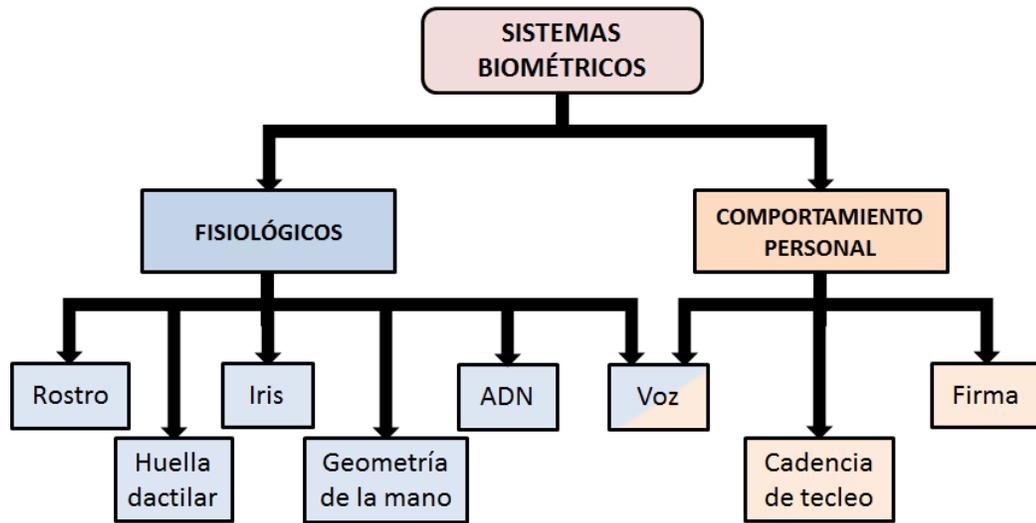


Figura 2.1. Clasificación de los sistemas biométricos.

Las propiedades de los sistemas biométricos basados en características de comportamiento personal (firma, cadencia del tecleo, forma de caminar, etc.) son distintas a las de los basados en características fisiológicas (rostro, huella dactilar, iris, ADN, etc.), marcando diferencias referentes a la aplicación y condiciones de uso. Sin embargo, los sistemas biométricos para el reconocimiento de voz presentan propiedades de ambas categorías, debido a que las características de la voz son tanto fisiológicas como de comportamiento.

Una de las propiedades más significativas de los sistemas biométricos basados en comportamiento personal es que para poder capturar la característica, el sistema requiere que el sujeto se encuentre en movimiento, es decir que realice una actividad la cual será analizada por el sistema. Por ejemplo, el usuario debe caminar frente a una cámara para el reconocimiento de forma de caminar, o teclear cierta combinación de caracteres para poder reconocer la cadencia del tecleo. Es importante mencionar que en contra parte a las características fisiológicas que se encuentran intrínsecamente en cualquier ser humano, las características de comportamiento personal el sujeto las aprende y adquiere con el tiempo.

Por otro lado, las características fisiológicas implican tomar una medición corporal del sujeto bajo análisis. Esta medición no necesariamente requiere una acción específica del sujeto. Por ejemplo, las características faciales pueden ser

12 | Capítulo II: Estado del Arte

adquiridas incluso sin conocimiento del usuario. De manera que, los sistemas biométricos que utilizan características fisiológicas se basan principalmente en una característica anatómica o física en lugar de un comportamiento aprendido.

En forma general, todos los sistemas biométricos deben desempeñar independientemente dos procesos distintos: registro y reconocimiento [3]. El registro, también llamado entrenamiento, consiste en almacenar la información biométrica de los posibles usuarios del sistema. Mientras que el reconocimiento se encarga de identificar o verificar al usuario utilizando la característica biométrica del sistema empleado.

Un sistema biométrico es esencialmente un sistema de reconocimiento de patrones. Por lo tanto, el proceso de un sistema biométrico en general puede resumirse en cuatro módulos principales: un módulo de captura; un módulo de extracción de características; un módulo de comparación y toma de decisiones (clasificación); y un módulo de base de datos [3]. Cada uno de estos módulos se describe a continuación.

- **Módulo de captura.**

Para este módulo es necesario un lector biométrico adecuado para obtener los datos biométricos en bruto de un individuo. Esta etapa define la interfaz usuario-máquina y es, por lo tanto, crucial para el rendimiento del sistema biométrico. Tomando en cuenta que la calidad de los datos en bruto se ve afectada por las características de la tecnología del lector que se utiliza.

- **Módulo de extracción de características.**

En este módulo los datos biométricos son procesados, y un conjunto de características discriminatorias sobresalientes es extraído para representar los rasgos más importantes de la identidad del individuo bajo análisis. Durante la fase de entrenamiento, este conjunto de características se almacena en la base de datos, el cual comúnmente se denomina como “modelo”.

- **Módulo de comparación y de toma de decisiones.**

Las características extraídas se comparan con los modelos almacenados para generar resultados coincidentes. En base al modo en el que esté operando el sistema dependerá el resultado que este ofrezca, ya sea una identidad en el modo de identificación o un valor lógico en el modo de verificación.

- **Módulo de base de datos del sistema.**

La base de datos actúa como repositorio de información biométrica. Durante el proceso de registro, se almacenan en la base de datos el conjunto de características extraídas de la muestra biométrica (el modelo), y en algunos casos, cierta información biográfica (nombre, número de identificación personal, dirección, etc.). Este módulo es indispensable, ya que los modelos se utilizarán para realizar el proceso de reconocimiento, finalidad de todo sistema biométrico.

Los términos comúnmente utilizados para definir las tareas de los sistemas biométricos llegan a ser confusos y en algunas ocasiones a utilizarse de manera errónea. Es por eso que a continuación se presenta la definición y diferencia entre los conceptos de: reconocimiento, identificación y verificación.

- **Reconocimiento.**

Este concepto debe utilizarse de manera general ya que no necesariamente implica verificación o identificación. De manera que este término generaliza la acción de identificar y verificar la identidad de un sujeto.

- **Identificación.**

Es la acción de establecer la identidad de un sujeto mediante la búsqueda en toda la base de datos del modelo correspondiente a esta persona, o en su defecto se determina que el sujeto no está inscrito en la base de datos del sistema.

- **Verificación.**

Es la acción de autenticar la identidad de un sujeto mediante la comparación de la característica biométrica capturada con su modelo, para esto inicialmente se debe proporcionar la identidad de dicho individuo, de manera que se rechaza o acepta la identidad proporcionada por el sujeto.

Las aplicaciones de los sistemas biométricos se pueden clasificar en 3 categorías: aplicaciones forenses, aplicaciones gubernamentales y aplicaciones comerciales [3, 4]. La Tabla 2.1 muestra las aplicaciones que pertenecen a cada una de estas categorías. Es importante mencionar que no todos los sistemas biométricos pueden aplicarse a los ejemplos mostrados en la tabla, ya que la aplicación depende del tipo de característica que el sistema utilice para reconocer y viceversa.

Tabla 2.1. Aplicaciones de los sistemas biométricos divididas en 3 categorías principales.

Forenses	Gubernamentales (Civiles)	Comerciales
Identificación del cadáveres	Seguridad social	Sistemas de inicio de sesión
Investigación criminal	Asistencia publica	Seguridad informática
Búsqueda de desaparecidos	Control de fronteras	Comercio electrónico
	Control de pasaportes	Control de acceso a internet
	Documentos de identidad	Cajeros automáticos (bancario)
	Licencias de conducir	Control de acceso físico
	Acreditación	Teléfonos móviles
		Asistente personal digital
		Manejo de historiales clínicos

A continuación se ahonda en los sistemas biométricos más representativos y normalmente utilizados (huella dactilar, firma, voz e iris). Y especialmente en el reconocimiento de rostros, tema principal del presente trabajo.

2.1.1 Reconocimiento de Huella Dactilar

El reconocimiento de huella dactilar es uno de los sistemas biométricos más antiguos (en 1963 se realizó el primer sistema automatizado), ampliamente utilizados y aceptados a nivel global [3, 4]. Una huella dactilar es la impresión dejada por las crestas papilares de un dedo de la mano al contacto con una superficie, la cual esta generalmente constituida por una serie de líneas oscuras que representan las crestas y una serie de espacios blancos que representan los valles. Está comprobado que las huellas dactilares son una característica única de las personas, de tal forma que es posible identificar a cada una por sus huellas dactilares.

Las huellas dactilares pueden obtenerse por dos medios. Uno es por medio de fotografías o entintados en papel, mientras que el otro medio, llamado “en vivo”, utiliza un sensor o escáner de huellas dactilares conectado directamente a la computadora que realizará el reconocimiento. La identificación con huellas dactilares está basada principalmente en procesar y reconocer “las minucias”, que refieren a la ubicación y dirección de las terminaciones de crestas, deltas y valles. La principal ventaja de este tipo de sistemas biométricos es que es muy difícil que las huellas dactilares cambien en un individuo, teniendo un alto grado de permanencia. Para más información de este tipo de sistemas consultar [4].

2.1.2 Reconocimiento de Firmas

El reconocimiento de firmas desde hace bastante tiempo se ha usado para la verificación de personas, específicamente en la validación de documentos oficiales tales como cheques bancarios, actas oficiales, etc. [3, 5]. Las firmas se han usado en los países occidentales desde hace más de un siglo, realizándose el reconocimiento de manera automatizada desde el año 1965. La firma es una característica biométrica del tipo de comportamiento personal, que normalmente es usada solo para la verificación de la identidad, puesto que esta característica biométrica no es práctica o muy precisa para la tarea de identificación. Esto debido a que se ve afectada por factores genéticos y físicos, que pueden voluntaria o involuntariamente modificarla. En este sentido la firma no tiene la misma permanencia que presentan la huella dactilar o el rostro.

Para la obtención de firmas existen básicamente dos métodos: el método dinámico que necesita de tabletas de digitalización u otras superficies que capturen el método de escritura usado para crear la firma; y el método estático que solamente utiliza una captura de la firma realizada sobre papel. Para realizar el reconocimiento por el método dinámico, se necesita procesar la información dinámica de la firma, es decir, la velocidad, aceleración, momentos gráficos, longitud total, cortes, longitud del primer trazo, orden de escritura en la firma, presión utilizada para escribir, etc. Mientras que para el reconocimiento estático de firmas, únicamente la imagen de la firma capturada está disponible para el proceso. Una de las ventajas de este tipo de sistemas es que la aceptación es casi total en la población, siendo un método ampliamente utilizado para la verificación de personas. Más información en [5].

2.1.3 Reconocimiento de Voz

El reconocimiento de voz es un sistema biométrico que se encarga de reconocer a una persona por medio del sonido generado por el aparato fonador humano (la voz) [3, 6]. La característica biométrica de la voz puede ser considerada en ambas categorías, fisiológica y de comportamiento personal. Las propiedades acústicas de la voz se catalogan como características anatómicas, tales como la forma de la boca del sujeto así como la longitud y calidad de las cuerdas vocales. Al mismo tiempo, la voz también contiene información de comportamiento personal como características de tono, volumen y ritmo (prosodia).

Este sistema biométrico necesita dispositivos de grabación de sonidos con los cuales obtener la información necesaria para poder actuar. Utilizando micrófonos autónomos, así como micrófonos incorporados en teléfonos celulares y otro tipo de dispositivos portátiles. Basándose en el tipo de información de entrada, el proceso para realizar el reconocimiento de voz es catalogado en dos formas. Usando un texto específico que el usuario debe recitar frente al micrófono (texto dependiente), o usando cualquier tipo de frases (texto independiente). Estos datos serán los que el sistema procese para realizar el reconocimiento. Una importante ventaja en este tipo de sistemas es que el reconocimiento de voz no es dependiente de un lenguaje en particular, ya que se puede entrenar el sistema para usuarios específicos. En [6] puede encontrar más información acerca de este tipo de sistemas.

2.1.4 Reconocimiento de Iris

El iris es un musculo dentro del ojo que regula el tamaño de la pupila, controlando la cantidad de luz que entra en el ojo. Específicamente, es la porción coloreada del ojo, basando su color en la cantidad del pigmento melatonina dentro del músculo. El iris se desarrolla durante crecimiento prenatal con un estricto proceso de formación y plegado de la membrana de tejido fino. Antes del nacimiento, ocurre la degeneración, dando por resultado la abertura de la pupila y los patrones únicos del iris [3, 7]. El iris fue sugerido como característica biométrica por el oftalmólogo Frank Burch en 1936, sin embargo el desarrollo de sistemas biométricos data de fechas más recientes [7].

Mientras la mayoría de las características biométricas tienen de 13 a 60 características distintas, se dice que el iris tiene 266 puntos únicos. Se cree que aunque genéticamente idénticos, el iris de cada ojo de un individuo es único y estructuralmente distinto, además de que sus características permanecen estables con el tiempo y el ambiente. Estas propiedades hacen del iris una característica biométrica ideal. Para obtener la información del iris, los sistemas biométricos necesitan cámaras con luz infrarroja para capturar una correcta imagen de esta característica. Durante los últimos años este sistema biométrico ha sido claramente establecido, los Emiratos Árabes Unidos lo utilizan como control fronterizo y la India como identificación única. Para más información referente a este sistema biométrico revisar [7].

2.2 Reconocimiento de Rostros

El reconocimiento de rostros es un sistema biométrico que consiste en utilizar el rostro de una persona para realizar la verificación o identificación. Debido a que el rostro es la forma más habitual utilizada por los seres humanos para reconocer a las personas, el reconocimiento de rostros es uno de los sistemas biométricos más ampliamente utilizados [8].

El primer sistema automatizado para el reconocimiento de rostros fue desarrollado por Takeo Kanade en su trabajo de tesis doctoral en 1973. A partir de esta fecha hubo un período de aproximadamente 15 años de reposo hasta que el trabajo de Sirovich y Kirby (1987) [9] marcó la pauta para posteriores estudios en el campo de procesamiento de información facial. Basado en él, el trabajo de Turk y Pentland (Eigenfaces, 1991) [10] revitalizó la investigación de reconocimiento de rostros [8].

El reconocimiento de rostro tiene algunas ventajas sobre otros sistemas biométricos como el caso de los basados en: huella dactilar, iris, retina y firmas, ya que no es necesario un equipo sofisticado para la captura de la característica biométrica, puesto que con una simple cámara se puede capturar la imagen del rostro que posteriormente será procesada. Otra ventaja del reconocimiento de rostros es que puede hacerse a larga distancia, utilizando un medio de comunicación remota como por ejemplo una cámara web o con cámaras incorporadas en teléfonos celulares. Además de que el reconocimiento puede llevarse a cabo con o sin la cooperación de la persona que será analizada (encubierto), presentando una gran ventaja en aplicaciones de seguridad y vigilancia [8, 11].

Estos hechos hacen que el reconocimiento de rostros sea uno de los sistemas biométricos con mayor aceptación entre los usuarios, y siga siendo un tema de investigación activa desde hace más de tres décadas. Por otro lado, haciendo una comparación, el proceso de reconocimiento de rostros de un ser humano puede ser muy sensible, ya que es posible realizarlo aún con cambios en la apariencia, como adición o sustracción de bello facial, lentes de contacto, lentes oscuros, envejecimiento, cambio de peso, etc. Siendo estos cambios un reto activo para los sistemas biométricos basados en el rostro [8].

Es por eso que el rendimiento de un sistema de reconocimiento de rostros depende en gran medida de una variedad de factores tales como la iluminación, pose, expresiones faciales, intervalo de edades, estilos de cabello, desgaste facial, oclusión parcial y movimiento presentado al momento de la captura. La Figura 2.2 muestra ejemplos de los problemas más comunes en los sistemas de reconocimiento de rostros: variaciones en la iluminación, oclusión parcial y rotación del rostro, ilustrados en (a), (b) y (c) respectivamente.

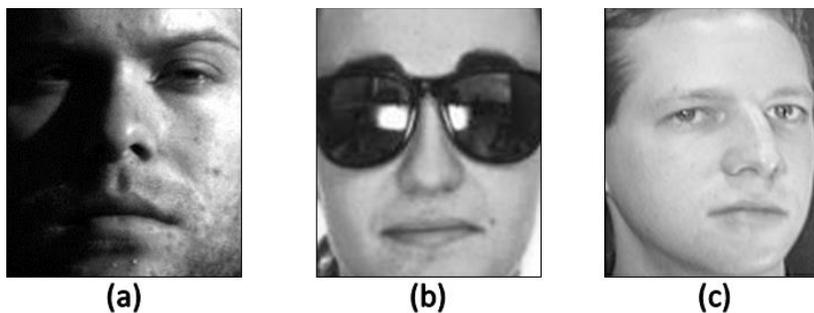


Figura 2.2. Ejemplo de problemas que afectan al reconocimiento de rostros. (a) Malas condiciones de iluminación. (b) Oclusión parcial (lentes oscuros). (c) Rotación.

Los cambios en las condiciones de iluminación representan uno de los problemas más persistentes en sistemas de reconocimiento de rostros, ya que no sólo se presenta en ambientes exteriores (al aire libre), sino también en ambientes interiores (lugares cerrados), esto debido a la forma 3D del rostro la cual produce sombras en función de la dirección de la iluminación al momento de la captura. Debido a esto, varios métodos se han propuesto en los últimos años, los cuales buscan reducir los problemas en cambios de la iluminación, enfocados principalmente en utilizar una etapa de pre-procesamiento en el sistema (información detallada en la sección 3.1 del presente trabajo).

La mayoría de los sistemas de reconocimiento de rostros, buscan resolver los problemas referentes a este sistema biométrico en la etapa de extracción de características. De tal manera que esta etapa es la responsable de estimar un conjunto de características del rostro que permita la reducción y posible eliminación de la influencia de variaciones relacionadas con: iluminación, pose, expresiones faciales, oclusión parcial, rotación, etc. Además de cumplir con su función principal de maximizar la diferencia entre los rostros de diferentes personas.

Por lo tanto, el desempeño de un sistema de reconocimiento de rostros depende del rendimiento del algoritmo utilizado en la etapa de extracción de características. Debido a esto, en las últimas décadas se han propuesto varios algoritmos de extracción de características, los cuales tienen la intención de cumplir con los requisitos de proporcionar una variabilidad intrapersonal más pequeña y una variabilidad interpersonal más grande. La mayoría de estos algoritmos están basados en transformadas de frecuencia y/o en algoritmos de compresión de datos (información detallada acerca de algoritmos de extracción de características en la sección 3.2 del presente trabajo).

El reconocimiento de personas por medio del rostro tiene una gran potencialidad en diversas aplicaciones, como puede ser en: tarjetas de identificación personal, control de acceso, seguridad, vigilancia, interacción humano-maquina, etc. Por ejemplo, en Estados Unidos esta tecnología ha empezado a crecer en la mayoría de los casinos, especialmente en mesas de Black Jack, donde se identifica a los clientes importantes y a los contadores de cartas; en Florida se utiliza el sistema de reconocimiento de rostro en los autos de la policía, donde los oficiales pueden inmediatamente reconocer a través de su base de datos a personas que tienen dos o más identidades para librarse de no pagar multas; otro famoso ejemplo donde se utilizó esta tecnología, fue en el Súper Bowl del 2001, donde se consiguió arrestar a más de diecinueve personas con antecedentes penales [8, 11].

Una importante aplicación de los sistemas de reconocimiento de rostros se dio en los Juegos Olímpicos de Beijing 2008. Donde se verificó la identidad del titular de un boleto (espectador) en las entradas del estadio nacional (inauguración y clausura). Cada boleto estaba asociado con un número de identificación único (RFID), además se requería que el titular previamente enviara un formulario de registro con una fotografía de dos pulgadas anexada a él. De manera que la foto del rostro era escaneada por el sistema (como etapa de entrenamiento) y el modelo se almacenaba en la base de datos (ligado al número de boleto). Posteriormente, en la entrada del estadio, el boleto era leído por un lector RFID, y la imagen del rostro era capturada por una cámara de vídeo, para así realizar una comparación entre la fotografía tomada en vivo y el modelo almacenado en la base de datos, con ello se obtenía un resultado de verificación permitiendo o denegando el acceso al estadio [12].

De manera comercial existen algunos productos que ofrecen la tecnología de este sistema biométrico al público en general. FaceGate [13] es un sistema de control de acceso que utiliza el reconocimiento de rostros junto con un código o una tarjeta para realizar la verificación del usuario. FaceKey [14] es un dispositivo que proporciona un sistema biométrico híbrido, ya que utiliza el reconocimiento de rostro junto con el reconocimiento de huella dactilar para permitir el acceso a los usuarios. Por otro lado, FaceVACS es una gama de productos de Cgnitec [15], los cuales ofrecen alternativas de reconocimiento de rostro en video, identificación de rostro para bases de datos, generación de fotografías de rostro para tarjetas de identificación y un SDK que incluye todo el software utilizado en los productos antes mencionados.

Por último, una de las aplicaciones para este sistema biométrico más ampliamente utilizadas hoy en día, es la de “etiquetado de fotos” en la red social Facebook. Esta aplicación consiste en la identificación, por medio del rostro, de amigos del usuario de Facebook en las fotos subidas por él mismo. El sistema de identificación de rostros utilizado en el proceso de etiquetado de fotos fue desarrollado por la empresa face.com, adquirida por Facebook en junio de 2012 [16], lo cual ha permitido un constante avance en la investigación y por ende una mejora de este sistema utilizado por la famosa red social.

2.2.1 Esquema General del Reconocimiento de Rostros

Los sistemas de reconocimiento de rostros, así como los demás sistemas biométricos, desempeñan dos procesos principales: entrenamiento y reconocimiento. En el proceso de reconocimiento se puede llevar a cabo la tarea de identificación o la de verificación, esto a partir de haber inicializado el sistema por medio de la fase de entrenamiento. De manera que, es fundamental tener presentes las principales diferencias entre las fases de identificación y verificación. En el primer caso, a partir de la imagen de rostro del usuario bajo análisis, el sistema se encarga de determinar la identidad de la persona de entre todos los modelos almacenados en la base de datos (obtenidos en la fase de entrenamiento). Por otro lado, en la fase de verificación es necesario obtener la identidad del individuo, proporcionada directamente por el usuario bajo análisis, de esta forma el sistema se encarga de determinar si la persona es quien dice ser, es decir, debe aceptar o rechazar la declaración de identidad otorgada por el usuario [17].

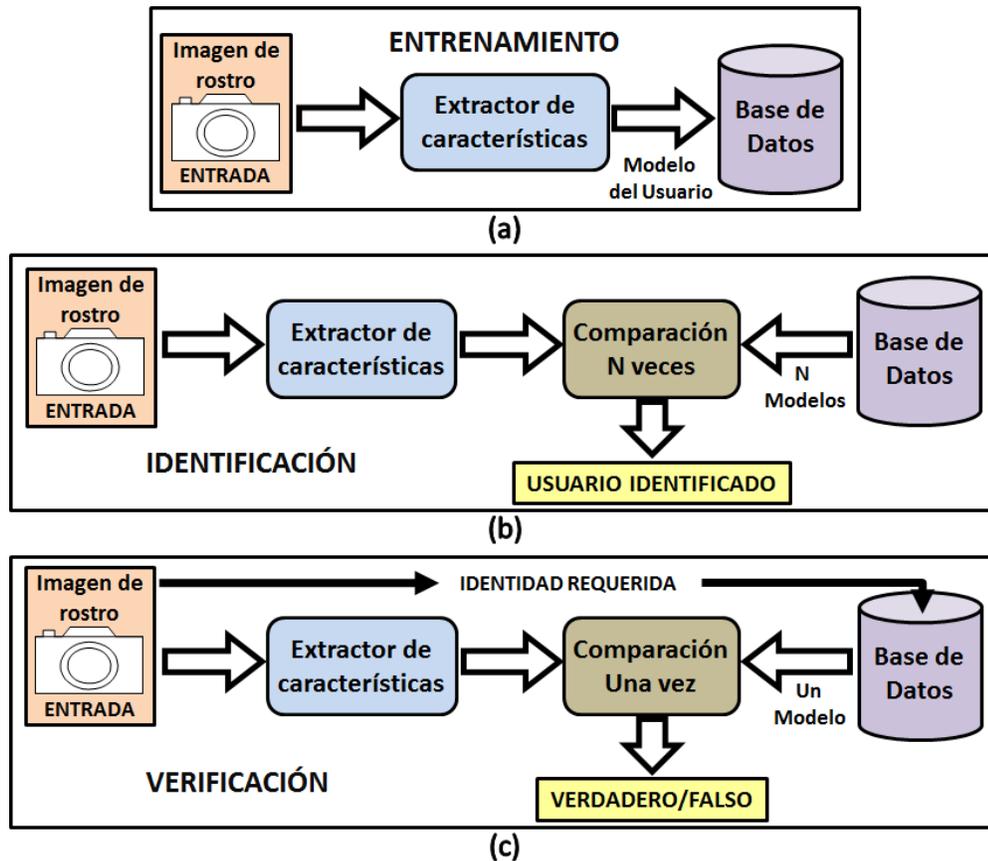


Figura 2.3. Diagrama a bloques general de un sistema de reconocimiento de rostros. (a) Fase de entrenamiento. (b) Fase de identificación. (c) Fase de verificación.

La Figura 2.3 muestra el diagrama a bloques general de un sistema de reconocimiento de rostros. (a) Ilustra la fase de entrenamiento, donde en primer lugar el sistema obtiene la imagen o imágenes del rostro de una persona, posteriormente, en la etapa de extracción de características se extrae un conjunto de características invariantes, con las cuales se forma un modelo que representa la identidad de la persona. En la fase de identificación (b), se proporciona solo una imagen de rostro a la etapa de extracción de características, con la que se genera un “modelo en vivo” del usuario (live-template), el cual se compara con todos los modelos de la base de datos, para determinar la identidad del sujeto bajo análisis. Por otro lado, durante la fase de verificación (c), es necesario que el usuario proporcione su identidad, con la cual el sistema elige el modelo relacionado a ella, para compararlo con el modelo en vivo generado con la imagen de entrada, de forma que el sistema determina si es verdadera o falsa la identidad proporcionada por el sujeto bajo análisis.

2.3 Reconocimiento de Expresiones Faciales (FER)

Desde la aparición de las primeras computadoras y robots, uno de los objetivos principales ha sido el de lograr una interfaz hombre-máquina compleja. Por lo tanto reconocer y clasificar de forma automatizada emociones y sentimientos humanos es una tarea indispensable para cumplir con este objetivo [18, 19]. El psicólogo A. Mehrabian en su estudio “Comunicación sin palabras” indicó que en la comunicación frente a frente entre seres humanos, solo el 7% de la información se transmite de manera lingüística, el 38% de manera paralingüística, y el 55% por medio de expresiones faciales [20]. Esto indica que las expresiones faciales desempeñan uno de los papeles más importantes en la comunicación humana, de modo que, las expresiones faciales y otro tipo de información no verbal son elementos indispensables para representar las emociones humanas [21]. Por esa razón, en las últimas dos décadas se han realizado avances de investigación relacionados con el reconocimiento de expresiones faciales (FER, por sus siglas en inglés) [22-29], el cual consiste en un conjunto de métodos automatizados para el reconocimiento de las expresiones faciales de los seres humanos.

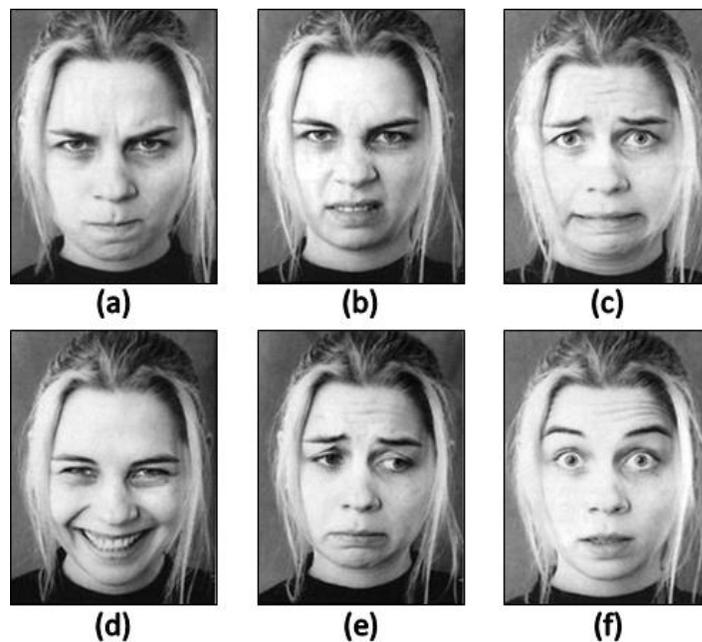


Figura 2.4. Ejemplo de las 6 expresiones básicas. (a) Ira. (b) Asco. (c) Miedo. (d) Felicidad. (e) Tristeza. (f) Sorpresa.

Según Darwin las expresiones faciales no son un atributo exclusivo de los seres humanos, sino que también se producen en la mayoría de los mamíferos y algunas otras especies de animales [30]. Una expresión facial se conforma por uno o más movimientos o posiciones de los músculos debajo de la piel del rostro. Los seres humanos poseemos más de 20 músculos en el rostro, los cuales participan para representar una expresión facial. Los psicólogos han definido las expresiones faciales humanas como un conjunto de seis expresiones básicas: ira, asco, miedo, felicidad, tristeza y sorpresa, además del estado neutral [21]. La Figura 2.4 muestra un ejemplo de cada una de estas 6 expresiones básicas.

En base a la definición y clasificación de las expresiones faciales humanas, muchos sistemas de FER utilizan el conjunto de 6 expresiones básicas. Existen muchas aplicaciones, como realidad virtual, entornos inteligentes, sistemas de videoconferencia, perfiles de usuario y estudios de satisfacción del cliente, que requieren sistemas de FER eficientes. Sin embargo, algunos problemas disminuyen el rendimiento de estos sistemas, los problemas que aquejan a FER son los mismos que los de reconocimiento de rostros. Por ejemplo, cambios de iluminación, pose, ángulo de la cámara de captura, oclusión parcial, y aunado a estos problemas se encuentra la intensidad de la expresión facial [22, 23].

La mayoría de los enfoques existentes sobre FER utilizan imágenes de rostros con expresiones faciales sin presencia de oclusión parcial y tomadas en condiciones controladas (con fondo uniforme y condiciones óptimas de iluminación). Sin embargo, este caso difícilmente se presenta en la vida real, siendo la oclusión parcial uno de los problemas más frecuentemente encontrados en los sistemas de FER [24, 25].

La oclusión parcial puede ser vista como ruido que perturba la imagen de la expresión facial en la etapa de extracción de características, causando pérdida de información importante y por ende un bajo rendimiento en sistemas de FER. Existen dos tipos de oclusión parcial: temporal y sistemática; oclusión temporal es cuando una parte de la cara se oscurece momentáneamente como resultado producido por sombras o un factor externo, mientras que la oclusión sistemática resulta cuando la persona está usando algo que cubre parte de su cara, como el caso de una bufanda,

lentes para sol o cubre bocas [26]. Por lo tanto, desarrollar un algoritmo de FER robusto a condiciones donde se presenten casos de oclusión parcial, se ha convertido en un importante tema de investigación en los últimos años [27-29].

La metodología para el desarrollo de sistemas de reconocimiento de expresiones faciales, así como para los sistemas biométricos, consta de 4 módulos fundamentales: captura, extracción de características, clasificación y base de datos. Y el procedimiento es muy similar al de los sistemas de reconocimiento de rostros, con la diferencia de que para los sistemas de FER no existe la fase de verificación. Esto debido a que la fase de reconocimiento opera solamente identificando la expresión facial del sujeto bajo análisis de entre las 6 expresiones básicas.

2.4 Conclusiones

En este capítulo se expusieron los conceptos básicos necesarios para una apropiada comprensión del tema principal de la presente tesis. De forma que los sistemas biométricos representan un tema imprescindible cuando se pretende hablar acerca del reconocimiento de rostros y expresiones faciales. Los sistemas biométricos más representativos son los basados en huella dactilar, firmas, voz, iris y específicamente rostro. El reconocimiento de rostro es uno de los sistemas biométricos más ampliamente aceptados por la población en general ya que basa su funcionamiento en la forma natural en la que los seres humanos reconocen a una persona. Por otro lado, el reconocimiento de expresiones faciales consiste en identificar la expresión facial mostrada por un rostro humano, la cual puede ser clasificada de entre un conjunto de seis expresiones básicas: ira, asco, miedo, felicidad, tristeza y sorpresa. Las expresiones faciales regularmente representan los sentimientos y emociones de los seres humanos, de forma que un sistema eficiente de reconocimiento de expresiones faciales contribuye a la interacción hombre-máquina. Los problemas que aquejan a los sistemas de reconocimiento de rostros y expresiones faciales son similares, y hoy en día no se ha logrado corregirlos a la perfección, por lo tanto estos temas siguen siendo un tema de investigación activa.

CAPÍTULO III

3. MARCO TEÓRICO

Este capítulo presenta los fundamentos teóricos relativos a este proyecto, con el fin de comprender el funcionamiento de los sistemas biométricos, específicamente el de reconocimiento de rostros. Tomando en cuenta que los métodos utilizados en este tipo de sistema biométrico también son aplicables a los sistemas de reconocimiento de expresiones faciales.

Asimismo, se describen detalladamente algunos algoritmos que permiten una posible solución al problema de cambios en la iluminación, que aqueja constantemente a los sistemas de reconocimiento de rostros. Estos algoritmos se basan en el mejoramiento de las imágenes de entrada, actuando en la etapa de pre-procesamiento. También se analizan tres métodos de extracción de características: Caras Propias, Fases Propias y Patrones Binarios Locales. Estudiando a fondo el uso del espectro de fase y el Análisis de Componentes Principales, utilizado por el algoritmo de Fases Propias, ya que una de las propuestas de este trabajo se fundamenta en base a estos dos procesos. Por último se analiza un algoritmo de clasificación llamado Máquinas de Soporte Vectorial, el cual será utilizado para todos los sistemas propuestos en la presente tesis.

3.1 Algoritmos de Pre-procesamiento

La etapa de pre-procesamiento en un sistema de reconocimiento de rostros se realiza con el objetivo de mejorar las imágenes de entrada, para que en la siguiente etapa (extracción de características) las características del rostro se obtengan con mayor facilidad y nitidez. Estas mejoras consisten principalmente en el ajuste del brillo, contraste, tamaño, resolución, eliminación de ruido, entre otras.

El pre-procesamiento está enfocado principalmente a combatir los problemas en cambios de iluminación en las imágenes de entrada. Una posible solución a estos problemas es el uso de algoritmos para el procesamiento de imágenes, que las aclaren o permitan cierta independencia a los cambios de iluminación [31]. Al realizar este tipo de mejoras en las imágenes, los sistemas de reconocimiento de rostros pueden obtener una mejora significativa en la precisión de sus resultados. Es por eso que en este trabajo se analizan 3 algoritmos para el procesamiento de imágenes: normalización del valor del pixel, ecualización del histograma (HE) y ecualización adaptativa del histograma con limitación de contraste (CLAHE); los cuales han sido utilizados en etapas de pre-procesamiento en sistemas de reconocimiento de rostros [32-37] e incluso en sistemas de reconocimiento de expresiones [38-40], obteniendo en general buenos resultados. El resto de esta sección describe detalladamente el funcionamiento de cada uno de los algoritmos antes mencionados.

3.1.1 Normalización del Valor del Pixel

La normalización del valor del pixel es un método sencillo y con un bajo costo computacional, el cual genera una cierta independencia de las características de la imagen, como lo son el brillo, el contraste y en algunas veces el fondo. La forma de aplicar la normalización del valor del pixel a una imagen consiste en dividir cada pixel que conforma la imagen entre una norma específica.

Este método ha sido aplicado en la etapa de pre-procesamiento en sistemas de reconocimiento de rostro propuestos en [32, 33]. En ambos casos el sistema ha sido propuesto utilizando el algoritmo de fases propias para la etapa de extracción de características, alcanzando un porcentaje de reconocimiento mayor al 95%. Por lo tanto es posible concluir que la normalización del valor de pixel presenta un buen

rendimiento cuando se utiliza en la etapa de pre-procesamiento para los sistemas de reconocimiento de rostro. Debido a las ventajas de este método, también ha sido usado como método de pre-procesamiento en sistemas de reconocimiento de expresiones faciales [38].

La normalización del valor del pixel para una imagen $I(x,y)$ de tamaño $N \times M$ se define por la siguiente ecuación:

$$I'(x,y) = \frac{I(x,y)}{\|I(x,y)\|}, \quad (3.1)$$

donde la norma es definida por:

$$\|I(x,y)\| = \left(\sum_{x=1}^M \sum_{y=1}^N I^2(x,y) \right)^{1/2}. \quad (3.2)$$

La Figura 3.1 muestra un ejemplo de la aplicación de este método a una imagen de rostro. En (a) se aplicó el método a una imagen en condiciones óptimas de iluminación, mientras que (b) a una con bajo brillo y contraste. De esta manera, se observa que el resultado de ambas pruebas refleja cierto grado de independencia a estos cambios de iluminación.

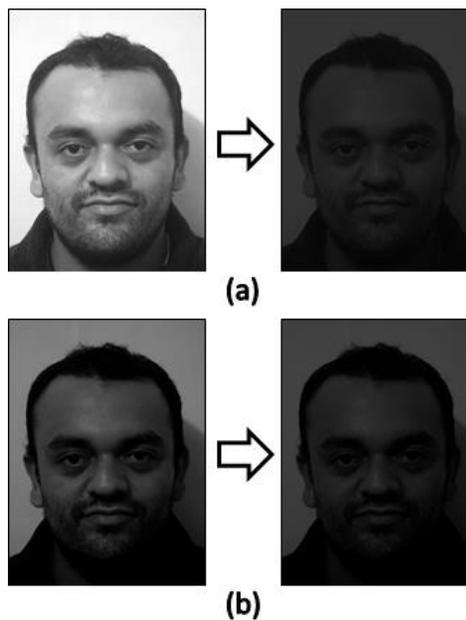


Figura 3.1. Ejemplo de la aplicación del método de normalización del valor del pixel a una imagen de rostro. (a) Aplicado a una imagen en condiciones óptimas de iluminación. (b) Aplicado a una imagen con bajo brillo y contraste.

3.1.2 Ecuación del Histograma

La ecualización del histograma (HE, por sus siglas en inglés) es un procedimiento de aclaración de imágenes. El objetivo de este método es ajustar el brillo y el contraste de una imagen a través de su histograma. Este método, a diferencia de la normalización del valor del pixel, mejora visiblemente la calidad de la imagen.

La aplicación de este método como pre-procesamiento en sistemas de reconocimiento de rostro ha sido usada frecuentemente. Por ejemplo, ha sido utilizado en conjunto con el algoritmo de análisis discriminante lineal (LDA, por sus siglas en inglés) en [34], y en conjunto con el algoritmo de fases propias en [35]. En ambos casos se obtienen resultados mayores al 90% de reconocimiento en el sistema. Siendo un método muy común para el aclaramiento de imágenes, este método también ha sido utilizado como algoritmo de pre-procesamiento en sistemas de reconocimiento de expresiones faciales [39, 40].

Para poder entender el proceso de la ecualización del histograma, es necesario definir el histograma de una imagen. El histograma es un gráfico que relaciona los niveles de intensidad de una imagen y el número de pixeles que poseen tal intensidad, la ecuación (3.3) expresa matemáticamente un histograma [41].

$$h(r_k) = n_k, \quad (3.3)$$

donde r_k es la k -ésima intensidad, y n_k es el número de pixeles con la intensidad r_k .

Por otra parte, la normalización del histograma se define matemáticamente por:

$$p(r_k) = \frac{n_k}{MN}, \quad (3.4)$$

donde M y N representan respectivamente el ancho y el alto de la imagen. Por lo tanto los valores del histograma normalizado $p(r_k)$, se encuentran en un rango de 0 y 1.

Un ejemplo de la representación gráfica de un histograma se muestra en la Figura 3.2. (a) Muestra una imagen en escala de grises [0-255], la cual está formada solamente por 3 diferentes niveles de gris. Mientras que (b) muestra el histograma

normalizado de la imagen (a). Con este ejemplo podemos observar que el histograma normalizado muestra la probabilidad de aparición de cada nivel de gris en relación a la cantidad total de los píxeles de la imagen, representando además una relación porcentual de la frecuencia de aparición de cada nivel de gris en la totalidad de la imagen (PDF). En este caso específico, el histograma muestra la presencia de solamente 3 niveles de gris con un porcentaje de frecuencia de aproximadamente 0.5, 0.4 y 0.1, los cuales corresponden a los colores negro, gris oscuro y gris claro de la imagen (a).

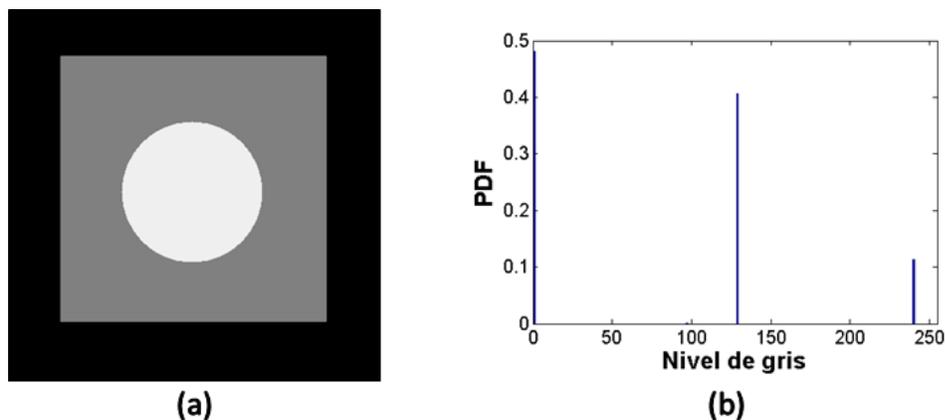


Figura 3.2. Ejemplo del histograma de una imagen. (a) Imagen en escala de grises. (b) Histograma normalizado de la imagen (a).

Entonces, la ecualización del histograma consiste en convertir el histograma de la imagen original a un histograma uniforme, produciendo como resultado un cambio en las condiciones de iluminación en esta. En otras palabras, consiste en tratar de que la distribución de la intensidad de los píxeles de la imagen sea totalmente uniforme. Esta distribución uniforme ideal, es posible alcanzarla solamente cuando se realiza la ecualización del histograma en tiempo continuo. Debido a que las imágenes digitales son evaluadas pixel por pixel es necesario procesarlas en tiempo discreto, por lo tanto solo se llega a una aproximación de la uniformidad en el histograma ecualizado.

Para explicar el proceso llevado a cabo en la ecualización del histograma en tiempo continuo, es necesario suponer que r es una variable que representa los niveles de gris de la imagen inicial, la cual estará constituida por píxeles vistos como cantidades continuas que han sido normalizadas de tal forma que pertenezcan al

intervalo $[0,1]$ representando al valor 0 el color negro y al 1 el blanco. De tal manera que para cada valor de r se define una variable s dada por:

$$s = T(r), \quad (3.5)$$

relación que proporciona un nivel de gris s para cada valor de r . Debe suponerse que la función de transformación cumple con las siguientes condiciones:

- $T(r)$ es una función monótonamente creciente en el intervalo $0 \leq r \leq L - 1$, donde L es el número máximo de niveles de gris.
- $0 \leq T(r) \leq L - 1$, $0 \leq r \leq L - 1$, de manera que $0 \leq s \leq L - 1$.

Con la ecuación (3.5) se define $T(r)$ en su manera inversa:

$$r = T^{-1}(s) \quad 0 \leq s \leq L - 1, \quad (3.6)$$

donde debe suponerse que $T^{-1}(s)$ también cumple con las condiciones antes expuestas.

El nivel de intensidad de una imagen puede verse como variables aleatorias en el intervalo $[0, L-1]$. Al tratarse de variables continuas, los niveles de gris originales y su transformación puede caracterizarse por Funciones de Densidad de Probabilidad (PDF), $p_r(r)$ y $p_s(s)$. De tal manera que la ecuación (3.7) expresa matemáticamente la ecualización del histograma en tiempo continuo [41].

$$s = T(r) = (L - 1) \int_0^r p_r(w) dw, \quad (3.7)$$

donde s es el resultado de la ecualización del histograma, $T(r)$ es la función de transformación continua, p_r es el histograma normalizado, y w es una variable de integración ficticia, a esta ecuación también se le conoce como Función de Distribución Acumulativa (CDF, por sus siglas en inglés).

La Figura 3.3 representa la ecualización del histograma en tiempo continuo, donde a partir de una imagen con un histograma inicial, se alcanza una distribución totalmente uniforme. (a) Muestra el histograma de una imagen en condiciones de iluminación desconocidas. Mientras que (b) representa el resultado de la ecualización del histograma (a), donde se logra una distribución uniforme ideal produciendo una iluminación ideal en la imagen de salida.

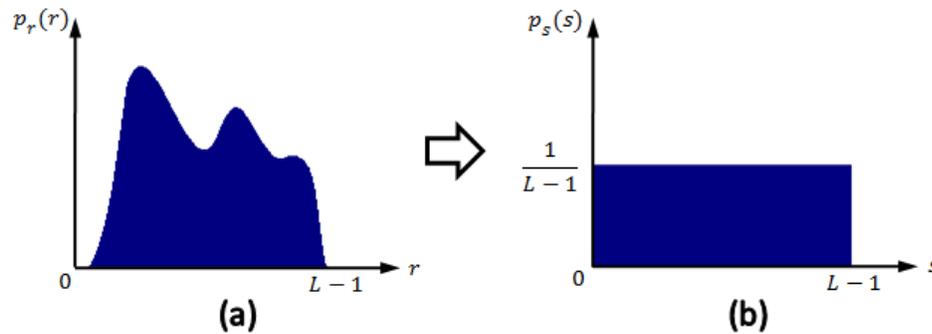


Figura 3.3. Ejemplo de la ecuación del histograma en tiempo continuo. (a) Histograma original. (b) Histograma ecualizado.

En el caso particular de imágenes digitales, como se mencionó previamente, el proceso de la ecualización del histograma debe expresarse en tiempo discreto. La principal razón es porque el análisis de una imagen se realiza pixel por pixel, siendo que no existen pixeles con valores decimales, por lo tanto una función que represente la relación que hay entre los pixeles no puede ser continua. De tal manera que la ecualización del histograma en tiempo discreto está definida por la ecuación (3.8).

$$s_k = T(r_k) = (L-1) \sum_{j=0}^k p_r(r_j) \quad k = 0, 1, 2, \dots, L-1, \quad (3.8)$$

donde s_k es la salida del histograma ecualizado, $T(r_k)$ representa la función de transformación, p_r el histograma normalizado, y k representa todos los niveles de gris en la escala L , utilizando la tradicional escala de grises $L = 256$. La Figura 3.4 muestra el resultado de aplicar la ecualización del histograma a la Figura 3.2.

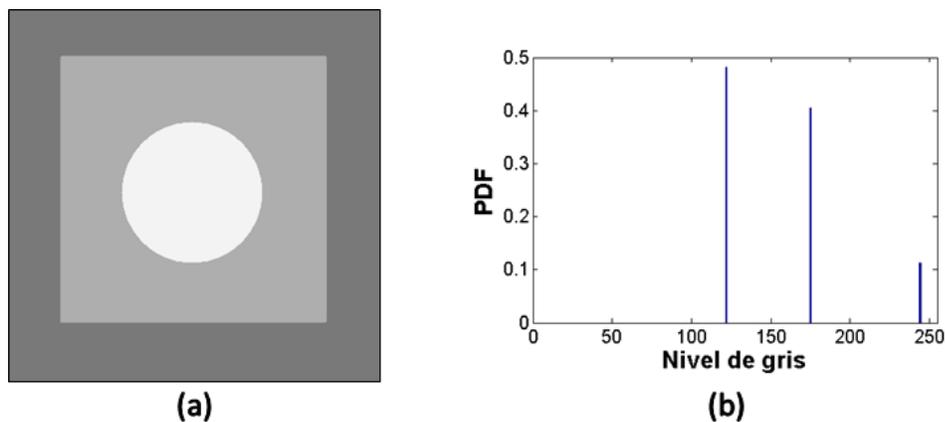


Figura 3.4. Resultado de la ecualización del histograma aplicado a la imagen de la Figura 3.2(a). (a) Imagen ecualizada. (b) Histograma ecualizado.

De la Figura 3.4 se puede observar la capacidad del método para modificar la iluminación en la imagen de salida. Donde a pesar de alcanzar una distribución uniforme en el histograma ecualizado (b), la imagen resultante (a) presenta un cambio significativo en la iluminación con respecto a la imagen original, ilustrada en la Figura 3.2(a). Obteniendo cierta uniformidad en la iluminación, es decir los niveles de gris que aparecen en la imagen tratan de ser lo más semejantes entre sí.

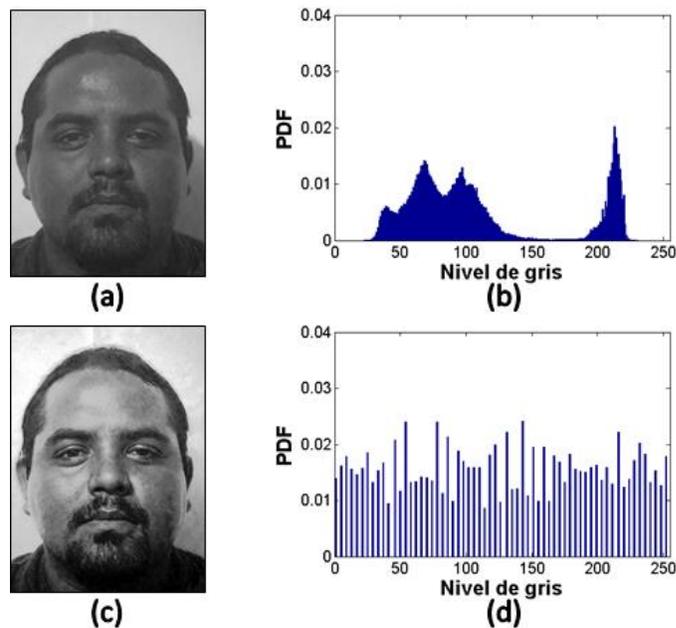


Figura 3.5. Ejemplo de la aplicación de la ecualización del histograma a una imagen de rostro. (a) Imagen del rostro original. (b) Histograma de la imagen (a). (c) Imagen resultante por la ecualización del histograma. (d) Histograma ecualizado.

En la Figura 3.5 se ilustra un ejemplo de la aplicación de la ecualización del histograma a una imagen de rostro. (a) Muestra la imagen de rostro original, la cual presenta un bajo contraste y bajo brillo acentuado del lado derecho, y su histograma normalizado se muestra en (b). Por otro lado, (c) muestra la imagen del rostro después de haber realizado la ecualización del histograma en (a), y por último (d) muestra el histograma de la imagen de rostro después de haber sido ecualizada. En este ejemplo se puede observar que el histograma ecualizado luce significativamente más uniforme. Por consiguiente, este cambio se ve claramente reflejado en la imagen resultante, ya que se obtuvo una imagen más clara, resaltando características faciales difíciles de distinguir en la imagen original.

3.1.3 CLAHE

Como se describió en la sección anterior, la ecualización del histograma es un método que mejora el brillo y el contraste de la imagen original. Este procedimiento comúnmente incrementa el contraste global de la imagen, suponiendo que se obtendrá una distribución totalmente uniforme en la intensidad de los píxeles y así la calidad de imagen será uniforme en todas las regiones de la misma. De manera práctica esta suposición no es válida, ya que una variación muy marcada entre dos regiones de la misma imagen afecta directamente al nivel en el que se va a uniformizar el histograma de manera global.

La ecualización adaptativa del histograma (AHE, por sus siglas en inglés) es una técnica alternativa a la ecualización del histograma que supone solucionar el problema de la diferencia de niveles de contraste que causan dos regiones de la misma imagen [42]. Esta técnica consiste en dividir la imagen en determinado número de regiones, para después aplicar la ecualización del histograma de manera independiente a cada una de ellas.

Incluso en algunos casos el método AHE no resuelve del todo el problema, esto ocurre cuando la distribución de la escala de grises se encuentra muy localizada. Es decir, con una frecuencia de píxeles alta solo en ciertas zonas de la escala, lo que no es conveniente para transformar imágenes de muy bajo contraste. En estos casos en la curva de análisis pueden existir segmentos con altas pendientes, problema que se puede resolver mediante la limitación del contraste para la ecualización del histograma. En consecuencia, S.M. Piezer *et al.* propusieron en [42] la combinación de la limitación de contraste junto con la ecualización adaptativa del histograma, resultando el método de ecualización adaptativa del histograma con limitación de contraste (CLAHE, por sus siglas en inglés) [43].

CLAHE es considerado un método eficiente para el aclaramiento de imágenes, ya que es capaz de incrementar el contraste en determinadas regiones de la imagen, y también de limitar el contraste en otras. Como consecuencia se produce una mayor definición en toda la imagen, además de acentuar significativamente detalles locales sin afectar el nivel de contraste global de la imagen, ya que este es limitado desde un principio.

El método CLAHE también ha sido utilizado en etapas de pre-procesamiento para sistemas de reconocimiento de rostros. En [36] se utiliza CLAHE en conjunto con el algoritmo de modelos ocultos de Markov (HMM, por sus siglas en inglés) y en [37] se utiliza para mejorar las imágenes infrarrojas de rostros. En ambos casos los sistemas que utilizan CLAHE obtuvieron resultados aceptables (en el mejor de los casos mayores al 90% de reconocimiento).

El procedimiento del método CLAHE se muestra en la Figura 3.6. Donde podemos observar que este método es un tanto complejo, de manera que es importante tomar en cuenta que este algoritmo de aclaración de imágenes tiene un costo computacional mayor a los analizados previamente. A grandes rasgos, el método consiste en, dividir la imagen de entrada en regiones del mismo tamaño, para después calcular el histograma independientemente de cada región, el cual debe limitarse usando un límite de contraste (β), con el que se redistribuirá el histograma para enseguida aplicar el método de ecualización del histograma en este nuevo “histograma redistribuido”, para finalmente unificar las regiones por medio de la interpolación bilineal.

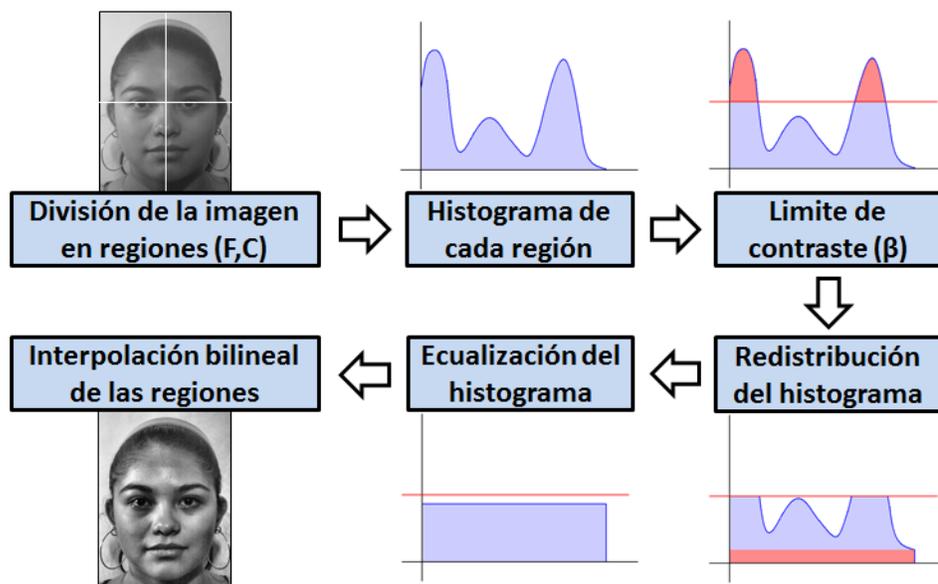


Figura 3.6. Procedimiento para la aplicación del método CLAHE.

Como primer paso, la imagen debe ser dividida en ventanas del mismo tamaño. La división debe realizarse tanto en el eje x, como en el eje y. De manera que

tomando en cuenta una imagen de entrada $I(x, y)$ de tamaño $M \times N$, esta se debe dividir en $F \times C$ ventanas de tamaño $m \times n$, donde $mF \times nC$ respeta el tamaño original de la imagen de entrada.

A cada una de las ventanas de tamaño $m \times n$ se les debe calcular el histograma de manera independiente, para este paso se debe utilizar la ecuación (3.3). Enseguida, el límite de corte debe ser calculado. Para esta función existen dos caminos diferentes. Uno está enfocado en ayudar a mejorar imágenes en condiciones extremas de iluminación, es decir con muy bajo contraste y muy bajo brillo, como el caso de imágenes médicas. Para este particular caso, el límite de corte se define por:

$$\beta = \frac{MN}{L} \left(1 + \frac{\alpha}{100} (s_{max} - 1) \right), \quad (3.9)$$

donde β es el límite de corte, MN representa el tamaño de la imagen, L es la escala de grises utilizada, α es el factor de corte, y s_{max} es el límite máximo de corte.

Por lo tanto, en este caso, el límite de corte se encuentra principalmente relacionado con el factor de corte. De manera que si $\alpha = 0$, el límite de corte sería exactamente igual a (MN/L) . Mientras que si $\alpha = 100$ el límite de corte se convertiría en $(s_{max} * MN/L)$, siendo este el máximo límite de corte permitido. De forma que, s_{max} debe asignarse de manera correcta dependiendo del tipo de imágenes que se requieran aclarar, obteniendo un valor adecuado solamente por medio de la experimentación.

Esta manera de calcular el límite de corte es utilizada en [44, 45], donde además se menciona que el valor ideal de s_{max} para imágenes de rayos X es igual a 4. Por otra parte, cuando no es necesario un límite de corte máximo, es decir cuando no se utilicen imágenes en condiciones extremas de iluminación, la forma de calcular el límite de corte consiste en normalizar el valor de β en un rango de [0,1]. Donde $\beta = 0$ no representa cambio alguno en la imagen de entrada, mientras que con $\beta = 1$ se obtiene un resultado similar al de la aplicación de AHE, es decir no se presenta un límite de corte en los histogramas ecualizados. Es importante mencionar que esta forma de definir el límite de corte es la más convencional, presentando un buen resultado en la mayoría de las imágenes [46].

En el siguiente paso del método, para la redistribución de los pixeles, previamente se necesitan recortar los pixeles que sean mayores al límite de corte, enseguida la redistribución es definida por:

$$k = \frac{\sum p}{L}, \quad (3.10)$$

donde p representa los pixeles que están sobre el límite de corte, y k es la redistribución de estos pixeles para cada nivel de gris.

Una vez obteniendo los histogramas redistribuidos en base al límite de corte, el siguiente paso es realizar la ecualización del histograma en cada uno de ellos, proceso definido por la ecuación (3.8). Por último, para unir nuevamente las regiones de la imagen, se utiliza la interpolación bilineal, definida por:

$$m(i) = a[bm_{--}(i) + (1-b)m_{+-}(i)] + (1-a)[bm_{-+}(i) + (1-b)m_{++}(i)], \quad (3.11)$$

donde:

$$a = \frac{y - y_-}{y_+ - y_-} \quad b = \frac{x - x_-}{x_+ - x_-}, \quad (3.12)$$

siendo m el pixel bajo análisis, (i) representa la intensidad del pixel, $m_{--}, m_{+-}, m_{-+}, m_{++}$ corresponden a los pixeles vecinos superior izquierdo, superior derecho, inferior izquierdo e inferior derecho respectivamente, y (x, y) representan las coordenadas en la imagen de dichos pixeles.

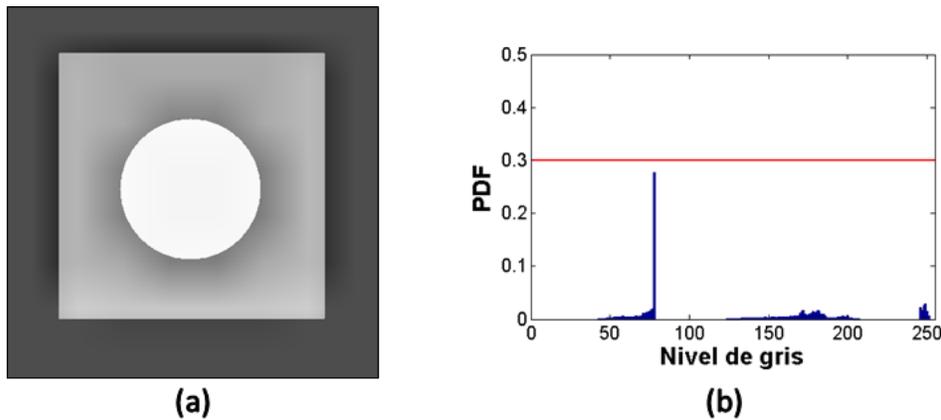


Figura 3.7. Resultado de aplicar CLAHE con $\beta=0.3$ y $(F,C)=2,2$ en la imagen de la Figura 3.2(a).
 (a) Imagen ecualizada. (b) Histograma ecualizado.

Retomando el ejemplo de la Figura 3.2, la Figura 3.7 ilustra el resultado de aplicar CLAHE en ella, usando como parámetros $\beta = 0.3$ y $(F, C) = (2, 2)$. (a) Muestra el resultado al aplicar CLAHE en la Figura 3.2(a), mientras que (b) muestra el histograma ecualizado por CLAHE. En esta figura podemos observar que el histograma se ecualizó bajo un límite de contraste de 0.3, marcado con una línea roja en la Figura 3.7(b). Por lo tanto, la imagen resultante presenta un cambio en la iluminación particularmente diferente al mostrado por el resultado de la aplicación de HE en la Figura 3.4.

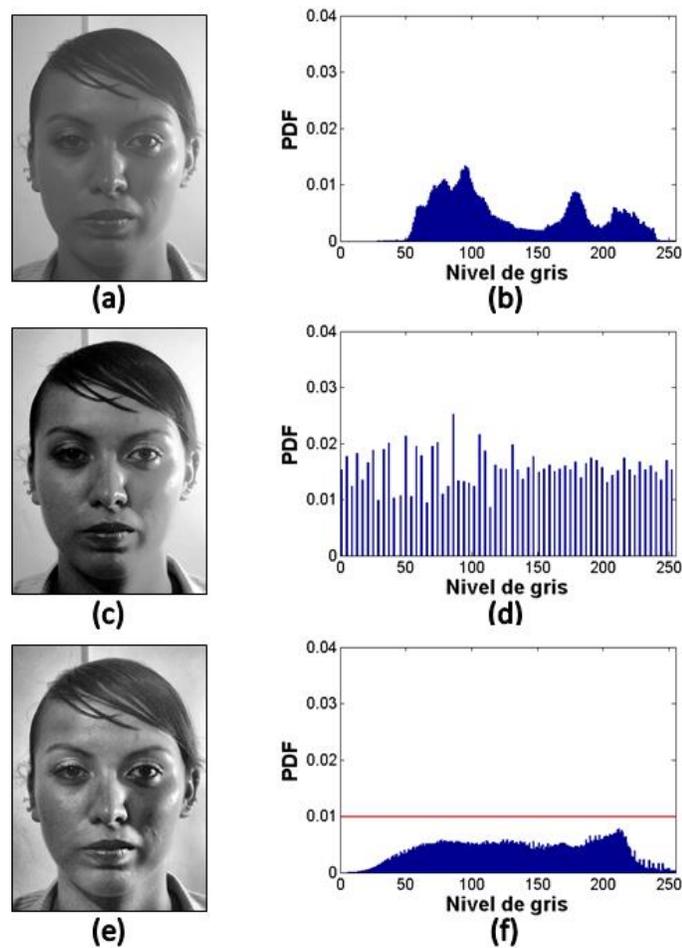


Figura 3.8. Diferencias entre la aplicación de HE y CLAHE a una imagen de rostro. (a) Imagen del rostro original. (b) Histograma de la imagen (a). (c) Imagen resultante por la ecualización del histograma (HE). (d) Histograma ecualizado por HE. (e) Imagen resultante por CLAHE(8,6) y $\beta=0.01$. (f) Histograma ecualizado por CLAHE.

La Figura 3.8 muestra las diferencias entre la aplicación del algoritmo de HE y CLAHE a una imagen de rostro. (a) Ilustra la imagen del rostro original bajo

condiciones de iluminación deficientes, (b) muestra el histograma de dicha imagen, (c) muestra el resultado de aplicar HE en (a), (d) muestra el histograma (b) después de haberle aplicado la HE, mientras que (e) muestra la imagen después aplicar CLAHE con $\beta = 0.01$ y $(F, C) = (8,6)$ en (a), por último, (f) muestra su histograma resultante. Con este ejemplo podemos observar los beneficios que ofrece el método de CLAHE para la aclaración de imágenes de rostros. Donde, la imagen original presenta serios problemas de iluminación, específicamente de alto brillo en la parte derecha del rostro, problema que persiste aun cuando se aplica la HE, mientras que en el resultado de CLAHE este problema de iluminación es prácticamente solucionado, puesto que gracias a la aplicación de este método los detalles de cada región de la imagen son resaltados significativamente.

Es importante mencionar que para un rendimiento óptimo en el uso de CLAHE, es necesario definir correctamente no solo el límite de corte, sino también el número de divisiones para la imagen de entrada. De manera que si se elige un mayor número de divisiones, los cambios en la iluminación serán más notorios, por otro lado, si no se utilizan las divisiones, $(F, C) = (1,1)$, el resultado sería el mismo al aplicar la HE con una limitación de contraste. Por lo tanto, la forma más simple de utilizar CLAHE es utilizando una división de $(F, C) = (2,2)$, es decir CLAHE(2,2).

3.2 Métodos de Extracción de Características

Como se mencionó anteriormente, la extracción de características es una etapa en los sistemas biométricos, que consiste en transformar un conjunto de atributos fisiológicos a un conjunto de atributos matemáticos distintivos y reproducibles. En el caso de los sistemas de reconocimiento de rostros, las funciones de esta etapa consisten principalmente en: reducir las dimensiones del espacio original de la imagen de entrada, eliminar la influencia de las variaciones no relacionadas a cambios de identidad, y maximizar la diferencia entre rostros de distintas personas.

Para cualquier sistema de reconocimiento de rostros la etapa de extracción de características juega un papel fundamental, ya que además de estimar un conjunto de parámetros que describan correctamente las imágenes de entrada, estos deben de permitir la minimización o posible eliminación de la influencia de las variaciones relacionadas con los cambios de iluminación, cambios en tipos de corte de cabello,

oclusión parcial, entre otras. Por lo tanto la eficiencia de un sistema de reconocimiento de rostros depende en gran medida del rendimiento del algoritmo utilizado para la etapa de extracción de características.

Uno de los primeros algoritmos usados en esta etapa para un sistema de reconocimiento de rostros está basado en el algoritmo de Sirovich y Kirby [9]. El cual utiliza los valores y vectores propios generados por el algoritmo de análisis de componentes principales (PCA, por sus siglas en inglés), razón por la cual fue nombrado método de Caras Propias (Eigenfaces, del idioma inglés) [10].

En las últimas décadas varios métodos de extracción de características han sido propuestos. Los basados en el principio de Caras Propias, que utilizan el método de PCA para caracterizar las imágenes de entrada [47, 48], el método de Fisherfaces [49, 50], que utiliza el algoritmo de análisis discriminante lineal (LDA) y el método de Laplacianfaces [51], que utiliza proyecciones locales. Todos ellos logrando tasas razonablemente altas de reconocimiento bajo ciertas condiciones en las pruebas.

Por otro lado, el uso de transformaciones de frecuencia en las imágenes como es el caso de la transformada discreta de Haar [52], han sido métodos de extracción de características ampliamente utilizados. En este campo tenemos también los casos de la transformada discreta de coseno [53-55], la transformada discreta de Gabor [56-60] y la transformada discreta de Wavelet [61-64], métodos que presentan tasas de reconocimiento mayores al 90%.

De entre los métodos basados en el principio de Caras Propias, el algoritmo de Fases Propias (Eigenphases, del idioma inglés) parece ser una alternativa atractiva, ya que proporciona mejores resultados que otros métodos [65, 66] con una complejidad de cálculo razonable. Es por eso que recientemente ha sido objeto de análisis, proponiendo combinaciones de este con diferentes algoritmos de pre-procesamiento y sugiriendo el uso de sub-bloques en su procesamiento [32, 33, 35].

Recientemente se ha introducido el algoritmo descriptor de texturas llamado patrones binarios locales (LBP, por sus siglas en inglés) [67] como un nuevo método de extracción de características para sistemas de reconocimiento de rostros [68, 69]. Durante los últimos años este algoritmo ha atraído la atención de varios investigadores, y gracias a sus principales ventajas se han propuesto diversas

modificaciones referentes a los sistemas reconocimiento de rostros [70-73]. Ventajas que ofrecen una invariancia a los cambios de niveles de gris y una eficiencia computacional, las cuales hacen que sea un método adecuado para el reconocimiento de rostros, permitiendo la exclusión de la etapa de pre-procesamiento.

Es por estas razones que en esta sección se analizan detalladamente los métodos de extracción de características de: Caras Propias (Eigenfaces), Fases Propias (Eigenphases) y Parones Binarios Locales (LBP).

3.2.1 Caras Propias (Eigenfaces)

El método de Caras Propias, propuesto en [10], está basado en la técnica para representar imágenes de rostros propuesta por Sirovich y Kirby [9]. La cual consiste en, a partir de un conjunto de imágenes de rostros, calcular el mejor sistema de coordenadas para la compresión de las imágenes, en el que cada coordenada es en realidad una imagen llamada Cara Propia (eigenface).

En términos matemáticos, el método de Caras Propias [10] busca encontrar los componentes principales de la distribución de rostros o de vectores propios de la matriz de covarianza del conjunto de imágenes de rostros, considerando una imagen como un punto en un espacio de alta dimensionalidad. Estos vectores propios se consideran como un conjunto de características que describen la variación entre las imágenes de rostros, es decir, cada imagen localizada contribuye con cada vector propio, tal que es posible visualizar este vector propio como un tipo de rostro fantasmal conocido como cara propia.

De manera general, este método requiere de dos fases: inicialización y reconocimiento. Para la fase de inicialización se requiere en primera instancia adquirir un conjunto inicial de imágenes de rostros (imágenes de entrenamiento). Enseguida se deben calcular las caras propias del conjunto de entrenamiento, preservando solamente las M mejores imágenes correspondientes a los valores propios más altos, las cuales definen el “espacio de rostro”. Finalmente, se debe calcular la correspondiente distribución del espacio de rostro M -dimensional para cada individuo conocido (imagen de entrenamiento), proyectando su imagen del rostro sobre el espacio de rostro.

Una vez teniendo inicializado el sistema, para la fase de reconocimiento, se debe calcular un conjunto de pesos de las M caras propias de la imagen de entrada (imagen de prueba). Seguidamente se determina si la imagen de prueba es un rostro, verificando si la imagen es lo suficientemente cercana al espacio de rostro. En base a determinar que la imagen es de un rostro, por último, se clasifica el patrón de pesos como un rostro ya sea conocido o desconocido.

Para el cálculo del método de Caras Propias, se debe tomar en cuenta que $\Gamma(x, y)$ representa una imagen de rostro de dimensiones $N \times N$, la cual puede ser considerada como un vector de dimensión N^2 . De forma que una imagen de un tamaño de 256×256 pixeles se convierte en un vector de tamaño 65 536, en otras palabras, un punto en un espacio de 65 536 dimensiones.

Dado un conjunto de imágenes de entrenamiento $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M$, el rostro promedio del conjunto se define como:

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n. \quad (3.13)$$

De forma que cada rostro se diferencia del promedio por el vector:

$$\Phi_n = \Gamma_n - \Psi \quad n=1,2,\dots,M. \quad (3.14)$$

Posteriormente, al conjunto de vectores obtenido por la ecuación (3.14) se le aplica el método de análisis de componentes principales (PCA), para lo que es necesario calcular su matriz de covarianza, definida por:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T, \quad (3.15)$$

donde:

$$A = [\Phi_1, \Phi_2, \dots, \Phi_M] \quad (3.16)$$

La matriz C es de dimensión $N^2 \times N^2$ y determinar los N^2 vectores propios y valores propios es una tarea imposible para tamaños típicos de imágenes, por esta razón, se necesita de un método factible para encontrar los vectores propios. Si el número de puntos de datos en el espacio de imagen es menor que la dimensión del espacio total ($M < N^2$), habrá únicamente $M - 1$, en lugar de N^2 vectores propios.

De esta forma, podemos considerar que los vectores propios v_i de $A^T A$ tales que:

$$A^T A v_i = \mu_i v_i. \quad (3.17)$$

Multiplicando ambos lados por A tenemos:

$$A A^T A v_i = \mu_i A v_i, \quad (3.18)$$

de donde se observa que $A v_i$ son los vectores propios de $A A^T$ y que $C = A A^T$.

Por lo tanto, estos vectores definen las combinaciones lineales de las M imágenes del conjunto de entrenamiento para formar las caras propias (u_l), definidas por:

$$u_l = \sum_{k=1}^M v_k^l \Phi_k \quad l = 1, 2, \dots, M. \quad (3.19)$$

Con este análisis los cálculos se reducen considerablemente, del orden del número de pixeles de las imágenes (N^2) al orden del número de imágenes en el conjunto de entrenamiento (M).

Una vez que las caras propias han sido calculadas, cada imagen es proyectada en el espacio de rostro mediante la siguiente ecuación:

$$\omega_k = u_k^T (\Gamma - \Psi) \quad k = 1, 2, \dots, M, \quad (3.20)$$

donde los pesos de esta proyección forman un vector definido por:

$$\Omega^T = [\omega_1, \omega_2, \dots, \omega_M], \quad (3.21)$$

que describe la contribución de cada cara propia en la imagen de rostro de entrada. En otras palabras, Ω representa el vector de características del rostro que aparece en la imagen de entrada del sistema.

Por último, el método más simple para determinar cuál de los rostros almacenados proporciona la mejor descripción de la imagen de entrada, es encontrando el rostro k que minimice la distancia Euclidiana:

$$\varepsilon_k^2 = \|\Omega - \Omega_k\|^2, \quad (3.22)$$

donde Ω_k es un vector que describe la k -ésima clase de rostro.

3.2.2 Fases Propias (Eigenphases)

El método de Fases Propias (Eigenphases, del idioma inglés), fue propuesto por Savvides *et al.* [48] como una mejora directa al algoritmo de Caras Propias. Este método está basado en el proceso utilizado por el algoritmo convencional de caras propias, con la principal diferencia de utilizar el espectro de fase de las imágenes de rostros.

De manera general, el método de Fases Propias, de la misma forma que el de Caras Propias, se divide en dos etapas fundamentales: la de inicialización (entrenamiento) y la de reconocimiento. Para la etapa de entrenamiento se requiere en primera instancia adquirir un conjunto inicial de imágenes de entrenamiento. Enseguida, se debe extraer el espectro de fase (utilizando la transformada discreta de Fourier) para poder calcular las fases propias del conjunto de entrenamiento (utilizando el algoritmo de PCA), para después almacenar solamente los M mejores espectros de fase correspondientes a los valores propios más altos, los cuales definen el “espacio de fases”. Finalmente, se debe calcular la correspondiente distribución del espacio de fase M -dimensional para cada individuo del conjunto de entrenamiento, proyectando su espectro de fase sobre el espacio de fases, a diferencia del método de Caras Propias que lo hace sobre el espacio de rostro.

Para la etapa de reconocimiento, se debe calcular un conjunto de pesos de las M fases propias de la imagen de prueba (utilizando previamente la transformada de Fourier en la imagen de entrada), de manera que el vector que contiene estos pesos es llamado “vector característico”. Por lo tanto, con el conjunto de vectores característicos de las imágenes de prueba se realizara la clasificación del sistema, utilizando un algoritmo específicamente para esta tarea, siendo el más sencillo el que utiliza la mínima distancia Euclideana.

En conclusión, el método de Fases Propias depende directamente de la utilización de la transformada discreta de Fourier, para la obtención de los espectros de fase, y del algoritmo de análisis de componentes principales (PCA), para la obtención de las fases propias. Es por eso que estos dos procesos se presentan individualmente en la siguientes subsecciones.

3.2.2.1 Extracción del Espectro de Fase

La propuesta de utilizar el espectro de fase en el método de Fases Propias, está fundamentada en el experimento de Oppenheim [74], donde se demuestra que la fase de la transformada de Fourier conserva información más relevante que la magnitud [75]. Por lo tanto, la información de la fase es la más importante en la representación de una señal de 2D en el dominio de Fourier.

En la Figura 3.9 se ilustra un ejemplo del experimento de Oppenheim. Donde en primer lugar, se aplica la transformada discreta de Fourier a dos imágenes de rostros diferentes, rostro 1 (a) y rostro 2 (e). De manera que, la imagen (d) fue reconstruida utilizando el espectro de fase del rostro 1 (c) y el espectro de magnitud del rostro 2 (f), mientras que la imagen (h) fue reconstruida usando el espectro de fase del rostro 2 (g) y el espectro de magnitud del rostro 1 (b). Por lo tanto, en este ejemplo se demuestra que el espectro de fase guarda la información más importante de las imágenes de rostros, ya que el rostro que más predomina en las imágenes reconstruidas pertenece al del espectro de fase utilizado en el proceso de reconstrucción.

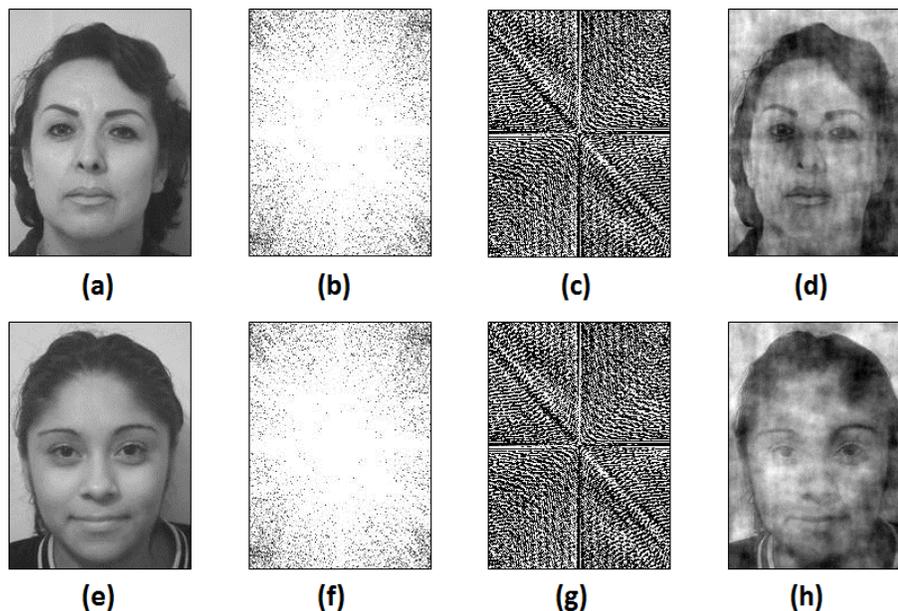


Figura 3.9. Experimento de Oppenheim. (a) Imagen del rostro 1. (b) Espectro de magnitud del rostro 1. (c) Espectro de fase del rostro 1. (d) Imagen reconstruida usando (c) y (f). (e) Imagen del rostro 2. (f) Espectro de magnitud del rostro 2. (g) Espectro de fase del rostro 2. (h) Imagen reconstruida usando (b) y (g).

Para la obtención del espectro de fase, es necesario aplicar la transformada de Fourier, definida por:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt. \quad (3.23)$$

Pero debido a que se están empleando imágenes digitales, la transformada de Fourier en tiempo continuo no es aplicable, de manera que se debe utilizar la transformada discreta de Fourier (DFT, por sus siglas en inglés). Teniendo en cuenta que las imágenes digitales son representadas por una señal 2-D, se debe utilizar la DFT en 2D. Entonces, considerando la entrada de una señal discreta 2-D $x[m, n]$ con dimensiones $M \times N$, denotamos su transformada de Fourier como $X[k, l]$, definida por:

$$X[k, l] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x[m, n] \exp \frac{-i2\pi km}{M} \exp \frac{-i2\pi l n}{N}. \quad (3.24)$$

Normalmente nos interesan solo los elementos reales de una función. Sin embargo, la transformada de Fourier de una función real es casi siempre compleja, por lo tanto, representando la transformada de Fourier como $F(u)$, tenemos:

$$F(u) = R(u) + I(u). \quad (3.25)$$

De esta manera la magnitud de la transformada de Fourier se calcula con:

$$|F(u)| = [R^2(u) + I^2(u)]^{\frac{1}{2}}. \quad (3.26)$$

Y la fase con:

$$\phi(u) = \arctan \left[\frac{I(u)}{R(u)} \right]. \quad (3.27)$$

Por lo tanto, tomando en cuenta una imagen de rostro $I(x, y)$ y su transformada de Fourier como $F(u, v)$, el espectro de fase de esta imagen se puede expresar de manera general como:

$$\phi(u, v) = \tan^{-1} \left[\frac{I(F(u, v))}{R(F(u, v))} \right]. \quad (3.28)$$

3.2.2.2 Análisis de Componentes Principales (PCA)

El método de Análisis de Componentes Principales (PCA, por sus siglas en inglés), es una herramienta estándar en el análisis de datos ampliamente utilizada, ya que es un método no paramétrico eficaz para extraer la información más relevante de un conjunto de datos [76].

Una de las principales ventajas del PCA es la reducción de la dimensionalidad de los datos extraídos sin mucha pérdida de información. Es decir, tiene la capacidad de comprimir la información del conjunto de datos original, reduciendo el número de dimensiones de este y finalmente obteniendo solo la información más importante.

Gracias a sus múltiples ventajas, el PCA es utilizado en muchas y diversas aplicaciones [77-79], siendo una de las más representativas la de extracción de las características de un conjunto de imágenes faciales para sistemas de reconocimiento de rostros [10, 48, 80].

La idea principal del PCA, en la extracción de características para sistemas de reconocimiento de rostros, consiste en encontrar el vector que mejor describa la distribución de las imágenes de rostros dentro del espacio total de la imagen. El conjunto de estos vectores define el sub-espacio de imágenes de rostro llamado “espacio de rostro” [10]. En el caso específico del método de Fases Propias, cada vector representa una combinación lineal de los espectros de fase de las imágenes de entrenamiento, y debido a que estos vectores son los vectores propios de la matriz de covarianza correspondiente a los espectros de fase del conjunto de entrenamiento, fueron llamados Fases Propias [48].

El procedimiento para calcular el PCA consiste en, inicialmente obtener el conjunto de datos a analizar:

$$x^i = [x_1^i, x_2^i, \dots, x_N^i]^T \quad i = 1, 2, \dots, P, \quad (3.29)$$

donde P representa el número de imágenes de entrenamiento. Enseguida se debe extraer el promedio de este conjunto de imágenes a cada imagen de entrenamiento como se muestra a continuación:

$$\bar{x}^i = x^i - m_i, \quad (3.30)$$

donde:

$$m_i = \frac{1}{P} \sum_{n=1}^P x_n^i. \quad (3.31)$$

Después, se deben combinar los vectores resultantes para formar una matriz a la que posteriormente se calculará su matriz de covarianza. Esta combinación se define por:

$$\bar{X} = [\bar{x}^1, \bar{x}^2, \dots, \bar{x}^P] \quad (3.32)$$

Y su matriz de covarianza se calcula con:

$$\Omega = \overline{XX^T}, \quad (3.33)$$

donde Ω es la matriz de covarianza, la cual tiene hasta P vectores propios asociados con los valores propios diferentes de cero, asumiendo que $P < N$.

Una vez obteniendo la matriz de covarianza, el paso siguiente es calcular los vectores y valores propios:

$$\Omega V = \Lambda V, \quad (3.34)$$

donde V representa los vectores propios, y Λ los valores propios, asociados a la matriz de covarianza (Ω).

Por último, se deben ordenar de mayor a menor todos los vectores propios $v_i \in V$ de acuerdo a su correspondiente valor propio $\lambda_i \in \Lambda$. Después de este ordenamiento se deben conservar solo los vectores propios con los valores propios diferentes de cero. Esta matriz de vectores propios es el sub-espacio V , donde cada columna de V es un vector propio:

$$V = [v_1, v_2, \dots, v_P] \quad (3.35)$$

En términos generales la matriz V representa la información dominante del conjunto de datos x . Es importante mencionar que $V < x$, es decir que, las dimensiones de la matriz de información dominante (características dominantes) es menor a la del conjunto de datos originales (imágenes de entrenamiento), debido a la eliminación de los vectores propios cuyos valores propios asociados eran igual a cero.

3.2.3 Patrones Binarios Locales (LBP)

El algoritmo de Patrones Binarios Locales (LBP, por sus siglas en inglés) propuesto por Ojala *et al.* [67], es uno de los mejores métodos descriptores de textura. Este algoritmo combina características de métodos estadísticos y estructurales del análisis de texturas, siendo muy eficaz para caracterizar la regularidad y aleatoriedad de los elementos texturales repetitivos de dichas imágenes. Características como la independencia en el promedio de la intensidad de los píxeles han ocasionado que LBP se haya propuesto como algoritmo extractor de características en sistemas de reconocimiento de rostros [68, 69].

La propuesta original para la aplicación de LBP en sistemas de reconocimiento de rostros emplea el histograma de las imágenes LBP (también llamadas “espectros de textura”) para formar el vector característico [68]. A partir de esta propuesta, se han sugerido varias modificaciones con el objetivo de incrementar el porcentaje de reconocimiento, las cuales están principalmente enfocadas a la utilización de sub regiones para el cálculo de LBP, utilización del LBP de manera holística, entre otras [70-73]. El grado de aceptación y popularidad de este método de extracción de características ha propiciado su implementación incluso en sistemas de reconocimiento de expresiones faciales, logrando buenos resultados [38, 81, 82].

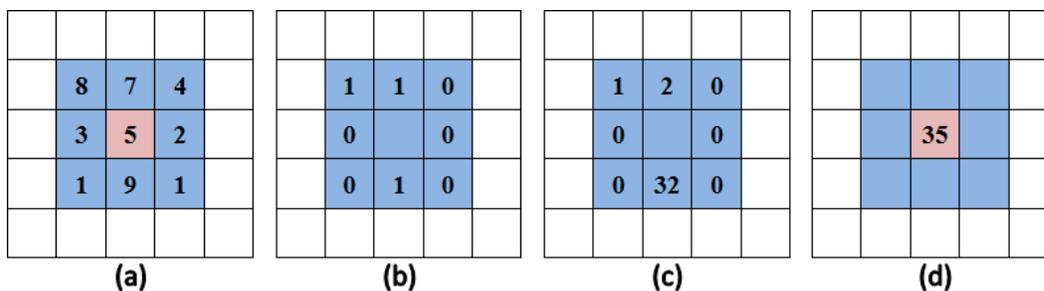


Figura 3.10. Proceso del método original de LBP. (a) Valores de los vecinos alrededor del píxel central. (b) Comparación de cada vecino con el píxel central. (c) Sustitución de cada valor de la comparación por su correspondiente valor 2^p . (d) Suma y sustitución del píxel central por el valor resultante.

El método original de LBP, ilustrado en la Figura 3.10, utiliza ventanas de 3x3 píxeles de una imagen, representando una vecindad alrededor del píxel central, como se muestra en (a), donde se usa el píxel central como umbral para comparar los valores de sus 8 vecinos. Los píxeles cuyos valores sean menores que el umbral

deben ser etiquetados con 0 y los que sean mayores al umbral deben etiquetarse con 1, como se muestra en (b). Enseguida, las etiquetas de los píxeles son multiplicadas por 2^P , donde $0 \leq P \leq 7$ representa la posición de cada píxel en la vecindad, como se observa en (d). Finalmente, se suman los valores resultantes para obtener la etiqueta del píxel central de esa vecindad, como se observa en (d). Este método produce 256 valores posibles para la etiqueta del píxel central. Este proceso se repite para toda la imagen, produciendo una matriz de etiquetas LBP (imagen LBP) que es usada para estimar el vector característico de la imagen del rostro. En la propuesta original [68], después de obtener la matriz de etiquetas LBP, el histograma de sus coeficientes es estimado, con el cual se representa el vector característico de la imagen bajo análisis.

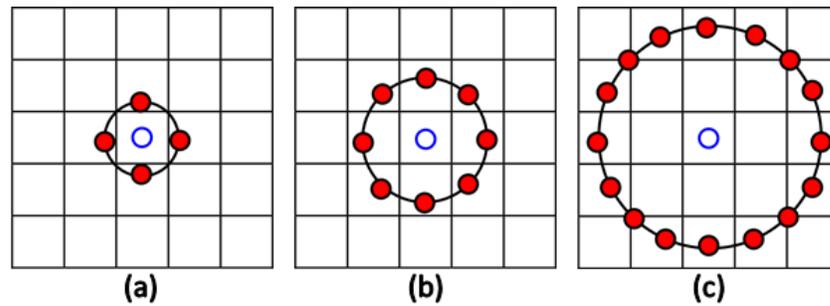


Figura 3.11. Ejemplos del método LBP con diferentes valores de P y R . (a) $P=4$ y $R=0.5$. (b) $P=8$ y $R=1$. (c) $P=16$ y $R=2$.

El método LBP también se puede obtener mediante el uso de una vecindad circular detonada por (P,R) , donde P representa el número de puntos de muestreo (vecinos) y R es el radio de la circunferencia. La Figura 3.11 muestra un ejemplo de tres variantes del algoritmo de LBP con vecindarios circulares. Es importante aclarar que, cuando el punto de muestreo no se encuentra en el centro de un píxel, su valor se obtiene mediante la interpolación bilineal. A continuación se explica el procedimiento que debe llevarse a cabo para la utilización del método LBP.

Considerando una imagen monocromática $I(x,y)$, donde g_c representa la intensidad del píxel central en una vecindad dada, cuyos píxeles se encuentran en una vecindad circular de radio R con P píxeles equidistantes con intensidades definidas por g_p , se tiene:

$$g_p = I(x_p, y_p) \quad p = 1, 2, \dots, P-1, \quad (3.36)$$

$$x_p = x_c + R \cos(2\pi p/P), \quad (3.37)$$

y:

$$y_p = y_c - R \sin(2\pi p/P), \quad (3.38)$$

son las coordenadas de los puntos de muestreo. Suponiendo que la textura local de la imagen $I(x, y)$ se caracteriza por la distribución conjunta de la intensidad de los $P - 1$ píxeles ($P > 0$):

$$T = t(g_c, g_0, g_1, \dots, g_{P-1}). \quad (3.39)$$

Sin pérdida de información, T puede expresarse restando g_c a cada valor de los píxeles g_p de la vecindad:

$$T = t(g_c, g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c). \quad (3.40)$$

Enseguida, suponiendo que el píxel central g_c es estadísticamente independiente de las diferencias $(g_p - g_c)$, la distribución T dada por la ecuación (3.40) se convierte en:

$$T \approx t(g_c) t(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c), \quad (3.41)$$

donde el primer factor $t(g_c)$ es la distribución de la intensidad sobre $I(x, y)$. Sin embargo, desde el punto de vista del análisis de los patrones de textura local, este factor no proporciona información útil. Por lo tanto la distribución T se puede aproximar por la distribución conjunta de las diferencias de la siguiente manera:

$$T \approx t(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c), \quad (3.42)$$

este operador puede ser utilizado para modelar la textura local. Por lo tanto, para obtener la invariancia con respecto a cualquier transformación de escala gris monótona, sólo los signos de las diferencias pueden ser considerados:

$$T \approx t(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{P-1} - g_c)), \quad (3.43)$$

donde:

$$s(x) = \begin{cases} 1, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0 \end{cases} \quad (3.44)$$

De esta manera, asignando un peso binomial de 2^P a cada componente $s(g_p - g_c)$, y transformando las diferencias de la vecindad en un único código LBP, alrededor de g_c , se deduce que:

$$LBP_{P,R}(g_c) = \sum_{p=1}^{P-1} s(g_p - g_c) 2^p. \quad (3.45)$$

En la práctica, esto significa que los signos de las diferencias en una vecindad se pueden interpretar como un número binario de P -bits, resultando en 2^P valores diferentes para el código LBP. De esta manera, $LBP_{P,R}(x, y)$ representa la matriz de etiquetas LBP (también llamada imagen LBP), de la imagen original $I(x, y)$.

Como se mencionó previamente, el algoritmo propuesto de LBP para la extracción de características en sistemas de reconocimiento de rostros, calcula el histograma de la matriz de etiquetas LBP, el cual es utilizado como vector característico [68]. El cálculo del histograma de una imagen LBP denotada como $LBP_{P,R}(x, y)$ se realiza de la siguiente forma:

$$LBPh_i = \sum_{x,y} w\{LBP_{P,R}(x, y), i\} \quad i = 0, 1, \dots, 2^P - 1, \quad (3.46)$$

donde:

$$w\{A, B\} = \begin{cases} 1, & \text{si } A = B \\ 0, & \text{otro caso} \end{cases} \quad (3.47)$$

donde LBPh representa el histograma de la imagen LBP.

Es importante aclarar que las dimensiones del histograma (LBPh) dependerán del número de puntos de muestro utilizados en el proceso del LBP. De manera que para una imagen LBP con $P=8$ y $R=1$, su histograma LBPh será un vector de dimensión 256. Mientras que, para una imagen LBP con $P=16$ y $R=2$ su histograma constara de 65 536 elementos.

3.3 Clasificador

Un clasificador es un algoritmo capaz de colocar, o en ciertos casos nombrar, un objeto (patrón) desconocido a una apropiada categoría conocida (clase), denominada acción de clasificación. Esta clasificación se realiza según las características del patrón, en base al modelado previamente realizado por el mismo algoritmo, denominada acción de aprendizaje. Por lo tanto, en el aprendizaje el

calificador se encarga de, a partir de un conjunto de objetos relacionados entre sí (conjunto de entrenamiento), generar un modelo para cada clase del sistema, con los cuales podrá predecir la clase de un conjunto de objetos desconocidos (conjunto de prueba).

Los clasificadores pueden dividirse principalmente en dos categorías, los supervisados y los no supervisados. Esta división se basa en la forma en que el clasificador lleva a cabo la tarea de aprendizaje, también llamada etapa de entrenamiento. En el aprendizaje supervisado se proporciona una etiqueta o un peso para cada patrón en el conjunto de entrenamiento, lo cual permite separar las clases según la etiqueta proporcionada. Por otro lado, en el aprendizaje no supervisado no se tiene este tipo de etiquetas, por lo que el sistema genera las clases a partir de las características de los patrones del conjunto de entrenamiento, basándose en la similitud entre ellos.

Los algoritmos de clasificación comúnmente utilizados en el campo de la biometría y reconocimiento de patrones son principalmente 3: Redes Neuronales Artificiales (RNA) [83], Modelo de Mezclas Gaussianas (GMM, por sus siglas en inglés) [84] y Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés) [85]. La mayoría de los sistemas de reconocimiento de rostros han utilizado alguno de estos tres algoritmos para la etapa de clasificación [31-36, 53-60].

En los últimos años se han realizado trabajos para comprobar qué algoritmo de clasificación presenta un mejor desempeño sobre los demás, siendo SVM el que ha obtenido mejores resultados en diferentes aplicaciones [86, 87]. Es por esta razón que se propone la utilización de este algoritmo para la etapa de clasificación del presente trabajo.

Es importante mencionar que los algoritmos de clasificación son una parte fundamental en los sistemas de reconocimiento de rostros, es por eso que la elección de un buen algoritmo de clasificación juega un papel importante para el desempeño general del sistema. Sin embargo, el rendimiento de los clasificadores está basado en la capacidad de descripción de los patrones de entrada. Por lo tanto, para obtener resultados satisfactorios en el sistema, es necesaria la combinación de patrones robustos en la descripción de las muestras y un clasificador con gran capacidad de generalización.

3.3.1 Máquinas de Soporte Vectorial (SVM)

El método de Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés) es un algoritmo de aprendizaje constructivo, propuesto por Vapnik [85], basado en el principio de la teoría de aprendizaje estadístico. En los últimos años, este algoritmo ha ganado gran popularidad en los sistemas de reconocimiento de patrones, principalmente por su alto desempeño basado en el principio de minimización del riesgo estructural (SRM, por sus siglas en inglés), el cual ha demostrado ser superior al principio de minimización del riesgo empírico (ERM, por sus siglas en inglés) utilizado por las redes neuronales artificiales [88-91].

La idea general del SVM consiste en crear un modelo en un espacio dimensional donde se representen los puntos de los patrones de entrenamiento y colocar un límite (hiperplano) que separe las categorías con un margen lo más amplio posible. Por lo tanto, cuando los patrones de prueba se colocan en correspondencia con dicho modelo, en función de su proximidad con el límite, pueden ser clasificados a una u otra categoría. Esta idea es fácil de representar cuando se tienen patrones de entrenamiento linealmente separables, como se muestra en la Figura 3.12.

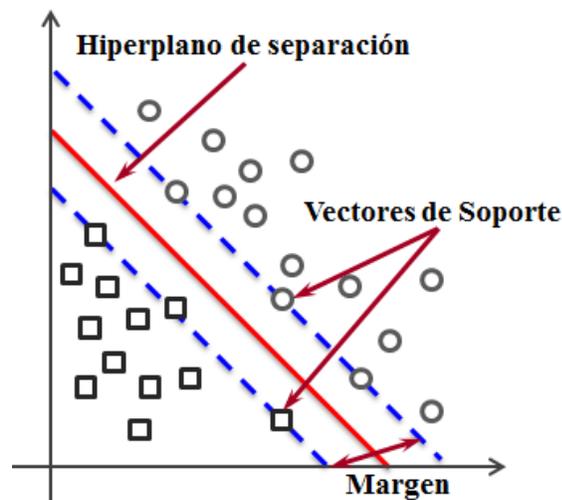


Figura 3.12. Ejemplo de la aplicación del SVM en un problema linealmente separable.

Cuando se requiere clasificar patrones que sean linealmente separables, se puede utilizar la forma lineal del SVM. El cual consiste en construir un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta, en busca de la separación óptima de los patrones de entrenamiento. Intuitivamente, existe un número

infinito de hiperplanos que logran separar las categorías; sin embargo, la separación óptima es alcanzada por el hiperplano que tenga la mayor distancia con los puntos de los patrones más cercanos a él (margen funcional). Por lo tanto, entre mayor sea el margen, menor será el error de generalización.

De forma que, los puntos de los vectores de una categoría se encontraran de un lado del hiperplano y los puntos de los vectores de otra categoría quedaran del otro lado del hiperplano. A los vectores formados por los puntos más cercanos al hiperplano se le llaman vectores de soporte y debido a esta cercanía, son conceptualmente los datos de los patrones más difíciles de clasificar.

Por lo tanto, el SVM debe encontrar el Hiperplano de Separación Óptima (OSH, por sus siglas en inglés), el cual es el clasificador lineal con el margen máximo para un determinado conjunto finito de patrones de entrenamiento. La definición matemática de este proceso se presenta a continuación.

Considérese un patrón \mathbf{x} n -dimensional, que tiene n número de coordenadas, $\mathbf{x} = (x_1, x_2, \dots, x_n)$, donde cada x_i es un número real, $x_i \in \mathfrak{R}$ para $i = 1, 2, \dots, n$. Cada patrón x_j pertenece a una categoría $y_j \in \{-1, +1\}$. Considérese también un conjunto de entrenamiento T de m patrones junto con su respectiva categoría, $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$. Finalmente, se debe considerar un espacio de producto escalar S , en el que están contenidos los patrones $(x, x_1, x_2, \dots, x_2) \in S$. Cualquier hiperplano en el espacio S se puede escribir como:

$$\{x \in S \mid w \cdot x + b = 0\} \quad w \in S, b \in \mathfrak{R}, \quad (3.48)$$

donde el producto punto $w \cdot x$ está definido por:

$$w \cdot x = \sum_{i=1}^n w_i x_i. \quad (3.49)$$

Un conjunto de patrones de entrenamiento es linealmente separable si existe al menos un clasificador lineal definido por el par (w, b) , el cual puede clasificar correctamente todos los patrones de entrenamiento. Este clasificador lineal es representado por el hiperplano H ($w \cdot x + b = 0$), el cual define una región para patrones de la categoría $+1$ ($w \cdot x + b > 0$) y otra región para patrones de la categoría -1 ($w \cdot x + b < 0$).

Después del entrenamiento, el clasificador está listo para predecir la membresía a una categoría para patrones nuevos (patrones de prueba). La categoría de los patrones x_k se determina con:

$$categoría(x_k) = \begin{cases} +1, & \text{si } w \cdot x_k + b > 0 \\ -1, & \text{si } w \cdot x_k + b < 0 \end{cases} \quad (3.50)$$

Por lo tanto, la clasificación de nuevos patrones depende solo del signo de la expresión $w \cdot x + b$. Para un conjunto de datos de entrenamiento, todos los hiperplanos de separación se pueden representar con la ecuación anterior.

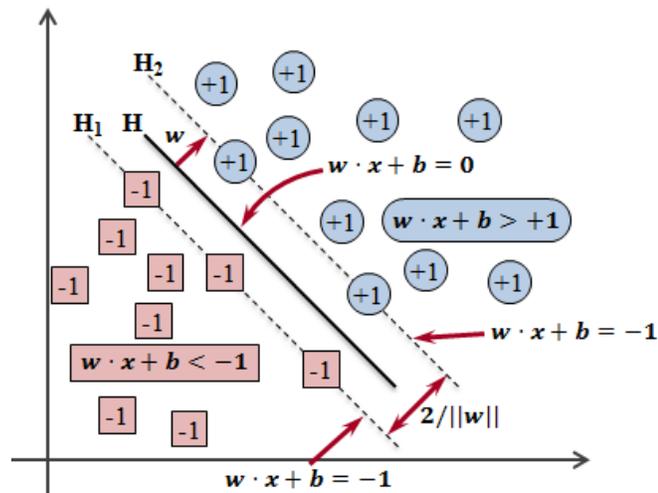


Figura 3.13. Ejemplo del Hiperplano de Separación Óptima (OSH).

En la Figura 3.13 se presenta un ejemplo del hiperplano de separación óptima (OSH). Donde se considera que: para todos los puntos del hiperplano H ($w \cdot x + b = 0$), la distancia entre el origen y el hiperplano H es $|b|/||w||$. Considerando que los patrones de la categoría -1 satisfacen la igualdad $w \cdot x + b = -1$, y definen el hiperplano H_1 ; la distancia entre el origen y el hiperplano H_1 es igual a $|-1 - b|/||w||$. De igual forma, los patrones de la categoría $+1$ satisfacen la igualdad $w \cdot x + b = +1$, y determinan el hiperplano H_2 ; la distancia entre el origen y el hiperplano H_2 es igual a $|+1 - b|/||w||$. Suponiendo que, los hiperplanos H , H_1 y H_2 son paralelos y ningún patrón de entrenamiento está ubicado entre los hiperplanos H_1 y H_2 ; basados en las consideraciones anteriores, la distancia entre hiperplanos (margen) H_1 y H_2 es $2/||w||$. A partir de estas consideraciones se deduce que la identificación del OSH se lleva a cabo mediante la maximización de $2/||w||$. Lo que

es equivalente a minimizar $\|w\|/2$. El problema de encontrar el OSH es representado por la identificación de (w, b) que satisfice:

$$\begin{cases} w \cdot x_i + b \geq +1, & \text{si } y_i = +1 \\ w \cdot x_i + b \leq -1, & \text{si } y_i = -1 \end{cases} \quad (3.51)$$

para el cual $\|w\|$ es mínimo.

En el caso de que los patrones a clasificar no sean linealmente separables, lo cual sucede en la mayoría de los problemas prácticos, los patrones de entrada se proyectan a un espacio dimensional mayor usando un conjunto de funciones base no lineales. Haciendo significativamente más fácil la localización del hiperplano óptimo. A este proceso se le conoce como SVM no lineal, y el algoritmo se utiliza de manera similar al SVM lineal, con la diferencia de que cada producto punto es reemplazado por la función base no lineal.

Las funciones base para el SVM no lineal, se conocen como funciones *kernel*. La función *kernel* se relaciona con la transformación $\varphi(x_i)$ definida por:

$$k(x_i, x_j) = \varphi(x_i)\varphi(x_j). \quad (3.52)$$

Equivalente a la distancian entre x_i y x_j medida en un espacio dimnensional mayor transformado por el mapeo φ . Algunos tipos de funciones *kernel* se describen a continuación:

- *Kernel* polinomial:

$$k(x_i, x_j) = (x_i^T x_j + c)^q, \quad (3.53)$$

donde c es una constante y q es un entero positivo.

- *Kernel* Gaussiano:

$$k(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{\sigma^2}\right). \quad (3.54)$$

- Expansión de Fourier:

$$k(x_i, x_j) = \frac{\sin(q + 1/2)(x_i - x_j)}{\sin(x_i - x_j)/2}. \quad (3.55)$$

La Figura 3.14 muestra un ejemplo de la aplicación del SVM no lineal, donde se realiza una proyección de un espacio bidimensional a uno tridimensional. (a) muestra los patrones de entrenamiento, los cuales no pueden ser separados linealmente, por lo que es necesario aplicar el método de SVM no lineal. (b) muestra la aplicación del SVM no lineal, donde utilizando la transformación ϕ , los patrones de entrenamiento fueron proyectados a un espacio tridimensional, en el cual es posible encontrar el hiperplano óptimo de separación y definir un margen entre los hiperplanos de los vectores de soporte.

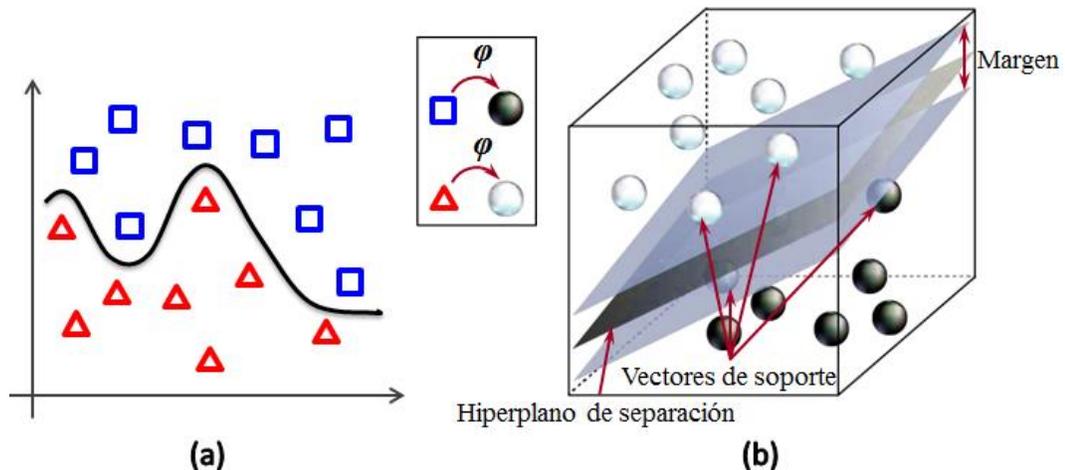


Figura 3.14. Ejemplo de la aplicación del SVM en un problema no separable linealmente. (a) Patrones no separables linealmente en un espacio bidimensional. (b) Obtención del hiperplano de separación de los patrones de (a), proyectados a un espacio tridimensional.

3.3.1.1 Esquema de uno contra todos (OVA)

Existe un problema particularmente interesante referente al método de SVM, y en general al proceso de clasificación, llamado clasificación multi-clase. En la práctica no siempre se tienen patrones de entrenamiento que sean relacionados a solo dos categorías, de tal forma que el problema de clasificación multi-clase se presenta cuando los patrones de entrenamiento están relacionados a N categorías.

El concepto es sencillo, una clasificación binomial se presenta cuando solo se requieren clasificar 2 clases, mientras que la multi-clase cuando se requieren clasificar más de dos. En la Figura 3.15 se ilustra la diferencia entre estos tipos de clasificación. (a) Muestra la clasificación binomial, separando los patrones de

entrenamiento en solo dos clases. Mientras que (b), muestra el tipo de clasificación multi-clase, específicamente de 4 clases, en este caso el SVM se encarga de separar el conjunto de patrones de entrenamiento en 4 diferentes clases.

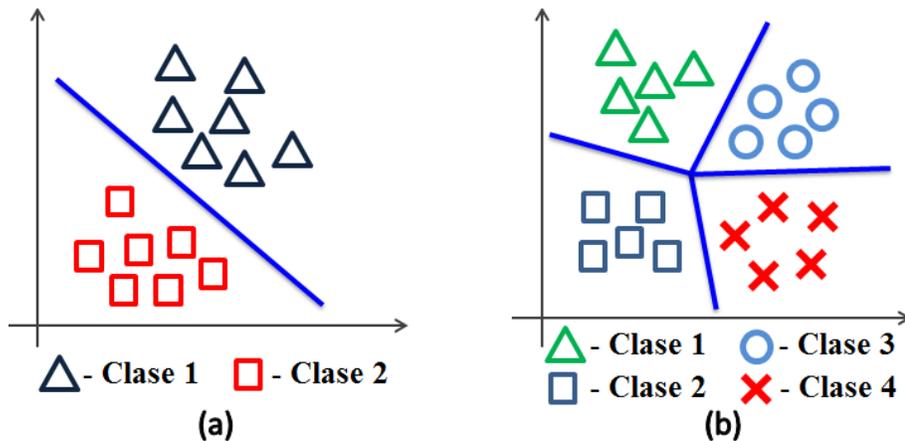


Figura 3.15. Diferencia entre clasificación binomial y clasificación multi-clase. (a) Ejemplo de clasificación binomial (solo 2 clases). (b) Ejemplo de clasificación multi-clase (4 clases).

Una posible solución para el problema de la multi-clasificación, radica en utilizar el enfoque de “uno contra todos” (OVA, por sus siglas en inglés) aplicado junto con el algoritmo de SVM. OVA ha sido catalogado como uno de los mejores métodos para solucionar el problema de la clasificación multi-clase, principalmente por la simplicidad de su aplicación y por su alto rendimiento [92].

Para poder aplicar el método de OVA, es necesario contar con un conjunto de patrones de entrenamiento relacionado a N clases diferentes entre sí. Basado en este conjunto, el enfoque de OVA consiste en entrenar N clasificadores binomiales diferentes, cada uno entrenado para distinguir los patrones de una sola clase contra los patrones de todas las clases restantes ($N-1$). Por otra parte, cuando se desea clasificar un nuevo patrón, se deben ejecutar los N clasificadores, de manera que la decisión final se toma eligiendo el valor positivo más alto de entre los resultados obtenidos por los N clasificadores.

La Figura 3.16 muestra el ejemplo de la aplicación del enfoque OVA en conjunto con el método de SVM, para el caso más simple de clasificación multi-clase, utilizando solo 3 clases.

En esta situación, se deben entrenar 3 SVMs diferentes, los cuales, obteniendo el hiperplano de separación óptimo, distinguirán las 3 clases individualmente, es decir cada uno realizará un entrenamiento binomial (como se muestra en la Figura 3.16). En el caso de que posteriormente se requiera clasificar un nuevo patrón, se deben de ejecutar los 3 SVMs con este patrón de entrada, obteniendo cada uno de ellos un resultado de esta clasificación. Para tomar la decisión final, será elegido el resultado con el valor positivo más alto.

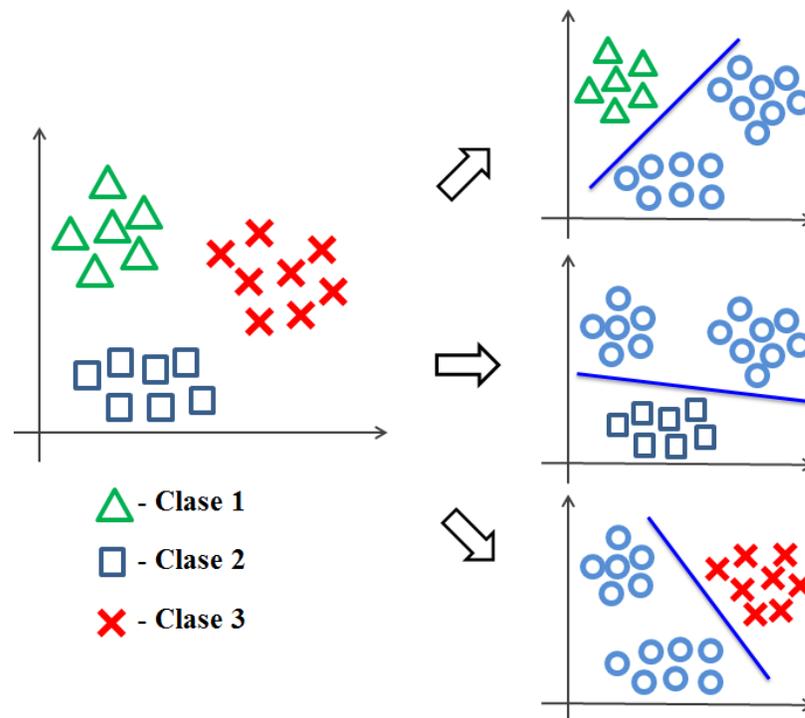


Figura 3.16. Ejemplo de la aplicación del algoritmo de SVM para un caso de multi-clasificación (3 clases) utilizando el método “uno contra todos” (OVA).

3.4 Conclusiones

En este capítulo se presentaron los fundamentos teóricos referentes a los algoritmos necesarios para el desarrollo de sistemas de reconocimiento de rostros, tomando en cuenta que la metodología para los sistemas de reconocimiento de expresiones faciales es similar a la de rostros, estos algoritmos pueden ser aplicables para ambos sistemas.

De acuerdo con lo expuesto en este capítulo, la metodología de los sistemas de reconocimiento de rostros y expresiones faciales puede dividirse en tres etapas: pre-procesamiento, extracción de características y clasificación. La etapa de pre-procesamiento se emplea para mejorar las imágenes de entrada, los algoritmos de normalización del valor del pixel, ecualización del histograma y CLAHE son algunos de los métodos que se han utilizado satisfactoriamente en esta etapa del proceso de reconocimiento de rostros y expresiones faciales. La etapa de extracción de características es fundamental en este tipo de sistemas, ya que es la encargada de extraer la información más relevante de la imagen del rostro. Además de, al mismo tiempo, disminuir o eliminar los problemas comunes que aquejan a estos sistemas. Para tal función fueron estudiados los métodos de Caras Propias, Fases Propias y LBP. Finalmente en la etapa de clasificación se analizó el algoritmo de SVM además de OVA, un método aplicable para SVM que soluciona el problema de la multi-clasificación, el cual se manifiesta tanto en los sistemas de reconocimiento de rostro como en los de expresiones faciales.

CAPÍTULO IV

4. SISTEMA PROPUESTO USANDO FASES PROPIAS

En este capítulo se plantea un sistema de reconocimiento de rostros, el cual para la tarea de extracción de características utiliza el algoritmo de Fases Propias basado en sub-bloques. Donde se propone una segmentación en sub-bloques de la imagen del rostro para poder extraer el espectro de fase de cada uno de ellos individualmente. Asimismo se presenta un análisis detallado para el tamaño de sub-bloque óptimo.

El sistema propuesto brinda la posibilidad de utilizar o no una etapa de pre-procesamiento, donde se analizan 3 algoritmos para este fin: normalización del valor del pixel, ecualización del histograma y ecualización adaptativa del histograma con limitación de contraste. Finalmente, para la etapa de clasificación el sistema utiliza el algoritmo de máquinas de soporte vectorial. Algoritmo que se encarga de realizar las tareas de entrenamiento, identificación y verificación del sistema.

4.1 Diagrama del Sistema

La Figura 4.1 muestra el diagrama a bloques del sistema propuesto. En donde se puede observar la opción de usar una etapa de pre-procesamiento con la cual se trata de reducir las variaciones de los pixeles en la imagen de entrada, producidas principalmente por cambios en las condiciones de iluminación. La extracción de características (basada en algoritmo de fases propias [48]) consiste en la segmentación de la imagen en sub-bloques, la extracción del espectro de fase y la aplicación del PCA para la estimación de los vectores característicos, los cuales se utilizaran para la etapa de clasificación. Por otra parte, la etapa de clasificación (usando SVM) puede realizar tanto la tarea de identificación como la de verificación. Ambos basados en la etapa de entrenamiento, donde se generan los modelos para cada sujeto que posteriormente se guardan en la base de datos de modelos. De esta manera, cuando el sistema es utilizado para llevar a cabo la tarea de identificación, se compara el modelo de la persona bajo análisis con todos los almacenados en la base de datos para así determinar la identidad de dicha persona, mientras que para llevar a cabo la tarea de verificación, el modelo del sujeto bajo análisis se compara solamente con el modelo de la persona que el sujeto afirma ser.

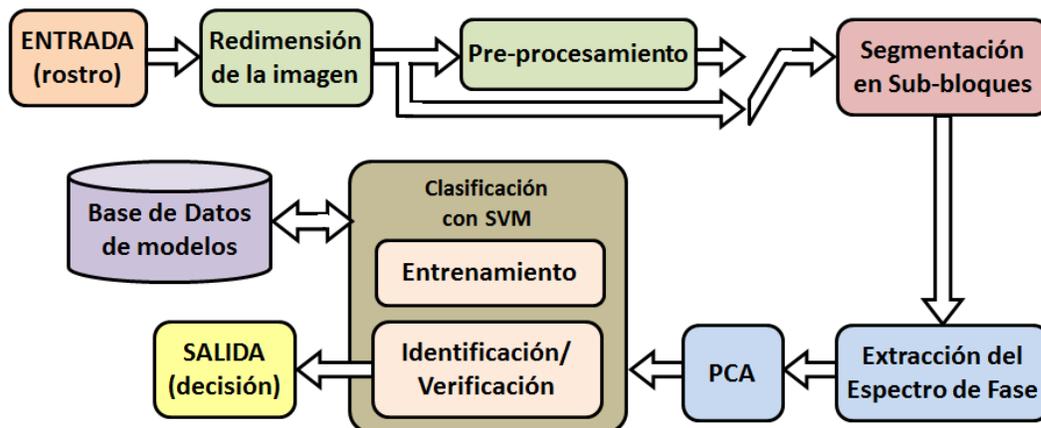


Figura 4.1. Diagrama a bloques del sistema propuesto de fases propias basado en sub-bloques.

A grandes rasgos el sistema propuesto de Fases Propias Basado en Sub-bloques consiste en: obtener la imagen del rostro de una persona como entrada para después redimensionar dicha imagen con el objetivo de trabajar con menos datos y así reducir el número de operaciones necesarias para calcular el PCA. La redimensión de la imagen está dada por la ecuación (4.1).

$$I(n, m) = I_{ent}(nR_s, mR_s). \quad (4.1)$$

Donde I_{ent} es la imagen bajo análisis, R_s es el factor de redimensionamiento e I es la imagen redimensionada. A continuación, la imagen redimensionada puede ser introducida a la etapa de pre-procesamiento donde se utiliza uno de los 3 algoritmos: normalización del valor del pixel, ecualización del histograma (HE) o CLAHE. Posteriormente la imagen bajo análisis es dividida en sub-bloques a los que se les extrae el espectro de fase individualmente. Después de la extracción del espectro de fase de cada sub-bloque se concatenan las matrices resultantes para poder aplicar el algoritmo de PCA, el cual es utilizado para una reducción dimensional y la formación del vector característico con la información más importante de la imagen bajo análisis. Finalmente, los vectores característicos de las imágenes de entrenamiento son utilizados para formar los modelos de cada persona, haciendo uso del algoritmo de SVM, y así poder ejecutar la tarea de identificación o verificación según sea el caso. La Figura 4.2 muestra la salida de cada una de las cinco primeras etapas del sistema propuesto. La imagen de entrada se muestra en (a), (b) muestra la imagen después del proceso de redimensión, (c) es el resultado obtenido después de aplicar el algoritmo de CLAHE a la imagen (b), este paso puede ser omitido si no se utiliza la etapa de pre-procesamiento. La imagen (d) es el resultado de la segmentación de la imagen en sub-bloques y (e) es la matriz obtenida por la concatenación del espectro de fase de cada sub-bloque de la imagen (d).

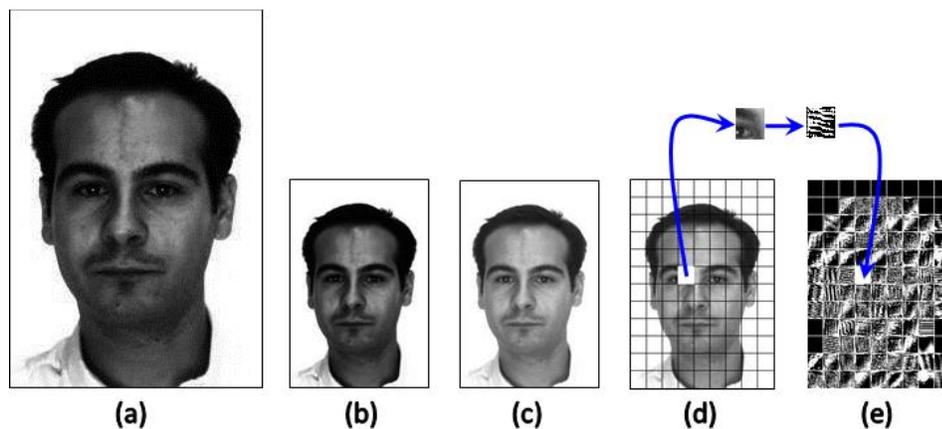


Figura 4.2. Resultado de la aplicación de las 5 primeras etapas del sistema propuesto. (a) Imagen original del rostro. (b) Imagen redimensionada. (c) Imagen después de usar el método CLAHE. (d) Imagen segmentada en sub-bloques. (e) Concatenación de los espectros de fase de cada sub-bloque de la imagen (D)

4.2 Etapa de Pre-procesamiento

El propósito de la etapa de pre-procesamiento es modificar la imagen redimensionada de tal manera que las variaciones producidas por las condiciones de iluminación sean lo más pequeñas posibles, con el objetivo de aumentar la precisión de la extracción de características en las siguiente etapa. Para esta etapa se utiliza uno de los 3 métodos descritos en la sección 3.1: normalización del valor del pixel, equalización del histograma (HE) y equalización adaptativa del histograma con limitación de contraste (CLAHE).

La forma de aplicación para los algoritmos de pre-procesamiento es como la descrita en la sección 3.1. Para la aplicación del algoritmo de normalización del valor del pixel, se utilizó la ecuación (3.1) a la imagen redimensionada. Por otro lado, con ayuda de la ecuación (3.8) se obtuvo la imagen resultante para el método de equalización del histograma. Mientras que para el caso de CLAHE se utilizó el procedimiento descrito en la sección 3.1.3 con los parámetros de $\beta = 0.01$ y $(F, C) = (2,2)$, es decir CLAHE(2,2) con un limite de corte de 0.01.

La Figura 4.3 muestra los resultados de aplicar los tres algoritmos disponibles para el pre-procesamiento a una imagen del rostro. La imagen original se muestra en (a), (b) muestra el resultado de aplicar la normalización del valor del pixel, por su parte (c) muestra la imagen resultante al aplicar la equalización del histograma, por ultimo (d) muestra la imagen después de utilizar CLAHE, con los parámetros mencionados anteriormente.

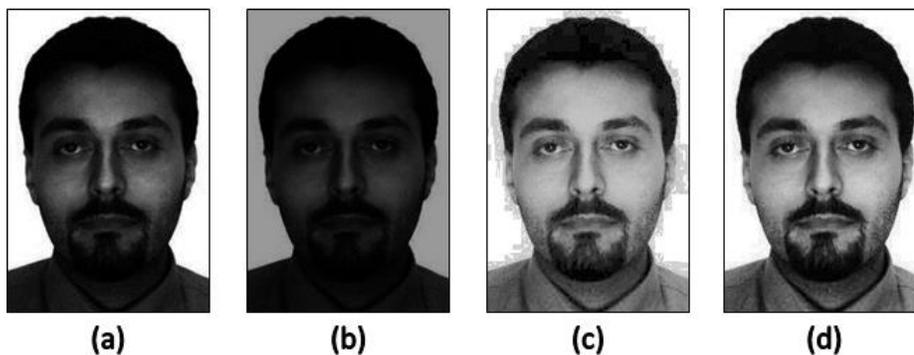


Figura 4.3. Resultado de la aplicación de los 3 algoritmos de pre-procesamiento a una imagen de rostro. (a) Imagen original. (b) Resultado de aplicar la normalización del valor del pixel en (a). (c) Resultado de aplicar HE en (a). (d) Resultado de aplicar CLAHE(2,2) en (a).

4.3 Segmentación en Sub-bloques

Para mejorar el desempeño del algoritmo de fases propias, antes de la extracción del espectro de fase, la imagen del rostro se divide en sub-bloques. Es importante mencionar que esta etapa se lleva a cabo después de la de pre-procesamiento, donde previamente ha sido redimensionada la imagen de entrada.

La segmentación se lleva a cabo dividiendo en $W \times H$ sub-bloques de tamaño $N \times N$ la imagen redimensionada, obtenida por la ecuación (4.1) (en caso de no utilizar la etapa de pre-procesamiento). En este trabajo se presenta un análisis completo de la utilización de 4 tamaños de sub-bloque diferentes, con dimensiones de $N \times N = 12 \times 12, 6 \times 6, 4 \times 4$ y 2×2 píxeles, además de usando la imagen completa (método convencional). La Figura 4.4 muestra un ejemplo de las 5 variaciones mencionadas: (a) imagen sin segmentación (método convencional), (b) segmentación en sub-bloques de 12×12 , (c) en sub-bloques de 6×6 , (d) en sub-bloques de 4×4 y (e) en sub-bloques de 2×2 .

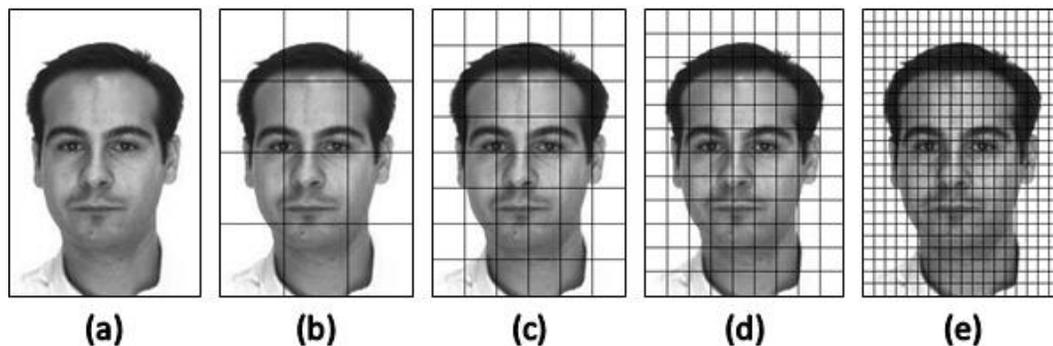


Figura 4.4. Ejemplo de los diferentes tamaños de sub-bloque. (a) Imagen redimensionada a un tamaño de 48×36 píxeles. (b) Segmentación de la imagen (a) en sub-bloques de tamaño $N=12$. (c) Resultado de la segmentación de (a) con $N=6$. (d) Imagen resultante de la segmentación de (a) con $N=4$. (e) Segmentación de (a) en sub-bloques con $N=2$.

Este enfoque ha sido utilizado en otros trabajos [32, 33, 35], usando solamente sub-bloques de tamaño: $12 \times 12, 6 \times 6$ y 3×3 píxeles. Sin embargo, a pesar de que es posible notar que usando el método basado en sub-bloques se mejora el rendimiento del algoritmo de fases propias convencional, ninguna de estas referencias proporciona un criterio para seleccionar el tamaño de sub-bloque óptimo. Es por eso que la siguiente subsección presenta un análisis referente a este tema.

4.3.1 Tamaño de Bloque Óptimo

Esta sección proporciona una explicación teórica del efecto de sub-bloques y prueba que el porcentaje de reconocimiento del algoritmo de fases propias debe mejorar cuando el tamaño de sub-bloque se hace más pequeño. Considerado un determinado sub-bloque $f(t_1, t_2)$, $0 \leq t_1 \leq NT, 0 \leq t_2 \leq MT$ y asumiendo que es una función separable, es decir $f(t_1, t_2) = f_1(t_1)f_2(t_2)$, donde $f_1(t_1)$ y $f_2(t_2)$ pueden representarse en términos de N o M intervalos de duración T como se muestra en la Figura 4.5. Entonces, la transformada de Fourier, ecuación (3.23) (descrita en la sección 3.2.2.1), del sub-bloque $f(t_1, t_2)$ está dada por la ecuación:

$$F(\omega_1, \omega_2) = \int_0^{NT} \int_0^{MT} f_1(t_1, t_2) e^{-j\omega_1 t_1} e^{-j\omega_2 t_2} dt_1 dt_2 = \int_0^{NT} f_1(t_1) e^{-j\omega_1 t_1} dt_1 \int_0^{MT} f_2(t_2) e^{-j\omega_2 t_2} dt_2, \quad (4.2)$$

entonces

$$F_1(\omega_1) = \int_0^{NT} f_1(t_1) e^{-j\omega_1 t_1} dt_1, \quad (4.3)$$

$$F_2(\omega_2) = \int_0^{MT} f_2(t_2) e^{-j\omega_2 t_2} dt_2, \quad (4.4)$$

donde N y M son el número de pixeles en cada fila y columna respectivamente, y T es la separación entre dos pixeles consecutivos para cualquier sub-bloque dado.

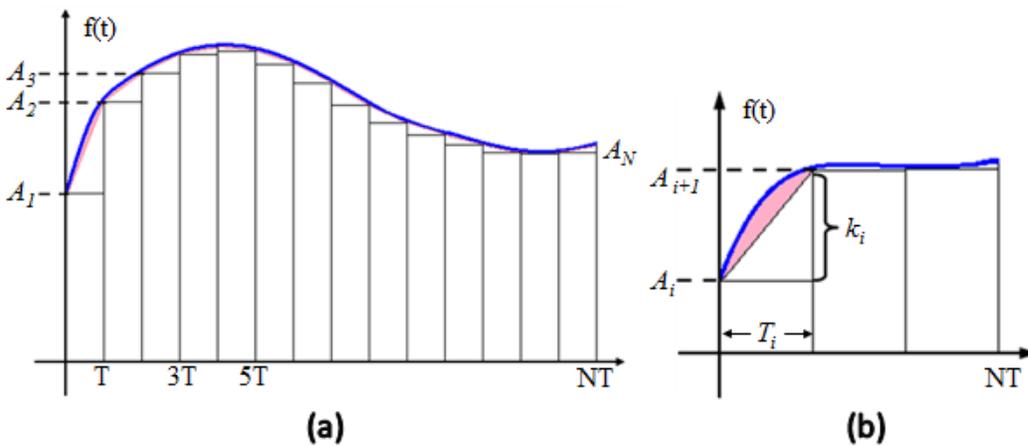


Figura 4.5. Representación por fragmentos de un sub-bloque de la imagen del rostro. (a) Representación de las filas (N) de un determinado sub-bloque $f(t_1, t_2)$. (b) Representación más detallada de (a).

De la Figura 4.5 se puede asumir que:

$$f_1(t_1) \begin{cases} \frac{k_0 A_0}{T} t + A_0, & 0 \leq t \leq T \\ \frac{k_1 A_1 (t - T)}{T} + A_1, & T \leq t \leq 2T \\ \frac{k_2 A_2 (t - 2T)}{T} + A_2, & 2T \leq t \leq 3T \\ \frac{k_{N-1} A_{N-1} (t - (N-1)T)}{T} + A_{N-1}, & (N-1)T \leq t \leq NT \end{cases} \quad (4.5)$$

De manera que la ecuación (4.3) puede aproximarse como

$$\begin{aligned} \int_0^{NT} f_1(t_1) e^{-j\omega_1 t_1} dt_1 &= \int_0^T \left(\frac{k_0 A_0}{T} t + A_0 \right) e^{-j\omega_1 t_1} dt_1 + \int_T^{2T} \left(\frac{k_1 A_1}{T} (t - T) + A_1 \right) e^{-j\omega_1 t_1} dt_1 \\ &+ \int_{nT}^{(n+1)T} \left(\frac{k_n A_n}{T} (t - nT) + A_n \right) e^{-j\omega_1 t_1} dt_1 \\ &+ \int_{(N-1)T}^{NT} \left(\frac{k_{N-1} A_{N-1}}{T} (t - (N-1)T) + A_{N-1} \right) e^{-j\omega_1 t_1} dt_1. \end{aligned} \quad (4.6)$$

Seguidamente considere:

$$\begin{aligned} \int_{nT}^{(n+1)T} \left(\frac{k_n A_n}{T} (t - nT) + A_n \right) e^{-j\omega_1 t_1} dt_1 & \quad n = 0, 1, \dots, N-1 \\ &= \int_{nT}^{(n+1)T} \frac{k_n A_n}{T} t e^{-j\omega_1 t_1} dt_1 - \int_{nT}^{(n+1)T} nT \frac{k_n A_n}{T} e^{-j\omega_1 t_1} dt_1 + \int_{nT}^{(n+1)T} A_n e^{-j\omega_1 t_1} dt_1 \\ &= \int_{nT}^{(n+1)T} \frac{k_n A_n}{T} t e^{-j\omega_1 t_1} dt_1 + \int_{nT}^{(n+1)T} (A_n - nk_n A_n) e^{-j\omega_1 t_1} dt_1. \end{aligned} \quad (4.7)$$

Por lo tanto:

$$\begin{aligned} \int_{nT}^{(n+1)T} \left(\frac{k_n A_n}{T} (t - nT) + A_n \right) e^{-j\omega_1 t_1} dt_1 & \quad n = 0, 1, \dots, N-1 \\ &= \int_{nT}^{(n+1)T} \left(\frac{k_n A_n}{T} t + A_n (1 - nk_n) \right) e^{-j\omega_1 t_1} dt_1. \end{aligned} \quad (4.8)$$

Sustituyendo (4.8) en (4.6) se obtiene:

$$\int_0^{NT} f_1(t_1) e^{-j\omega_1 t_1} dt_1 = \sum_{n=0}^{N-1} \int_{nT}^{(n+1)T} \left(\frac{k_n A_n}{T} t_1 + A_n (1 - nk_n) \right) e^{-j\omega_1 t_1} dt_1. \quad (4.9)$$

Enseguida, considerando los valores de las filas $F_1(\omega_1)$ y suponiendo que:

$$G_n(\omega_1) = \int_{nT}^{(n+1)T} \left(\frac{k_n A_n}{T} t_1 + A_n(1 - nk_n) \right) e^{-j\omega_1 t_1} dt_1 = \int_{nT}^{(n+1)T} \frac{k_n A_n}{T} t_1 e^{-j\omega_1 t_1} dt_1 + \int_{nT}^{(n+1)T} A_n(1 - nk_n) e^{-j\omega_1 t_1} dt_1. \quad (4.10)$$

Después de integrar y hacer algunas sustituciones, además de asumir (sin pérdida de generalidad) que $T = 1$, (4.10) se convierte en:

$$\begin{aligned} G_n(\omega_1) = & k_n A_n \left(\frac{(n+1)}{\omega_1} \sin((n+1)\omega_1) - \frac{n}{\omega_1} \sin(n\omega_1) + \frac{1}{\omega_1^2} \cos((n+1)\omega_1) - \frac{1}{\omega_1^2} \cos(n\omega_1) \right) \\ & + 2A_n(1 - nk_n) \sin\left(\frac{\omega_1}{2}\right) \cos\left((n + \frac{1}{2})\omega_1\right) + \\ & jk_n A_n \left(\frac{(n+1)}{\omega_1} \cos((n+1)\omega_1) - \frac{n}{\omega_1} \cos(n\omega_1) - \frac{1}{\omega_1^2} \sin((n+1)\omega_1) + \frac{1}{\omega_1^2} \sin(n\omega_1) \right) \\ & - j2A_n(1 - nk_n) \sin\left(\frac{\omega_1}{2}\right) \sin\left((n + \frac{1}{2})\omega_1\right) \end{aligned} \quad (4.11)$$

Seguidamente, tomando en cuenta que:

$$F_1(\omega_1) = G_n(\omega_1) = \sum_{n=0}^{N-1} (G_n^R(\omega_1) + jG_n^I(\omega_1)), \quad (4.12)$$

obtenemos:

$$\begin{aligned} G_n^R(\omega_1) = \text{Re}\{G_n(\omega_1)\} = & k_n A_n \left(\frac{(n+1)}{\omega_1} \sin((n+1)\omega_1) - \frac{n}{\omega_1} \sin(n\omega_1) + \frac{1}{\omega_1^2} \cos((n+1)\omega_1) - \frac{1}{\omega_1^2} \cos(n\omega_1) \right) \\ & + 2A_n(1 - nk_n) \sin\left(\frac{\omega_1}{2}\right) \cos\left((n + \frac{1}{2})\omega_1\right), \end{aligned} \quad (4.13)$$

$$\begin{aligned} G_n^I(\omega_1) = \text{Im}\{G_n(\omega_1)\} = & k_n A_n \left(\frac{(n+1)}{\omega_1} \cos((n+1)\omega_1) - \frac{n}{\omega_1} \cos(n\omega_1) - \frac{1}{\omega_1^2} \sin((n+1)\omega_1) + \frac{1}{\omega_1^2} \sin(n\omega_1) \right) \\ & - 2A_n(1 - nk_n) \sin\left(\frac{\omega_1}{2}\right) \sin\left((n + \frac{1}{2})\omega_1\right). \end{aligned} \quad (4.14)$$

De manera similar, se deduce que la transformada de Fourier para los valores de las columnas $F_2(\omega_2)$ se representa de la siguiente forma:

$$F_2(\omega_2) = G_m(\omega_2) = \sum_{m=0}^{M-1} (G_m^R(\omega_2) + jG_m^I(\omega_2)) \quad (4.15)$$

$$\begin{aligned} G_m^R(\omega_2) &= \text{Re}\{G_m(\omega_2)\} = k_m A_m \\ &\left(\frac{(m+1)}{\omega_2} \sin((m+1)\omega_2) - \frac{m}{\omega_2} \sin(m\omega_2) + \frac{1}{\omega_2^2} \cos((m+1)\omega_2) - \frac{1}{\omega_2^2} \cos(m\omega_2) \right) \\ &+ 2A_m(1 - mk_m) \sin\left(\frac{\omega_2}{2}\right) \cos\left((m + \frac{1}{2})\omega_2\right). \end{aligned} \quad (4.16)$$

$$\begin{aligned} G_m^I(\omega_2) &= \text{Im}\{G_m(\omega_2)\} = k_m A_m \\ &\left(\frac{(m+1)}{\omega_2} \cos((m+1)\omega_2) - \frac{m}{\omega_2} \cos(m\omega_2) - \frac{1}{\omega_2^2} \sin((m+1)\omega_2) + \frac{1}{\omega_2^2} \sin(m\omega_2) \right) \\ &- 2A_m(1 - mk_m) \sin\left(\frac{\omega_2}{2}\right) \sin\left((m + \frac{1}{2})\omega_2\right). \end{aligned} \quad (4.17)$$

Entonces, de las ecuaciones (4.2), (4.12) y (4.15) tenemos que:

$$F(\omega_1, \omega_2) = F_1(\omega_1)F_2(\omega_2) = \sum_{n=0}^{N-1} (G_n^R(\omega_1) + jG_n^I(\omega_1)) \sum_{m=0}^{M-1} (G_m^R(\omega_2) + jG_m^I(\omega_2)) \quad (4.18)$$

Finalmente, de la ecuación (4.18) la fase de $F(\omega_1, \omega_2)$ es definida por:

$$\theta(\omega_1, \omega_2) = \tan^{-1} \frac{\sum_{n=0}^{N-1} G_n^I(\omega_1)}{\sum_{n=0}^{N-1} G_n^R(\omega_1)} + \tan^{-1} \frac{\sum_{m=0}^{M-1} G_m^I(\omega_2)}{\sum_{m=0}^{M-1} G_m^R(\omega_2)}. \quad (4.19)$$

Por lo tanto, de la ecuación (4.19) se deduce que, si el tamaño del bloque usado para la extracción de características contienen solo 2 elementos, es decir si $N = 1$ el efecto de la magnitud de los pixeles A_n y A_m en las ecuaciones (4.13)-(4.19) desaparece, dependiendo solamente de los valores de k_n y k_m , de tal forma que la fase de la transformada de Fourier del sub-bloque bajo análisis se convierte en:

$$\begin{aligned} \theta(\omega_1, \omega_2) &= \tan^{-1} \left[\frac{(\omega_1 k_n \cos(\omega_1) - k_n \sin(\omega_1) - 2\omega_1^2 \sin^2(\omega_1/2))}{(\omega_1 k_n \sin(\omega_1) + k_n \cos(\omega_1) - \omega_1^2 \sin(\omega_1) - k_n)} \right] \\ &+ \tan^{-1} \left[\frac{(\omega_2 k_m \cos(\omega_2) - k_m \sin(\omega_2) - 2\omega_2^2 \sin^2(\omega_2/2))}{(\omega_2 k_m \sin(\omega_2) + k_m \cos(\omega_2) - \omega_2^2 \sin(\omega_2) - k_m)} \right]. \end{aligned} \quad (4.20)$$

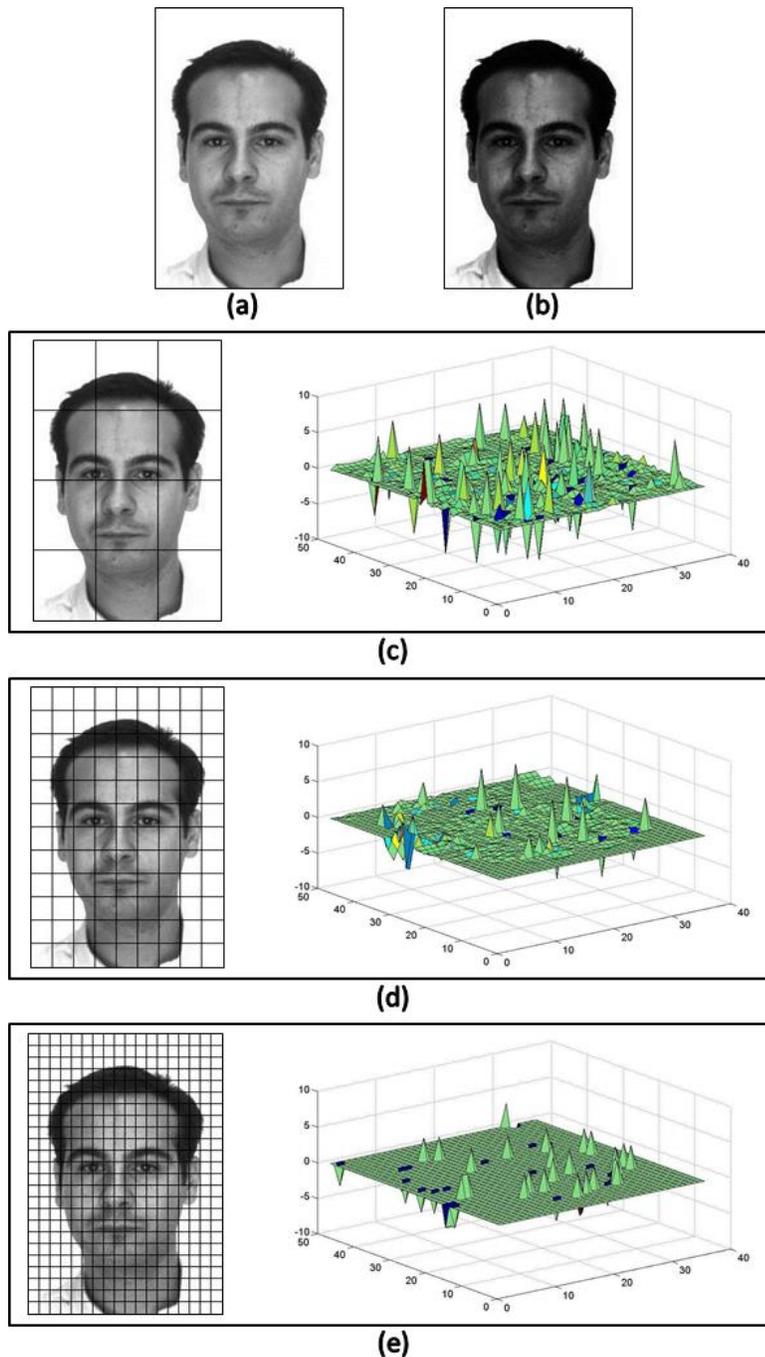


Figura 4.6. Ejemplo de la diferencia de los espectros de fase utilizando la segmentación en sub-bloques de dos imágenes con diferentes condiciones de iluminación. (a) Imagen del rostro con iluminación estable. (b) Imagen del rostro con poca iluminación. (c) Diferencia del espectro de fase de las imágenes (a) y (b), obtenido utilizando sub-bloques de 12x12 píxeles. (d) Diferencia del espectro de fase usando sub-bloques de 4x4 píxeles. (e) Diferencia del espectro de fase obtenido con sub-bloques de 2x2 píxeles.

Por lo tanto, de la ecuación (4.20) podemos concluir que el tamaño de bloque óptimo que proporciona el mejor rendimiento del algoritmo de Fases Propias Basado en Sub-bloques, pese a cambios en la iluminación de las imágenes, es el de 2×2 píxeles, esto se corrobora en la Figura 4.6.

La Figura 4.6 muestra un experimento en el que se obtiene el espectro de fase de dos imágenes del mismo sujeto pero con diferentes condiciones de iluminación. Entonces estos dos espectros de fase se restan, con la finalidad de observar la diferencia que hay entre ambos, basado en el tamaño de sub-bloque utilizado para la obtención de estos. De la figura podemos observar que (a) es la imagen del rostro del sujeto en condiciones de iluminación estables; (b) es la segunda imagen del mismo sujeto pero con baja iluminación; (c), (d) y (e) muestran la diferencia entre ambos espectros de fase, obtenidos dividiendo la imagen en sub-bloques de 12×12 , 4×4 y 2×2 píxeles respectivamente. Al observar las diferencias de ambos espectros de fase podemos deducir que al reducir el tamaño del sub-bloque obtenemos un mayor parecido entre los espectros de fase de ambas imágenes, cuestión que en las siguientes etapas ayudará enormemente al sistema a poder discriminar los cambios de iluminación entre imágenes del mismo sujeto.

Por otro lado la Figura 4.7 muestra la diferencia de los espectros de fase de dos sujetos distintos en condiciones de iluminación estables. (a) Muestra la imagen del rostro del sujeto 1, (b) la del sujeto 2 y (c) la gráfica de la diferencia de los espectros de fase de ambas imágenes usando sub-bloques de 2×2 . Con esto podemos observar que a pesar de usar el tamaño de bloque óptimo, cuando las imágenes son de sujetos diferentes los espectros de fase son ampliamente diferenciables.

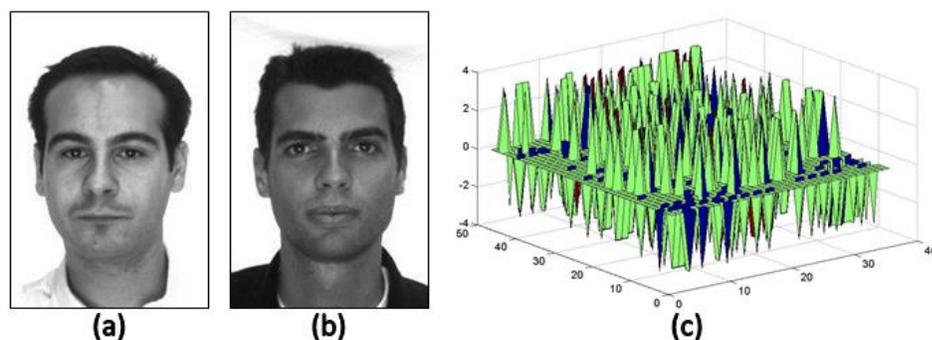


Figura 4.7. Ejemplo de la diferencia de los espectros de fase de dos sujetos distintos usando un tamaño de sub-bloque de 2×2 . (a) Imagen del sujeto 1. (b) Imagen del sujeto 2. (c) Diferencia del espectro de fase entre la imagen (a) y (b) utilizando sub-bloques de 2×2 píxeles (óptimo).

4.4 Extracción del Espectro de Fase

Como se mencionó en la sección 3.2.2, el algoritmo de fases propias se basa en la extracción del espectro de fase de la imagen de entrada. En el sistema propuesto esta extracción se lleva a cabo después de la segmentación en sub-bloques de la imagen. Segmentación que, como se explicó anteriormente, consiste en dividir la imagen del rostro en $W \times H$ bloques de $N \times N$ píxeles, donde $NW \times NH$ es el tamaño de la imagen. Es importante mencionar que esta segmentación se realiza a la imagen redimensionada después de aplicar o no algún tipo de pre-procesamiento.

El espectro de fase de cada sub-bloque se obtiene con la ecuación (4.19), y considerando la ecuación (3.27), de manera más general se puede expresar como:

$$\varphi_{j,k} = \tan^{-1} \left(\frac{\text{Im}(FFT(I_{j,k}(m,n)))}{\text{Re}(FFT(I_{j,k}(m,n)))} \right) \quad (4.21)$$

$$m, n = 1, 2, \dots, N, \quad j = 1, 2, \dots, W, \quad k = 1, 2, \dots, H,$$

donde $I_{j,k}(m,n)$ representa el (j,k) -ésimo sub-bloque de la imagen obtenida después de la etapa de segmentación. Enseguida, los espectros de fase de todos los sub-bloques de la imagen son concatenados:

$$\varphi = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} & \dots & \varphi_{1,H} \\ \varphi_{2,1} & \varphi_{2,2} & \dots & \varphi_{2,H} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{W,1} & \varphi_{W,2} & \dots & \varphi_{W,H} \end{bmatrix}, \quad (4.22)$$

donde φ es la matriz del espectro de fase que se utilizará en la etapa de conformación del vector característico, es decir el espectro de fase de todos los sub-bloques de la imagen concatenados.

La Figura 4.8 muestra el proceso de extracción del espectro de fase usando la segmentación en sub-bloques, donde (a) y (b) muestran la matriz del espectro de fase (φ) usando sub-bloques de tamaño 12×12 y 6×6 respectivamente. En esta figura se puede observar en el lado izquierdo el resultado de la segmentación y en el lado derecho el resultado de la concatenación de los espectros de fase de cada sub-bloque segmentado. Es importante notar que en (b) se observan cambios en el espectro de fase solo cuando el sub-bloque utilizado contenía únicamente fondo de la imagen.

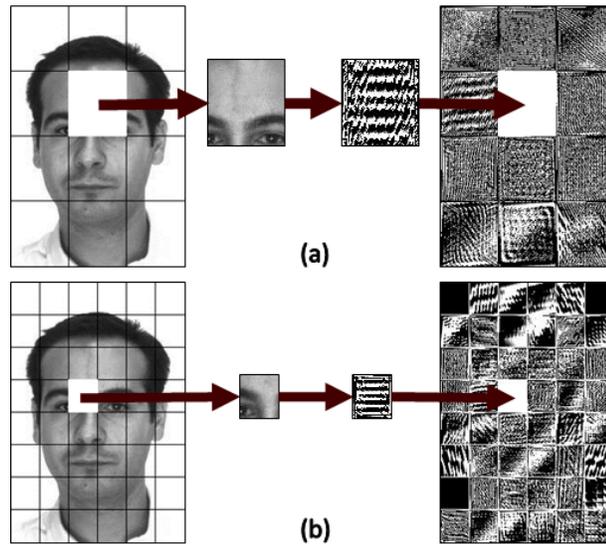


Figura 4.8. Ejemplo de la extracción del espectro de fase de la misma imagen usando diferente tamaño de sub-bloque. (a) Obtención del espectro de fase usando sub-bloques de 12x12. (b) Obtención del espectro de fase usando sub-bloques de 6x6.

4.5 Conformación del Vector Característico usando PCA

Para la etapa de extracción de características el algoritmo de fases propias además de utilizar el espectro de fase de la imagen, necesita aplicar el algoritmo de análisis de componentes principales (PCA) para poder reducir la dimensionalidad de la información proporcionada por las imágenes de entrenamiento y finalmente conformar los vectores característicos.

El procedimiento utilizado para la conformación de los vectores característicos aplicando el método de PCA se muestra en la Figura 4.9. El primer paso del proceso es convertir en vectores columna los espectros de fase de todas las imágenes de entrenamiento, estos vectores cuentan con un tamaño de $Mr = W \times H \times N^2 - 1$. Este paso se expresa de la siguiente manera:

$$\phi_i = [\phi_0 \quad \phi_1 \quad \dots \quad \phi_{Mr}]^T \quad i = 1, 2, \dots, Q, \quad (4.23)$$

donde ϕ_i es el vector columna de la i -ésima imagen de entrenamiento, y Q es el número total de estas. Los elementos del vector columna están definidos por:

$$\phi_r = \varphi_{p,q} \quad p = jN + m, \quad q = kN + n, \quad 0 \leq m, n \leq N - 1. \quad (4.24)$$

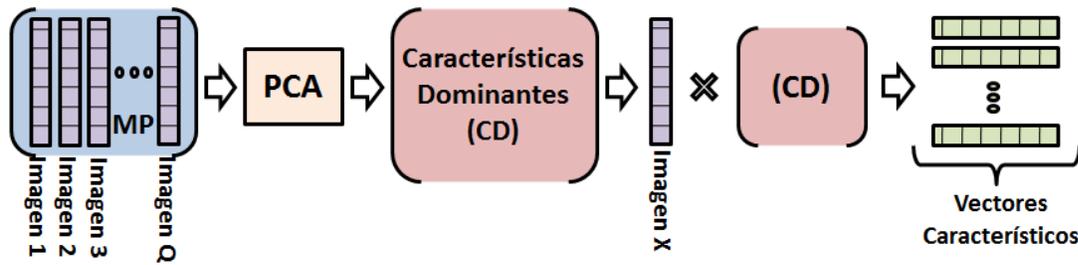


Figura 4.9. Conformación de vectores característicos usando PCA.

Después de obtener los vectores columna, es necesario formar una matriz principal (MP) con todos los vectores de las imágenes de entrenamiento. MP se expresa de la siguiente manera:

$$MP = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,Q} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{Mr,1} & \phi_{Mr,2} & \cdots & \phi_{Mr,Q} \end{bmatrix}. \quad (4.25)$$

Enseguida, el método PCA se aplica a la matriz MP (este proceso es descrito en la sección 3.2.2.2). Como resultado obtendremos una matriz de menor dimensionalidad que MP, llamada matriz de características dominantes (CD) la cual contiene la información más importante de todo el conjunto de imágenes de entrenamiento. Por último, los vectores característicos se forman por medio de la multiplicación del vector columna de cada imagen por la matriz CD, definidos por:

$$Y = \phi CD, \quad (4.26)$$

donde Y representa el vector característico de la imagen bajo análisis, ya sea imagen de entrenamiento o de prueba.

4.6 Clasificación usando SVM

El método de Máquinas de Soporte Vectorial (SVM), estudiado en el capítulo 3 sección 3.3.1, es el encargado de llevar acabo la tarea de clasificación para el sistema propuesto. Esta tarea está dividida en 2 modalidades: fase de entrenamiento y fase de reconocimiento (incluyendo tanto la tarea de identificación como la de verificación).

Para la función de entrenamiento el SVM se encarga de generar los modelos de cada persona, los cuales deben ser almacenados en una base de datos, ya que serán necesarios para la realización de las siguientes etapas. La función de identificación se lleva a cabo utilizando todos los modelos almacenados en la base de datos y el SVM se encarga de dictaminar a cuál de ellos corresponde la imagen bajo análisis. Para la función de verificación se utiliza solo un modelo, el correspondiente al que indicó la persona bajo análisis, en este caso el SVM se encarga de decidir si la imagen de esta persona corresponde o no al modelo seleccionado.

En este trabajo, la librería LIBSVM [93] fue utilizada para realizar las funciones del SVM. Los parámetros utilizados para las tareas de entrenamiento, identificación y verificación son las siguientes: tipo de kernel polinomial: $(\gamma * u' * v + coef0)^{grado}$, donde $coef0 = 1$, $\gamma = 1$, $costo = 100$, y los demás parámetros se utilizan con los valores por defecto asignados por dicha librería.

Esta sección está dividida en correspondencia a las 3 modalidades en las que puede aplicarse el método de SVM: etapa de entrenamiento, etapa de identificación y etapa de verificación.

4.6.1 Etapa de Entrenamiento

En esta etapa, la función del SVM es la de generar los modelos de cada persona, para ello es necesario proporcionarle los datos necesarios que serán asociados a esta. En otras palabras el SVM necesita los vectores característicos de todas las imágenes de entrenamiento para poder generar el modelo de cada persona. El proceso para realizar el entrenamiento de una persona y poder generar el modelo de esta se muestra en la Figura 4.10.

De la Figura 4.10 podemos observar que x_1, x_2, \dots, x_l son los vectores característicos obtenidos de las imágenes de entrenamiento de la persona asociada al modelo que será generado, y_1, y_2, \dots, y_{Q-l} son los vectores característicos de las imágenes de entrenamiento del resto de personas (a las cuales posteriormente se les asociará un modelo), l es el número de imágenes de entrenamiento por persona, las etiquetas sirven para diferenciar una clase de otra, y λ es el modelo generado.

Para este proceso se utilizó el SVM de la forma uno contra todos (OVA). Es decir, la clasificación se realiza generalizando los datos en solamente dos clases, donde la clase etiquetada con 1 representa la clase de la persona bajo análisis y la etiquetada con -1 la clase a la cual se quiere diferenciar. Por lo tanto el modelo λ es generado asociando los vectores característicos etiquetados con 1. Es importante mencionar que al utilizar este tipo de entrenamiento el proceso previamente descrito debe ser realizado L veces, para así obtener L modelos, donde L corresponde al número total de las personas que se quiere clasificar $L=Q/l$. Al final de esta etapa los L modelos son almacenados en una base de datos, la cual será utilizada en las dos etapas siguientes.

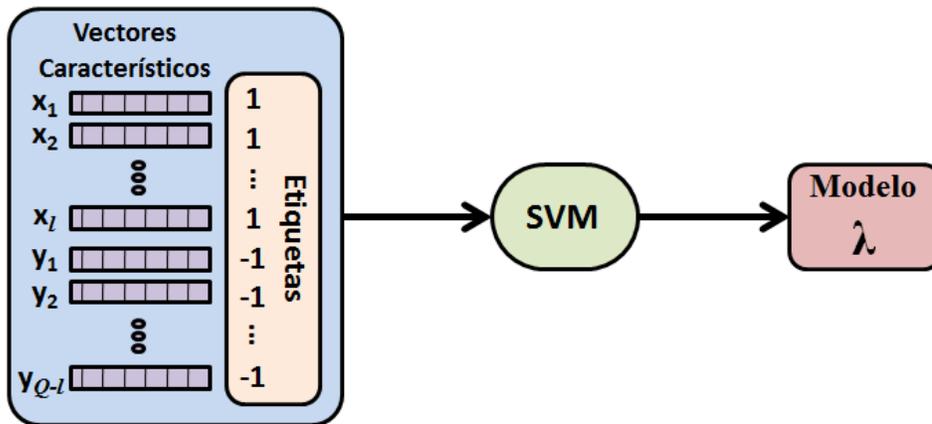


Figura 4.10. Esquema de la etapa de entrenamiento usando SVM.

4.6.2 Etapa de Identificación

La función principal del SVM en esta etapa es la de decidir a qué persona corresponde la imagen bajo análisis. En otras palabras, a qué modelo corresponde el vector característico de la imagen de prueba analizada. El método utilizado para tomar esta decisión se muestra en la Figura 4.11.

En la Figura 4.11 podemos observar que el vector característico de prueba es clasificado por el SVM con respecto a L modelos, para así obtener un valor de coincidencia S por cada clasificación. Con todos los valores de S se forma una vector del cual se elegirá su elemento con máximo valor, siendo este el índice del modelo al que el clasificador seleccionará como ganador. Este proceso es definido por:

$$\Psi = \max\{S_1, S_2, S_3, \dots, S_L\}, \quad (4.27)$$

donde Ψ es el índice del modelo ganador, el cual revela la identidad de la persona a quien pertenece la imagen de prueba representada por el vector característico que entró al SVM. S es un vector de tamaño L formado por los valores de coincidencia obtenidos después de la clasificación realizada por el SVM con todos los modelos generados en la etapa de entrenamiento. λ_Ψ corresponde al modelo de la persona Ψ y L es el número total de los modelos generados en la etapa de entrenamiento, es decir el número total de personas utilizadas para el entrenamiento. Por lo tanto la identidad del modelo λ_Ψ es la que se toma como la decisión final del sistema.

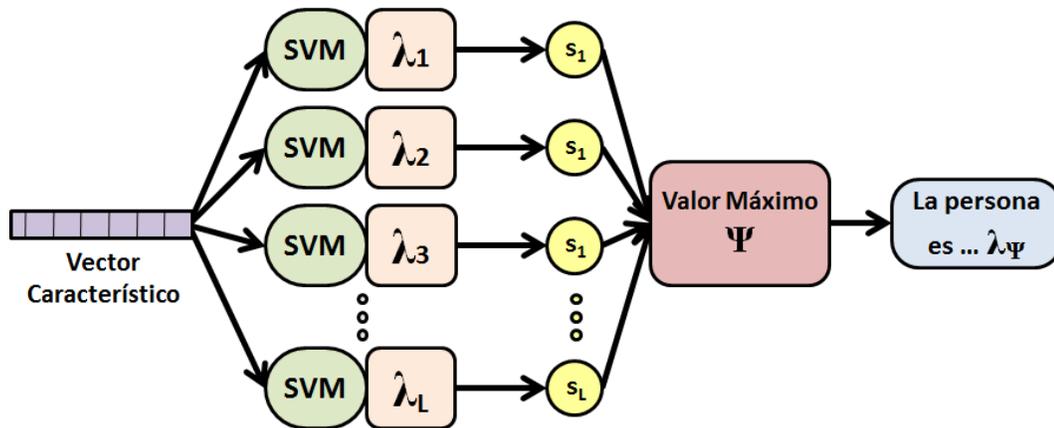


Figura 4.11. Esquema de la etapa de identificación usando SVM.

4.6.3 Etapa de Verificación

En esta etapa, el SVM se encarga de decidir si la identidad que la persona declaró es verdadera o no. Para este tipo de reconocimiento la persona bajo análisis debe declarar su identidad antes de iniciar el proceso del sistema, de esta manera el sistema se encargará de aceptar o rechazar esta declaración. En general, el SVM debe utilizar solo el modelo que la persona bajo análisis dijo ser y decidir si el vector característico de la imagen de esta persona corresponde o no a dicho modelo. La Figura 4.12 muestra el proceso que se lleva a cabo para tomar esta decisión.

En la Figura 4.12 podemos observar que X es la identidad que la persona bajo análisis proporcionó, S es el valor de coincidencia obtenido al clasificar el vector característico de la imagen de la persona bajo análisis con el modelo λ_X , y Th es el umbral del sistema.

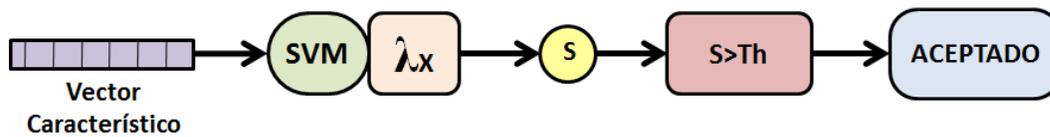


Figura 4.12. Esquema de la etapa de verificación usando SVM.

Por lo tanto, para poder tomar la decisión de aceptar o rechazar la identidad declarada por la persona bajo análisis, es necesario que el valor de coincidencia S sea mayor al umbral Th . Es por eso que el valor de Th toma protagonismo en este tipo de reconocimiento.

Es importante mencionar que los valores de S se encuentran en un rango de entre 0 y 1, donde 1 representa una coincidencia perfecta entre el vector característico bajo análisis y el modelo utilizado para la clasificación, mientras que 0 representa una coincidencia nula. Por consiguiente, el valor de Th también se encuentra entre el rango de 0 y 1. Es por eso que si $Th=1$ el sistema rechazaría todas las declaraciones de identidad que fueran recibidas, mientras que $Th=0$ significa que el sistema aceptaría todas las declaraciones de identidad, sin importar si son verdaderas o falsas.

4.7 Conclusiones

En este capítulo se presentó la metodología necesaria para la implementación del sistema propuesto de Fases Propias con sub-bloques de tamaño óptimo, el cual representa una de las principales aportaciones del presente trabajo. Este sistema brinda la posibilidad de utilizar o descartar la etapa de pre-procesamiento, ya que se espera que el algoritmo de extracción de características sea capaz de resolver los problemas que usualmente ataca la etapa de pre-procesamiento. La parte más relevante del sistema propuesto aquí presentado es el análisis realizado para definir el tamaño de sub-bloque óptimo, donde se concluye que a menor tamaño se obtiene cierta independencia de los cambios de nivel en los píxeles, por lo tanto se sugiere un tamaño óptimo de 2×2 píxeles. Por último, para la implementación de la etapa de clasificación se utiliza el algoritmo de SVM, que para su entrenamiento y clasificación utiliza el método OVA.

CAPÍTULO V

5. SISTEMA PROPUESTO USANDO LBP

En el presente capítulo se propone un sistema de reconocimiento de rostros basado en el algoritmo de patrones binarios locales (LBP, por sus siglas en inglés). El cual se utiliza en la etapa de extracción de características, sin embargo, a diferencia del método convencional, no se utiliza el histograma de la imagen LBP para la estimación del vector característico, sino que para este caso se aplica el algoritmo de análisis de componentes principales (PCA, por sus siglas en inglés) directamente a los coeficientes obtenidos por LBP. Asimismo, se propone un nuevo algoritmo de caracterización llamado: patrones binarios por ventanas (WBP, por sus siglas en inglés). El cual utiliza la teoría básica del método LBP, tomando el pixel central del análisis para caracterizar una ventana. En este caso, por los beneficios que otorga el método WBP que consisten en redimensionar la imagen original al tiempo de extraer sus características, es posible prescindir del algoritmo PCA, culminando en una variante de este método propuesto, denominada DI-WBP. Por último, se utilizan las máquinas de soporte vectorial para la etapa de clasificación del sistema, de la misma forma que se utilizó en el algoritmo propuesto de fases propias (entrenamiento, identificación y verificación).

5.1 Diagrama del Sistema

La Figura 5.1 muestra el diagrama a bloques del sistema propuesto basado en el algoritmo de patrones binarios locales (LBP). La propuesta de este sistema consiste en usar tres modificaciones del método LBP en la etapa de extracción de características.

El primer método propuesto utiliza directamente los coeficientes de LBP junto con el PCA para obtener el vector característico, mientras que en el segundo esquema (WBP, patrones binarios por ventanas), en primer lugar la imagen del rostro se divide en ventanas de 3x3 píxeles, las cuales son caracterizadas solamente por el coeficiente de LBP correspondiente al píxel central de cada ventana. Es importante mencionar que en este método no se utiliza la etapa de redimensión de la imagen, ya que este enfoque reduce aproximadamente 9 veces el tamaño del vector característico. Enseguida, se utiliza el PCA para reducir aún más los vectores característicos, debido a que el PCA proporciona una reducción en el espacio dimensional después de su aplicación. En el tercer método (DI-WBP, patrones binarios por ventanas con redimensión) primeramente la imagen del rostro debe ser redimensionada, a continuación, para obtener el vector de características el método WBP descrito anteriormente es aplicado a esta imagen, proporcionando la ventaja de no utilizar el PCA. Por último, para todos los métodos de extracción de características propuestos se utiliza el método de máquinas de soporte vectorial (SVM) para realizar las tareas de entrenamiento, identificación y verificación (de la misma forma en que se utilizó en el capítulo anterior).

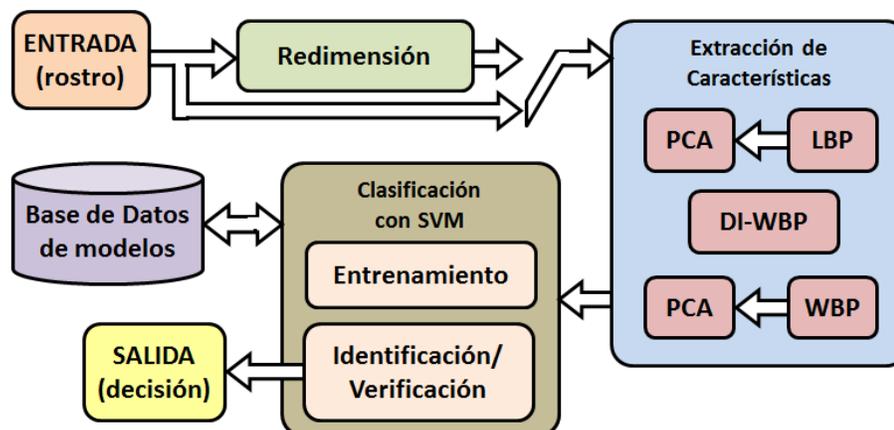


Figura 5.1. Diagrama a bloques del sistema propuesto basado en LBP.

5.2 Extracción de características usando LBP

Como se mencionó en el capítulo 3 sección 3.2.3, el método LBP ha sido utilizado para la extracción de características en muchos sistemas de reconocimiento de rostros [70-73]. Para esta función el LBP ha sido usado de diferentes formas, siempre con el objetivo de obtener un vector característico que pueda representar correctamente la información facial de la imagen de entrada, siendo el histograma de la imagen LBP el método más utilizado para esta tarea. A este método, propuesto en [68], lo llamaremos LBPh, haciendo referencia al nombre: histograma de patrones locales binarios. LBPh es el método convencional que tomaremos como base de comparación para los algoritmos aquí propuestos.

La Figura 5.2 muestra el proceso que lleva a cabo el método LBPh para la obtención del vector característico. En esta figura podemos observar que los pasos del método son: redimensión de la imagen, generación de la imagen LBP (imagen formada con los coeficientes obtenidos al aplicar el algoritmo LBP en la imagen redimensionada) y por último la obtención del histograma de la imagen LBP, definido por la ecuación (3.46) (proceso descrito en la sección 3.2.3).

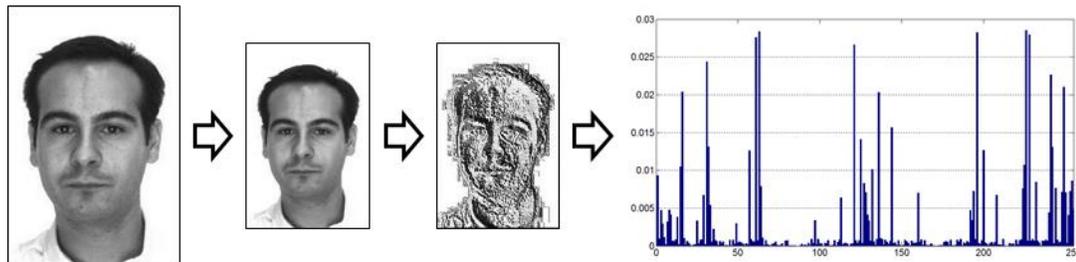


Figura 5.2. Procedimiento de la extracción de características llevada a cabo por el método LBPh.

Un factor importante que se debe tomar en cuenta al utilizar el algoritmo de LBP para extraer las características en un sistema de reconocimiento de rostro, son los puntos de muestra y el tamaño de radio para la aplicación del mismo. Es decir, los valores que se asignaran a (P, R) en la ecuación (3.45), formula general para la aplicación de LBP.

La Figura 5.3 muestra un ejemplo de la aplicación del método LBP a una imagen de rostro variando el valor del radio. (a) muestra la imagen original, (b), (c) y

(d) muestran el resultado de aplicar el algoritmo a la imagen (a) con las variaciones de radio: LBP(8,1), LBP(8,2) y LBP(8,3) respectivamente.

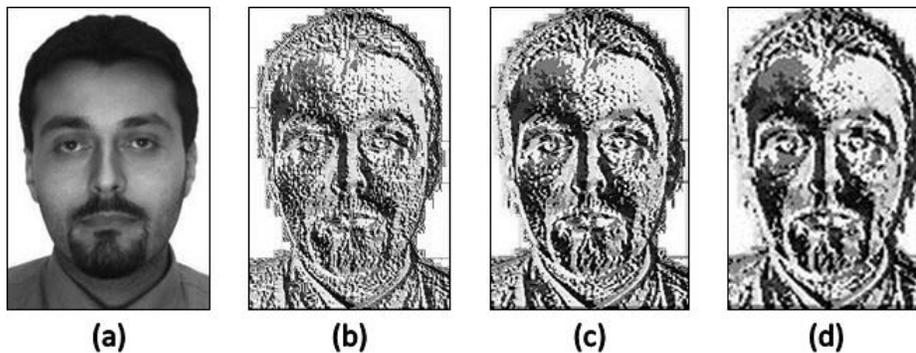


Figura 5.3. Comparación entre imágenes LBP obtenidas con diferente radio. (a) Imagen original. (b) Imagen obtenida con LBP(8,1). (c) Imagen obtenida con LBP(8,2). (d) Imagen obtenida con LBP(8,3)

De la Figura 5.3 podemos observar que mientras el valor del radio se va haciendo más grande, la imagen LBP va mostrando los rasgos faciales más acentuados. Esto es debido a que el algoritmo de LBP toma más píxeles para caracterizar un solo punto. Es importante mencionar que estas tres variaciones de radio serán utilizadas para la evaluación de los sistemas basados en LBP. La siguiente subsección detalla el sistema propuesto basado en LBP, el cual utiliza el algoritmo de PCA para la obtención de los vectores característicos.

5.2.1 LBP usando PCA (LBPC)

El primer método propuesto de este capítulo está basado en la aplicación del LBP en la imagen redimensionada y la utilización de PCA para la formación de los vectores característicos. El PCA se utiliza directamente a los coeficientes obtenidos por LBP, es por eso que a este método lo llamaremos LBPC, haciendo referencia al nombre: coeficiente de patrones locales binarios.

La Figura 5.4 muestra el procedimiento para la obtención de los vectores característicos del método LBPC. El cual consiste en, redimensionar la imagen de entrada, para enseguida aplicar el método LBP y así generar la imagen LBP formada con los coeficientes obtenidos por este algoritmo, por último se aplica el PCA a estas imágenes LBP para la conformación de los vectores característicos.

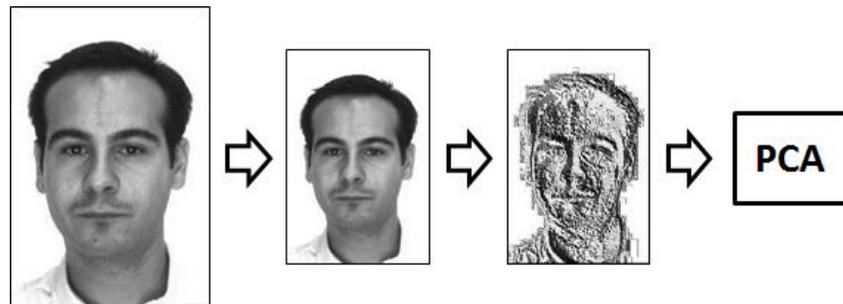


Figura 5.4. Procedimiento de la extracción de características llevada a cabo por el método LBPc.

El procedimiento para la aplicación del PCA es el mismo que el utilizado en el capítulo anterior, solo que esta vez se utilizan las imágenes LBP en vez de los espectros de fase. La Figura 5.5 muestra este procedimiento. En ella se puede observar que las imágenes LBP deben ser convertidas en vectores columna para formar la matriz principal (MP) a la cual se aplicara el PCA. La formación de los vectores característicos consiste en multiplicar la matriz obtenida de la aplicación del PCA (CD) por el vector columna de la imagen LBP bajo análisis.

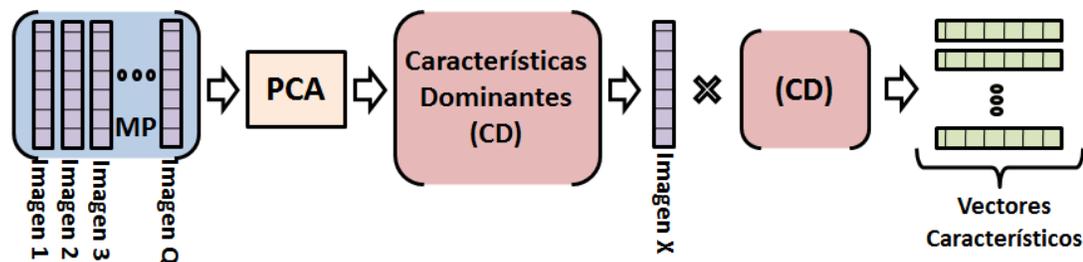


Figura 5.5. Conformación de vectores característicos usando PCA.

5.3 Extracción de características usando WBP

Uno de los problemas que aqueja al método LBP para el caso del reconocimiento de rostros es la complejidad computacional. Debido a que, cuando las imágenes de entrada son muy grandes el proceso del algoritmo es muy tardado, requiriendo un mayor costo computacional, es por eso que las imágenes deben ser redimensionadas antes de aplicar este método. Es por esta razón que el presente trabajo propone un método basado en los beneficios del LBP, en el cual se realiza una redimensión y caracterización simultáneamente.

Este método es llamado patrones binarios por ventanas (WBP, por sus siglas en ingles). Con el objetivo de reducir su complejidad computacional, WBP fue propuesto en base a la forma más básica del método LBP, que es donde se utilizan los parámetros de $P=8$ y $R=1$. En otras palabras, LBP con una vecindad de 3x3 pixeles.

El método WBP consiste en dividir la imagen de entrada en $L \times M$ ventanas no traslapadas de 3x3 pixeles, donde $3L \times 3M$ es el tamaño de la imagen de entrada. Basándose en la ecuación (3.45) que define a LBP (descrita en la sección 3.2.3), podemos definir el método WBP como:

$$WBP(j, k) = \sum_{p=1}^{P-1} s(g_p - g_c) 2^p \quad j = 1, 2, \dots, L \quad k = 1, 2, \dots, M, \quad (5.1)$$

de esta manera, similar a la ecuación (3.45), tenemos:

$$g_p = I_{j,k}(x_p, y_p), \quad p = 1, 2, \dots, P - 1, \quad (5.2)$$

donde $I_{j,k}(x, y)$ representa la (j, k) -ésima ventana de 3x3 pixeles de la imagen de entrada $I(x, y)$, y g_c es el pixel central de la misma ventana. Por lo tanto, el tamaño de la imagen $WBP(j, k)$ es de $L \times M$.

Un ejemplo de la implementación del método WBP se muestra en la Figura 5.6. Donde (a) representa los valores de la imagen de entrada, (b) el resultado de la comparación de los pixeles vecinos con el pixel central, (c) la sustitución de la comparación por sus valores en 2^p y (d) muestra la matriz resultante de la suma de los valores anteriores, es decir la imagen WBP. De esta forma podemos observar que el tamaño de la matriz resultante es 1/9 del tamaño de la imagen original.

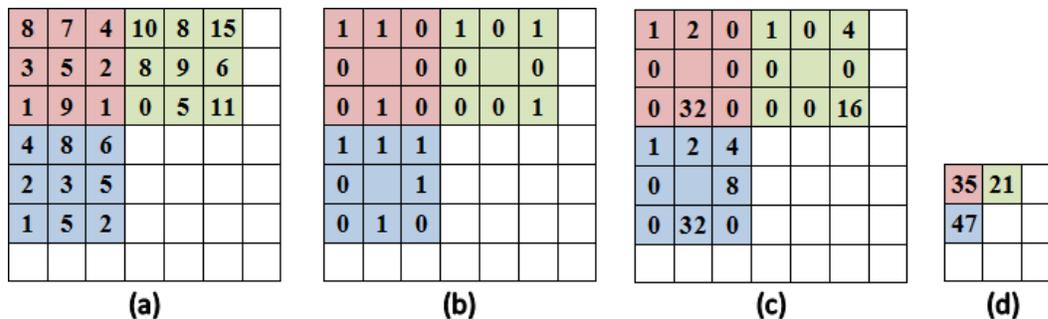


Figura 5.6. Ejemplo de la implementación de WBP. (a) Valores de la imagen de entrada, donde cada ventana es marcada con diferente color. (b) Comparación de cada vecino g_p con el valor central g_c . (c) Sustitución de cada valor por su correspondiente 2^p . (d) Resultado de la suma y formación de la nueva matriz WBP.

La Figura 5.7 muestra un ejemplo donde se aplican los métodos de LBP y WBP a una imagen de rostro, (a) es la imagen de rostro original, (b) es el resultado de aplicar LBP(8,1) a la imagen (a) y (c) es la imagen resultante después de usar WBP en (a). Con esta figura podemos observar que la imagen WBP, de la misma forma que la imagen LBP, resalta los rasgos característicos del rostro original además de que reduce su tamaño significativamente. Por esta razón se espera un buen desempeño del método WBP como extractor de características en los sistemas de reconocimiento de rostros.

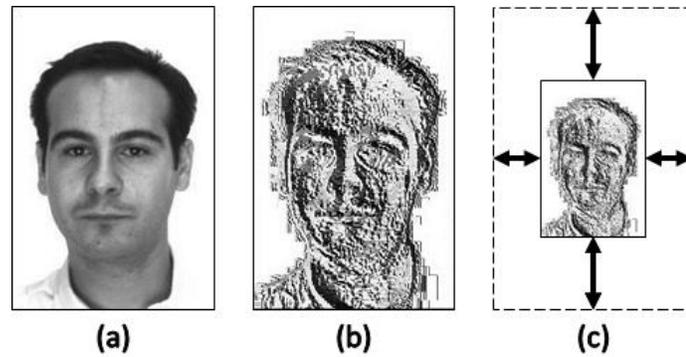


Figura 5.7. Diferencia entre los resultados de los métodos LBP(8,1) y WBP. (a) Imagen original. (b) Resultado de aplicar LBP(8,1) a la imagen (a). (c) Resultado de aplicar WBP en (a).

Para poder utilizar el método WBP como extractor de características es necesario definir el procedimiento para la obtención de los vectores característicos. Este procedimiento es mostrado en la Figura 5.8. El cual a grandes rasgos consiste en: dividir la imagen de entrada en ventanas de 3x3, utilizar el método WBP con la ayuda de la ecuación (5.1) y aplicar el PCA a las imágenes WBP previamente obtenidas. El PCA se utiliza en el procedimiento con la finalidad de reducir la dimensionalidad de los vectores característicos.

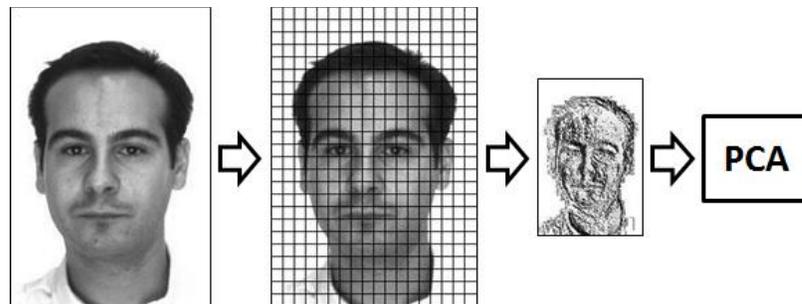


Figura 5.8. Procedimiento de la extracción de características llevada a cabo por el método WBP.

El PCA se aplica de la misma forma que en el método LBPc, siguiendo el procedimiento mostrado en la Figura 5.5. Para ello, es necesario convertir las imágenes WBP de entrenamiento en vectores columna, definidos por:

$$\delta_i = [WBP_0 \quad WBP_1 \quad \dots \quad WBP_{Mr}]^T \quad i=1,2,\dots,Q, \quad (5.3)$$

donde δ_i es el vector columna de la i -ésima imagen WBP, $Mr = LxM - 1$, y Q es el número total de imágenes de entrenamiento. De esta manera, la matriz principal se define como:

$$MP = \begin{bmatrix} \delta_{1,1} & \delta_{1,2} & \dots & \delta_{1,Q} \\ \delta_{2,1} & \delta_{2,2} & \dots & \delta_{2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{Mr,1} & \delta_{Mr,2} & \dots & \delta_{Mr,Q} \end{bmatrix}. \quad (5.4)$$

Por lo tanto, para generar los vectores característicos, es necesario multiplicar los vectores columna obtenidos con (5.3) por la matriz C.D. obtenida al aplicar el método de PCA en (5.4).

5.3.1 WBP con redimensión de la imagen (DI-WBP)

Con la implementación del método WBP es posible disminuir significativamente el costo computacional del sistema de reconocimiento de rostros, ya que no es necesario utilizar una etapa de redimensionamiento previa a este proceso. Pero por otro lado, para la estimación de los vectores característicos es aun necesaria la implementación del método PCA, el cual es un método con alto grado de complejidad computacional. Tomando en cuenta estos factores, y con el objetivo de disminuir el costo computacional en la etapa de extracción de características, se propone una variante del método WBP en la que no sea necesaria la implementación del algoritmo de PCA.

Para esta propuesta es necesaria la inclusión de la etapa de redimensionamiento, la cual se había omitido en el método WBP. Es por eso que el nombre de este nuevo método de extracción de características es DI-WBP, haciendo referencia a: “imagen redimensionada WBP” (proveniente de sus siglas en inglés Decimated Image WBP).

La Figura 5.9 muestra el proceso para la extracción de características usando el método DI-WBP. El cual consiste en: redimensionar la imagen de entrada, dividir la imagen redimensionada en ventanas de 3x3 pixeles, aplicar el método WBP en la imagen redimensionada, y finalmente obtener el vector característico.

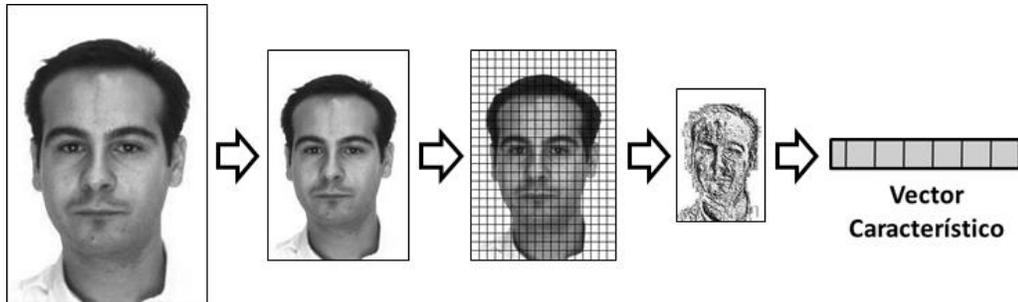


Figura 5.9. Procedimiento para la etapa de extracción de características llevada a cabo por el método DI-WBP.

Para generar los vectores característicos en el método DI-WBP es necesario convertir las matrices resultantes de este método en vectores fila, ya que estos vectores serán utilizados por el SVM en la etapa de clasificación.

Es importante mencionar que el método DI-WBP incluye la etapa de redimensionamiento, pero por otro lado omite la implementación del algoritmo de PCA, siendo esta propuesta la que ofrece el nivel de complejidad computacional más bajo del presente trabajo.

5.4 Clasificación usando SVM

Para la etapa de clasificación se utiliza el método de máquinas de soporte vectorial (SVM). Esta etapa se llevaba a cabo idénticamente que la descrita en el capítulo anterior.

Por lo tanto, para conocer el procedimiento utilizado en las fases de entrenamiento, identificación y verificación, es necesario dirigirse a la sección 4.6, página 74 del presente trabajo.

5.5 Conclusiones

El sistema propuesto presentado en este capítulo ofrece la posibilidad de elegir tres métodos de extracción de características, todos relacionados con el algoritmo de LBP. El primero de ellos (LBPC) presenta la particularidad de utilizar directamente los coeficientes de LBP caracterizados con ayuda de PCA, en contraparte al método convencional que utiliza el histograma de dichos coeficientes (LBPh). La principal propuesta de este sistema es el método denominado WBP, el cual reduce la complejidad computacional del algoritmo LBP, ya que caracteriza la imagen al mismo tiempo que la redimensiona, de manera que los datos ingresados al PCA son de menor dimensión preservando la capacidad de caracterización del LBP. En base a este nuevo método de clasificación y tomando en cuenta los beneficios que este proporciona, se propone una variante llamada DI-WBP. La cual es capaz de prescindir de la herramienta PCA, presentando una disminución en la complejidad del algoritmo muy significativa, generando vectores característicos 81 veces más pequeños que los proporcionados por el método LBPC. Finalmente el método de clasificación utilizado en el sistema propuesto es el mismo que el del capítulo anterior, usando SVM para las fases de entrenamiento, identificación y verificación.

CAPÍTULO VI

6. SISTEMA PROPUESTO PARA RECONOCIMIENTO DE EXPRESIONES FACIALES

Este capítulo presenta un sistema para el reconocimiento de expresiones faciales propuesto en base a una segmentación del rostro en 4 regiones fundamentales y al método de Fases Propias basado con sub-bloques de tamaño óptimo (propuesto en el capítulo 4). Donde la principal peculiaridad de este sistema, reside en el proceso de estimación de los vectores característicos, basándose en los beneficios otorgados por la utilización de las 4 regiones del rostro independientemente.

El sistema propuesto utiliza el algoritmo de máquinas de soporte vectorial para la etapa de clasificación. El cual se utiliza para clasificar 6 clases, una por cada expresión facial básica: ira, asco, miedo, felicidad, tristeza y sorpresa, de entre las cuales el SVM tomará la decisión final en la etapa de reconocimiento. Dentro de esta etapa se propone un nuevo método de decisión, llamado “Enfoque del Valor Modal”, el cual hace uso de las decisiones tomadas por tres o más clasificadores para elegir el valor modal y unificar dichas decisiones.

6.1 Diagrama del Sistema

El diagrama a bloques del sistema propuesto para el reconocimiento de expresiones faciales (FER, por sus siglas en inglés) se muestra en la Figura 6.1. En la etapa de segmentación del rostro se utiliza la propuesta de [39] para la segmentación del rostro en 4 regiones fundamentales: ojos-cejas, frente, boca y nariz. La etapa de extracción de características utiliza el algoritmo de fases propias basado en sub-bloques con tamaño de bloque óptimo (propuesto en el capítulo 4), con la peculiaridad que se utiliza independientemente en cada una de las regiones fundamentales del rostro. La primera propuesta de este sistema reside en la conformación del vector característico, donde utilizando los vectores característicos obtenidos a partir de las diferentes regiones faciales se genera el vector característico final, el cual representa la información más importante de una expresión facial. De esta manera se tiene la posibilidad de utilizar menos de 4 regiones faciales para la caracterización del rostro. Finalmente en la etapa de clasificación se utiliza el algoritmo de SVM, el cual se encarga de clasificar el vector característico de la imagen bajo análisis dentro de las 6 expresiones básicas (ira, asco, miedo, felicidad, tristeza y sorpresa). Después de la decisión tomada por el SVM y dependiendo del número de regiones utilizadas en el proceso, el sistema propuesto brinda la posibilidad de usar el método de decisión “enfoco del valor modal”. El cual consiste en unificar las decisiones tomadas por varios clasificadores usando el valor modal de estos resultados, siendo esta la segunda propuesta de este sistema de reconocimiento de expresiones faciales.

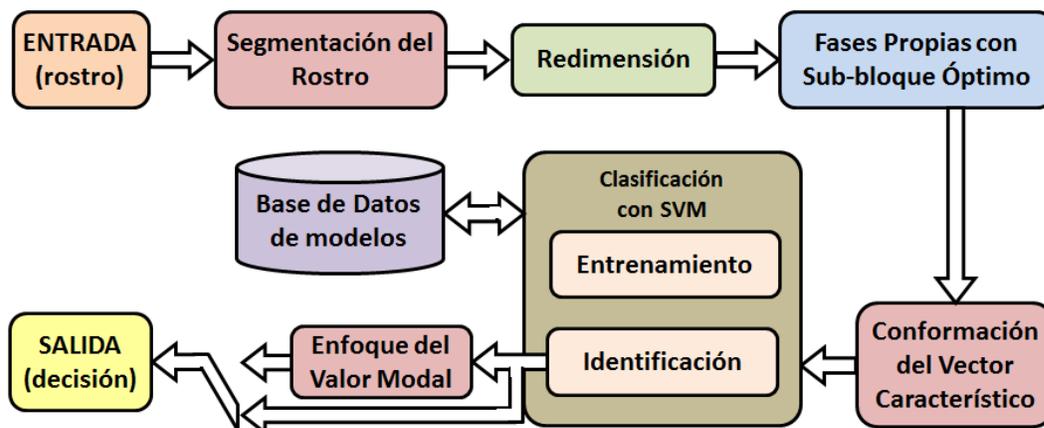


Figura 6.1. Diagrama a bloques del sistema propuesto para el reconocimiento de expresiones faciales.

6.2 Segmentación del Rostro

Las imágenes de rostro entran directamente a esta primera etapa, que consiste en la segmentación de 4 regiones fundamentales, las cuales contienen información de los ojos-cejas, frente, boca y nariz. Estas regiones fundamentales fueron propuestas en [39], con el objetivo de generalizar una expresión facial sin necesidad de utilizar toda la información del rostro. Debido a que en este tipo de sistemas la identidad de la persona no es importante, lo primordial es reconocer la expresión facial que está mostrando el rostro de esa persona.

El proceso para la segmentación, en primer lugar utiliza un robusto algoritmo para la detección de rostros [94], con el cual se detecta eficientemente la posición de los ojos. Enseguida, tomando en cuenta esta posición se obtiene la distancia que existe entre ambos iris (DI, distancia medida en píxeles). En base a esta distancia y a la posición de los ojos, se encuentra la localización de la región de la boca, con la que es posible localizar el resto de las regiones [39].

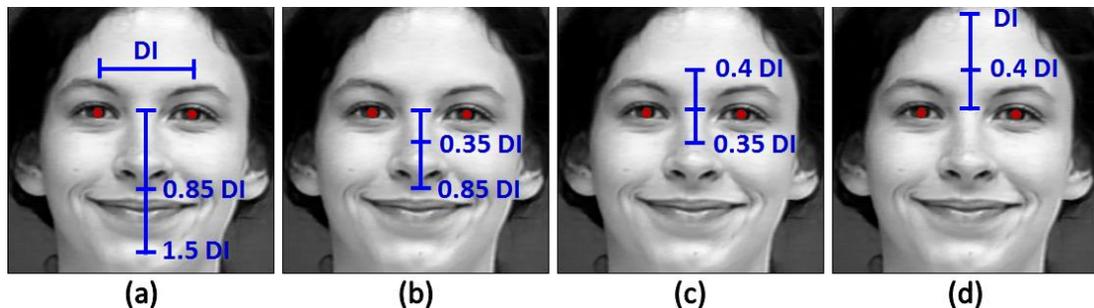


Figura 6.2. Localización de las 4 regiones fundamentales del rostro basada en la distancia entre iris (DI) y la posición de los ojos. (a) Localización de la región de la boca. (b) Localización de la región de la nariz. (c) Localización de la región de los ojos-cejas. (d) Localización de la región de la frente.

La Figura 6.2 muestra las relaciones necesarias para la localización de las 4 regiones fundamentales, a partir de la posición de los ojos y la distancia entre iris (DI). (a) Muestra la relación fundamental, donde se propone que la parte inferior de la región de la boca se encuentra a $1.5DI$ desde la posición de los ojos hacia el fondo de la imagen, mientras que la parte superior de la misma se encuentra a $0.85DI$. (b) muestra la relación para definir la región de la nariz, siendo $0.85DI$ la parte inferior y $0.35DI$ la parte superior de dicha región. Por otro lado, la región de los ojos-cejas es

definida por (c), donde se observa que la parte inferior de la región se encuentra a $0.35DI$ desde la posición de los ojos al fondo de la imagen y la parte superior a $0.4DI$ desde la misma posición pero hacia la parte superior de la imagen. Por último, la región de la frente se localiza a $0.4DI$ y DI desde la posición de los ojos hacia la parte superior de la imagen. Una vez definida la localización de las 4 regiones, se procede a recortar cada una de ellas, respetando las dimensiones preestablecidas para este fin.

Las dimensiones de cada región del rostro también son definidas en relación a DI . El tamaño para las regiones de la boca, nariz, ojos-cejas y frente es de $1.2DI(\text{ancho}) \times 0.65DI(\text{alto})$, $1.7DI \times 0.5DI$, $2DI \times 0.75DI$ y $DI \times 0.6DI$ respectivamente. La Figura 6.3 muestra un ejemplo de la segmentación de una imagen del rostro en estas 4 regiones. Del lado izquierdo de la figura se puede apreciar la imagen del rostro completa, solamente marcada con la localización de cada una de las regiones. Por otra parte, del lado derecho se observan las 4 regiones ya segmentadas, indicando las dimensiones de cada una de ellas.

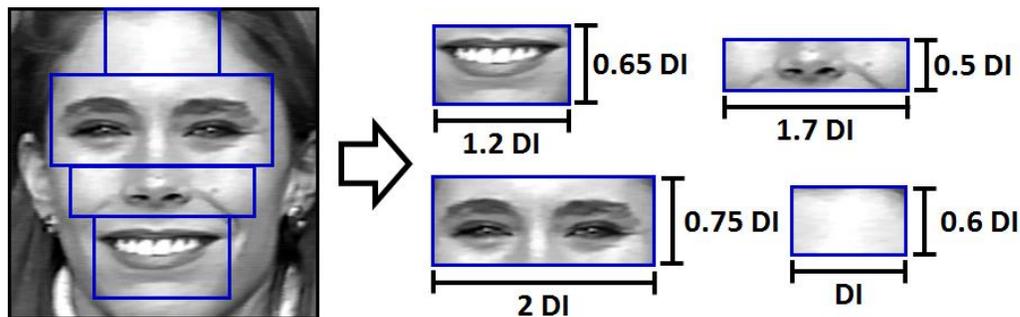


Figura 6.3. Ejemplo de la segmentación del rostro y dimensiones de las 4 regiones fundamentales: boca, nariz, ojos-cejas y frente.

La etapa de segmentación del rostro es parte vital en el sistema propuesto, ya que brinda la posibilidad no solo de excluir algunas partes del rostro en caso de la presencia de oclusión parcial, sino también de evaluar la contribución individualmente de cada región facial en el reconocimiento de expresiones faciales. Lo cual permite obtener más de una decisión por cada imagen de entrada y con ello utilizar el método propuesto de enfoque del valor modal que se explicará en las siguientes secciones. Por lo tanto, el desempeño del sistema depende de una eficiente segmentación de estas 4 regiones, es por eso que vale la pena mencionar que la propuesta fundamental de este sistema no consiste en la segmentación aquí descrita, siendo opción utilizar otro método de segmentación si la situación así lo requiere.

6.3 Extracción de Características usando Fases Propias

Para la etapa de extracción de características se propone la utilización del algoritmo de fases propias basado en sub-bloques con tamaño de bloque óptimo, es decir usando bloques de 2x2 pixeles para la extracción del espectro de fase. Siendo esta, una forma de evaluar el algoritmo propuesto del capítulo 4 en un escenario diferente para el cual originalmente fue propuesto, en este caso para el reconocimiento de expresiones faciales.

La Figura 6.4 muestra el proceso de aplicación del algoritmo de fases propias basado en sub-bloques de 2x2. El proceso consiste primero en redimensionar la imagen de entrada, que en este caso corresponde a una región del rostro. Después de la redimensión se aplica la segmentación en sub-bloques de 2x2 pixeles, para poder obtener el espectro de fase de cada sub-bloque y conformar una matriz con estos espectros de fase obtenidos. Para finalmente aplicar el método de PCA. Este proceso se explica detalladamente en las secciones 4.3 y 4.4 del capítulo 4 del presente trabajo. La única diferencia en este proceso consiste en la aplicación del PCA.

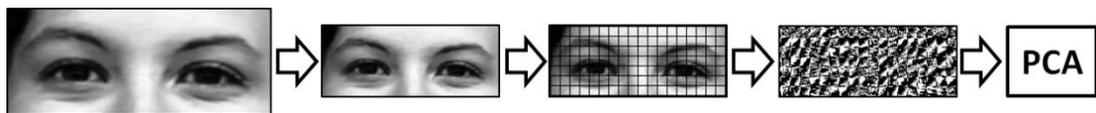


Figura 6.4. Procedimiento del algoritmo de fases propias con tamaño de bloque óptimo para el sistema de reconocimiento de expresiones faciales.

Esta etapa se encarga solamente de obtener la matriz de características dominantes (CD), es decir no se obtienen los vectores característicos en este punto del sistema. La Figura 6.5 muestra el proceso que se lleva a cabo para la obtención de dicha matriz. Donde, en esta ocasión se obtendrán 4, una por cada región del rostro.

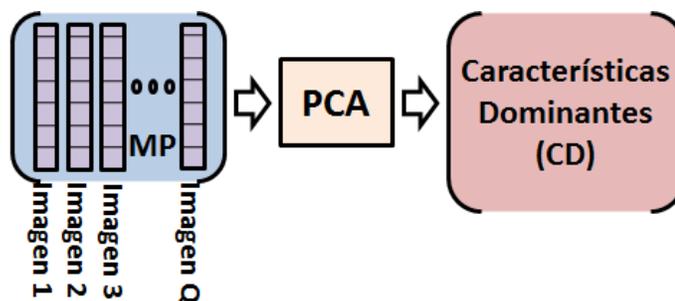


Figura 6.5. Proceso de aplicación del PCA para FER.

6.4 Conformación del Vector Característico

Los vectores característicos se generan al multiplicar la matriz de características dominantes (CD) por cada vector columna de la imagen bajo análisis (procedimiento descrito en la sección 4.5). Es importante mencionar que de manera similar a la etapa anterior, este proceso se aplica a cada región del rostro de forma independiente, siendo cada vector característico definido por:

$$y_j = \phi_j CD_j \quad j = 1, 2, \dots, 4, \quad (6.1)$$

donde y_j representa el vector característico de la j -ésima región del rostro (ojos-cejas, frente, boca o nariz), ϕ es el vector columna de la región bajo análisis y CD representa la matriz de características dominantes de la misma. De esta manera el vector característico final se forma concatenando los vectores característicos de cada región del rostro:

$$Y = \bigcup_{j=1}^4 y_j, \quad (6.2)$$

siendo Y el vector característico final, el cual representa la información más importante de la expresión facial bajo análisis. Información obtenida individualmente por las 4 regiones fundamentales del rostro: ojos-cejas, frente, boca y nariz. Regiones específicamente definidas para poder caracterizar eficientemente los rasgos más importantes de una expresión facial mostrada en una imagen de rostro.

La Figura 6.6 muestra el proceso para la obtención del vector característico final, tomando en cuenta los vectores característicos individuales de cada una de las 4 regiones fundamentales del rostro. En esta figura se puede observar cómo se deben obtener previamente los vectores característicos de cada región del rostro para finalmente conformar el vector característico final (Y). Los subíndices o , f , b , y n representan las regiones del rostro de ojos-cejas, frente, boca y nariz respectivamente. De manera que ϕ_o representa el vector columna obtenido de la región de ojos-cejas, y CD_o la matriz de características dominantes de la misma región. Por lo tanto la primera línea de la Figura 6.6 representa la conformación del vector característico de la región de ojos-cejas (y_o).

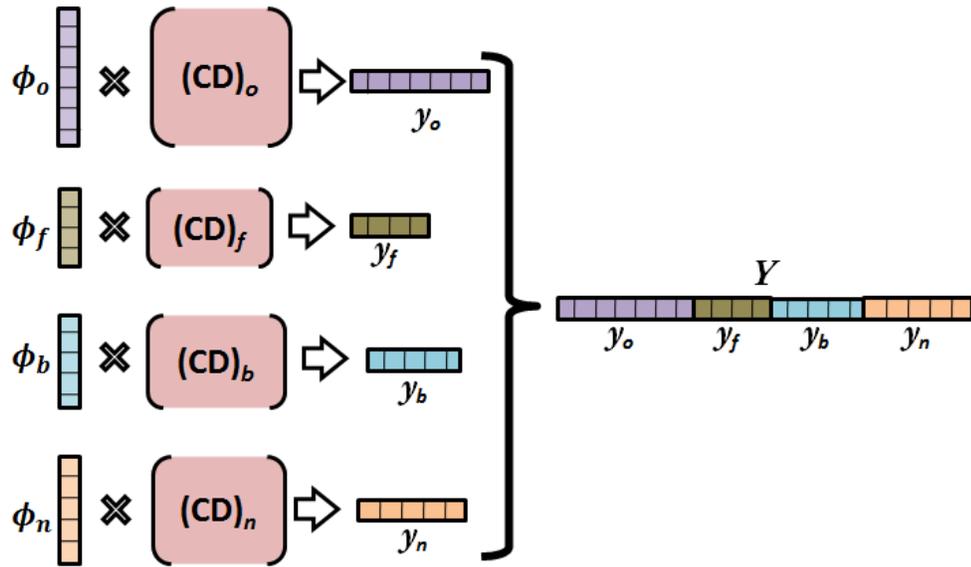


Figura 6.6. Proceso de conformación del vector característico final utilizando los vectores característicos de las 4 regiones fundamentales: ojos-cejas (y_o), frente (y_f), boca (y_b), y nariz (y_n).

Es importante mencionar que el vector característico final depende del número de regiones fundamentales del rostro utilizadas. De manera que para representar una imagen del rostro hasta 4 vectores característicos individuales pueden conformar su vector característico final, aunque en el caso de oclusiones parciales en la imagen, el número podría ser inferior a 4. Por lo tanto, utilizando una imagen de entrada que tenga disponibles las 4 regiones fundamentales del rostro, las combinaciones de los posibles vectores característicos finales pueden ser hasta 15. La Figura 6.7 muestra ejemplos de posibles vectores característicos finales obtenidos de una imagen de entrada que fue segmentada en las 4 regiones fundamentales del rostro. (a), (b), (c) y (d) muestran vectores característicos finales utilizando combinaciones de 4, 3, 2 y 1 vectores característicos individuales respectivamente. Nótese que el orden de concatenación de estos no se altera, respetando siempre el orden de ojos-cejas, frente, boca y nariz, solamente eliminando la región que no se encuentre disponible.

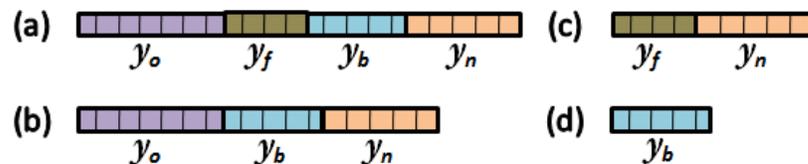


Figura 6.7. Ejemplo de diferentes vectores característicos finales obtenidos de una misma imagen de entrada. (a) Usando las 4 regiones. (b) Usando las regiones de ojos-cejas, boca y nariz. (c) Usando las regiones de frente y nariz. (d) Usando solo la región de la boca.

6.5 Clasificación usando SVM

Para la etapa de clasificación, de la misma manera que en los anteriores sistemas propuestos, se utiliza el método de máquinas de soporte vectorial (SVM) con los parámetros descritos en la sección 4.6.

En la fase de entrenamiento el SVM recibe todos los vectores característicos finales de las imágenes de entrenamiento, pero esta vez el proceso descrito en la sección 4.6.1 solo se repetirá 6 veces, formando solamente 6 modelos diferentes, uno por cada expresión facial básica: ira, asco, miedo, felicidad, tristeza y sorpresa. De tal manera que de la Figura 4.10, $L = 6$ (representando una constante en este tipo de sistemas), y l representa el número de imágenes de entrenamiento utilizadas para cada expresión facial. Es importante aclarar que en este caso el número de imágenes de entrenamiento no necesariamente debe ser el mismo para cada expresión facial, y se recomienda utilizar imágenes de diferentes sujetos para cada conjunto de imágenes de entrenamiento por expresión facial. Al final de la fase de entrenamiento los 6 modelos obtenidos deben ser almacenados en una base de datos para poder utilizarlos en la fase de reconocimiento.

Para la fase de reconocimiento, el SVM se utiliza de manera muy similar a la etapa de identificación descrita en la sección 4.6.2, con la diferencia de que esta vez el sistema solo debe clasificar con respecto a 6 modelos. En otras palabras el sistema debe identificar a cuál de las 6 expresiones facial básicas corresponde el vector característico bajo análisis.

La fase de reconocimiento consiste en ingresar el vector característico final bajo análisis a SVM y clasificarlo con cada uno de los 6 modelos previamente obtenidos ($\lambda_{1...6}$) de esta manera se obtendrán 6 valores de coincidencia diferentes, de entre los cuales se debe elegir el valor mayor, siendo este el índice del modelo al que el clasificador seleccionará como ganador. Este proceso es definido por:

$$\Psi = \max\{S_1 \ S_2 \ S_3 \ S_4 \ S_5 \ S_6\}, \quad (6.3)$$

donde Ψ representa el índice del modelo ganador, el cual revela la expresión facial que está mostrando la imagen de la persona bajo análisis. S_i es el valor de coincidencia obtenido al clasificar el vector característico final con el modelo λ_i .

Por lo tanto la expresión facial correspondiente al modelo λ_{Ψ} es la que el sistema toma como decisión final. La Figura 6.8 muestra el método utilizado para tomar la decisión final del sistema a partir de un vector característico final. Cabe señalar que este vector característico representa la información del rostro necesaria para describir la expresión facial de la imagen bajo análisis, utilizando de 1 a 4 regiones faciales para su proceso de generación.

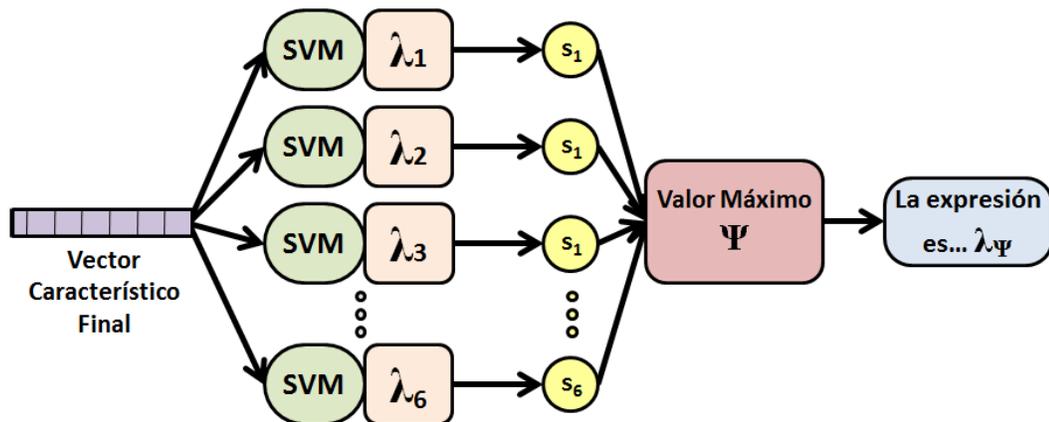


Figura 6.8. Esquema de la etapa de reconocimiento para FER usando SVM.

6.6 Enfoque del Valor Modal

De acuerdo al método de extracción de características propuesto en este sistema, es posible caracterizar en más de una forma una sola imagen del rostro. Esto se debe a la posibilidad que nos brinda la utilización de 4 regiones faciales independientemente, pudiendo concatenar sus vectores característicos entre sí y obtener hasta un máximo de 15 combinaciones a partir de estas 4 regiones fundamentales. Por cada una de estas posibles representaciones de la imagen, se entrena y clasifica el SVM individualmente, por lo tanto obtenemos varias decisiones independientes de una sola imagen de entrada. Es por eso que una propuesta para este sistema está dirigida a la unificación de dichas decisiones, utilizando el método del enfoque del valor modal. Un método de decisión que ayudará a mejorar el reconocimiento general del sistema.

El método del enfoque del valor modal consiste en tomar como decisión final la decisión más frecuente (valor modal) obtenida por un cierto grupo de clasificadores, cada uno asociado a una sola imagen de entrada.

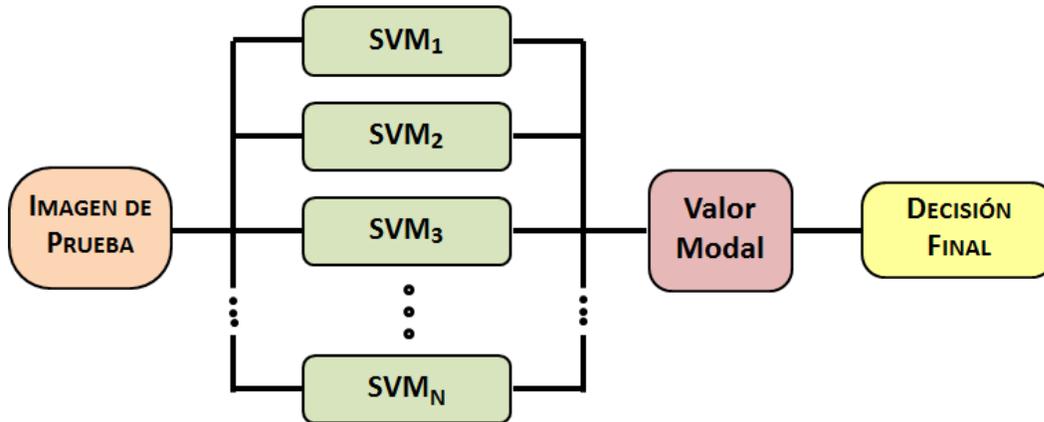


Figura 6.9. Diagrama del método de decisión: Enfoque del Valor Modal.

La Figura 6.9 muestra el procedimiento para aplicar el método de enfoque del valor modal. De esta figura podemos observar que más de un SVM es aplicado a una sola imagen de prueba, donde N depende del número de clasificadores utilizados en el enfoque, los cuales a su vez dependen de las combinaciones de regiones faciales empleadas en la etapa de conformación del vector característico. Con la finalidad de tomar la decisión final, se selecciona el valor modal de las decisiones individuales de todos los clasificadores utilizados en el enfoque. Es importante mencionar que para poder aplicar este método de decisión, al menos 3 clasificadores deben ser utilizados, ya que con solo dos decisiones no es posible calcular correctamente el valor modal.

Tabla 6.1. Ejemplo de la aplicación del enfoque del valor modal.

Muestra	1	2	3	4	Resultado
Expresión	Ira	Ira	Miedo	Miedo	--
Imagen					--
SVM ₁	Ira	Asco	Miedo	Miedo	3/4
SVM ₂	Ira	Ira	Miedo	Asco	3/4
SVM ₃	Miedo	Ira	Felicidad	Miedo	2/4
Modal	Ira	Ira	Miedo	Miedo	4/4

La Tabla 6.1 muestra un ejemplo de la aplicación del método del enfoque del valor modal utilizando 3 clasificadores y 4 imágenes de prueba. En este ejemplo, dos de las imágenes de prueba están mostrando la expresión de ira y las otras dos la expresión de miedo. Ninguno de los 3 clasificadores utilizados: SVM₁, SVM₂ y SVM₃ obtuvieron resultados de reconocimiento del 100%; los aciertos de cada clasificador son 3/4, 3/4 y 2/4 respectivamente. Sin embargo, la unificación de las decisiones de estos 3 clasificadores mediante la utilización de los valores modales por cada imagen de prueba, lleva al 100% de reconocimiento, obteniendo 4/4 aciertos como se muestra en la última fila de la Tabla 6.1. De esta manera al aplicar el método del enfoque del valor modal se puede mejorar el reconocimiento general del sistema.

En el caso potencial de que un cierto número de clasificadores provean el mismo número de decisiones positivas, los valores de coincidencia de las decisiones de estos clasificadores son promediados y se toma como decisión final el resultado con el promedio más alto. La Tabla 6.2 muestra un ejemplo de este caso en especial, donde se utilizan 4 clasificadores para tomar la decisión final de una imagen de prueba que está mostrando la expresión facial de ira. En este caso, la decisión de los clasificadores SVM₁ y SVM₂ es ira, mientras que la decisión de SVM₃ y SVM₄ es miedo. Sin embargo, teniendo en cuenta la media entre ellos el valor más alto fue proporcionado por el promedio entre los valores de coincidencia de SVM₁ y SVM₂, por lo tanto la decisión final tomada en este ejemplo es la expresión facial de ira.

Tabla 6.2. Caso especial para la aplicación del enfoque del valor modal, cuando un cierto número de clasificadores proveen la misma decisión.

Expresión	Ira		
Imagen			
SVM ₁	Ira	0.82	media
SVM ₂	Ira	0.89	0.855
SVM ₃	Miedo	0.75	media
SVM ₄	Miedo	0.80	0.775
Decisión Final	Ira		

6.7 Conclusiones

En este capítulo se presentó la propuesta referente a los sistemas de reconocimiento de expresiones faciales, la cual reside en la utilización del método de Fases Propias con sub-bloques de tamaño óptimo y el método de decisión denominado Enfoque del Valor Modal. El sistema propuesto basa su funcionamiento en la segmentación del rostro en 4 regiones fundamentales, las cuales fueron definidas para poder caracterizar correctamente una expresión facial, enfocándose en las secciones de ojos-cejas, frente, boca y nariz. En base a estas regiones faciales, el algoritmo de Fases Propias con sub-bloques de tamaño óptimo es aplicado, ofreciendo la posibilidad de generar vectores característicos independientes por cada región facial, incluyendo las posibles combinaciones de estos. De esta manera se obtendrán múltiples resultados en base a una sola imagen de rostro, es aquí donde el Enfoque del Valor Modal actúa, eligiendo el valor modal como decisión final del sistema.

CAPÍTULO VII

7. RESULTADOS EXPERIMENTALES

En este capítulo se presenta una extensa evaluación de los resultados experimentales obtenidos por los tres sistemas propuestos: Reconocimiento de Rostros usando Fases Propias, donde la principal propuesta es el uso de sub-bloques de tamaño óptimo; Reconocimiento de Rostros usando LBP, donde se proponen los métodos de WBP y DI-WBP que presentan una disminución significativa en la complejidad computacional de LBP; y Reconocimiento de Expresiones Faciales, donde se plantea el uso del algoritmo de Fases Propias con sub-bloques de tamaño óptimo y se propone un nuevo método de decisión denominado Enfoque del Valor Modal.

En la primera parte del capítulo se presentan las bases de datos que fueron utilizadas para la obtención de los resultados que aquí se presentan. Asimismo, se detallan las modificaciones que fueron realizadas en dichas bases para una evaluación apropiada de los sistemas propuestos. Vale la pena mencionar que en cada sección de este capítulo se presenta la forma común de evaluar los resultados, es decir se describe la forma de medir el rendimiento de un sistema para identificación, verificación y reconocimiento de expresiones faciales.

7.1 Bases de Datos

Para la evaluación de los sistemas propuestos en este trabajo se utilizaron dos bases de datos. La base de datos “AR Face Database” [95] para los dos sistemas propuestos de reconocimiento de rostros y la base de datos “Cohn-Kanade Database” [96] para el sistema propuesto de reconocimiento de expresiones faciales. Ambas bases fueron modificadas para poder evaluar apropiadamente cada uno de los sistemas propuestos, estas modificaciones así como las especificaciones de las bases de datos se presentan a continuación.

7.1.1 Base de Datos AR

La base de datos AR [95] fue creada en la universidad estatal de Ohio, esta contiene más de 3000 imágenes a color de rostros de 120 personas (65 hombres y 55 mujeres). El tamaño de las imágenes es de 192x288 pixeles y fueron capturadas en un estricto ambiente controlado con el cual se obtuvieron vistas frontales de rostros con diferentes expresiones faciales (sorpresa, felicidad y enojo), condiciones de iluminación y oclusión parcial (usando lentes oscuros y bufanda). Cada sujeto de la base de datos participó en dos sesiones con 14 días de diferencia, es decir la mitad de la base de datos fue tomada 14 días después de la primera sesión.

La base de datos incluye 26 imágenes por sujeto, 13 fueron tomadas en la primera sesión y 13 en la segunda. El orden y las especificaciones de las imágenes son como se muestra a continuación:

1. Expresión neutral.
2. Felicidad.
3. Enojo.
4. Sorpresa.
5. Luz izquierda encendida.
6. Luz derecha encendida.
7. Ambas luces encendidas.
8. Usando lentes oscuros.
9. Usando lentes oscuros y luz izquierda encendida.
10. Usando lentes oscuros y luz derecha encendida.

11. Usando bufanda.
12. Usando bufanda y luz izquierda encendida.
13. Usando bufanda y luz derecha encendida.
14. – 26. Segunda sesión en las mismas condiciones de 1 a 13.

La Figura 7.1 muestra dos sujetos que forman parte de la base de datos (el primero en la lista de hombre y la primera en la de mujeres) con sus respectivas 26 imágenes, la primera fila muestra las 13 imágenes capturadas en la primera sesión y la segunda fila muestra las imágenes capturadas en la segunda sesión.



Figura 7.1. Ejemplo de las imágenes de dos sujetos de la base de datos AR.

7.1.1.1 Transformación de Iluminación

Para poder evaluar apropiadamente los algoritmos propuestos de reconocimiento de rostros la base de datos AR fue modificada. La modificación consiste en expandir el número de imágenes con el objetivo de obtener un rango más amplio de cambios de iluminación. Esta expansión se ve reflejada obteniendo 5 imágenes con diferentes tipos de iluminación por cada imagen original.

Es importante aclarar que para la correcta utilización de los algoritmos propuestos en las capítulos 4 y 5, las imágenes que se introducen al sistema deben ser en escala de grises [0-255]. Es por eso que todas las imágenes resultantes de la transformación de iluminación se encuentran en escala de grises.

Para realizar la transformación de la iluminación en las imágenes originales de la base de datos se utilizó la transformada gamma, definida por:

$$s = cr^\gamma, \quad (7.1)$$

donde s es el valor de intensidad de gris de salida, c es una constante, r es el valor de intensidad de gris de entrada y γ es el factor gamma. Cuando el factor gamma es menor que uno, la mayor parte del rango de valores de entrada se mapea a valores altos (blancos) del rango de salida, lo que provoca que la imagen se aclare. Mientras que si el factor gamma es mayor a uno, la mayor parte del rango de valores de entrada se mapea a valores bajos (oscuros) del rango de salida, provocando que la imagen se oscurezca.

Como se mencionó anteriormente, de una imagen original se obtuvieron 5 imágenes con variaciones en la iluminación, estas variaciones se diferencian entre sí con respecto al factor gamma utilizado y al método de transformación. Las 5 imágenes resultantes a estas variaciones de iluminación se enlistan a continuación:

1. En escala de grises.
2. Oscura (con un factor gamma de $\gamma = 2.2$).
3. Clara (con un factor gamma de $\gamma = 0.5$).
4. Mitad izquierda del rostro oscura.
5. Mitad derecha del rostro oscura.



Figura 7.2. Ejemplo de las 5 variaciones de iluminación aplicadas a la base de datos AR.

La Figura 7.2 muestra las 5 variaciones de iluminación aplicadas a dos imágenes de la base de datos. En la figura se puede observar una imagen original de la base de datos AR seguida de las 5 variaciones obtenidas con las transformaciones antes mencionadas.

7.1.1.2 Imágenes Utilizadas para las Pruebas

Después de haber modificado la base de datos original, se obtuvieron como resultado 130 imágenes por persona (26x5) formando una base de datos AR extendida con un total de 15 600 imágenes. Para las pruebas realizadas con los algoritmos propuestos de reconocimiento de rostros solo se utilizaron 100 imágenes por persona, excluyendo las imágenes que incluyen el uso de la bufanda como oclusión parcial. Es decir, en este trabajo se utilizó una base de datos AR extendida con un total de 12 000 imágenes.



Figura 7.3. Ejemplo de las imágenes del set A.

Las imágenes incluidas en esta base de datos extendida fueron divididas en dos grupos: el set A y el set B. El set A cuenta con 70 imágenes por personas que incluyen cambios en la iluminación y diferentes expresiones faciales. Por otro lado, el set B cuenta con 30 imágenes por persona que incluyen cambios en la iluminación y el uso de lentes oscuros como oclusión parcial. La Figura 7.3 muestra las 70 imágenes del set A y la Figura 7.4 muestra las 30 imágenes del set B de un sujeto de la base de datos.

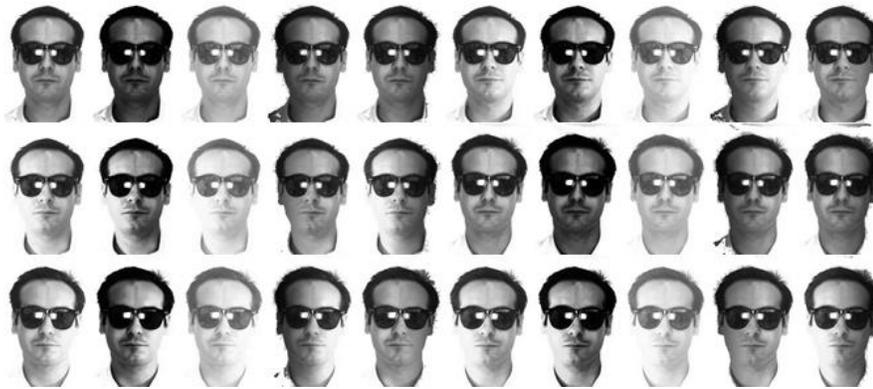


Figura 7.4. Ejemplo de las imágenes del set B.

7.1.2 Base de Datos Cohn-Kanade

La base de datos Cohn-Kanade [96] fue creada gracias a un proyecto interdisciplinario en el cual intervinieron los departamentos de psicología y robótica de dos universidad de Pittsburg. La base de datos contiene más de 400 secuencias de imágenes de 97 personas ente 18 y 30 años de edad (65 por ciento mujeres y 45 hombres). Cada secuencia de imágenes inicia con una expresión neutral y termina con una expresión pico de entre las 6 expresiones básicas (ira, asco, miedo, felicidad tristeza y sorpresa).

Cada secuencia de imágenes fue capturada en un estricto ambiente controlado con una resolución de 640x480 pixeles en escala de grises. Cada sujeto de la base de datos fue instruido para realizar cada una de las 6 expresiones básicas frente a la cámara, en condiciones óptimas de iluminación y con un fondo uniforme. Es importante mencionar que para este trabajo solo se utilizaron las expresiones pico de cada secuencia.

Un ejemplo de las imágenes con expresiones pico se muestra en la Figura 7.5. En esta figura se muestran las 6 expresiones básicas desarrolladas por 3 sujetos de la base de datos.



Figura 7.5. Ejemplo de imágenes pico de la base de datos Cohn-Kanade.

Para poder utilizar correctamente el algoritmo propuesto de reconocimiento de expresiones faciales, las imágenes pico de la base de datos fueron recortadas obteniendo únicamente la sección del rostro. Este recorte se realizó utilizando el algoritmo Viola-Jones [97] para la detección de rostros. La Figura 7.6 muestra el resultado de recortar la parte del rostro de las imágenes mostradas en la Figura 7.5. Después de recortar todas las imágenes con expresiones pico de la base de datos, estas fueron normalizadas a un tamaño de 280x280 píxeles, por lo tanto las 6 imágenes mostradas en la Figura 7.6 poseen estas dimensiones.



Figura 7.6. Ejemplo de la sección del rostro recortada de la base de datos Cohn-Kanade.

7.1.2.1 Simulación de la Oclusión

Con el objetivo de atacar el problema de la oclusión parcial en sistemas de reconocimiento de rostros, imágenes que presentaran diferentes tipos de oclusión eran necesarias. Actualmente no existe una base de datos de expresiones faciales estándar que contenga diferente tipo, tamaño o posición de oclusión parcial, por lo tanto, cuatro diferentes tipos de oclusión parcial fueron simulados utilizando la base de datos Cohn-Kanade. Los cuatro tipos de oclusión parcial que se utilizaron en este trabajo son: oclusión de la mitad izquierda del rostro, oclusión de la mitad derecha del rostro, oclusión de los ojos-cejas y oclusión de la boca.

La principal motivación para aplicar estos cuatro tipos de oclusión viene de situaciones de la vida diaria. Por ejemplo, el uso de lentes oscuros, bufandas, cubre bocas, algunos tipos de corte de cabello y sombras en las imágenes. La Figura 7.7 muestra los cuatro diferentes tipos de oclusión aplicados a un sujeto de la base de datos que ejecuta las 6 expresiones básicas, de arriba hacia abajo se muestra la oclusión de la mitad izquierda del rostro, de la mitad derecha, de los ojos-cejas y de la boca. Estas imágenes fueron generadas superponiendo gráficamente una región negra en todas las imágenes con expresiones pico de base de datos Cohn-Kanade.

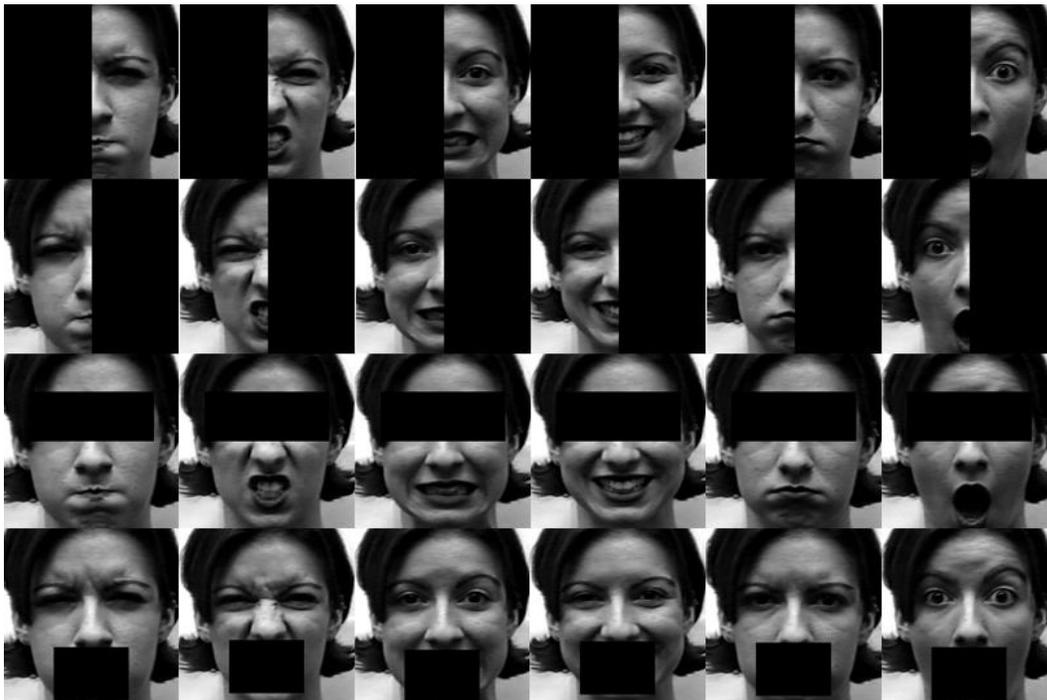


Figura 7.7. Ejemplo de los cuatro tipos de oclusión aplicados a la base de datos.

Es importante destacar que los tipos de oclusión presentados en este trabajo son más críticos que los ocurridos en la vida diaria. Por ejemplo, como se muestra en la Figura 7.7, en el tipo de oclusión de la mitad del rostro, el cual emula la oclusión creada por algunos tipos de corte de cabello y especialmente por sombras generadas gracias a una deficiente iluminación, la mitad del rostro fue completamente sustituida por una región negra lo cual no ocurre en las situaciones previamente mencionadas. Así mismo para emular el uso de lentes oscuros la parte de los ojos y las cejas fue completamente ocluida, siendo que en la realidad unos lentes no causan ese tamaño de oclusión. Finalmente la oclusión de la boca se podría comparar con el uso de algún tipo de cubre bocas. Por lo tanto debido a que los tipos de oclusión utilizados en este trabajo son más críticos que los de la vida diaria, se espera que si el sistema propuesto funciona bien con ellos lo hará aún mejor con los tipos de oclusión de la vida diaria.

7.1.2.2 Resolviendo la Oclusión de la Mitad del Rostro

Para los casos de oclusión de la mitad izquierda o derecha del rostro el algoritmo propuesto para el reconocimiento de expresiones fáciles no se puede aplicar directamente. Esto debido a que una parte vital del algoritmo es la segmentación del rostro (descrita en la sección 6.2) y para llevar a cabo este proceso se necesitan ambas partes del rostro. Ya que por ejemplo, no sería funcional segmentar la sección de la boca si la mitad de esta sección va estar completamente ocluida, y esta situación ocurriría para las cuatro regiones que se necesitan segmentar. Por lo tanto, en el presente trabajo, se generan “imágenes espejo” ocupando la mitad del rostro que no se encuentra ocluido, así se trabajaría con un rostro completo en contraste a usar solo la mitad del mismo.

Usando las imágenes espejo, es posible aplicar una correcta segmentación del rostro, como se menciona en la sección 6.2, debido a que prácticamente se está trabajando con un rostro sin oclusión. De esta forma se intenta resolver el problema de la oclusión de la mitad del rostro y de esta forma poder incrementar la tasa de reconocimiento.

El proceso para la obtención de las imágenes espejo consiste en recortar la sección de la imagen no ocluida para después obtener la imagen espejo de esta y finalmente concatenar estas dos mitades de rostro, de esta manera obtenemos una

imagen del rostro completo solucionando el problema de oclusión de la mitad del rostro.

La Figura 7.8 muestra dos imágenes espejo y el proceso de obtención de ellas, en la primera fila se muestra primero una imagen con la mitad izquierda del rostro ocluida, enseguida se muestra la parte del rostro sin oclusión junto con la imagen espejo de esta, y finalmente se muestra la imagen compuesta por ambas mitades. La segunda fila de la figura muestra el mismo proceso pero usando una imagen con la mitad derecha del rostro ocluida.

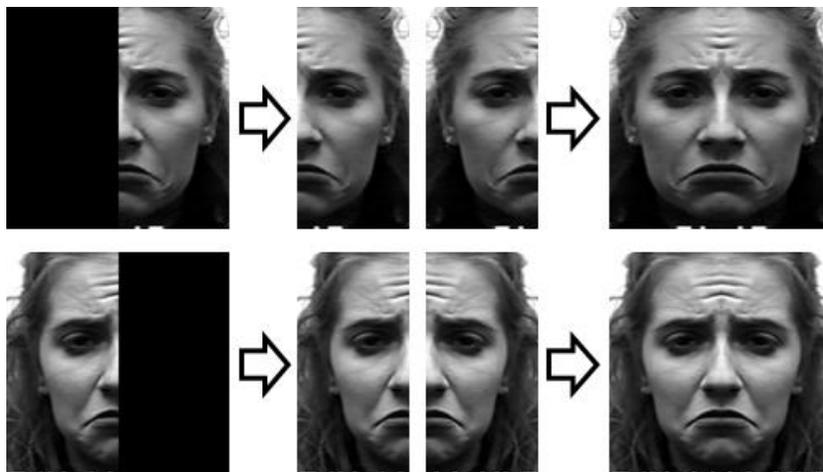


Figura 7.8. Proceso para la obtención de las imágenes espejo.

Este proceso se utilizó para las dos formas de oclusión de mitad del rostro (izquierda y derecha) y para todas las imágenes con expresiones pico a las que se les aplicaron estos tipos de oclusión.

7.1.2.3 Imágenes Utilizadas para las Pruebas

Para las pruebas realizadas con el sistema propuesto para el reconocimiento de expresiones faciales se utilizaron 300 imágenes con expresiones pico de la base de datos Cohn-Kanade. Las imágenes utilizadas cuentan con un tamaño de 280x280 píxeles y después de la segmentación del rostro detallada en la sección 6.2 el tamaño de las regiones de ojos-cejas, frente, boca y nariz fueron normalizadas a 200x80, 100x65, 140x80 y 175x50 píxeles respectivamente.

Debido a que no todos los sujetos de la base de datos realizaron las 6 expresiones básicas, el número de imágenes por cada expresión facial no es la misma. Por lo tanto, la Tabla 7.1 muestra el número de imágenes por cada una de las expresiones básicas.

Tabla 7.1. Número de imágenes por cada expresión facial.

Expresión	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
# imágenes	30	34	47	70	54	65

Las pruebas del sistema propuesto para reconocimiento de expresiones faciales se dividen en dos, sin oclusión y con oclusión. Para las pruebas sin oclusión solo se utilizó un set de imágenes (el set de 300 imágenes sin oclusión) y para las pruebas con oclusión se utilizaron cuatro sets, cada uno correspondiente a los cuatro tipos de oclusión antes mencionados.



Figura 7.9. Ejemplo de los 5 sets de imágenes utilizados para las pruebas.

En general se utilizaron 5 diferentes sets de pruebas para la evaluación del sistema propuesto. La Figura 7.9 muestra un sujeto de la base de datos realizando las 6 expresiones básicas para los 5 diferentes sets utilizados en las pruebas, de arriba hacia abajo los sets mostrados en esta figura se describen a continuación:

1. Sin oclusión.
2. Oclusión de la mitad izquierda del rostro. Donde se utilizan imágenes espejo de la mitad derecha del rostro.
3. Oclusión de la mitad derecha del rostro. Donde se utilizan imágenes espejo de la mitad izquierda del rostro.
4. Oclusión de ojos-cejas.
5. Oclusión de la boca.

7.2 Numero Óptimo de Imágenes de Entrenamiento

Para el caso de los sistemas propuestos de reconocimiento de rostros se realizó una prueba usando diferente número de imágenes de entrenamiento. La Figura 7.10 muestra los resultados de esta prueba, en la que se utilizaron de 1 a 14 imágenes de entrenamiento para el método convencional de Fases Propias y para los sistemas propuestos de Fases Propias con Sub-bloque de 2x2, LBPC(8,1) y WBP.

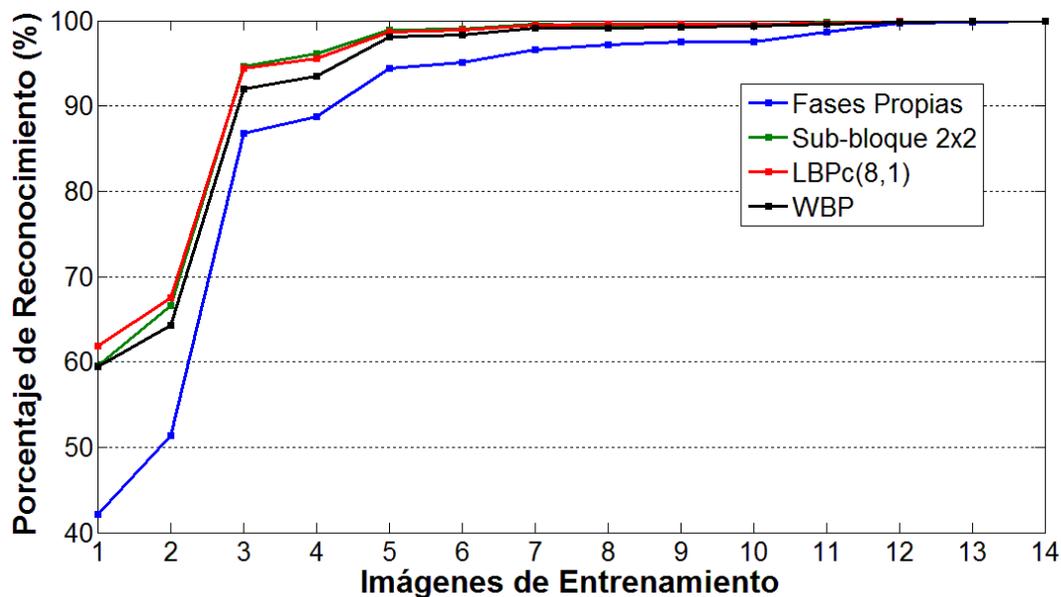


Figura 7.10. Resultados usando diferente número de imágenes de entrenamiento.

Es importante mencionar que para todas las pruebas realizadas con los algoritmos propuestos para reconocimiento de rostros se utilizaron imágenes sin oclusión, es decir solamente imágenes del set A (Figura 7.3). Es por eso que se utilizaron hasta 14 imágenes de entrenamiento porque dejando de lado las modificaciones de iluminación solo son 14 diferentes imágenes por persona.

Se puede observar en la Figura 7.10 que los sistemas analizados en esta prueba presentan una mejora en la tasa de reconocimiento cuando se utilizan más de 2 imágenes de entrenamiento, más del 90% de reconocimiento para los sistemas propuestos y cerca del 85% para el método convencional de fases propias. Otro factor que se puede observar es que con 6 o 7 imágenes de entrenamiento los sistemas propuestos alcanzan un muy buen porcentaje de reconocimiento, mientras que el método convencional necesita alrededor de 12 imágenes para alcanzar un porcentaje de reconocimiento similar. En general con 7 imágenes de entrenamiento los sistemas propuestos alcanzan un rendimiento estable. Por lo tanto 7 es el número óptimo de imágenes de entrenamiento y el utilizado en todas las pruebas realizadas por los métodos de reconocimiento de rostros. La Figura 7.11 muestra las 7 imágenes que se utilizaron para el entrenamiento de uno de los 120 sujetos de la base de datos.



Figura 7.11. Ejemplo de las imágenes de entrenamiento utilizadas para una persona.

7.3 Resultados de Identificación

En esta sección se evalúan individualmente los resultados de identificación de los sistemas propuestos para el reconocimiento de rostros. Es decir, del sistema propuesto usando fases propias y del sistema propuesto usando LBP y WBP. Finalmente se presenta una comparación de los mejores resultados de identificación aquí presentados.

Los resultados de identificación son evaluados usando dos métodos: la tasa media de reconocimiento o porcentaje de reconocimiento general y el ranking. El porcentaje de reconocimiento general es el promedio de todos los aciertos obtenidos por el sistema, en otras palabras es el promedio de las imágenes que pudo reconocer correctamente entre todas las imágenes de prueba. Este valor se encuentra entre un rango de 0 al 100%. Por otro lado el ranking N es la probabilidad de que la persona ligada a la imagen bajo análisis se encuentre dentro un grupo de N personas las cuales tengan los valores más altos de coincidencia (valores numéricos obtenidos por la SVM), no importando que la imagen sea asociada a la persona incorrecta, lo importante es que la persona correcta este dentro de este grupo. N puede tener valores entre 1 y n , donde n es el número total de personas en la base de datos. De manera que el ranking 10 es la probabilidad de que la persona de la imagen analizada se encuentre dentro del grupo de las 10 personas con mayor valor de coincidencia, es decir dentro de las 10 personas que el sistema eligió como las más parecidas a la imagen. En contraparte, el resultado de ranking 1 es la probabilidad de que la persona de la imagen sea la persona con mayor valor de coincidencia, en este caso el resultado es idéntico al porcentaje de reconocimiento general del sistema.

Las gráficas de ranking muestran el porcentaje de reconocimiento por cada valor del ranking. En el eje de las abscisas se muestra el valor del ranking y en el eje de las ordenadas el porcentaje de reconocimiento.

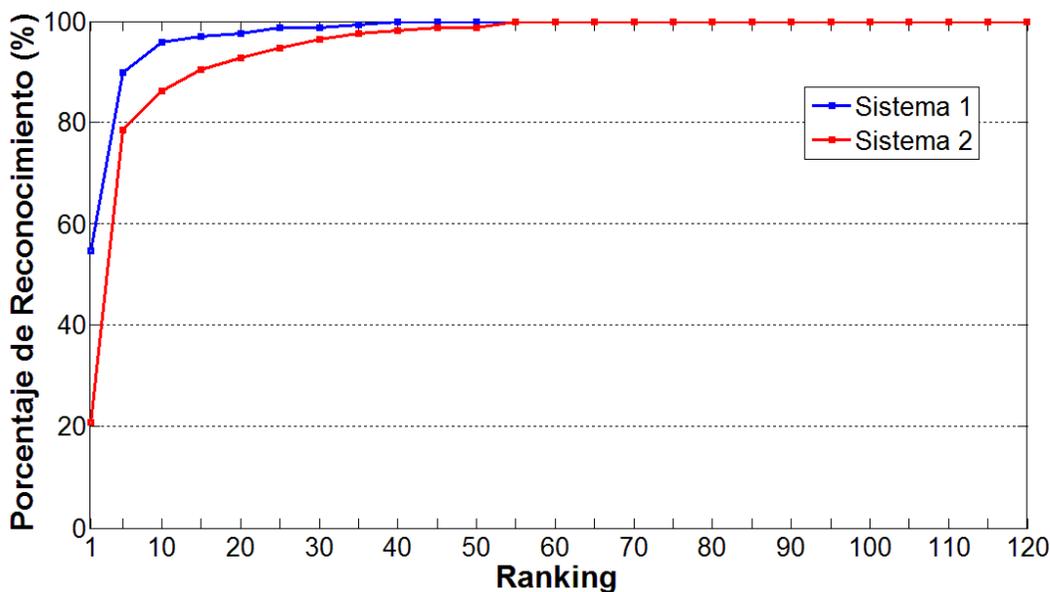


Figura 7.12. Ejemplo de una gráfica de ranking.

La Figura 7.12 muestra el ejemplo de una gráfica de ranking, en la cual se presentan los resultados de dos sistemas. En ella se puede observar que el porcentaje de reconocimiento general es de aproximadamente 55% y 20% para el sistema 1 y 2 respectivamente, resultados de ranking 1. En cambio, cuando se usa el ranking 5 este porcentaje se incrementa notablemente, cerca de 90% para el sistema 1 y 80% para el 2. Esto se debe a que las personas ligadas a las imágenes de pruebas se encontraban dentro del top 5 de personas con los valores más altos de coincidencia. Es por eso que los resultados se acercan al 100% conforme el ranking se acerca a 120, ya que la base de datos utilizada en este trabajo solo cuenta 120 personas.

El porcentaje de reconocimiento nos sirve para evaluar el rendimiento general del sistema, de tal manera que entre más cerca se encuentre del 100% mejor rendimiento presenta. Por otra parte las gráficas de ranking nos sirven para analizar en qué valor de ranking el sistema presenta un buen rendimiento, este tipo de evaluación es utilizado en situaciones donde no es estrictamente necesario encontrar la identidad del sujeto sino que con un grupo de posibles sospechosos es suficiente. A continuación se presentan los resultados de los algoritmos propuestos para el reconocimiento de rostros usando estos dos métodos de evaluación

7.3.1 Sistema Propuesto usando Fases Propias

El desempeño del sistema fue evaluado usando 4 variantes diferentes: sin pre-procesamiento, con pre-procesamiento usando la normalización del valor del pixel, ecualización del histograma (HE) y CLAHE. Adicionalmente se utilizaron 5 diferentes tamaños de sub-bloque: imagen completa, donde el espectro de fase se obtiene de toda la imagen como fue propuesto en el algoritmo convencional de fases propias [48]; y el método basado en sub-bloques con tamaños de 12x12, 6x6, 4x4 y 2x2 pixeles, donde el espectro de fase es obtenido independientemente por cada sub-bloque.

La Tabla 7.2 muestra los resultados de identificación (porcentaje de reconocimiento y derivación estándar) de la prueba donde solo se utilizaron imágenes del set A (Figura 7.3). Los resultados muestran que para las 4 variantes del sistema, usando un tamaño de sub-bloque de 2x2 pixeles se obtiene el porcentaje de reconocimiento más alto y la derivación estándar más baja, alcanzando alrededor de 95.5% de reconocimiento, siendo CLAHE 2x2 el mejor resultado con 99.62%.

Tabla 7.2. Resultados de las 4 variaciones y los 5 tamaños de sub-bloques de las pruebas usando Fases Propias realizadas con el set A de la base de datos (sin oclusión).

Tamaño de sub-bloque	Sin pre-proc.		Normalización		HE		CLAHE	
	%	σ	%	σ	%	σ	%	σ
Imagen completa	96.60	4.81	96.60	4.81	96.74	4.90	96.50	4.71
12 x 12	97.88	3.65	97.89	3.72	97.44	3.87	97.42	4.47
6 x 6	99.00	2.19	98.98	2.32	98.56	3.12	99.01	2.27
4 x 4	99.15	2.11	99.10	2.21	98.90	2.51	99.06	2.42
2 x 2 (óptimo)	99.56	1.45	99.57	1.46	99.42	1.55	99.62	1.40

Así mismo, de la Tabla 7.2 podemos observar que en todos los casos, el uso del método basado en sub-bloques mejora el desempeño del método convencional de fases propias (usando la imagen completa). Particularmente el uso del método basado en sub-bloques que utiliza el tamaño de bloque óptimo proporciona una mejora de aproximadamente el 3% con respecto al método convencional.

Las Figura 7.13 - 7.16 muestran las gráficas de ranking para las 4 variaciones del sistema propuesto, usando los 5 tamaños de sub-bloque. En estas gráficas se observa que el método con tamaño de bloque óptimo alcanza un reconocimiento del 100% con el ranking 10 mientras que el convencional lo hace después del ranking 70.

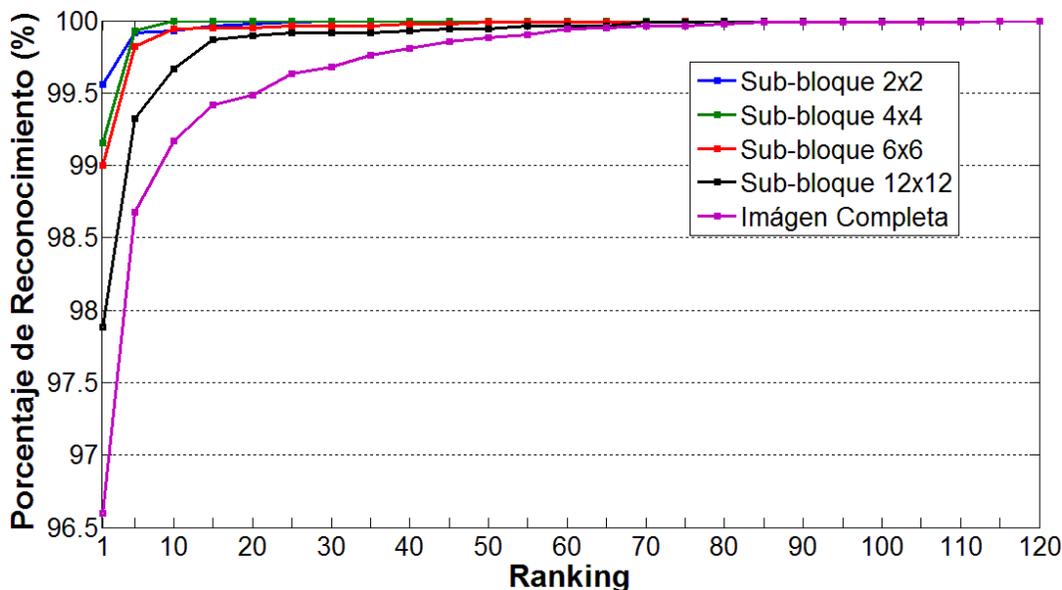


Figura 7.13. Gráfica de ranking del sistema propuesto sin pre-procesamiento usando las imágenes del set A (sin oclusión).

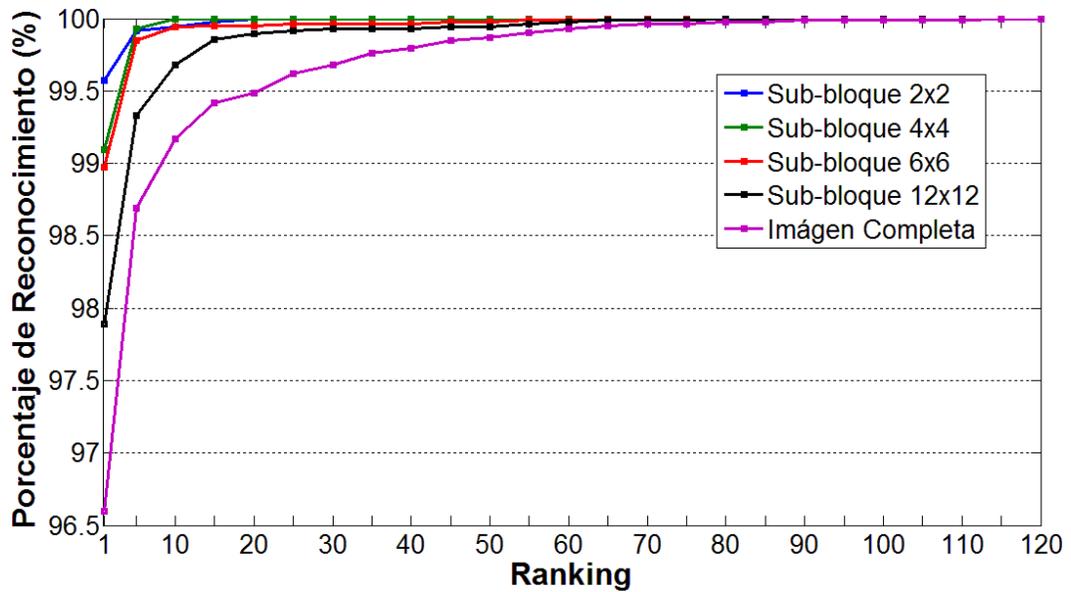


Figura 7.14. Gráfica de ranking del sistema propuesto usando normalización del valor del pixel y las imágenes del set A (sin oclusión).

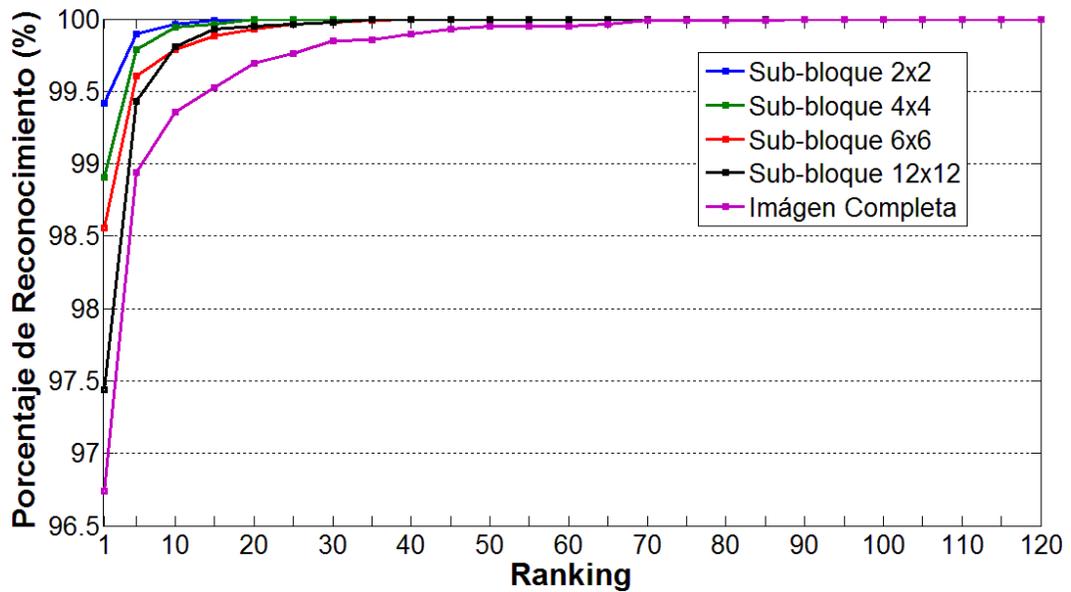


Figura 7.15. Gráfica de ranking del sistema propuesto usando ecualización del histograma (HE) y las imágenes del set A (sin oclusión).

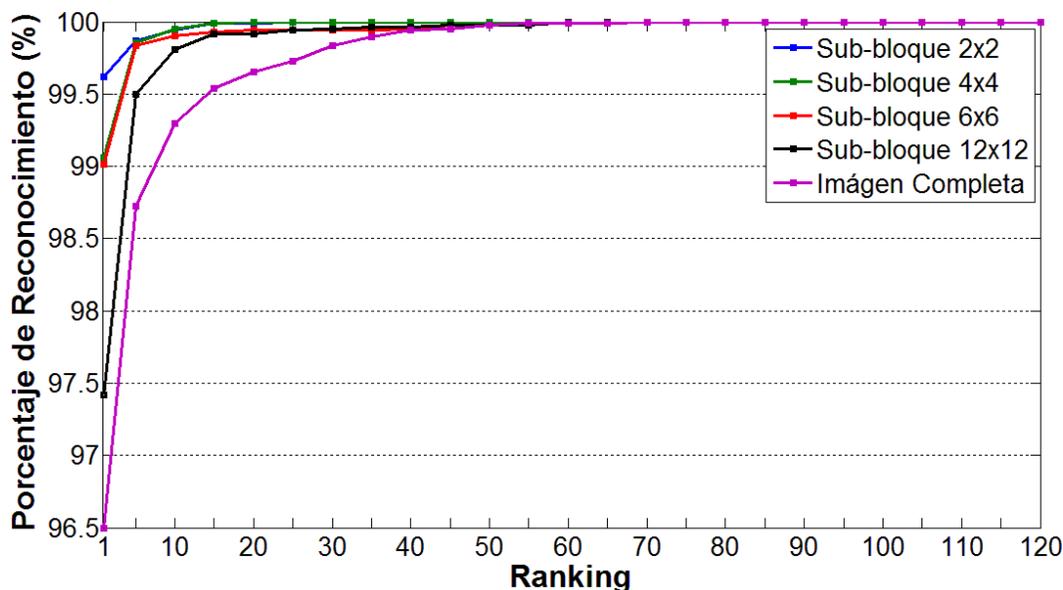


Figura 7.16. Gráfica de ranking del sistema propuesto usando CLAHE y las imágenes del set A (sin oclusión).

Las pruebas realizadas con imágenes en presencia de oclusión parcial se llevaron a cabo con las 4 variaciones del sistema y los 5 tamaños de sub-bloque previamente descritos. Es importante mencionar que para esta prueba se utilizaron las imágenes del set B de la base de datos, pero el entrenamiento se realizó usando las mismas 7 imágenes sin oclusión usadas en la prueba anterior. Por lo tanto esta prueba se asemeja más a una aplicación real, ya que las imágenes de entrenamiento se usan sin oclusión, mientras que todas las imágenes de prueba presentan oclusión parcial como se muestra en la Figura 7.4. La Tabla 7.3 muestra los resultados de identificación de esta prueba (porcentaje de reconocimiento y derivación estándar), utilizando solo imágenes del set B.

Tabla 7.3. Resultados de las 4 variaciones y los 5 tamaños de sub-bloques de las pruebas usando Fases Propias realizadas con el set B de la base de datos (con oclusión).

Tamaño de sub-bloque	Sin pre-proc.		Normalización		HE		CLAHE	
	%	σ	%	σ	%	σ	%	σ
Imagen completa	85.44	19.49	85.44	19.44	83.17	19.43	84.81	19.87
12 x 12	87.83	17.22	88.00	17.12	83.31	18.56	86.58	16.52
6 x 6	93.08	15.22	93.03	15.19	90.89	15.37	92.50	14.59
4 x 4	95.00	12.05	94.92	11.93	94.00	11.71	95.17	11.80
2 x 2 (óptimo)	97.19	9.56	97.33	9.21	96.56	10.70	97.03	9.84

Los resultados de la Tabla 7.3 muestran que usando imágenes con oclusión parcial el método basado en sub-bloques funciona mejor que el algoritmo original de fases propias. Sin embargo, esta vez el incremento en el porcentaje de reconocimiento es mayor, ya que el método que utiliza el tamaño de bloque óptimo proporciona un reconocimiento de identificación alrededor de 97%, mejorando significativamente al método convencional, con cerca del 12% de mejoría.

Las 4 variaciones del sistema presentan aproximadamente el mismo rendimiento cuando se utiliza el tamaño de bloque óptimo, aun así el método de Normalización 2x2 alcanza el mejor resultado con un 97.33% de reconocimiento. Es importante mencionar que este resultado es apenas 2% menor que el mejor resultado obtenido en las pruebas sin oclusión, una diferencia mucho menor al 11% de decremento que presenta el método convencional.

Las Figura 7.17 - 7.20 muestran las gráficas de ranking para las variaciones del sistema en presencia de oclusión parcial. En estas gráficas es aún más evidente el mejor rendimiento del algoritmo con tamaño de bloque óptimo en contra del algoritmo convencional. En esta prueba el método convencional necesita más del ranking 110 para alcanzar el 100% de reconocimiento, mientras que el método con bloque óptimo lo logra usando el ranking 25 aproximadamente.

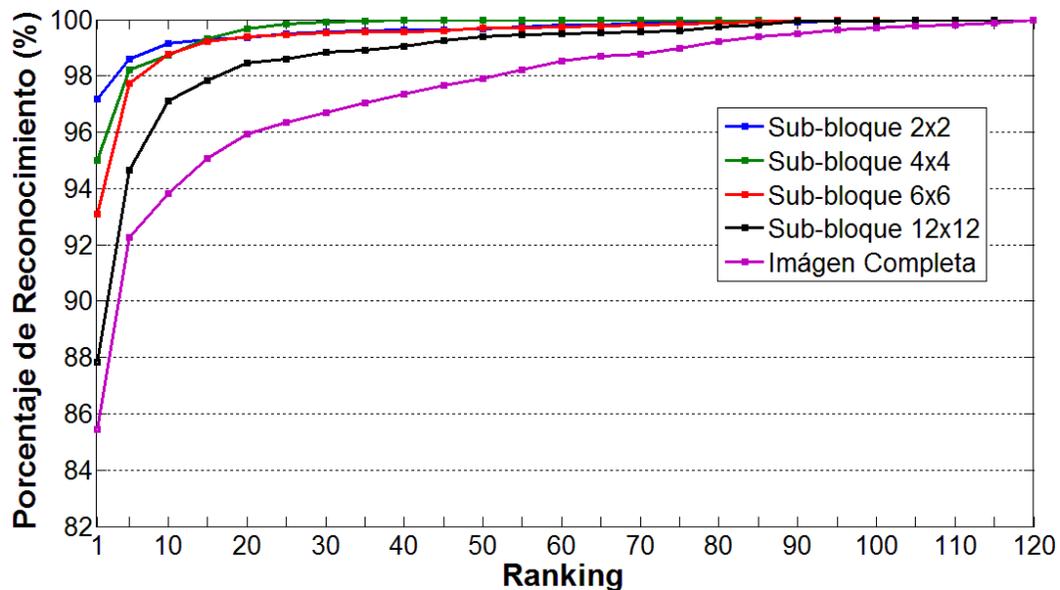


Figura 7.17. Gráfica de ranking del sistema propuesto sin pre-procesamiento usando las imágenes del set B (con oclusión).

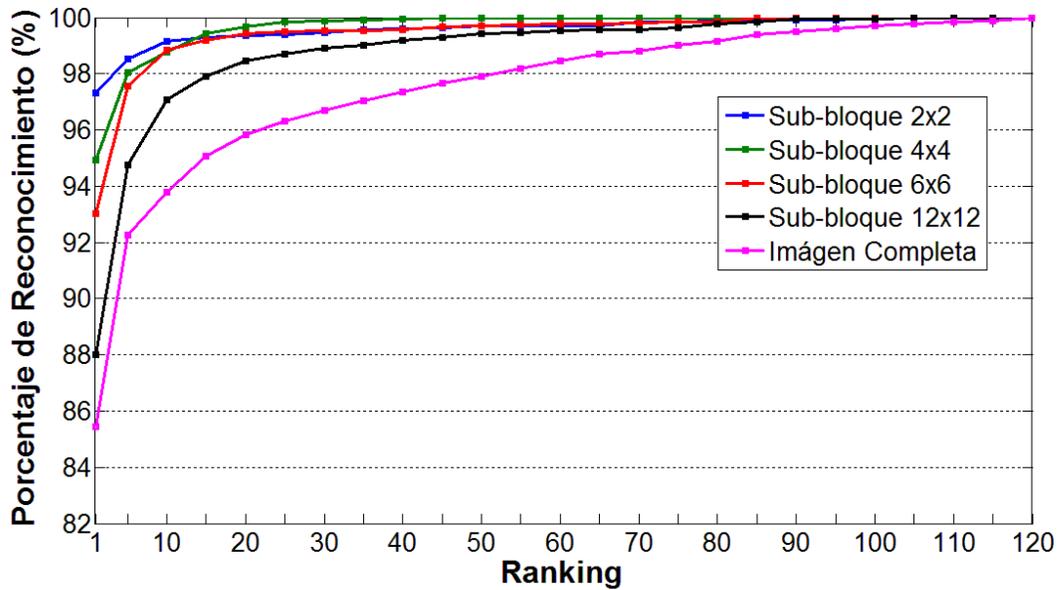


Figura 7.18. Gráfica de ranking del sistema propuesto usando normalización del valor del pixel y las imágenes del set B (con oclusión).

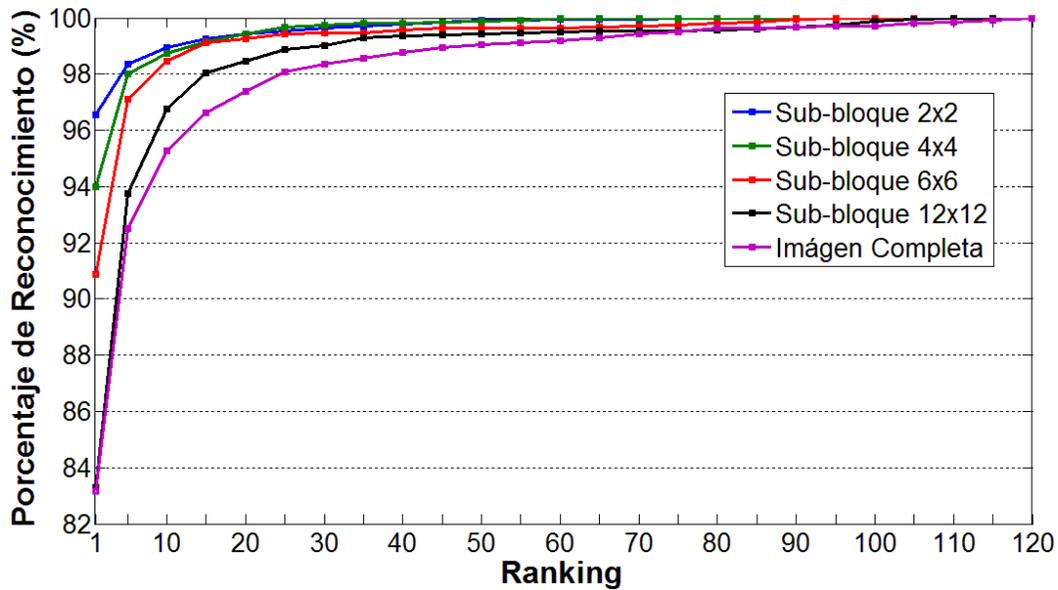


Figura 7.19. Gráfica de ranking del sistema propuesto usando ecualización del histograma (HE) y las imágenes del set B (con oclusión).

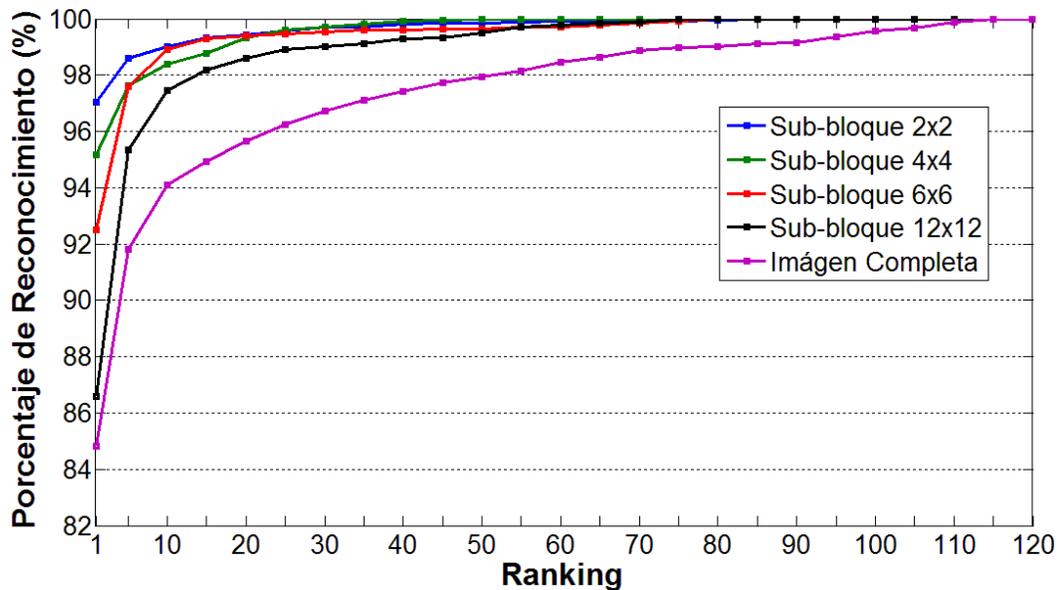


Figura 7.20. Gráfica de ranking del sistema propuesto usando CLAHE y las imágenes del set B (con oclusión).

7.3.2 Sistema Propuesto usando LBP y WBP

El sistema propuesto fue evaluado con sus 4 variaciones: LBPh, donde se utilizan los histogramas de la imagen LBP como propone el método convencional [68]; LBPC, que utiliza el PCA directamente a los coeficientes de la imagen LBP; WBP, donde se utiliza la idea original de LBPC aplicado a ventanas, en vez de píxeles individuales, y directamente a la imagen original, sin redimensión; y DI-WBP, que utiliza el método WBP con imágenes redimensionadas, excluyendo el uso del PCA. Adicionalmente, para las variaciones de LBPh y LBPC se utilizaron 3 tipos de raído: (8,1), (8,2) y (8,3). Esto con la finalidad de mostrar el desempeño del algoritmo propuesto discriminando los parámetros del LBP. Así mismo, el sistema fue evaluado con imágenes en presencia de oclusión parcial, utilizando para las pruebas imágenes tanto del set A como del set B de la base de datos.

La Tabla 7.4 muestra los resultados de identificación (porcentaje de reconocimiento y derivación estándar) para todas las variaciones del sistema, usando las imágenes del set A y del set B. Es importante mencionar que para todas las pruebas solo imágenes sin oclusión fueron utilizadas en el entrenamiento, utilizando nuevamente las 7 imágenes de entrenamiento por persona descritas en la sección 7.2.

Tabla 7.4. Resultados de las pruebas de identificación usando las variantes del sistema basado en LBP realizadas con ambos sets de la base de datos.

Variante del sistema	Set A (sin oclusión)		Set B (con oclusión)	
	%	σ	%	σ
LBPh(8,1)	72.93	9.78	53.56	17.90
LBPh(8,2)	82.94	8.08	68.08	17.82
LBPh(8,3)	86.83	8.06	73.17	17.46
LBPh(8,1)	99.51	1.65	95.69	12.83
LBPh(8,2)	99.49	1.63	95.86	12.13
LBPh(8,3)	99.55	1.56	96.22	11.19
WBP	99.10	2.39	95.39	12.38
DI-WBP	98.42	2.91	90.69	15.59

En la Tabla 7.4 se puede observar que para la prueba utilizando imágenes del set A (sin oclusión) los métodos propuestos logran un mayor porcentaje de reconocimiento que el método convencional LBPh, alcanzando alrededor del 99% de reconocimiento. Superando en aproximadamente 13% al método convencional que usa radio 3 y en más de 25% al que usa radio 1. También es posible notar que las variaciones de LBPh y WBP proporcionan resultados muy similares (mayor al 99% de reconocimiento), por otro lado el método DI-WBP presenta un resultado de aproximadamente 1% menor, debido a que en esta situación la cantidad de coeficientes de LBP es 81 veces más pequeña. Es importante mencionar que el mejor resultado de esta prueba es el proporcionado por el método LBPh(8,3) con 99.55% de reconocimiento.

En cuanto a las pruebas usando imágenes con presencia de oclusión parcial (set B) la Tabla 7.4 muestra que el rendimiento de los métodos propuestos es mucho mejor que los convencionales, superando en más del 20% al método convencional de LBPh(8,3) y en más del 40% al de LBPh(8,1), siendo esta una mejora bastante significativa. Una vez más los resultados presentados por los métodos de LBPh y WBP son muy parecidos, pero esta vez el porcentaje de reconocimiento del método DI-WBP es menor que el de ellos en aproximadamente 5%. Nuevamente el método de LBPh(8,3) proporciona el mejor porcentaje de reconocimiento alcanzando 96.22%, siendo usadas imágenes con oclusión para las pruebas. Este porcentaje es apenas 3% menor al alcanzado por las pruebas sin oclusión, mostrando que el algoritmo presenta un buen rendimiento aun en presencia del problema de la oclusión parcial.

La Figura 7.21 muestra la gráfica de ranking para los resultados obtenidos con el método convencional LBPh usando imágenes sin oclusión (set A). En ella se aprecia que el método LBPh(8,3) muestra un mejor rendimiento en todos los posibles niveles del ranking y alcanza el 100% aproximadamente en el ranking 55.

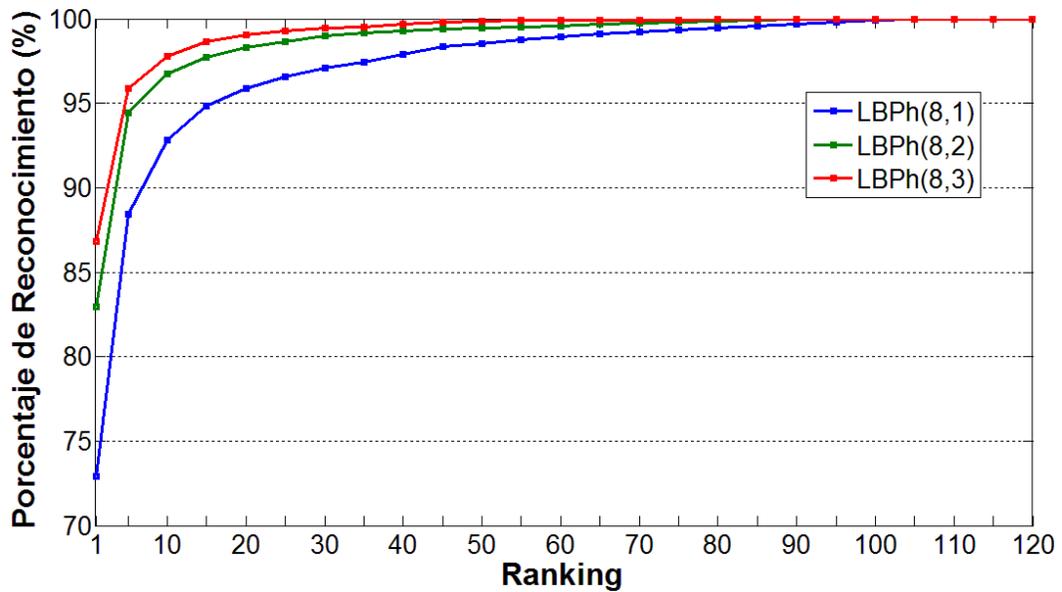


Figura 7.21. Gráfica de ranking con las variaciones del método LBPh usando las imágenes del set A (sin oclusión).

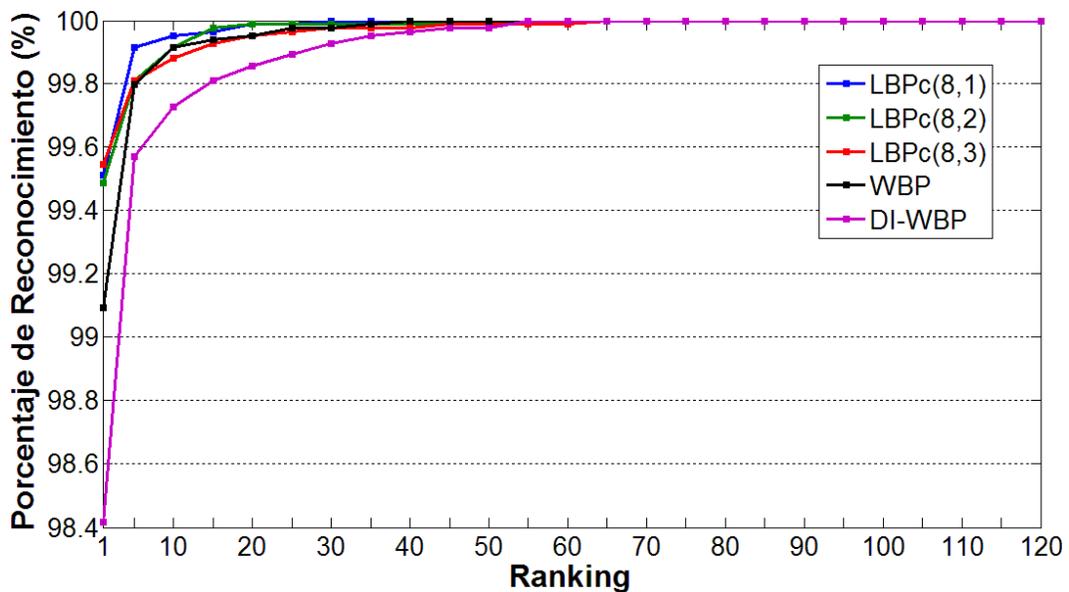


Figura 7.22. Gráfica de ranking con las variaciones de los métodos LBPh, WBP y DI-WBP usando las imágenes del set A (sin oclusión).

Por otro lado la Figura 7.22 muestra la gráfica de ranking con los resultados de los métodos propuestos usando imágenes sin oclusión. En esta grafica se puede observar que la relación de los resultados del ranking no es igual en todos los niveles. Por ejemplo, en el ranking 1 el mejor porcentaje de reconocimiento es el proporcionado por el método LBPC(8,3), mientras que en el ranking 5 por el método LBPC(8,1) y en el ranking 15 por LBPC(8,2). También se puede observar que el método de LBPC(8,1) logra el 100% de reconocimiento en el ranking 30, seguido por los métodos WBP y LBPC(8,2) que lo logran en el nivel 40 y 45 respectivamente. Es importante mencionar que el método de DI-WBP alcanza el 100% de reconocimiento en el ranking 55 aproximadamente 10 niveles antes que el método LBPC(8,3) el cual lo logra en el ranking 65. Esto demuestra que, al menos en las pruebas de ranking, los métodos de WBP y DI-WBP presentan un mejor rendimiento que la mayoría de los métodos de LBPC.

La gráfica de ranking con los resultados de las variaciones del método convencional LBPh usando imágenes con oclusión (set B) se muestra en la Figura 7.23. De la misma manera que en la prueba sin oclusión, el método convencional LBPh(8,3) presenta un mejor rendimiento en todos los niveles del ranking y alcanza el 100% aproximadamente en el ranking 75.

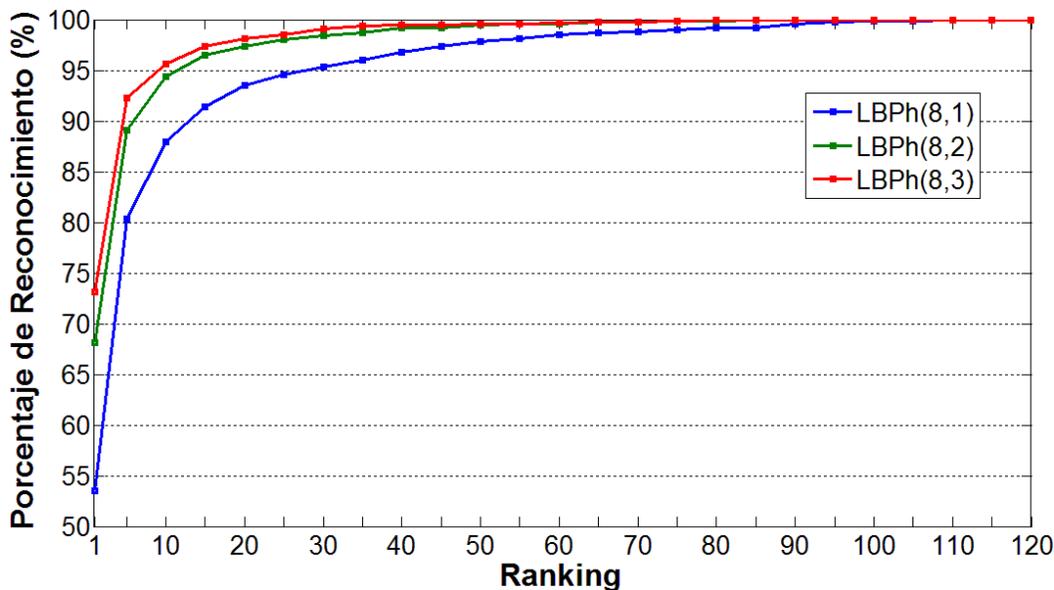


Figura 7.23. Gráfica de ranking con las variaciones del método LBPh usando las imágenes del set B (con oclusión).

En la Figura 7.24 se muestra la gráfica de ranking de los métodos propuestos usando imágenes con oclusión. En esta gráfica se puede observar que los métodos de WBP y DI-WBP presentan un mejor rendimiento que los de LBPC, puesto que a partir del ranking 20 estos métodos logran un mejor porcentaje de reconocimiento, siendo DI-WBP el cual alcanza el 100% con el menor ranking.

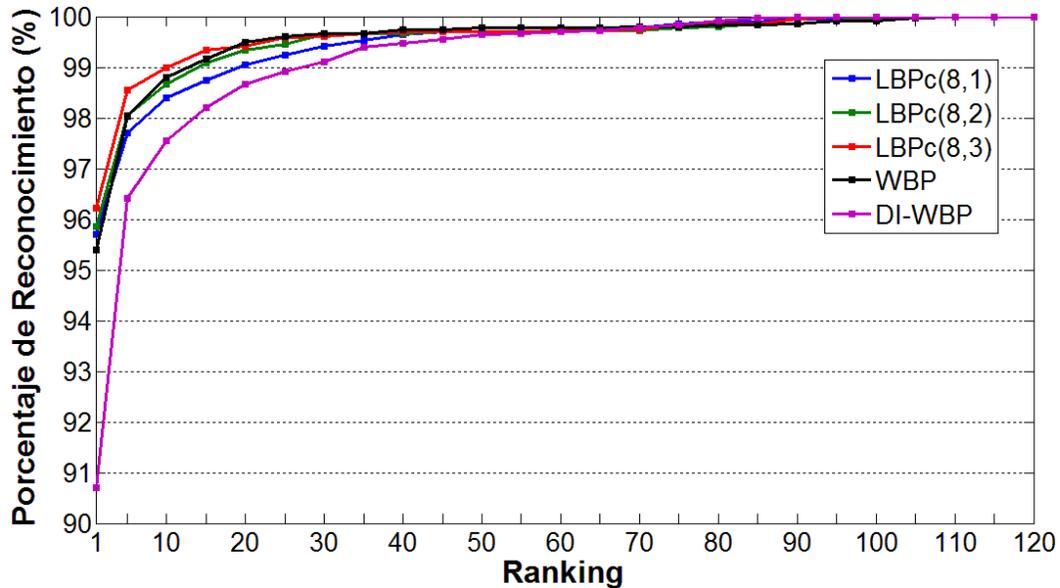


Figura 7.24. Gráfica de ranking con las variaciones de los métodos LBPC, WBP y DI-WBP usando las imágenes del set B (con oclusión).

7.3.3 Comparación de los 5 Mejores Resultados

En esta sección se presenta una comparación de los 5 mejores resultados de identificación obtenidos en este trabajo. La Tabla 7.5 muestra estos resultados, que fueron obtenidos por los métodos: CLAHE(2x2), Normalización del valor del pixel(2x2), Fases Propias sin pre-procesamiento(2x2), LBPC(8,3) y WBP.

Tabla 7.5. Mejores 5 resultados de identificación utilizando ambos sets de la base de datos.

Método	Set A (sin oclusión)		Set B (con oclusión)	
	%	σ	%	σ
CLAHE(2x2)	99.62	1.40	97.03	9.84
Normalización(2x2)	99.57	1.46	97.33	9.21
Sin pre-proc.(2x2)	99.56	1.45	97.19	9.56
LBPC(8,3)	99.55	1.56	96.22	11.19
WBP	99.10	2.39	95.39	12.38

En la Tabla 7.5 se puede observar que para las pruebas sin oclusión los 5 resultados superan el 99%, siendo CLAHE(2x2) el que presenta el mejor porcentaje de reconocimiento con 99.62% y la más baja derivación estándar con 1.4, siendo este el mejor resultado de identificación obtenido en este trabajo. Además, es posible observar que los métodos propuestos usando Fases Propias presentan un mejor rendimiento que los basados en LBP.

Para las pruebas con oclusión, la Tabla 7.5 muestra que los 5 resultados superan el 95% y el mejor es el obtenido por Normalización(2x2), con 97.33% de reconocimiento y 9.21 de derivación estándar. Cabe mencionar que los métodos de Fases Propias que utilizan el tamaño de bloque óptimo (2x2 pixeles) son los que, en este trabajo presentan un mejor desempeño, ya que con la presencia de oclusión en las imágenes de prueba solo reducen su porcentaje de reconocimiento un 2% aproximadamente, esto se puede ver claramente en la Tabla 7.5.

Las Figura 7.25 - 7.26 muestran las gráficas de ranking de los 5 mejores resultados usando imágenes del set A y del set B respectivamente. En ellas se puede observar que el método CLAHE(2x2) es el que presenta un mejor rendimiento. Asimismo, WBP es el método que aumenta más su porcentaje de reconocimiento conforme el valor de ranking se incrementa.

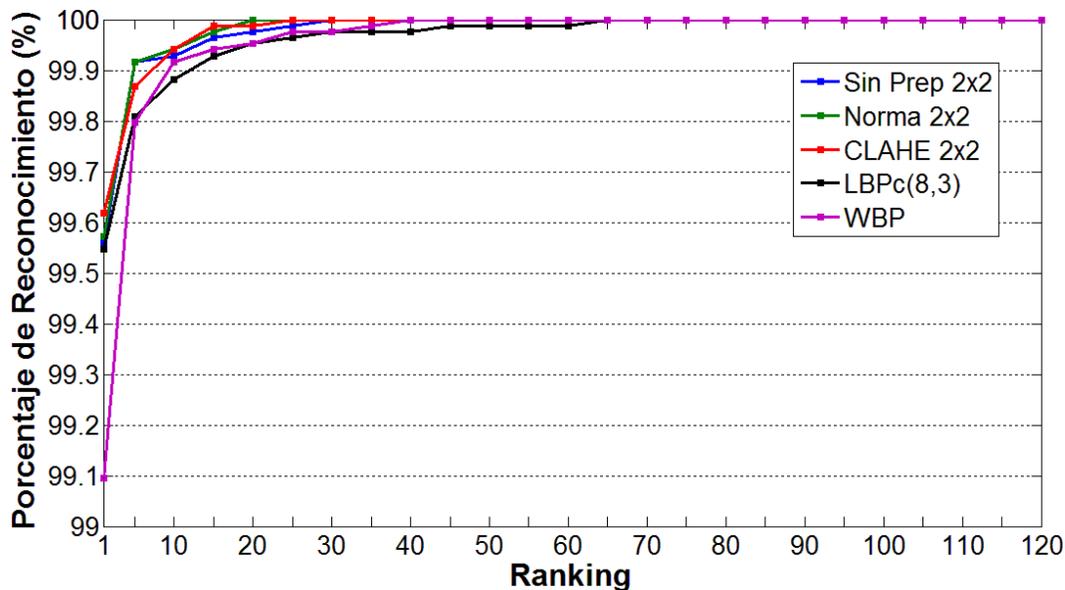


Figura 7.25. Gráfica de ranking con los 5 mejores resultados de identificación usando las imágenes del set A (sin oclusión).

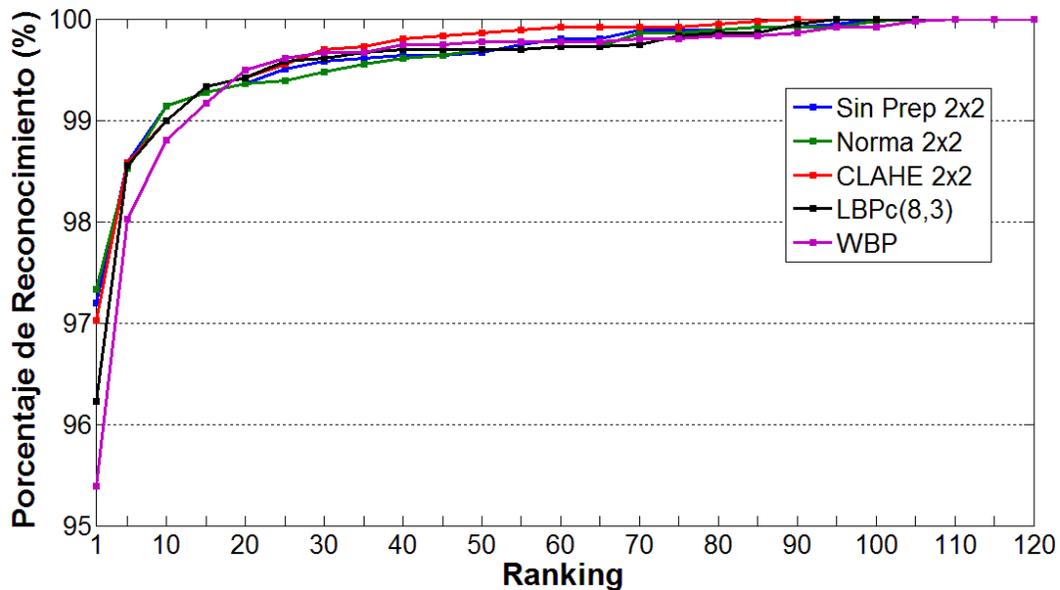


Figura 7.26. Gráfica de ranking con los 5 mejores resultados de identificación usando las imágenes del set B (con oclusión).

7.4 Resultados de Verificación

Esta sección presenta la evaluación de los resultados de verificación tanto del sistema propuesto usando Fases Propias como del sistema propuesto usando LBP y WBP. Asimismo, al final de esta sección se presenta una comparación de los mejores resultados de verificación obtenidos en este trabajo.

Primeramente es necesario describir la forma de evaluación para este tipo de resultados. A diferencia de los resultados de identificación que toman en cuenta solamente el porcentaje de acierto, para los resultados de verificación el punto más importante a tomar en cuenta es el porcentaje de error.

Para el caso de la verificación el porcentaje de error se divide en dos: el porcentaje de falsa aceptación y el de falso rechazo. El porcentaje de falsa aceptación es el promedio de veces en las que se produce una falsa aceptación, lo que ocurre cuando un individuo dice ser la persona que no es y esto es erróneamente aceptado por el sistema. El porcentaje de falso rechazo es el promedio de veces en las que se produce un falso rechazo, esto ocurre cuando un individuo proporciona su identidad y el sistema erróneamente rechaza esta declaración.

Otra diferencia entre las pruebas de identificación y verificación es el uso de un umbral para el sistema. En las pruebas de identificación no es estrictamente necesario el uso de un umbral, porque el sistema simplemente toma la decisión en base al valor de coincidencia más alto, es decir el sistema elige a la persona de la base de datos más parecida a la imagen analizada. Sin embargo, para las pruebas de verificación el umbral es esencial, por ejemplo, si el sujeto en cuestión proporciona su verdadera identidad y el valor de coincidencia de la imagen analizada es menor al del umbral se produciría un falso rechazo, por otro lado si el sujeto en cuestión dice ser una persona que no es y el valor de coincidencia de la imagen analizada es mayor al del umbral se produciría una falsa aceptación. Es importante mencionar que estos porcentajes de error están estrechamente ligados el uno con el otro, es decir, dependiendo del umbral elegido el porcentaje de falsa aceptación disminuirá pero el porcentaje de falso rechazo aumentará y viceversa. Por lo tanto los porcentajes de falsa aceptación y falso rechazo dependen completamente del umbral del sistema,

Por consiguiente, la evaluación de resultados de verificación se basa en presentar el porcentaje de falsa aceptación y falso rechazo, además del umbral utilizado para la obtención de estos porcentajes. Es por eso que en este trabajo se muestran los resultados de dos formas diferentes. La primera es presentando el punto de equilibrio de los porcentajes de error junto con el valor del umbral donde esto ocurre, en otras palabras consiste en mostrar el valor del umbral y el porcentaje de error donde la falsa aceptación y el falso rechazo tienen el mismo valor. La segunda, por otro lado, consiste en presentar las gráficas de verificación, que muestran el porcentaje de falsa aceptación y el falso rechazo, donde se puede observar el comportamiento de ambos porcentajes de error en un rango con diferentes umbrales. La Figura 7.27 muestra el ejemplo de una gráfica de verificación.

En la Figura 7.27 se puede observar que el falso rechazo se incrementa cuando el umbral tiende a 1 y la falsa aceptación cuando el umbral tiende a 0. Por lo tanto, la parte más importante de la gráfica es la mostrada en el recuadro de acercamiento, donde se puede observar el punto en el que se intersectan la falsa aceptación y el falso rechazo, el porcentaje de error y el umbral de este punto se muestran en la Tabla 7.6

Tabla 7.6. Punto de equilibrio del ejemplo de la Figura 7.27.

Método	% error	umbral
Ejemplo de la Fig. 7.27	0.42	0.564

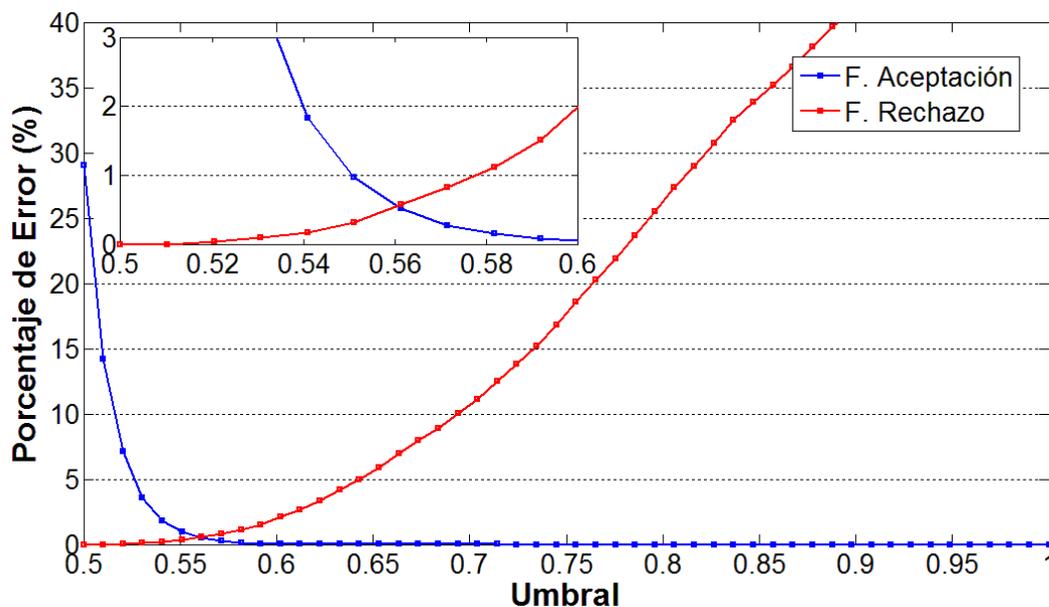


Figura 7.27. Ejemplo de una gráfica de verificación.

Los datos presentados en la Tabla 7.6 representan el valor del umbral (0.564) en el que la falsa aceptación y el falso rechazo logran el mismo valor (0.42%), es decir, el punto de equilibrio. Pero si se requiere específicamente un porcentaje de error más bajo, simplemente se debe elegir otro umbral de los presentados en la Figura 7.27. Por ejemplo si se necesita el porcentaje de falsa aceptación lo más bajo posible, se podría utilizar el umbral de 0.59, ya que en este punto la falsa aceptación es de aproximadamente 0%, pero hay que tomar en cuenta que el falso rechazo se incrementaría a 1.5%. Es por eso que, la elección del umbral óptimo para el sistema depende completamente de la aplicación del mismo.

7.4.1 Sistema Propuesto usando Fases Propias

En esta sección se presentan los resultados de verificación obtenidos por las 4 variaciones del sistema usando Fases Propias con los 5 diferentes tamaños de sub-bloque, descritos en la sección 7.3.1. Vale la pena mencionar que estos resultados se obtuvieron usando únicamente imágenes sin oclusión (set A), ya que en la aplicación los sistemas en modo de verificación difícilmente presentan el problema de oclusión parcial. La Tabla 7.7 presenta los resultados descritos previamente, esta muestra el porcentaje de error cuando la falsa aceptación y el falso rechazo tienen el mismo valor, además del umbral donde esto ocurre.

Tabla 7.7. Resultados de verificación para todas las variaciones del sistema usando Fases Propias, porcentaje de error (F. aceptación = F. rechazo) y umbral.

Tamaño de sub-bloque	Sin pre-proc.		Normalización		HE		CLAHE	
	% error	umbral	% error	umbral	% error	umbral	% error	umbral
Imagen completa	2.12	0.516	2.06	0.517	2.58	0.516	2.33	0.515
12 x 12	1.02	0.537	1.04	0.537	1.17	0.535	1.02	0.534
6 x 6	0.45	0.561	0.47	0.560	0.69	0.553	0.51	0.559
4 x 4	0.41	0.564	0.40	0.564	0.55	0.560	0.42	0.564
2 x 2 (óptimo)	0.23	0.591	0.23	0.592	0.31	0.586	0.23	0.593

En la Tabla 7.7 se puede observar que los valores de error más bajos son los obtenidos cuando se usa el tamaño de bloque óptimo, específicamente los de Sin Pre-procesamiento(2x2), Normalización(2x2) y CLAHE(2x2), siendo aproximadamente 2% más bajos que los resultados proporcionados por el método convencional (usando la imagen completa). Además es posible notar que entre una variación y otra no difieren significativamente los resultados. Es decir, no importa el tipo de pre-procesamiento que se utilice, sino el tamaño de sub-bloque.

Las Figura 7.28 - 7.47 muestran el rendimiento tanto de los métodos convencionales como de los propuestos para la tarea de verificación con diferentes valores de umbral, todos los posibles entre el rango de [0.5 - 1].

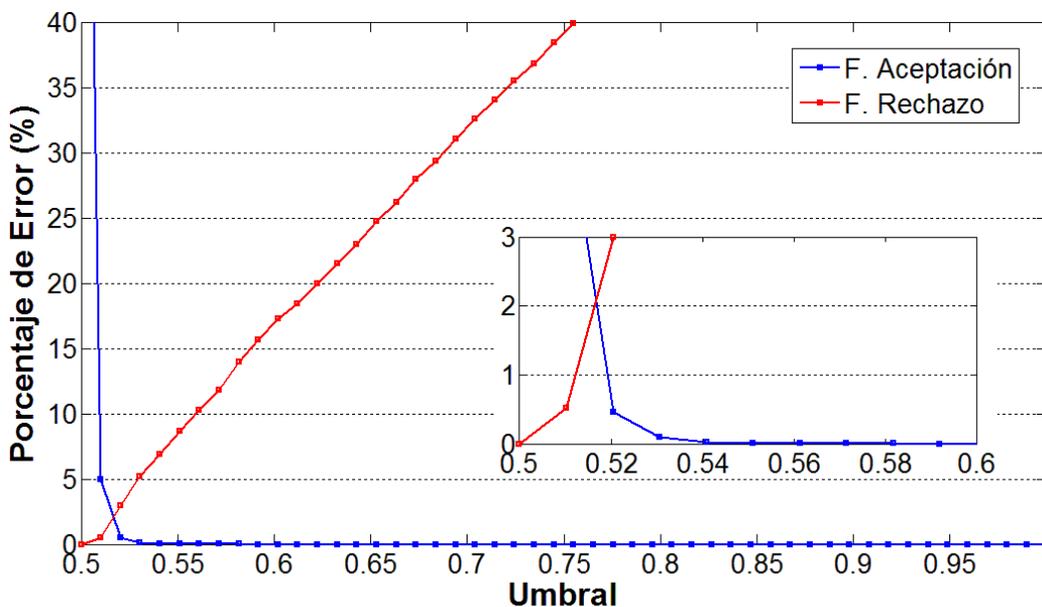


Figura 7.28. Gráfica de verificación para el método sin pre-procesamiento usando como tamaño de sub-bloque la imagen completa (convencional).

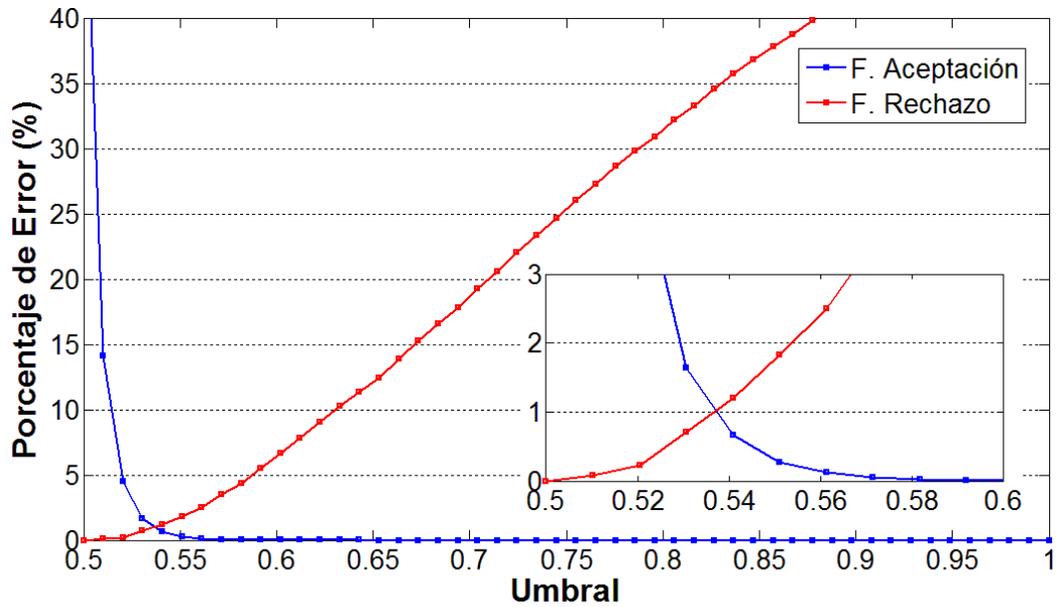


Figura 7.29. Gráfica de verificación para el método sin pre-procesamiento usando un tamaño de sub-bloque de 12x12 píxeles.

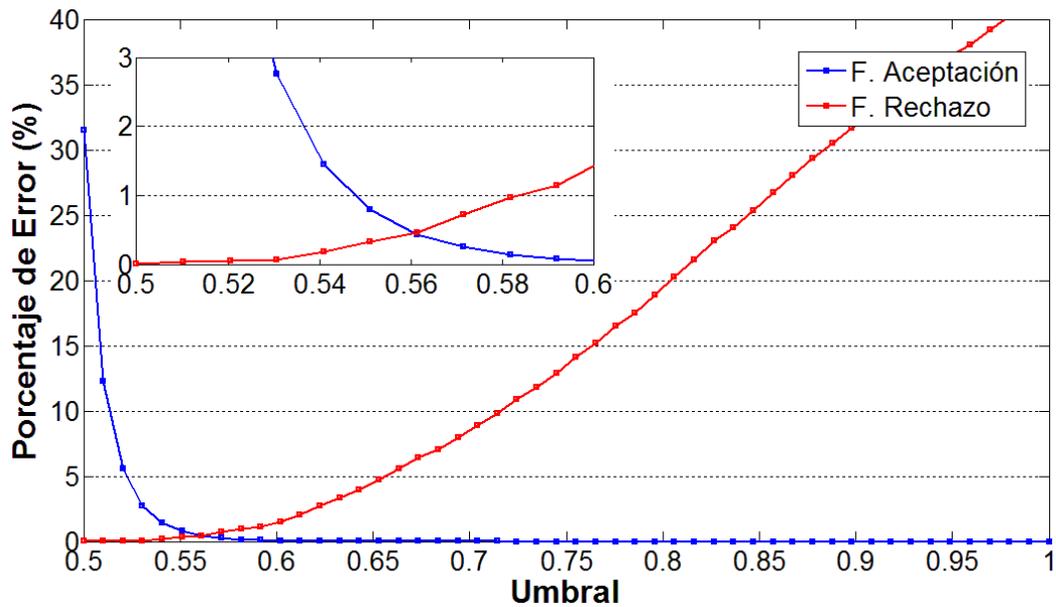


Figura 7.30. Gráfica de verificación para el método sin pre-procesamiento usando un tamaño de sub-bloque de 6x6 píxeles.

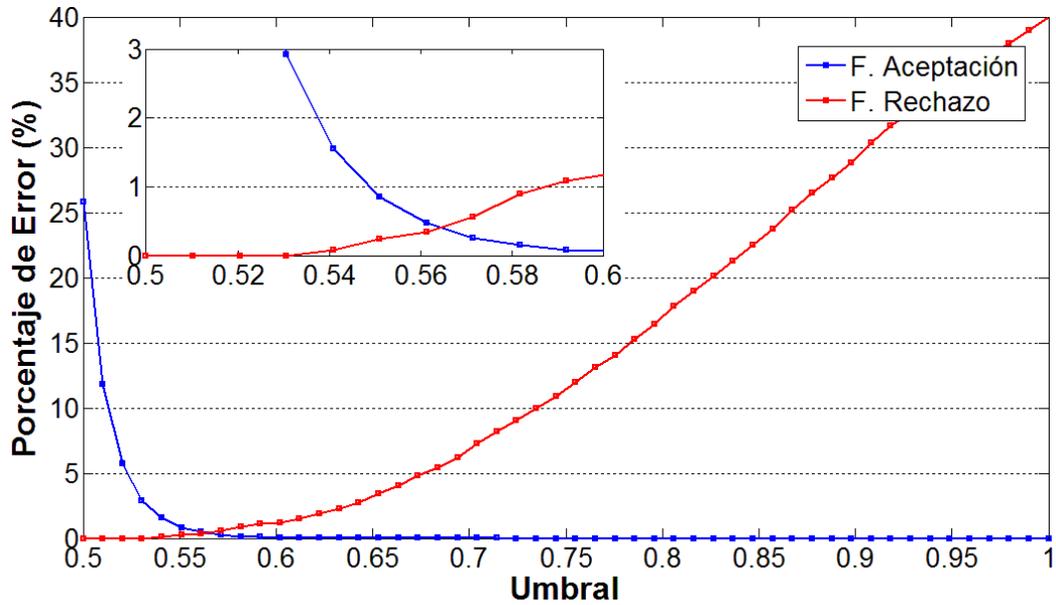


Figura 7.31. Gráfica de verificación para el método sin pre-procesamiento usando un tamaño de sub-bloque de 4x4 píxeles.

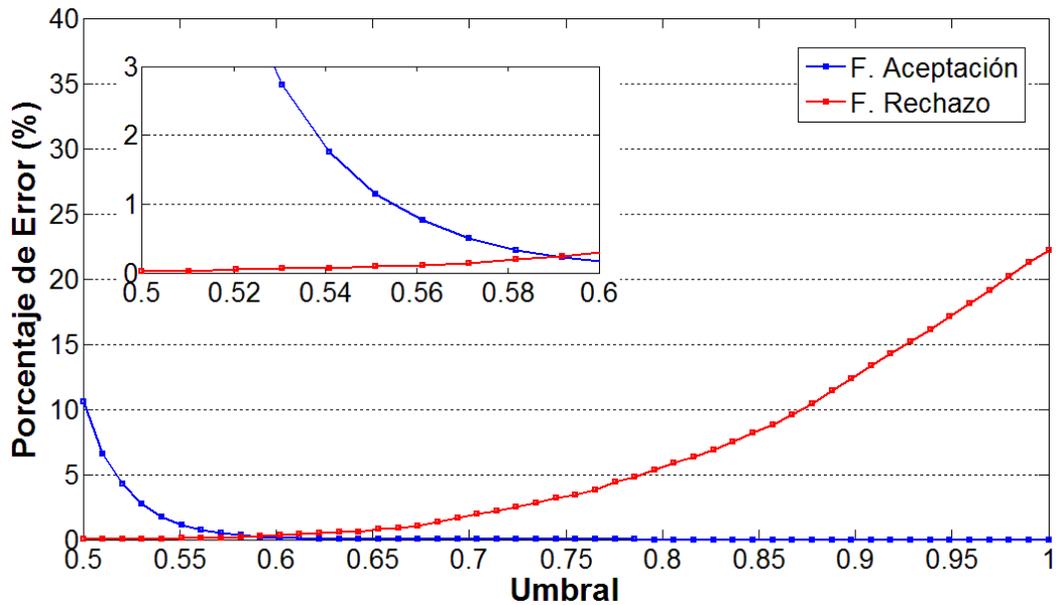


Figura 7.32. Gráfica de verificación para el método sin pre-procesamiento usando un tamaño de sub-bloque de 2x2 píxeles (óptimo).

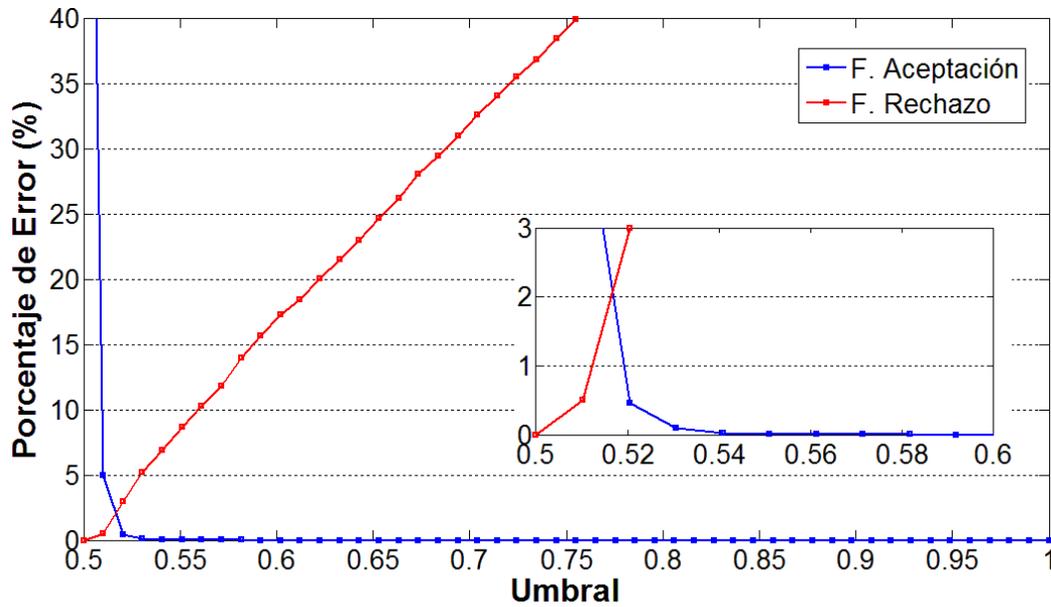


Figura 7.33 Gráfica de verificación para el método de normalización del valor del pixel usando como tamaño de sub-bloque la imagen completa (convencional).

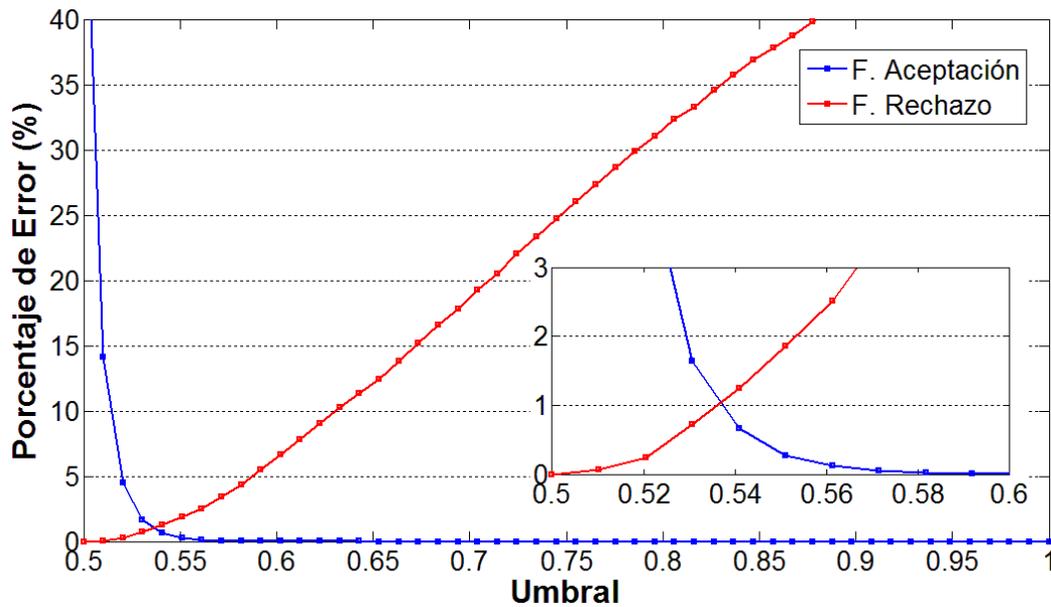


Figura 7.34. Gráfica de verificación para el método de normalización del valor del pixel usando un tamaño de sub-bloque de 12x12 pixeles.

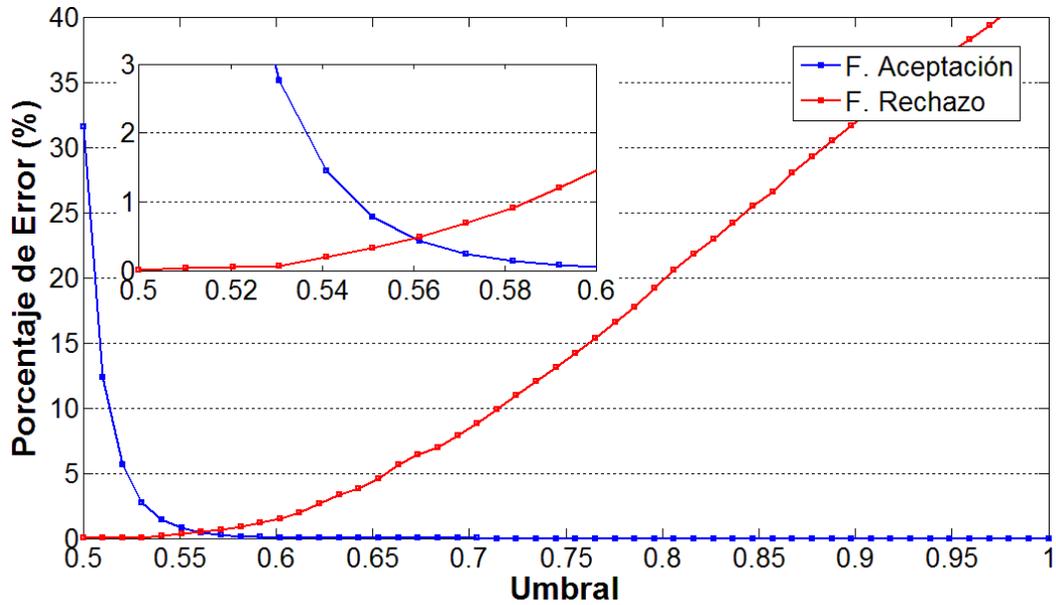


Figura 7.35. Gráfica de verificación para el método de normalización del valor del pixel usando un tamaño de sub-bloque de 6x6 píxeles.

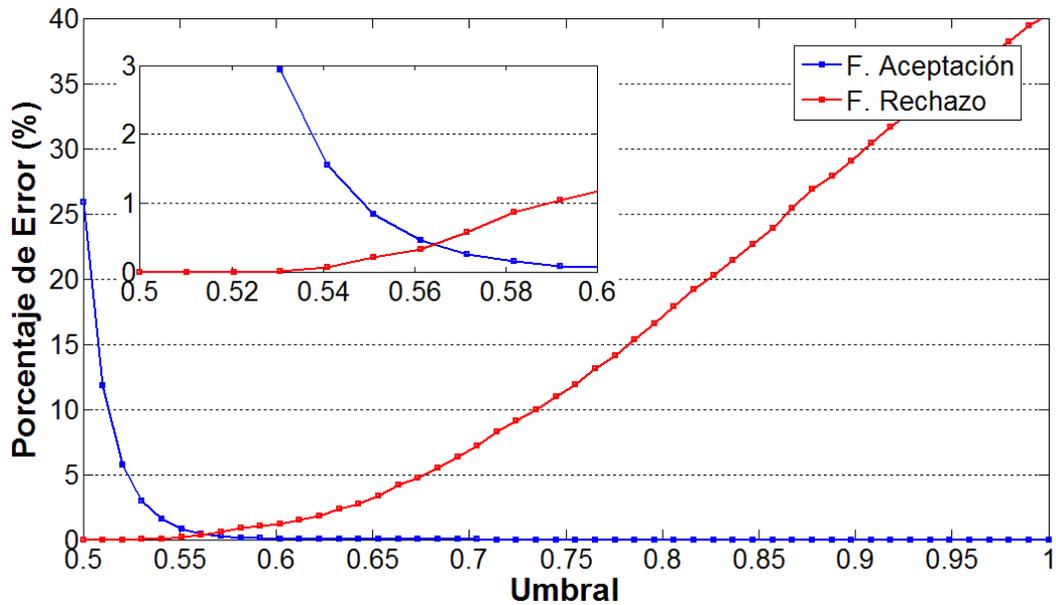


Figura 7.36. Gráfica de verificación para el método de normalización del valor del pixel usando un tamaño de sub-bloque de 4x4 píxeles.

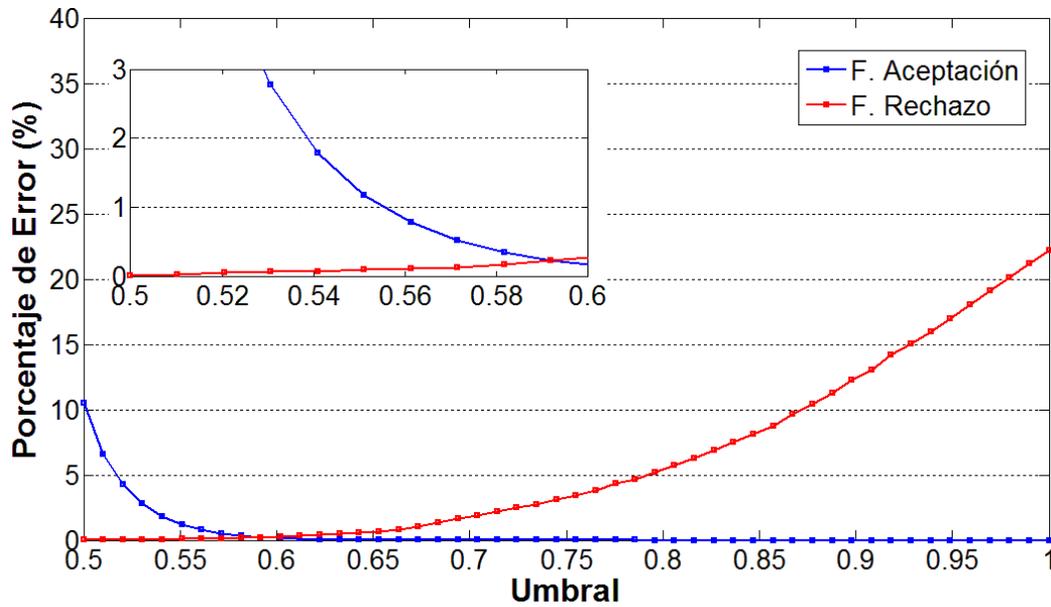


Figura 7.37. Gráfica de verificación para el método de normalización del valor del pixel usando un tamaño de sub-bloque de 2x2 pixeles (óptimo).

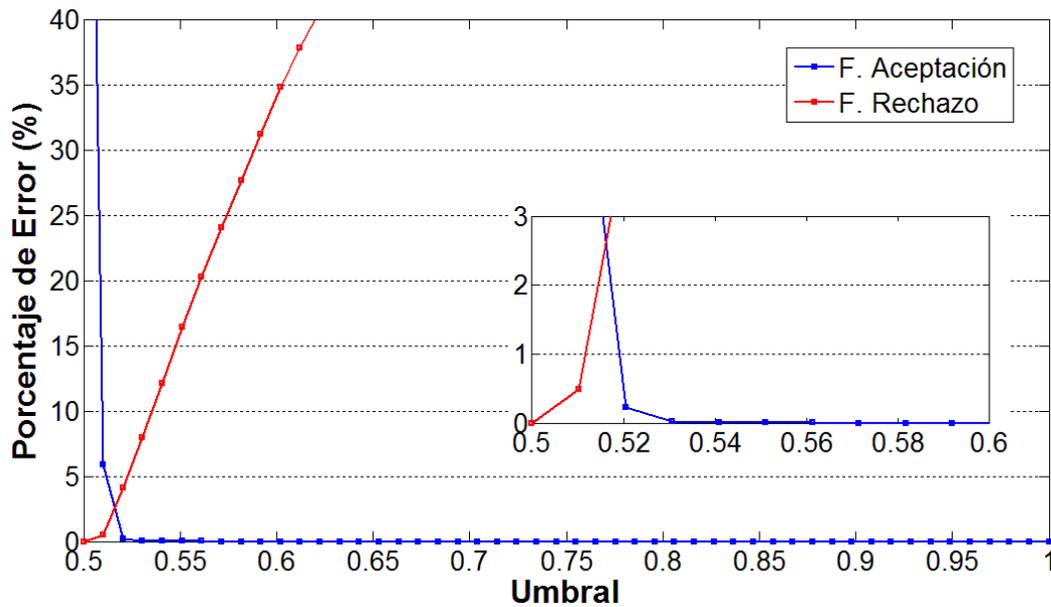


Figura 7.38. Gráfica de verificación para el método de ecualización del histograma (HE) usando como tamaño de sub-bloque la imagen completa (convencional).

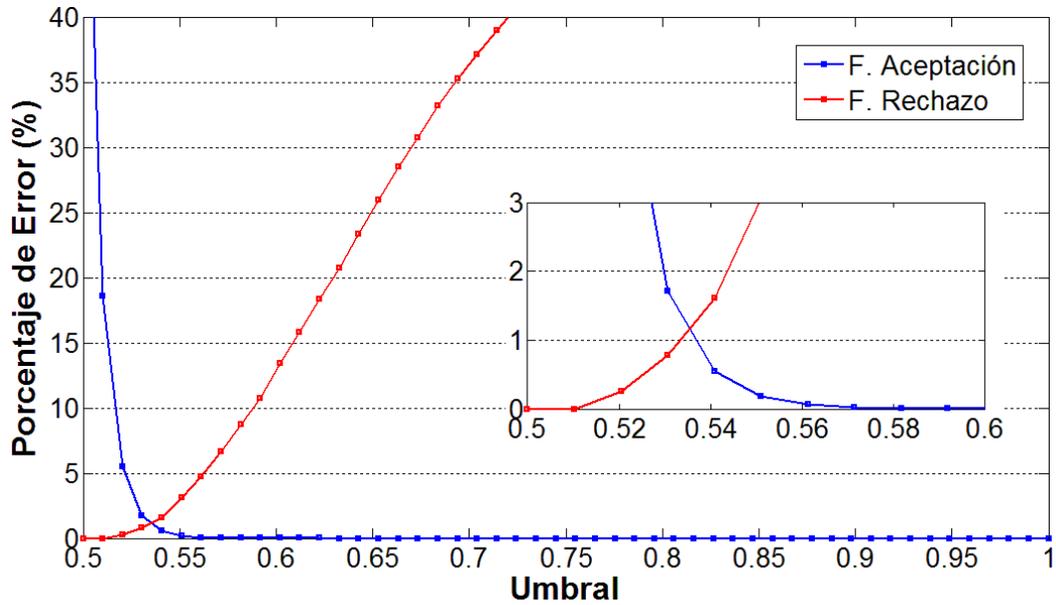


Figura 7.39. Gráfica de verificación para el método de ecualización del histograma (HE) usando un tamaño de sub-bloque de 12x12 píxeles.

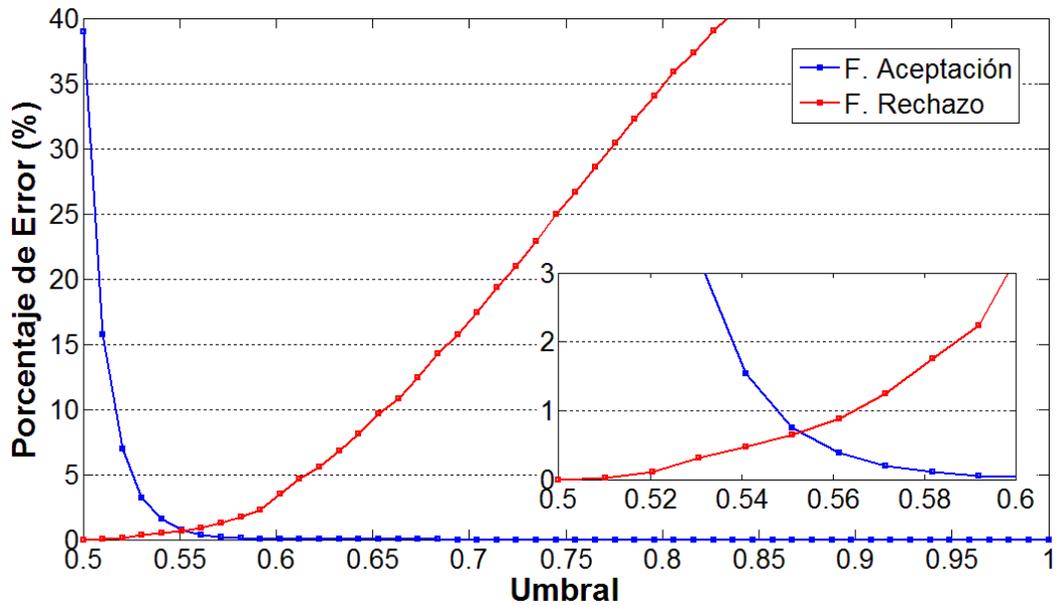


Figura 7.40. Gráfica de verificación para el método de ecualización del histograma (HE) usando un tamaño de sub-bloque de 6x6 píxeles.

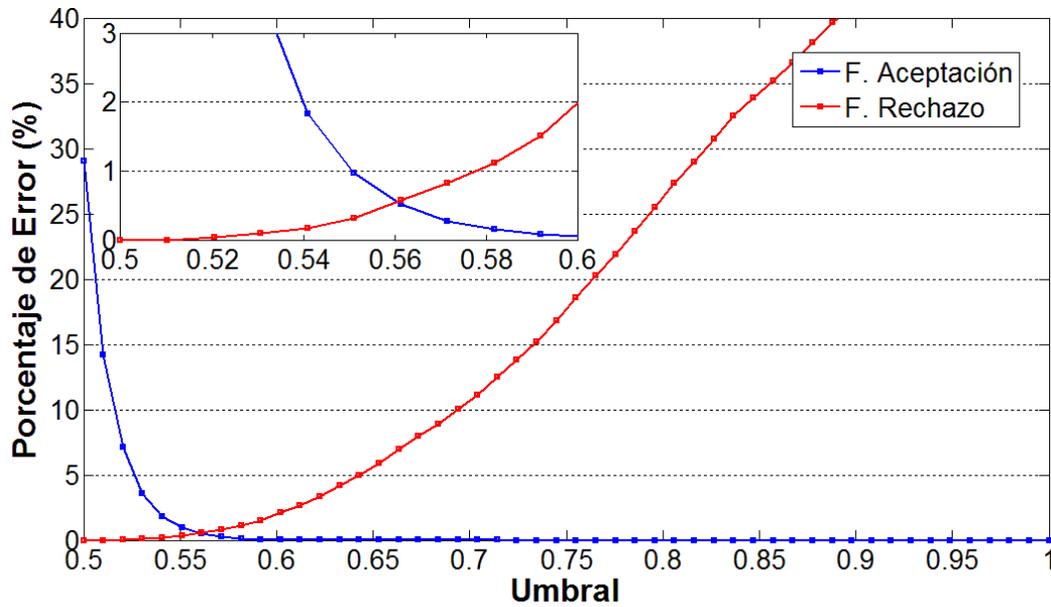


Figura 7.41. Gráfica de verificación para el método de ecualización del histograma (HE) usando un tamaño de sub-bloque de 4x4 píxeles.

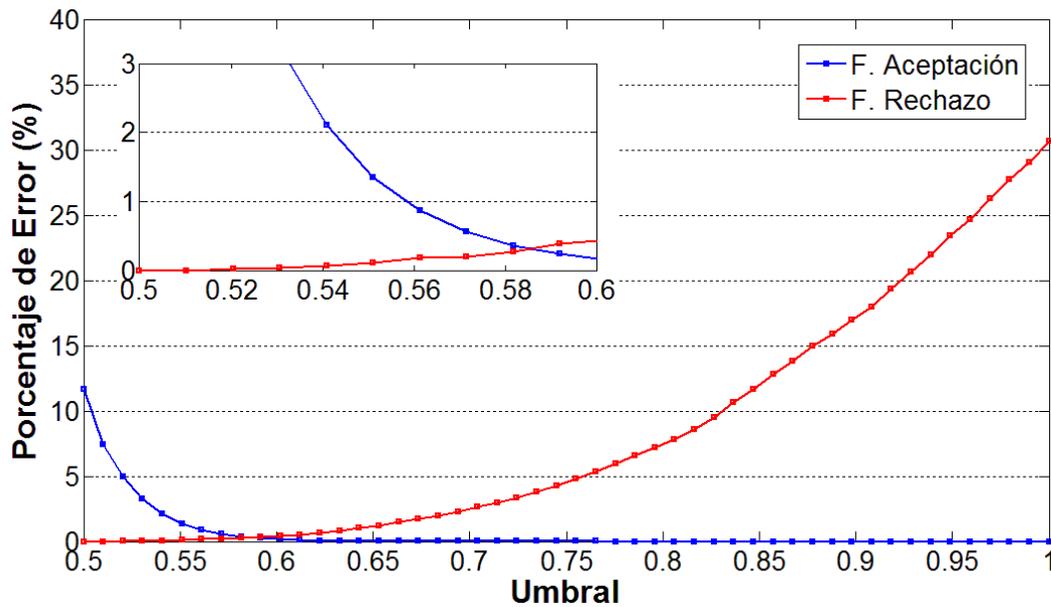


Figura 7.42. Gráfica de verificación para el método de ecualización del histograma (HE) usando un tamaño de sub-bloque de 2x2 píxeles (óptimo).

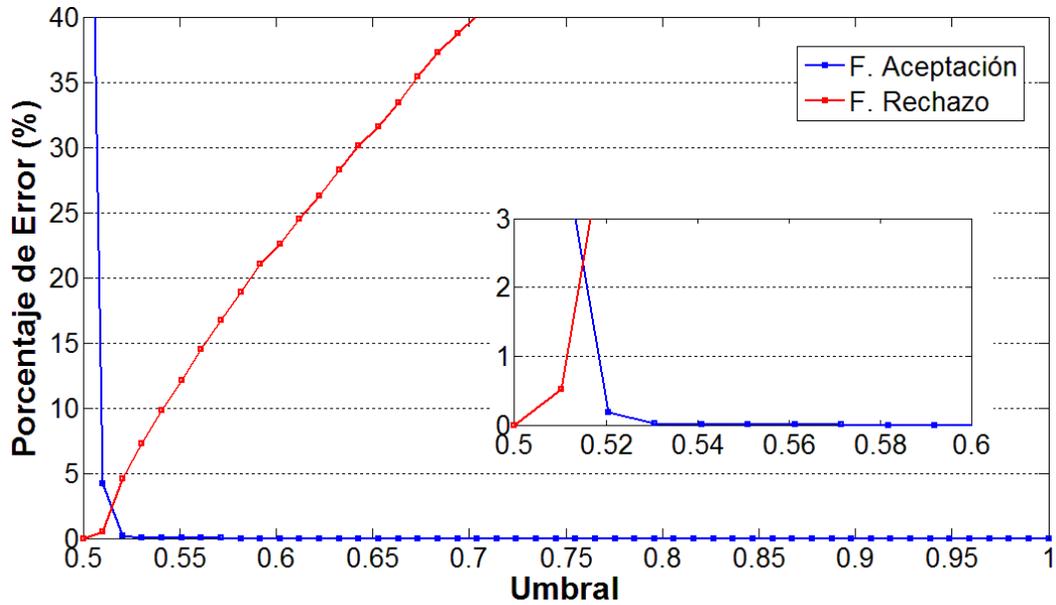


Figura 7.43. Gráfica de verificación para el método CLAHE usando como tamaño de sub-bloque la imagen completa (convencional).

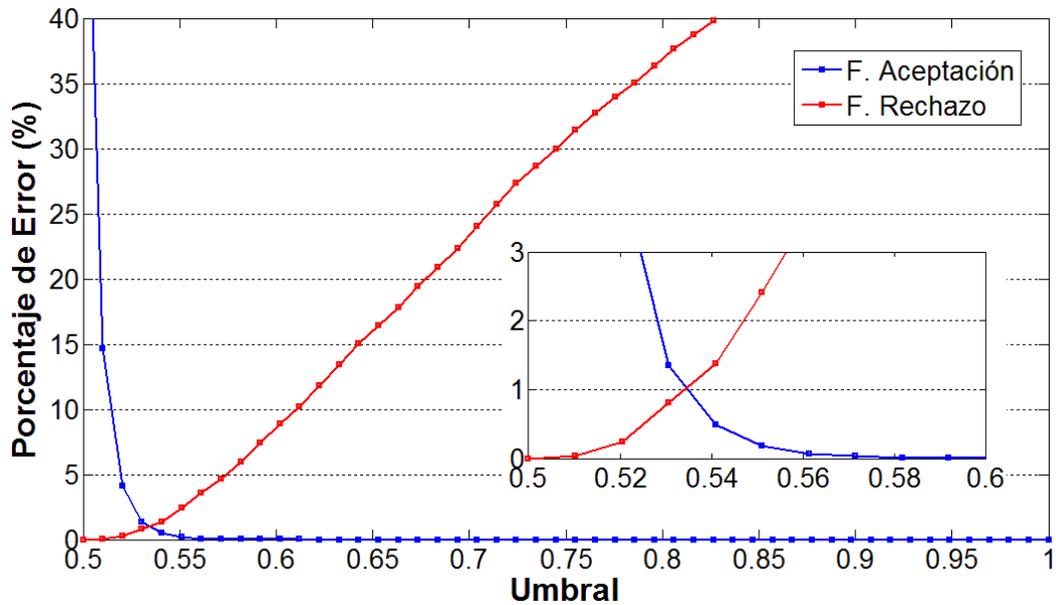


Figura 7.44. Gráfica de verificación para el método CLAHE usando un tamaño de sub-bloque de 12x12 píxeles.

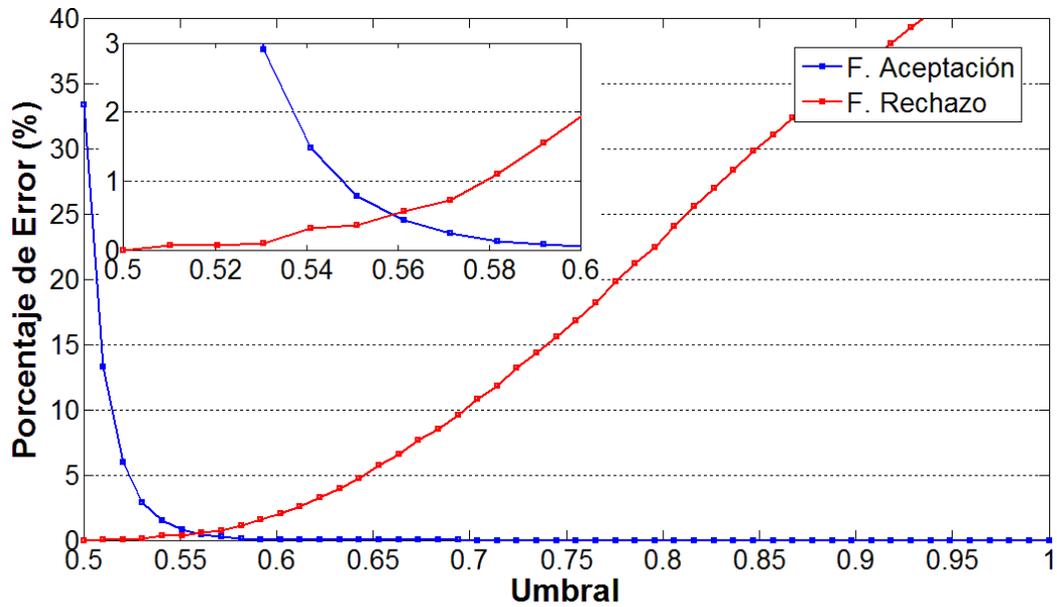


Figura 7.45. Gráfica de verificación para el método CLAHE usando un tamaño de sub-bloque de 6x6 píxeles.

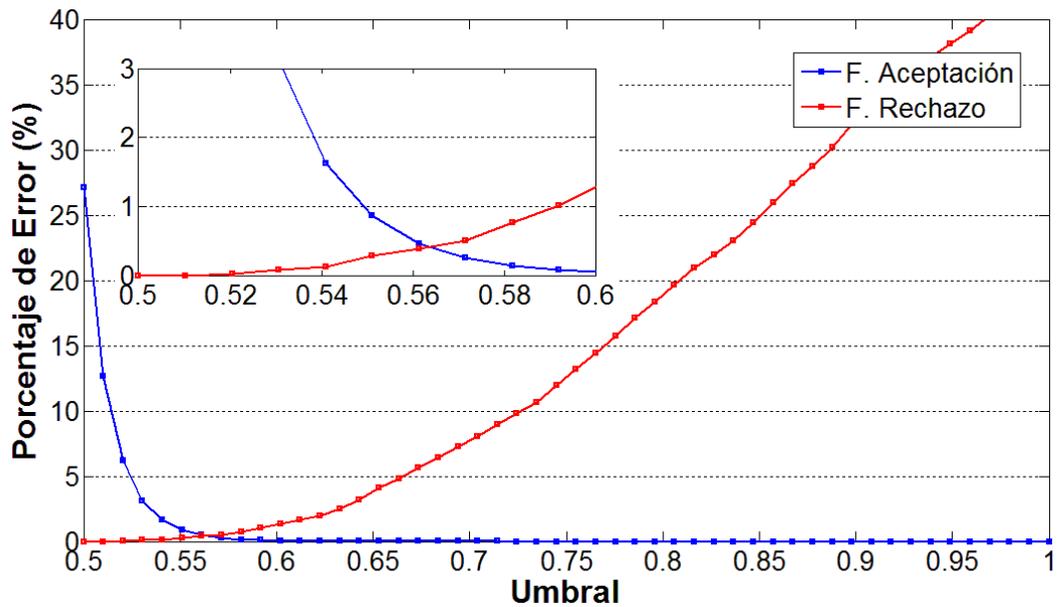


Figura 7.46. Gráfica de verificación para el método CLAHE usando un tamaño de sub-bloque de 4x4 píxeles.

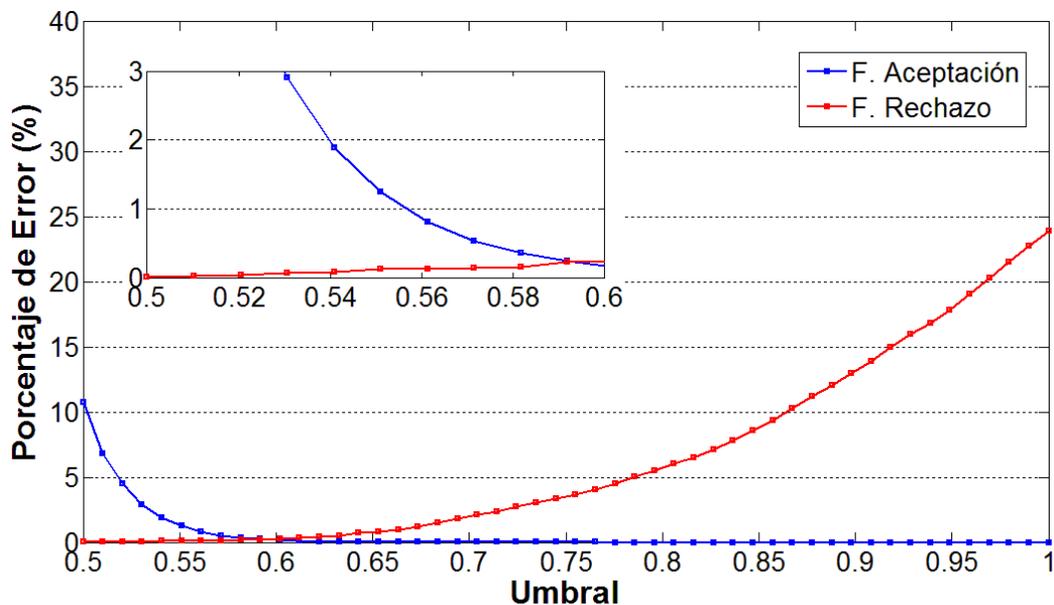


Figura 7.47. Gráfica de verificación para el método CLAHE usando un tamaño de sub-bloque de 2x2 píxeles (óptimo).

De las gráficas de verificación podemos observar que en todos los casos cuando se utiliza el tamaño de bloque óptimo se obtiene un mejor comportamiento en los porcentajes de falsa aceptación y falso rechazo. Debido a que cuando se utiliza este tamaño de sub-bloque es posible obtener simultáneamente una falsa aceptación y un falso rechazo menor al 0.5%. Este desempeño no se puede obtener usando los métodos convencionales (usando la imagen completa como tamaño de sub-bloque) puesto que, si se necesita una falsa aceptación menor a 0.5% el falso rechazo sería mayor al 2.5%, y si el falso rechazo se necesita menor al 0.5% la falsa aceptación se incrementa a más del 5%. Por lo tanto aun con diferentes valores de umbral los sistemas propuestos presentan un mejor rendimiento que el convencional.

7.4.2 Sistema Propuesto usando LBP y WBP

Los resultados de verificación para todas las variantes de los sistemas basados en LBP son evaluados en esta sección. Estos resultados fueron obtenidos utilizando las 4 variaciones y los 3 tipos de radio descritos en la sección 7.3.2. En estas pruebas solo se utilizaron imágenes sin oclusión (set A), de la misma forma que las realizadas para los sistemas usando Fases Propias (sección anterior).

Tabla 7.8. Resultados de verificación para todas las variaciones del sistema basado en LBP, porcentaje de error (F. aceptación = F. rechazo) y umbral.

Variante	% error	umbral
LBPh(8,1)	6.91	0.424
LBPh(8,2)	4.24	0.499
LBPh(8,3)	3.24	0.535
LBPc(8,1)	0.22	0.610
LBPc(8,2)	0.28	0.604
LBPc(8,3)	0.26	0.614
WBP	0.37	0.590
DI-WBP	0.65	0.609

La Tabla 7.8 muestra el porcentaje de error y el umbral de cada uno de los métodos basados en LBP donde la falsa aceptación y el falso rechazo tienen el mismo valor. En esta tabla se puede observar que las variaciones de los sistemas propuestos (LBPc, WBP y DI-WBP) presentan un mejor rendimiento que las convencionales (LBPh), ya que el porcentaje de error proporcionado por los sistemas propuestos es menor al 0.7%, en cambio para el mejor de los casos las variaciones del método convencional proporcionan un porcentaje mayor al 3%.

El mejor resultado mostrado en la Tabla 7.8 es el obtenido por el método LBPc(8,1) con un error del 0.22% para un umbral de 0.61, una diferencia de más del 6.5% con el de su símil del método convencional, LBPh(8,1). Con estas pruebas se observa que los resultados de los métodos basados en LBPc son ligeramente mejores que los basados en WBP. Sin embargo los resultados basados en WBP superan considerablemente a los obtenidos por el método convencional.

Las Figura 7.48 - 7.55 muestran el porcentaje de falsa aceptación y falso rechazo para diferentes valores de umbral, específicamente del rango [0.5 - 1], de todas las variaciones de los métodos convencionales y propuestos. Cabe mencionar que para el caso de los métodos LBPh(8,1) y LBPh(8,2), donde el punto de intersección entre la falsa aceptación y el falso rechazo se encuentra fuera del rango [0.5 - 1], el recuadro de acercamiento muestra claramente el punto de equilibrio ya que se toma como rango para este acercamiento los valores de umbral de [0.4 – 0.6] y de 0 a 9% para el rango de error.

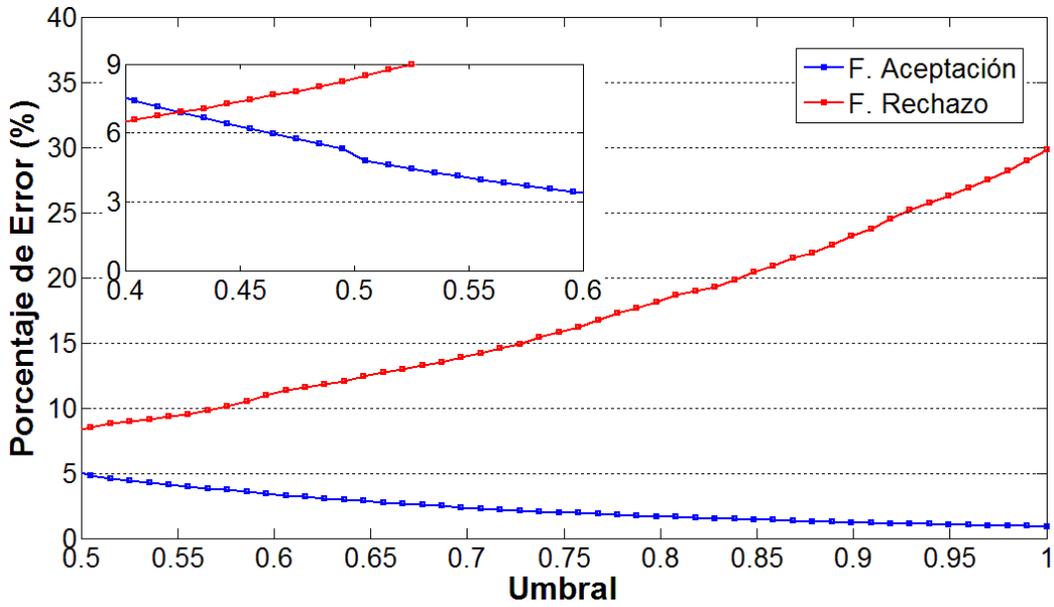


Figura 7.48. Gráfica de verificación para el método LBPh(8,1).

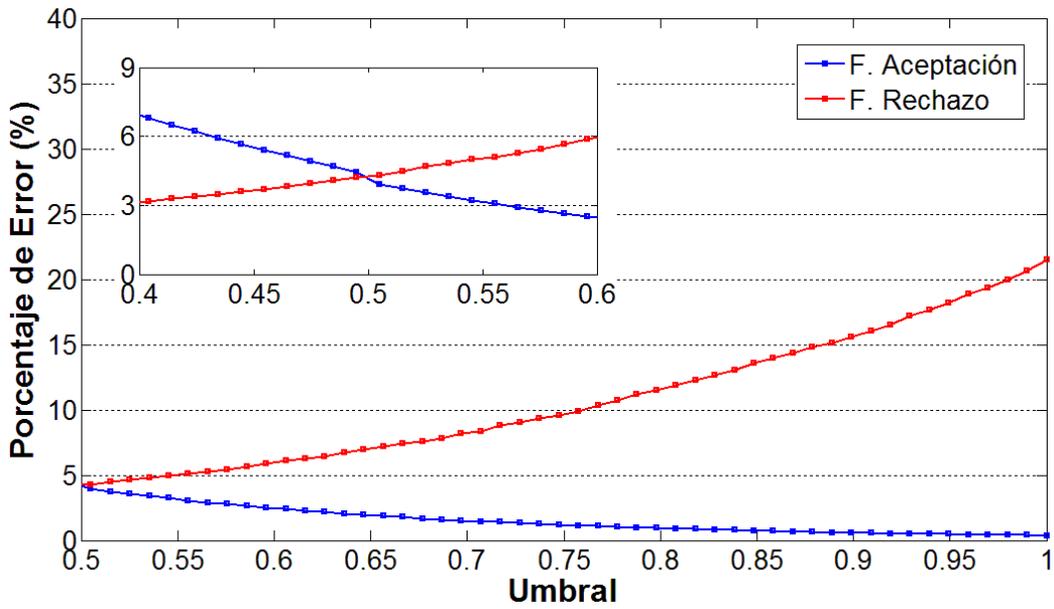


Figura 7.49. Gráfica de verificación para el método LBPh(8,2).

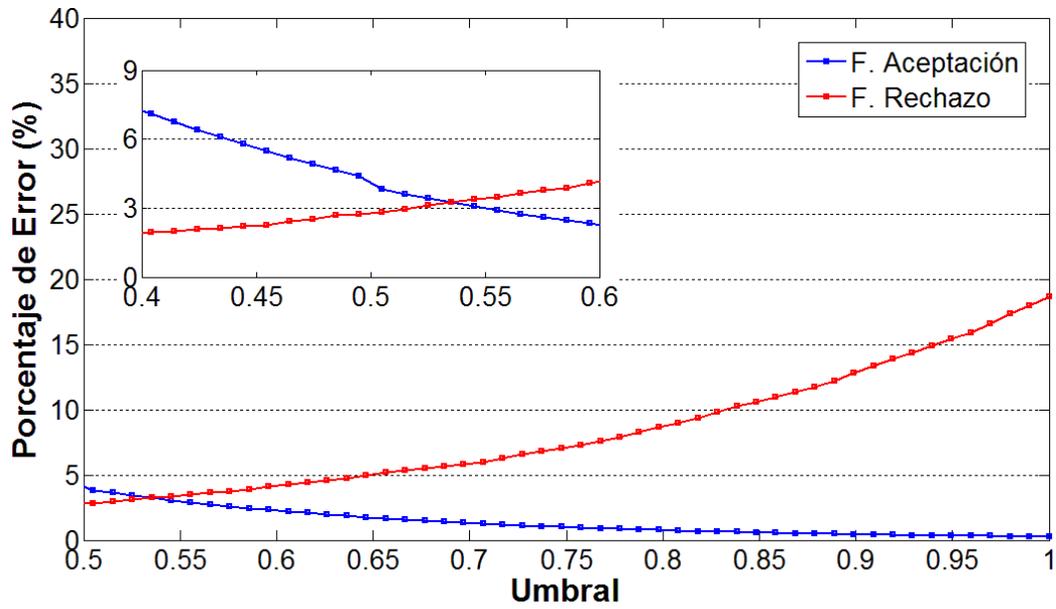


Figura 7.50. Gráfica de verificación para el método LBPh(8,3).

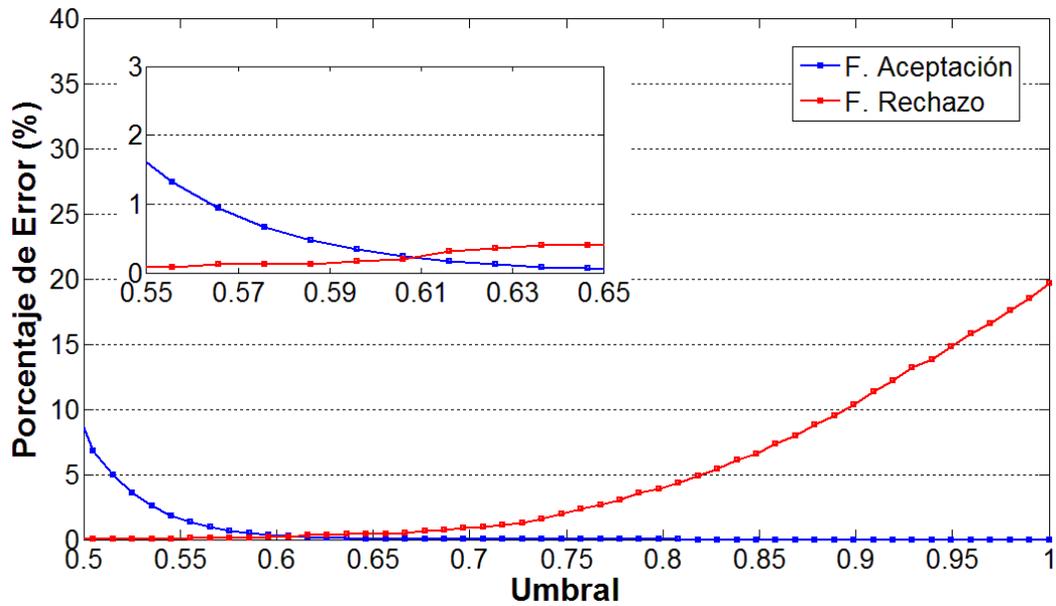


Figura 7.51. Gráfica de verificación para el método LBPh(8,1).

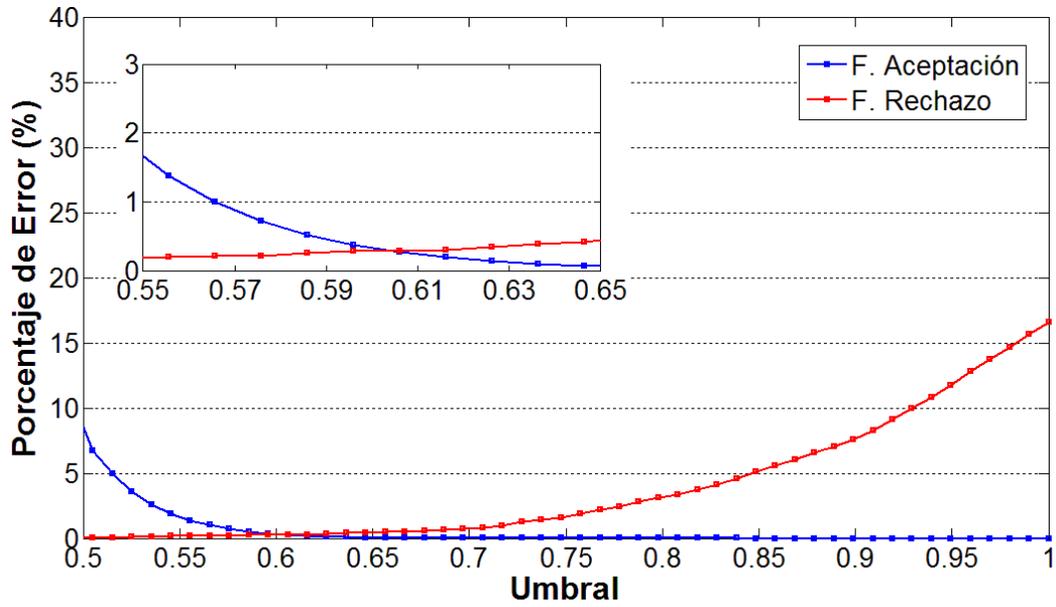


Figura 7.52. Gráfica de verificación para el método LBPC(8,2).

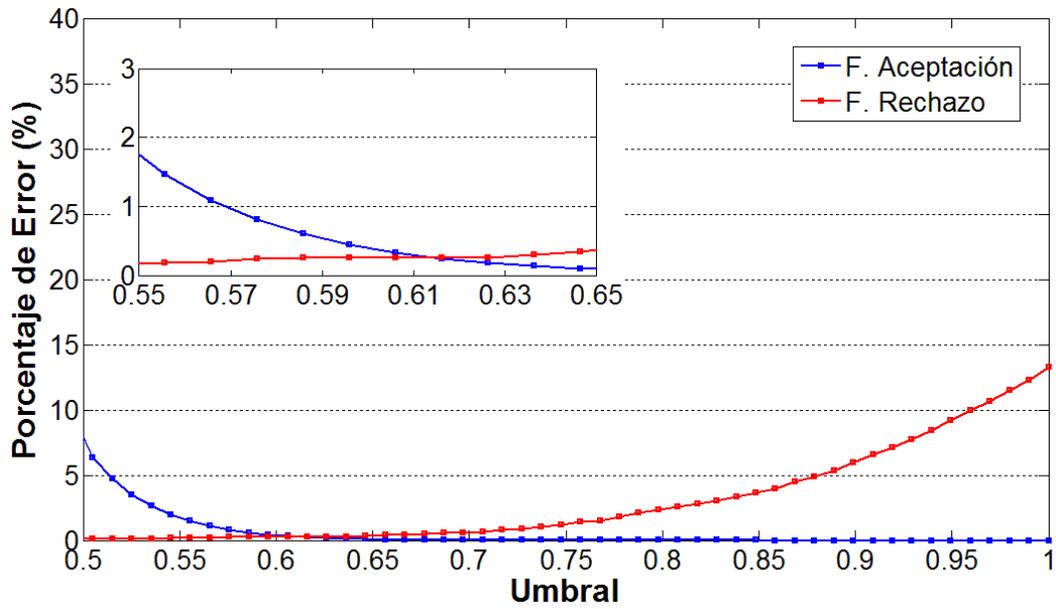


Figura 7.53. Gráfica de verificación para el método LBPC(8,3).

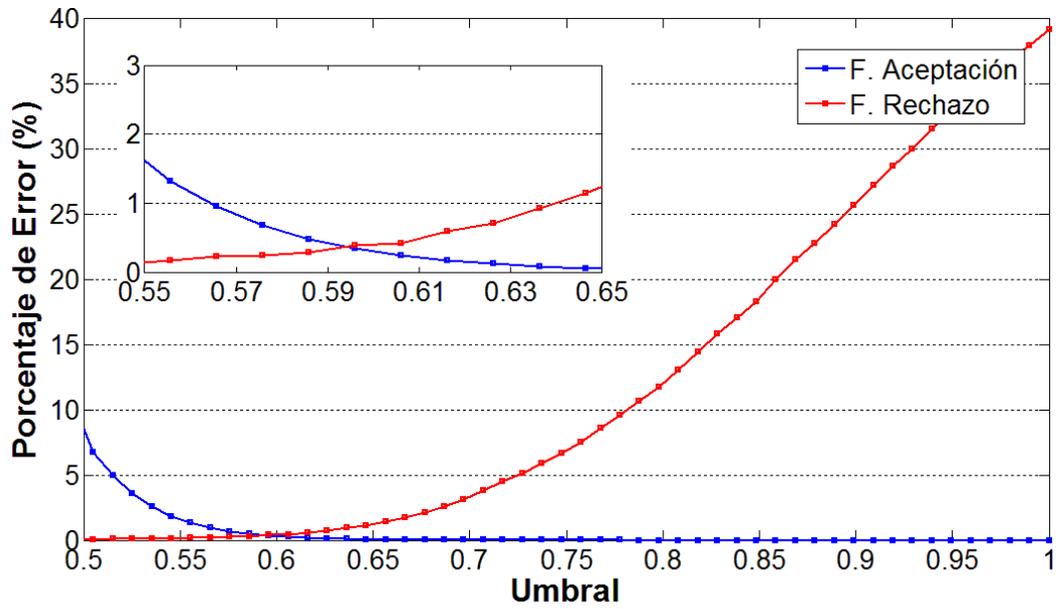


Figura 7.54. Gráfica de verificación para el método WBP.

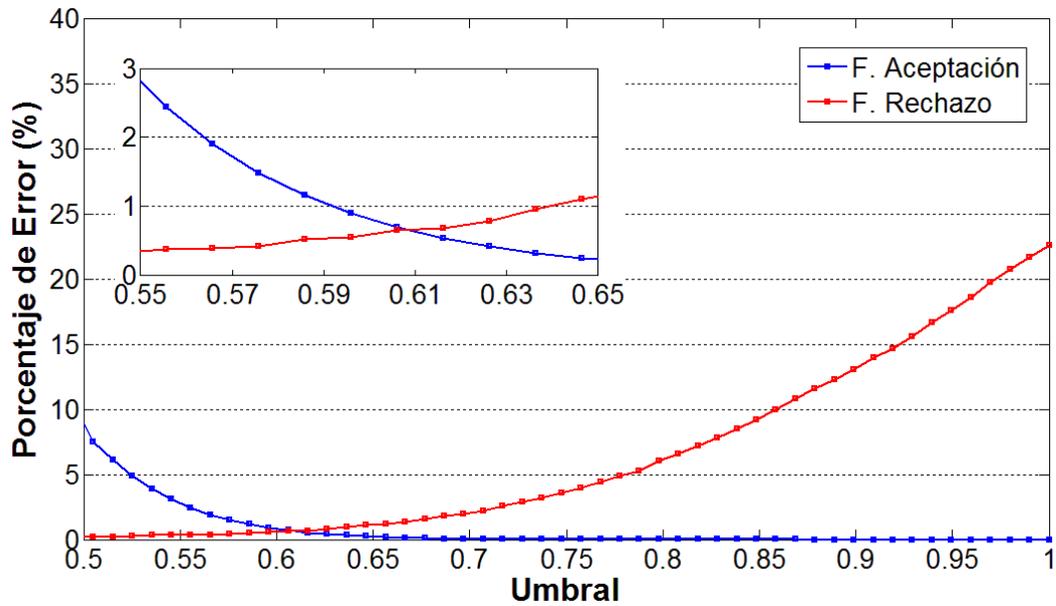


Figura 7.55. Gráfica de verificación para el método DI-WBP.

7.4.3 Comparación de Resultados

Los mejores resultados de verificación obtenidos por los diferentes sistemas propuestos (usando Fases Propias y basados en LBP) se presentan en esta sección. La Tabla 7.9 muestra los mejores resultados de las variaciones de los métodos LBPC, Fases Propias Basados en Sub-bloques y WBP, donde el mejor resultado es el obtenido por el método LBPC(8,1) con un porcentaje de error del 0.22% para un umbral de 0.61. Mostrando que LBPC(8,1) es el método que presenta el mejor rendimiento de verificación de entre todos los analizados en este trabajo.

Tabla 7.9. Mejores resultados de verificación por parte de los sistemas propuestos.

Método	% error	umbral
LBPC(8,1)	0.22	0.610
CLAHE(2x2)	0.23	0.593
Normalización(2x2)	0.23	0.592
Sin pre-proc.(2x2)	0.23	0.591
WBP	0.37	0.590

Es importante mencionar que todos los resultados de la Tabla 7.9 se encuentran por debajo del 0.5% de error, donde tienen el mismo valor tanto la falsa aceptación como el falso rechazo. Otra característica resaltante de esta tabla es que los resultados obtenidos por los métodos de fases propias que usan el tamaño de bloque óptimo son casi idénticos al mejor resultado obtenido en este trabajo, solo 0.01% de diferencia. Demostrando así, la efectividad de los algoritmos de fases propias usando sub-bloques de 2x2 para las pruebas de verificación.

7.5 Resultados de Reconocimiento de Expresiones Faciales

Esta sección presenta los resultados del sistema propuesto para expresiones faciales (FER). Donde las principales dos propuestas son: una modificación en la etapa de extracción de características, específicamente en la conformación del vector característico; y un nuevo método de decisión llamado enfoque del valor modal. Todo esto partiendo de la propuesta de usar una etapa de segmentación del rostro [39], con el objetivo de trabajar específicamente con la información que es más relevante para los sistemas de expresiones faciales.

7.5.1 Resultados Sin Oclusión

Estas pruebas se realizaron con las 300 imágenes pico sin oclusión descritas en la sección 7.1.2.3. La primera prueba realizada en estas condiciones fue la del método de Fases Propias Basado en Sub-bloques de 2x2 (FPBS), el cual obtuvo un porcentaje general de reconocimiento de 78.33%. Con objetivo de analizar su comportamiento en un sistema de FER y tomarlo como punto de comparación, el algoritmo se utilizó directamente como fue propuesto en el capítulo 4, sin utilizar la segmentación del rostro ni el método de enfoque del valor modal. La Tabla 7.11 muestra su matriz de confusión, donde se observa que el método reconoce con cierta facilidad la expresión de sorpresa pero con serias dificultades las de miedo e ira.

Tabla 7.11. Matriz de confusión del método FPBS.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	63.3	6.7	6.7	3.3	20.0	0.0
Asco	8.8	73.5	2.9	5.9	8.8	0.0
Miedo	0.0	2.1	55.3	29.8	8.5	4.3
Felicidad	1.4	1.4	7.1	90.0	0.0	0.0
Tristeza	13.0	1.9	3.7	0.0	75.9	5.6
Sorpresa	0.0	0.0	1.5	0.0	4.6	93.8

La segunda prueba fue realizada utilizando la etapa de segmentación del rostro pero aún sin el método de enfoque del valor modal. De esta manera se obtuvieron resultados individuales del sistema propuesto usando cada una de las 4 regiones segmentadas, además de sus posibles combinaciones (combinaciones realizadas como se detalla en la sección 6.4). La Tabla 7.12 muestra los resultados del sistema propuesto cuando solo se utiliza una región del rostro para realizar el reconocimiento, marcando con letra negrita el mejor resultado. Siendo la boca la región del rostro que proporciona el mejor reconocimiento general, con casi 80%, su matriz de confusión se muestra en la Tabla 7.13.

Tabla 7.12. Resultados de reconocimiento del sistema utilizando solo una región del rostro.

Abrev.	Región	%
O	Ojos	53.33
F	Frente	28.67
B	Boca	79.33
N	Nariz	61.00

Tabla 7.13. Matriz de confusión del sistema propuesto utilizando solo la región de la boca (B).

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	70.0	3.3	6.7	3.3	13.3	3.3
Asco	5.9	61.8	2.9	5.9	8.8	14.7
Miedo	4.3	4.3	59.6	21.3	2.1	8.5
Felicidad	0.0	1.4	11.4	87.1	0.0	0.0
Tristeza	3.7	7.4	1.9	0.0	85.2	1.9
Sorpresa	0.0	1.5	4.6	0.0	0.0	93.8

Los resultados del sistema utilizando las combinaciones de 2 regiones del rostro se muestran en la Tabla 7.14 en donde se puede observar que la combinación de las regiones de ojos y boca alcanza un reconocimiento del 86%, un resultado aceptable tomando en cuenta que solo se están utilizando dos regiones del rostro. Resultado que es significativamente mayor que el obtenido por FPBS (78.33%), en el que se utiliza el rostro completo, excluyendo la etapa de segmentación. La Tabla 7.15 muestra la matriz de confusión para el resultado de la prueba utilizando la combinación de Ojos-Boca.

Tabla 7.14. Resultados de reconocimiento del sistema utilizando dos regiones del rostro.

Abrev.	Combinación	%
OF	Ojos-Frente	56.67
OB	Ojos-Boca	86.00
ON	Ojos-Nariz	69.33
FB	Frente-Boca	79.33
FN	Frente-Nariz	61.67
BN	Boca-Nariz	83.00

Tabla 7.15. Matriz de confusión del sistema propuesto utilizando la combinación de Ojos-Boca (OB).

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	83.3	6.7	0.0	0.0	6.7	3.3
Asco	5.9	82.4	2.9	2.9	2.9	2.9
Miedo	0.0	6.4	66.0	19.1	4.3	4.3
Felicidad	0.0	1.4	5.7	91.4	0.0	1.4
Tristeza	1.9	3.7	1.9	0.0	90.7	1.9
Sorpresa	0.0	0.0	3.1	0.0	3.1	93.8

La Tabla 7.16 muestra los resultados del sistema cuando se utilizan las combinaciones de 3 regiones del rostro. El mejor resultado es el obtenido por la combinación de Ojos-Boca-Nariz, alcanzando 87.67% de reconocimiento, su matriz de confusión se muestra en la Tabla 7.17.

Tabla 7.16. Resultados de reconocimiento del sistema utilizando tres regiones del rostro.

Abrev.	Combinación	%
OFB	Ojos-Frente-Boca	85.00
OFN	Ojos-Frente-Nariz	70.33
OBN	Ojos-Boca-Nariz	87.67
FBN	Frente-Boca-Nariz	82.67

Tabla 7.17. Matriz de confusión del sistema propuesto utilizando la combinación de Ojos-Boca-Nariz (OBN).

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	70.0	6.7	6.7	0.0	16.7	0.0
Asco	5.9	85.3	8.8	0.0	0.0	0.0
Miedo	0.0	4.3	72.3	14.9	4.3	4.3
Felicidad	0.0	0.0	5.7	94.3	0.0	0.0
Tristeza	0.0	0.0	3.7	0.0	96.3	0.0
Sorpresa	0.0	0.0	3.1	0.0	3.1	93.8

Finalmente la combinación de las 4 regiones del rostro ojos, frente, boca y nariz (OFBN) alcanzo un reconocimiento general de 86.67% y su matriz de confusión se muestra en la Tabla 7.18. La Figura 7.56 muestra una gráfica con todos los resultados de FER obtenidos hasta este punto, esto en función de poder comparar los diferentes resultados obtenidos con todas las posibles combinaciones de las 4 regiones del rostro.

Tabla 7.18. Matriz de confusión del sistema propuesto utilizando la combinación de las 4 regiones del rostro (OFBN).

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	73.3	6.7	6.7	0.0	13.3	0.0
Asco	11.8	76.5	2.9	2.9	2.9	2.9
Miedo	0.0	0.0	72.3	19.1	4.3	4.3
Felicidad	0.0	0.0	4.3	95.7	0.0	0.0
Tristeza	1.9	0.0	3.7	0.0	94.4	0.0
Sorpresa	0.0	0.0	4.6	1.5	1.5	92.3

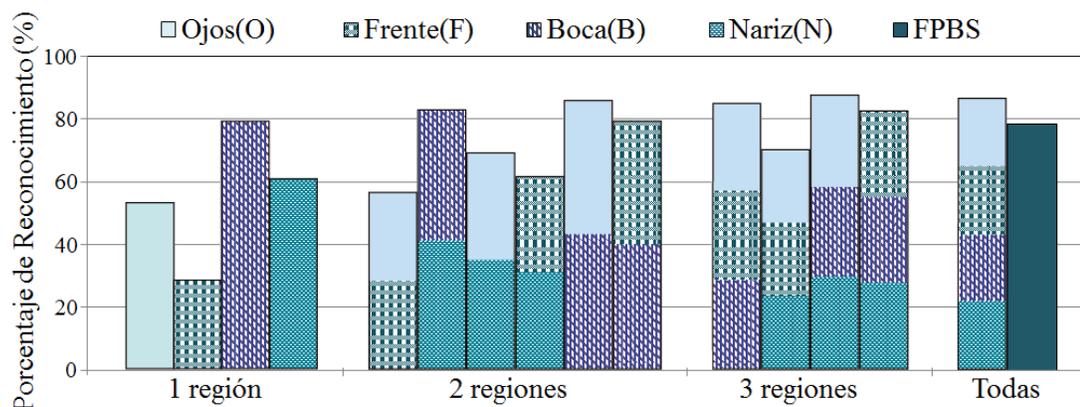


Figura 7.56. Resultados de todas las posibles combinaciones utilizando las 4 regiones del rostro.

De la Figura 7.56 podemos observar que el mejor resultado es el obtenido por la combinación de 3 regiones OBN (87.7%), siendo aún mejor que los obtenidos por FPBS (78.3%) y OFBN (86.7%). Otro punto importante que podemos ver en esta figura es que las combinaciones que utilizan la región de la boca proporcionan el mayor porcentaje de reconocimiento independientemente del número de las regiones faciales utilizadas en el proceso. Mientras tanto, cuando no se emplea la región de la boca, las combinaciones de las otras regiones no proporcionan porcentaje de reconocimiento competitivo.

La siguiente prueba fue utilizando el método de enfoque del valor modal, descrito en la sección 6.6. En este método, varios SVMs asociados a diferentes combinaciones de regiones faciales fueron ejecutados en paralelo, obteniendo un resultado a partir de cada clasificador, y tomando el valor modal para así unificar todos los resultados individuales. El número de clasificadores utilizados en esta prueba fue de 16, 15 a partir de las combinaciones de las 4 regiones faciales y uno donde se excluye la etapa de segmentación del rostro (FPBS). Es importante mencionar que, al menos, 3 clasificadores se requieren para encontrar el valor modal.

Los mejores resultados obtenidos utilizando el método de enfoque del valor modal se muestran en la Tabla 7.19. Es posible ver que el mejor resultado alcanza el 92% de reconocimiento, obtenido utilizando 4 clasificadores, las combinaciones utilizadas en este caso fueron: Ojos-Boca (OB), Frente-Boca (FB), las cuatro regiones (OFBN) y sin segmentación facial (FPBS).

El resultado de la combinación OB–FB–OFBN–FPBS es el que proporciona el más alto nivel de reconocimiento en este trabajo, superando en casi un 15% el porcentaje de reconocimiento de FPBS y alrededor de 5% al mejor resultado obtenido cuando se utiliza sólo un clasificador, proporcionado por la combinación de Ojos-Boca-Nariz (OBN).

Tabla 7.19. Resultados de reconocimiento del sistema utilizando dos regiones del rostro.

#SVMs	Combinación	%
4	OB – FB – OFBN – FPBS	92.00
4	B – FB – OBN – OFBN	91.67
6	O – N – OB – FB – OBN – OFBN	90.00
4	B – OB – OBN – OFBN	89.33
3	FB – OBN – FPBS	88.00

La Tabla 7.20 muestra la matriz de confusión del mejor resultado obtenido para FER, alcanzado por la combinación OB–FB–OFBN–FPBS, usando el método del enfoque del valor modal. En ella se puede observar que sorpresa, tristeza, felicidad y asco son expresiones fáciles de reconocer para el sistema, mientras que ira y miedo no lo son. Asimismo, el porcentaje de reconocimiento de sorpresa es el más alto y miedo es el más bajo, con un 98.5% y 76.6% respectivamente. El problema del sistema en reconocer la expresión de miedo está relacionado con la expresión de felicidad, ya que en el 17% de los casos el sistema propuesto confunde la expresión de miedo con la de felicidad.

Tabla 7.20. Matriz de confusión del mejor resultado obtenido para FER, usando el método de enfoque del valor modal con la combinación OB–FB–OFBN–FPBS.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	83.3	0.0	3.3	3.3	10.0	0.0
Asco	0.0	88.2	5.9	2.9	0.0	2.9
Miedo	0.0	0.0	76.6	17.0	6.4	0.0
Felicidad	0.0	0.0	2.9	97.1	0.0	0.0
Tristeza	0.0	0.0	0.0	1.9	98.1	0.0
Sorpresa	0.0	0.0	1.5	0.0	0.0	98.5

7.5.2 Resultados Con Oclusión

El algoritmo propuesto fue evaluado usando cuatro tipos de oclusión parcial: oclusión de la mitad izquierda del rostro, de la mitad derecha, de los ojos-cejas y de la boca, como se describen en la sección 7.1.2.1 y se muestran en la Figura 7.7.

Para poder resolver de manera más eficiente el problema de oclusión, 4 nuevas regiones faciales fueron incluidas a las descritas en la sección 6.2, esto con la finalidad de tener más combinaciones posibles para cuando se utilice el método del enfoque del valor modal. Las regiones faciales anexadas son: ojo izquierdo, ojo derecho, mitad izquierda del rostro y mitad derecha. Para las regiones de un solo ojo, la segmentación se realizó recortando la mitad de la región de los ojos-cejas y para las regiones de la mitad del rostro, se recortó la mitad de la imagen del rostro completo.

Tabla 7.21. Resultados de reconocimiento del sistema utilizando las 4 regiones anexadas.

Abrev.	Región	%
Oi	Ojo izquierdo	61.00
Od	Ojo derecho	50.33
Ri	Rostro izquierdo	75.00
Rd	Rostro derecho	79.33

Para determinar la contribución de estas nuevas regiones en el reconocimiento de expresiones faciales, cada uno de ellas se utilizó de forma independiente y los resultados de esta evaluación se muestran en la Tabla 7.21. Estos resultados pueden ser comprobables con los de la Tabla 7.12, ya que prácticamente solo se está utilizando una región del rostro para realizar el reconocimiento. Siendo 79.33% el mejor resultado de entre estas cuatro nuevas regiones, proporcionado por la región de la mitad derecha del rostro (Rd), resultado idéntico al obtenido por la región de la boca.

Para cada tipo de oclusión parcial este trabajo propone una solución diferente. Para el caso específico de la oclusión de la mitad del rostro, se emplea la solución de utilizar imágenes espejo en las que la imagen reconstruida es segmentada en las 8 regiones descritas anteriormente, obteniendo resultados individuales que se unifican con el uso del método del enfoque del valor modal. Por otro lado, el problema de oclusión de los ojos-cejas se soluciona simplemente utilizando las combinaciones de las regiones faciales que no se encuentran ocluidas: frente (F), boca (B) y nariz (N).

La misma solución se emplea para la oclusión de la boca, utilizando las combinaciones de las regiones de ojos-cejas (O), ojo izquierdo (Oi), ojo derecho (Od), frente (F) y nariz (N).

Los porcentajes de reconocimiento de las combinaciones con los mejores resultados obtenidos para cada tipo de oclusión se muestran en la Tabla 7.22. Es importante mencionar que estos fueron los mejores resultados usando el método del enfoque del valor modal, siendo 4 clasificadores los ideales para esta tarea.

Tabla 7.22. Mejores resultados utilizando para cada tipo de oclusión utilizando el método del enfoque del valor modal.

Tipo de oclusión	Combinación	%
Mitad izquierda	OBN – OFBN – Ri – Rd	87.00
Mitad derecha	OBN – OFBN – Ri – Rd	83.33
Ojos	B – FB – BN – FBN	87.67
Boca	N – ON – FN – OFN	75.33

Las Tabla 7.23 - 7.26 muestran las matrices de confusión para cada uno de los resultados presentados en la Tabla 7.22. En estas tablas se puede observar que el problema de reconocimiento de la expresión de miedo sigue existiendo, puesto que aparece en los 4 casos de oclusión, ya que el sistema confunde esta expresión con la de felicidad. Por otra parte al observar la matriz de confusión correspondiente al problema de la oclusión de ojos-cejas podemos notar que las expresiones de ira y miedo son las que el sistema tiene más dificultades para reconocer. Mientras tanto, para el caso de la oclusión de la boca el sistema presenta un bajo rendimiento en la labor de reconocer las expresiones de miedo, ira y tristeza.

Tabla 7.23. Matriz de confusión para la combinación OBN–OFBN– Ri–Rd con imágenes en presencia de oclusión de la mitad izquierda del rostro.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	60.0	6.7	6.7	16.7	10.0	0.0
Asco	0.0	85.3	5.9	0.0	5.9	2.9
Miedo	0.0	0.0	83.0	8.5	4.3	4.3
Felicidad	0.0	0.0	4.3	94.3	1.4	0.0
Tristeza	0.0	0.0	1.9	3.7	83.3	11.1
Sorpresa	0.0	0.0	0.0	1.5	0.0	98.5

Tabla 7.24. Matriz de confusión para la combinación OBN–OFBN– Ri–Rd con imágenes en presencia de oclusión de la mitad derecha del rostro.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	70.0	3.3	6.7	13.3	6.7	0.0
Asco	0.0	91.2	0.0	2.9	2.9	2.9
Miedo	0.0	4.3	76.6	10.6	4.3	4.3
Felicidad	0.0	0.0	7.1	87.1	5.7	0.0
Tristeza	3.7	5.6	5.6	5.6	70.4	9.3
Sorpresa	0.0	0.0	3.1	0.0	0.0	96.9

Tabla 7.25. Matriz de confusión para la combinación B–FB–BN–FBN con imágenes en presencia de oclusión de los ojos-cejas.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	70.0	3.3	0.0	10.0	16.7	0.0
Asco	0.0	91.2	2.9	2.9	0.0	2.9
Miedo	0.0	2.1	66.0	23.4	6.4	2.1
Felicidad	0.0	1.4	2.9	95.7	0.0	0.0
Tristeza	1.9	0.0	3.7	0.0	94.4	0.0
Sorpresa	0.0	1.5	3.1	0.0	0.0	95.4

Tabla 7.26. Matriz de confusión para la combinación N–ON–FN–OFN con imágenes en presencia de oclusión de la boca.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	60.0	6.7	0.0	6.7	23.3	3.3
Asco	0.0	88.2	5.9	2.9	2.9	0.0
Miedo	0.0	4.3	29.8	34.0	19.1	12.8
Felicidad	0.0	0.0	0.0	95.7	1.4	2.9
Tristeza	3.7	1.9	3.7	5.6	70.4	14.8
Sorpresa	0.0	1.5	1.5	3.1	3.1	90.8

Con el objetivo de tener un punto de referencia con el cual comparar los resultados obtenidos por el método propuesto, se realizó una prueba utilizando los 4 sets de imágenes con oclusión para el método de FPBS. En esta prueba se utilizaron las imágenes con oclusión mostradas en la Figura 7.7. La Tabla 7.27 muestra los resultados de esta prueba. Donde se observa que en general los resultados son muy bajos, ninguno alcanza más del 50% de reconocimiento.

Tabla 7.27. Resultados del método FPBS para cada uno de los tipos de oclusión.

Tipo de oclusión	%
Mitad izquierda	33.67
Mitad derecha	50.00
Ojos	39.33
Boca	45.67

Las Tabla 7.28 - 7.25 muestran las matrices de confusión para cada uno de los resultados presentados en la Tabla 7.27. Donde se pueden observar más claramente las deficiencias que presentan estos resultados. Especialmente el obtenido con oclusión de la mitad izquierda del rostro, ya que dos de las expresiones faciales son confundidas en su totalidad, presentando 0% de reconocimiento para miedo y sorpresa. Los bajos resultados presentados en esta prueba, suponen que la oclusión simulada en este trabajo afecta demasiado al rendimiento del método FPBS, ya que se utilizan mascarar negras que podrían afectar al momento de la extracción de características y de la clasificación.

Tabla 7.28. Matriz de confusión del método FPBS con imágenes en presencia de oclusión de la mitad izquierda del rostro.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	90.0	0.0	0.0	3.3	6.7	0.0
Asco	76.5	5.9	0.0	11.8	5.9	0.0
Miedo	66.0	0.0	0.0	31.9	2.1	0.0
Felicidad	18.6	0.0	0.0	80.0	1.4	0.0
Tristeza	66.7	0.0	0.0	3.7	29.6	0.0
Sorpresa	75.4	0.0	0.0	15.4	9.2	0.0

Tabla 7.29. Matriz de confusión del método FPBS con imágenes en presencia de oclusión de la mitad derecha del rostro.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	0.0	60.0	10.0	3.3	6.7	20.0
Asco	0.0	70.6	5.9	0.0	2.9	20.6
Miedo	0.0	29.8	46.8	12.8	2.1	8.5
Felicidad	0.0	25.7	25.7	44.3	0.0	4.3
Tristeza	0.0	25.9	7.4	1.9	22.2	42.6
Sorpresa	0.0	3.1	3.1	0.0	0.0	93.8

Tabla 7.30. Matriz de confusión del método FPBS con imágenes en presencia de oclusión de los ojos-cejas.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	33.3	0.0	0.0	0.0	66.7	0.0
Asco	14.7	0.0	2.9	5.9	76.5	0.0
Miedo	12.8	0.0	46.8	8.5	31.9	0.0
Felicidad	14.3	0.0	22.9	44.3	18.6	0.0
Tristeza	9.3	0.0	0.0	0.0	90.7	0.0
Sorpresa	6.2	0.0	1.5	0.0	83.1	9.2

Tabla 7.31. Matriz de confusión del método FPBS con imágenes en presencia de oclusión de la boca.

(%)	Ira	Asco	Miedo	Felicidad	Tristeza	Sorpresa
Ira	6.7	3.3	40.0	10.0	0.0	40.0
Asco	0.0	55.9	8.8	5.9	0.0	29.4
Miedo	0.0	6.4	31.9	6.4	0.0	55.3
Felicidad	0.0	2.9	5.7	52.9	0.0	38.6
Tristeza	0.0	1.9	20.4	1.9	0.0	75.9
Sorpresa	0.0	0.0	1.5	0.0	0.0	98.5

La Figura 7.57 muestra la comparación de los resultados presentados en esta sección. Donde en todos los casos el método propuesto mejora los resultados del método de FPBS. La mejora más significativa se obtiene para la oclusión de la mitad izquierda del rostro, con cerca de 55% de aumento. Además, se puede observar que el problema de oclusión en los ojos-cejas se puede solucionar con mayor eficiencia que el presentado por la oclusión de la boca.

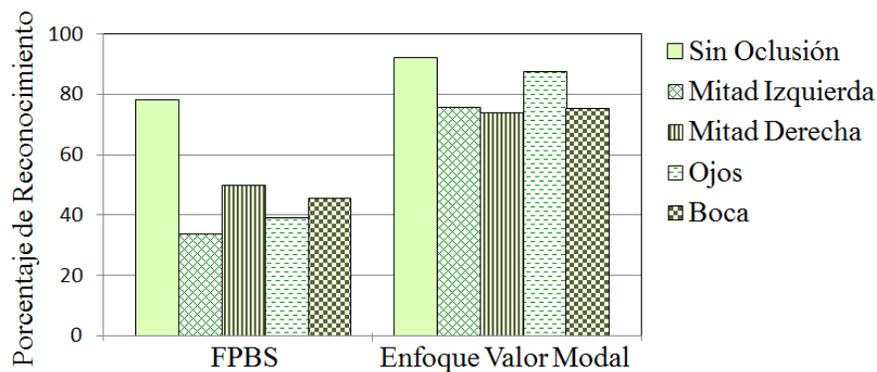


Figura 7.57. Comparación de los resultados del método propuesto con los de FPBS.

7.6 Conclusiones

En la primera parte de este capítulo se presentan las modificaciones realizadas a la base de datos AR la cual se utiliza para evaluar los sistemas propuestos de reconocimiento de rostros, y la base de datos Cohn-Kanade que se usa para evaluar el sistema propuesto de reconocimiento de expresiones faciales.

La primera prueba presentada en este capítulo está enfocada a encontrar el número óptimo de imágenes de entrenamiento para los sistemas de reconocimiento de rostros. La cual consiste en evaluar los métodos de Fases Propias convencional, Fases Propias con sub-bloques de 2×2 , LBPC(8,1) y WBP con diferente número de imágenes de entrenamiento, donde se concluye que con 7 imágenes de entrenamiento se obtienen resultados óptimos en general para los sistemas de reconocimiento de rostros analizados.

En base a los resultados de identificación y verificación podemos concluir que el mejor sistema presentado en este trabajo es el método propuesto de Fases Propias con sub-bloques de tamaño óptimo. Debido a que el Top 3 de resultados de identificación fueron los obtenidos por CLAHE(2×2), Normalización(2×2) y Fases Propias(2×2), alcanzando más del 99% de reconocimiento; y en cuanto a resultados de verificación estos métodos se encuentran dentro del Top 5, con porcentajes de error menores al 0.3%. Una particularidad de estos resultados que vale la pena mencionar, es referente al método DI-WBP el cual obtiene resultados apenas 1% menor que los alcanzados por el Top 5, presentando un rendimiento sobresaliente tomando en cuenta que es el método con menor costo computacional de los presentados en esta tesis.

Con respecto a los resultados del sistema de reconocimiento de expresiones faciales, se concluye que el uso del método de Fases Propias con sub-bloques de tamaño óptimo presenta un rendimiento notable cuando se utiliza una segmentación del rostro. Además, el uso del Enfoque del Valor Modal mejora los resultados obtenidos por sistemas de reconocimiento de expresiones faciales que utilicen más de un clasificador basados en distintas regiones faciales. Destacando que el rendimiento del sistema propuesto se mantiene estable aun con la presencia de oclusión parcial en estado crítico.

CAPÍTULO VIII

8. CONCLUSIONES GENERALES Y TRABAJO A FUTURO

En este capítulo se presentan las conclusiones generales a las que se llegaron luego de la culminación de este trabajo de investigación. Asimismo, se enlistan algunas de las posibles líneas de continuación concernientes a los temas de reconocimiento de rostros y reconocimiento de expresiones faciales, temas principales de este trabajo de tesis.

8.1 Conclusiones Generales

En este trabajo de tesis se analizaron diferentes técnicas para el desarrollo de sistemas de reconocimiento de rostros y de expresiones faciales. Las propuestas planteadas pueden clasificarse en 3 categorías: enfocadas al sistema de reconocimiento de rostros usando Fases Propias, enfocadas al sistema de reconocimiento de rostros usando LBP, y enfocadas al sistema de reconocimiento de expresiones faciales.

Referente al sistema de reconocimiento de rostros usando el método de Fases Propias, en base al análisis teórico del efecto de la utilización de sub-bloques para el método de Fases Propias, se concluye que el rendimiento del sistema aumenta cuando el tamaño de sub-bloque disminuye. Los resultados teóricos y prácticos demuestran que con el uso de un sub-bloque de tamaño de 2x2 píxeles se obtienen los mejores resultados, siendo este el tamaño de sub-bloque óptimo. De esta manera, al utilizar un sub-bloque de tamaño óptimo, el rendimiento del algoritmo de Fases Propias se vuelve casi independiente a los cambios de iluminación en las imágenes de entrada, así como a la presencia de expresiones faciales e incluso de la oclusión parcial.

Los resultados experimentales muestran que se obtiene un porcentaje de reconocimiento más alto y una desviación estándar más baja cuando el tamaño de sub-bloque va disminuyendo, como se estimaba de manera teórica. El método propuesto de Fases Propias con sub-bloques de 2x2 píxeles proporciona un porcentaje de reconocimiento de 99.6% mejorando en 3% al del método convencional. Mejora que persiste incluso en presencia de oclusión parcial, ya que el porcentaje de reconocimiento, con respecto a los resultados sin oclusión, solo se degrada en aproximadamente 2% en comparación con el 11% del método convencional.

Asimismo, en la evaluación de resultados para la fase de verificación, el método propuesto con el tamaño de sub-bloque de 2x2 píxeles muestra que es capaz de conseguir, al mismo tiempo, una falsa aceptación y un falso rechazo menor al 0.5%, mientras que para lograr una falsa aceptación inferior al 0.5% con el método convencional, el falso rechazo aumenta a 2.5%, y si se desea un falso rechazo inferior al 0.5 %, la falsa aceptación falsa se vuelve superior al 5%.

Por estas razones podemos concluir que el sistema propuesto denominado Fases propias con sub-bloques de tamaño óptimo supera al método convencional de Fases Propias tanto para la tarea de identificación como la de verificación.

Referente a los métodos propuestos para el reconocimiento de rostros usando el método de LBP, primeramente este trabajo presenta un análisis de los resultados usando los métodos convencionales de LBPh y LBPC, donde se demuestra que al usar directamente los coeficientes de LBP (LBPC) se logran mejores resultados que al utilizar los histogramas del mismo (LBPh). Además, se compara el rendimiento de estos métodos en su forma más sencilla (8 vecinos y radio igual 1), con variaciones más complejas (P,R igual a 8,2 y 8,3), demostrando que el uso de la forma LBP(8,1) proporciona buenos resultados en comparación a las más complejas, siendo esta una buena alternativa al tratar de reducir costos computacionales cuando se usa LBP.

Asimismo, la evaluación de resultados muestra que los métodos WBP y DI-WBP proporciona un porcentaje de reconocimiento bastante aceptable, cercano al proporcionado por LBPC(8,1). Sin embargo, en la mayoría de los casos el porcentaje de reconocimiento de DI-WBP es ligeramente menor que el de LBPC(8,1) y WBP, esto debido a que su complejidad de cálculo es significativamente menor, proporcionando un vector de características 81 veces más pequeño. Es importante mencionar que incluso con la presencia de oclusión parcial los algoritmos propuestos presentan un rendimiento estable, con porcentajes de reconocimiento de 95.4% y 90.7% para los métodos de WBP y DI-WBP respectivamente.

Referente al sistema de reconocimiento de expresiones faciales, este trabajo plantea la utilización del método de Fases Propias con sub-bloques de tamaño óptimo utilizando una segmentación del rostro en cuatro regiones faciales. De forma que varias combinaciones de estas regiones faciales son posibles, lo que resulta en la utilización de diferentes clasificadores, correspondientes a cada combinación. En base a esto, se propuso el Enfoque del Valor Modal, el cual cumple con el objetivo de unificar los resultados obtenidos a partir de diferentes los clasificadores. Basándose en los resultados experimentales se puede concluir que el uso de la segmentación facial mejora el porcentaje de reconocimiento en comparación a utilizar la imagen completa.

El mejor resultado obtenido en este enfoque fue proporcionado por la combinación OB–FB–OFBN–FPBS la cual utiliza el método del Enfoque del Valor Modal, alcanzando un porcentaje de reconocimiento del 92%. Asimismo, a partir de los resultados obtenidos por las combinaciones de las regiones faciales utilizando 1, 2, 3 y las 4 regiones, se puede concluir que la boca es la parte más importante del rostro y la frente la más irrelevante para el reconocimiento de expresiones faciales.

La Figura 8.1 ilustra la importancia de la boca para los sistemas de reconocimiento de expresiones faciales. Donde se presentan las regiones de frente y boca de un sujeto que se encuentra mostrando diferentes expresiones faciales, (a) felicidad, (b) sorpresa y (c) tristeza. De esta manera, podemos observar que los cambios entre las expresiones faciales son bastante significativos para la región de la boca, mientras que los presentados por la frente son prácticamente indistinguibles.

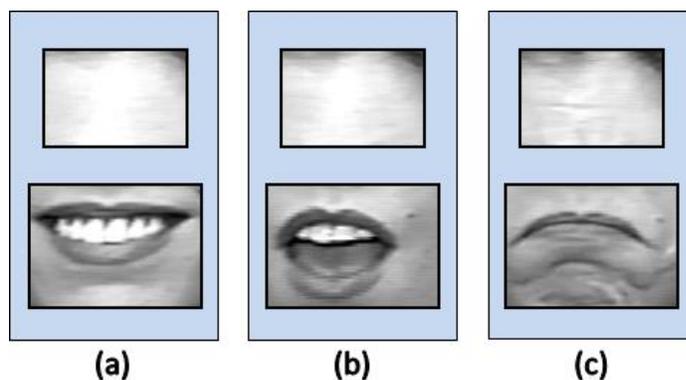


Figura 8.1. Ejemplo de la importancia de la boca y frente para poder reconocer una expresión facial. (a) Frente y boca mostrando la expresión de felicidad. (b) Frente y boca mostrando la expresión de sorpresa. (c) Frente y boca mostrando la expresión de tristeza.

Otra ventaja del método propuesto para FER, es que incluso con sólo una región del rostro es posible realizar el reconocimiento, alcanzando aproximadamente un 80% de reconocimiento cuando solo la región de la boca está disponible. Donde se puede concluir que con una sola región facial es posible realizar el reconocimiento de expresiones faciales. Este hecho toma vital importancia cuando varias regiones del rostro no se encuentran disponibles, en el caso de oclusión parcial. De esta forma también se puede concluir que el sistema propuesto resuelve de cierta manera el problema de la oclusión parcial en las imágenes de entrada.

De manera general se concluye que el objetivo principal de este trabajo de tesis ha sido satisfactoriamente alcanzado, ya que se consiguió desarrollar un par de sistemas para el reconocimiento de rostros (basado en fases propias y basado en LBP), además de un sistema de reconocimiento de expresiones, los cuales presentan un rendimiento sobresaliente aun en presencia de problemas como cambios de iluminación, tamaño de las imágenes y oclusión parcial.

8.2 Trabajo a Futuro

En base a los métodos analizados y a los sistemas propuestos en este trabajo de tesis, algunas de las posibles líneas de continuación concernientes a los temas de reconocimiento de rostros y reconocimiento de expresiones faciales se enlistan a continuación.

- Reconocimiento de rostros en condiciones no controladas incluyendo una etapa de detección de rostros.
- Segmentación automática de para sistemas de reconocimiento de expresiones faciales.
- Segmentación basada en el movimiento de los músculos del rostro y realizar el reconocimiento de expresiones faciales en secuencias de video.
- Modificación del Enfoque del Valor Moral, empleando lógica difusa, o los valores de coincidencia otorgados por los clasificadores.

REFERENCIAS

- [1] S. Y. Kung, M.-W. Mak, and S.-H. Lin, *Biometric Authentication: a machine learning approach*. New York: Prentice Hall Professional Technical Reference, 2005.
- [2] H. M. El-Bakry and N. Mastorakis, "Personal identification through biometric technology," in *9th WSEAS International Conference on Applied Informatics and Communications (AIC'09), Moscow, Russia, 2009*, pp. 325-340.
- [3] A. K. Jain, P. J. Flynn, and A. A. Ross, *Handbook of Biometrics*. US: Springer, 2008.
- [4] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. London: Springer, 2009.
- [5] O. Henniger, D. Muramatsu, T. Matsumoto, I. Yoshimura, and M. Yoshimura, "Signature Recognition," in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds., ed US: Springer, 2009, pp. 1196-1205.
- [6] A. Rosenberg, F. Bimbot, and S. Parthasarathy, "Overview of Speaker Recognition," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., ed Berlin: Springer, 2008, pp. 725-741.
- [7] M. Burge and K. Bowyer, *Handbook of Iris Recognition*. London: Springer, 2013.
- [8] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*: Springer, 2011.
- [9] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *JOSA A*, vol. 4, pp. 519-524, 1987.
- [10] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, pp. 71-86, 1991.
- [11] R. Chellappa, P. Sinha, and P. J. Phillips, "Face recognition by computers and humans," *Computer*, vol. 43, pp. 46-55, 2010.
- [12] (2008). *Facial Recognition Technology Safeguards Beijing Olympics* [Online]. Available: http://english.cas.ac.cn/Ne/CN/200909/t20090923_42959.shtml
- [13] (November, 2013), FaceGate. [Online]. Available: http://www.premierelect.co.uk/biometric-access-control-face-readers-c-31_139.html
- [14] (November, 2013), FaceKey. [Online]. Available: <http://www.facekey.com>
- [15] (November, 2013), Cognitec. [Online]. Available: <http://www.cognitec-systems.de>

- [16] A. Tsotsis. (2012). *Facebook Scoops Up Face.com For \$55-60M To Bolster Its Facial Recognition Tech* [Online]. Available: <http://techcrunch.com/2012/06/18/facebook-scoops-up-face-com-for-100m-to-bolster-its-facial-recognition-tech>
- [17] Y. Gao and M. K. Leung, "Face recognition using line edge map," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 764-779, 2002.
- [18] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Facial Expression Recognition," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds., ed: Springer, 2011, pp. 487-519.
- [19] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 1424-1445, 2000.
- [20] A. Mehrabian, "Communication Without Words," *Psychology Today*, vol. 2, pp. 53-56, 1968.
- [21] P. Ekman and W. V. Firesen, *Emotion in the Human Face* New Jersey: Prentice Hall, 1975.
- [22] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial Expression Analysis," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds., ed: Springer, 2005, pp. 247-275.
- [23] B. Fasel and J. Luetin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, pp. 259-275, 2003.
- [24] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," *Pattern Recognition*, vol. 45, pp. 80-91, 2012.
- [25] S. Guo and Q. Ruan, "Facial expression recognition using local binary covariance matrices," in *Wireless, Mobile & Multimedia Networks (ICWMMN 2011), 4th IET International Conference on*, 2011, pp. 237-242.
- [26] H. Towner and M. Slater, "Reconstruction and recognition of occluded facial expressions using PCA," in *Affective Computing and Intelligent Interaction*, ed: Springer, 2007, pp. 36-47.
- [27] I. Kotsia, I. Buciu, and I. Pitas, "An analysis of facial expression recognition under partial facial image occlusion," *Image and Vision Computing*, vol. 26, pp. 1052-1067, 2008.
- [28] L. Zhang, D. Tjondronegoro, and V. Chandran, "Toward a more robust facial expression recognition in occluded images using randomly sampled Gabor based templates," in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, 2011, pp. 1-6.
- [29] Y. Miyakoshi and S. Kato, "Facial emotion detection considering partial occlusion of face using Bayesian network," in *Computers & Informatics (ISCI), 2011 IEEE Symposium on*, 2011, pp. 96-101.
- [30] C. Darwin, *The Expression of The Emotions in Man and Animals* vol. 526. London (1872): University of Chicago Press, 1965.
- [31] J. Ruiz-del-Solar and J. Quinteros, "Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of

- different pre-processing approaches," *Pattern Recognition Letters*, vol. 29, pp. 1966-1979, 2008.
- [32] J. Olivares-Mercado, K. Hotta, H. Takahashi, H. Perez-Meana, and G. Sanchez-Perez, "Face recognition based on the phase spectrum of local normalized image," in *Artificial Intelligence, 2008. MICAI'08. Seventh Mexican International Conference on*, 2008, pp. 123-127.
- [33] J. Olivares-Mercado, K. Hotta, H. Takahashi, M. Nakano-Miyatake, K. Toscano-Medina, and H. Perez-Meana, "Improving the eigenphase method for face recognition," *IEICE Electronics Express*, vol. 6, pp. 1112-1117, 2009.
- [34] V. Štruc and N. Pavešić, "Image normalization techniques for robust face recognition," in *Proceedings of the 8th WSEAS international conference on Signal processing, robotics and automation*, 2009, pp. 155-160.
- [35] K. Ramirez-Gutierrez, D. Cruz-Perez, J. Olivares-Mercado, M. Nakano-Miyatake, and H. Perez-Meana, "A face recognition algorithm using eigenphases and histogram equalization," *International Journal of Computers*, vol. 5, pp. 34-41, 2011.
- [36] P. M. Corcoran and C. Iancu, "Automatic Face Recognition System for Hidden Markov Model Techniques," *Face Recognition Volume 2, Intech Publishing*, 2011.
- [37] N. D. Kalka, T. Bourlai, B. Cukic, and L. Hornak, "Cross-spectral face recognition in heterogeneous environments: A case study on matching visible to short-wave infrared imagery," in *Biometrics (IJCB), 2011 International Joint Conference on*, 2011, pp. 1-8.
- [38] Y. Luo, C.-m. Wu, and Y. Zhang, "Facial expression recognition based on fusion feature of PCA and LBP with SVM," *Optik-International Journal for Light and Electron Optics*, 2012.
- [39] Z. Li, J.-i. Imai, and M. Kaneko, "Facial expression recognition using facial-component-based bag of words and PHOG descriptors," *Information and Media Technologies*, vol. 5, pp. 1003-1009, 2010.
- [40] Z. Li, J.-i. Imai, and M. Kaneko, "Facial-component-based bag of words and PHOG descriptor for facial expression recognition," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, 2009, pp. 1353-1358.
- [41] R. C. Gonzalez and R. E. Woods, in *Digital Image Processing*, Third ed: Pearson Prentice Hall, 2008, pp. 120-128.
- [42] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, pp. 355-368, 1987.
- [43] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," in *Graphics Gems: IV*. vol. 4, P. S. Heckbert, Ed., ed: Morgan Kaufmann, 1994.
- [44] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, pp. 35-44, 2004.

- [45] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, pp. 193-200, 1998.
- [46] Z. Xu, X. Liu, and N. Ji, "Fog removal from color images using contrast limited adaptive histogram equalization," in *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*, 2009, pp. 1-5.
- [47] R. Gottumukkal and V. K. Asari, "An improved face recognition technique based on modular PCA approach," *Pattern Recognition Letters*, vol. 25, pp. 429-436, 2004.
- [48] M. Savvides, B. V. Kumar, and P. K. Khosla, "Eigenphases vs eigenfaces," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, pp. oo810-oo813 Vol. 3.
- [49] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, pp. 711-720, 1997.
- [50] M. Koc and A. Barkana, "A new solution to one sample problem in face recognition using FLDA," *Applied Mathematics and Computation*, vol. 217, pp. 10368-10376, 2011.
- [51] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, pp. 328-340, 2005.
- [52] P. Mali, B. Chaudhuri, and D. Dutta Majumder, "Properties and some fast algorithms of the Haar transform in image processing and pattern recognition," *Pattern Recognition Letters*, vol. 2, pp. 319-327, 1984.
- [53] M. Sharkas, "Application of DCT blocks with principal component analysis for face recognition," in *5th WSEAS International Conference on Signal, Speech and Image Processing*, 2005, pp. 107-111.
- [54] S. Dabbaghchian, M. P. Ghaemmaghami, and A. Aghagolzadeh, "Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology," *Pattern Recognition*, vol. 43, pp. 1431-1440, 2010.
- [55] P. Lee and G.-S. Liu, "An efficient algorithm for the 2-D discrete cosine transform," *Signal processing*, vol. 55, pp. 221-239, 1996.
- [56] J. Olivares-Mercado, G. Sanchez-Perez, M. Nakano-Miyatake, and H. Perez-Meana, "Feature extraction and face verification using Gabor and Gaussian mixture models," in *MICAI 2007: Advances in Artificial Intelligence*, ed: Springer, 2007, pp. 769-778.
- [57] G. Aguilar-Torres, K. Toscano-Medina, G. Sanchez-Perez, M. Nakano-Miyatake, and H. Perez-Meana, "Eigenface-Gabor algorithm for feature extraction in face recognition," *International Journal of Computers*, vol. 3, pp. 20-30, 2009.

-
-
- [58] L.-L. Huang, A. Shimizu, and H. Kobatake, "Robust face detection using Gabor filter features," *Pattern Recognition Letters*, vol. 26, pp. 1641-1649, 2005.
- [59] R. Thiyagarajan, S. Arulselvi, and G. Sainarayanan, "Gabor feature based classification using statistical models for face recognition," *Procedia Computer Science*, vol. 2, pp. 83-93, 2010.
- [60] H. Qin, L. Qin, L. Xue, and C. Yu, "Gabor-based weighted region covariance matrix for face recognition," *Electronics letters*, vol. 48, pp. 992-993, 2012.
- [61] H. Hu, "Variable lighting face recognition using discrete wavelet transform," *Pattern Recognition Letters*, vol. 32, pp. 1526-1534, 2011.
- [62] D.-Q. Dai and H. Yan, "Wavelets and face recognition," *Face recognition*, pp. 59-74, 2007.
- [63] A. Eleyan, H. Özkaramanli, and H. Demirel, "Complex wavelet transform-based face recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 195, 2008.
- [64] K. Delac, M. Grgic, and S. Grgic, "Face recognition in JPEG and JPEG2000 compressed domain," *Image and Vision Computing*, vol. 27, pp. 1108-1120, 2009.
- [65] S. Ribarić and M. Maračić, "Eigenphase-based face recognition: a comparison of phase-information extraction methods," in *Nineteenth International Electrotechnical and Computing Science Conference ERK 2010*, 2010.
- [66] G. Aguilar, J. Olivares, G. Sánchez, H. Pérez, and E. Escamilla, "Face Recognition Using Frequency Domain Feature Extraction Methods," *Face Analysis, Modeling and Recognition Systems*, p. 35, 2011.
- [67] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, 1994, pp. 582-585.
- [68] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer Vision-ECCV 2004*, ed: Springer, 2004, pp. 469-481.
- [69] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, pp. 2037-2041, 2006.
- [70] W. Xia, S. Yin, and P. Ouyang, "A High Precision Feature Based on LBP and Gabor Theory for Face Recognition," *Sensors*, vol. 13, pp. 4499-4513, 2013.
- [71] D. Maturana, D. Mery, and A. Soto, "Face recognition with decision tree-based local binary patterns," in *Computer Vision-ACCV 2010*, ed: Springer, 2011, pp. 618-629.
- [72] B. Yang and S. Chen, "A Comparative Study on Local Binary Pattern (LBP) based Face Recognition: LBP Histogram versus LBP Image," *Neurocomputing*, 2013.
- [73] C. Kang, S. Liao, S. Xiang, and C. Pan, "Kernel sparse representation with local patterns for face recognition," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 3009-3012.

- [74] M. Hayes, J. Lim, and A. Oppenheim, "Signal reconstruction from phase or magnitude," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, pp. 672-680, 1980.
- [75] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, pp. 529-541, 1981.
- [76] J. Shlens, "A Tutorial on Principal Component Analysis Center for Neural Science," *New York University New York City, NY*, pp. 10003-6603.
- [77] Y. X. Zhang, "Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis," *Talanta*, vol. 73, pp. 68-75, 2007.
- [78] T. Howley, M. G. Madden, M.-L. O'Connell, and A. G. Ryder, "The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data," *Knowledge-Based Systems*, vol. 19, pp. 363-370, 2006.
- [79] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowledge-Based Systems*, vol. 24, pp. 1024-1032, 2011.
- [80] G.-C. Luh and C.-Y. Lin, "PCA based immune networks for human face recognition," *Applied Soft Computing*, vol. 11, pp. 1743-1752, 2011.
- [81] X. Zhao and S. Zhang, "Facial expression recognition based on local binary patterns and kernel discriminant isomap," *Sensors*, vol. 11, pp. 9573-9588, 2011.
- [82] S. Zhang, X. Zhao, and B. Lei, "Robust facial expression recognition via compressive sensing," *Sensors*, vol. 12, pp. 3747-3761, 2012.
- [83] R. Rojas, *Neural Networks: A Systematic Introduction*: Springer, 1996.
- [84] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *Speech and Audio Processing, IEEE Transactions on*, vol. 3, pp. 72-83, 1995.
- [85] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in neural information processing systems*, pp. 281-287, 1997.
- [86] J. Olivares-Mercado, G. Aguilar-Torres, K. Toscano-Medina, M. Nakano-Miyatake, and H. Perez-Meana, "GMM vs SVM for Face Recognition and Face Verification."
- [87] J. Ren, "ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging," *Knowledge-Based Systems*, vol. 26, pp. 144-153, 2012.
- [88] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, and P. A. Torres-Carrasquillo, "Support vector machines for speaker and language recognition," *Computer Speech & Language*, vol. 20, pp. 210-229, 2006.
- [89] G. Costantini, D. Casali, and T. Massimiliano, "An SVM based classification method for EEG signals," in *Proceedings of the 14th WSEAS international conference on Circuits*, 2010, pp. 107-109.

- [90] W. Li, L. Liu, and W. Gong, "Multi-objective uniform design as a SVM model selection tool for face recognition," *Expert Systems with Applications*, vol. 38, pp. 6689-6695, 2011.
- [91] J. J. Rodríguez, C. J. Alonso, and J. A. Maestro, "Support vector machines of interval-based features for time series classification," *Knowledge-Based Systems*, vol. 18, pp. 171-178, 2005.
- [92] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *The Journal of Machine Learning Research*, vol. 5, pp. 101-141, 2004.
- [93] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, p. 27, 2011.
- [94] D. Vukadinovic and M. Pantic, "Fully automatic facial feature point detection using Gabor feature based boosted classifiers," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, 2005, pp. 1692-1698.
- [95] A. M. Martinez, "The AR face database," *CVC Technical Report*, vol. 24, 1998.
- [96] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 46-53.
- [97] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, pp. 137-154, 2004.

ANEXO

PUBLICACIONES

Revistas indexadas JCR:

- “A sub-block-based eigenphases algorithm with optimum sub-block size,” G. Benitez-Garcia, J. Olivares-Mercado, G. Sanchez-Perez, M. Nakano-Miyatake, and H. Perez-Meana. *Knowledge-Based Systems*, vol. 37, pp. 415-426, 2012.
- “New Window Binary Pattern Approach Based on LBP for Face Recognition,” G. Benitez-Garcia, J. Olivares-Mercado, G. Sanchez-Perez, M. Nakano-Miyatake, and H. Perez-Meana. *Sensor*. (aceptado con revisiones menores).
- “Facial Expression Recognition Based on Facial Region Segmentation and Modal Value Approach,” G. Benitez-Garcia, G. Sanchez-Perez, H. Perez-Meana, K. Takahashi and M. Kaneko. *IEICE Transactions on Fundamentals*. (aceptado para su publicación).

Revistas indexadas:

- “HALFTONING: REVIEW AND ANALYSIS,” F. Pelcastre, L. Peregrina, M. García, A. Ochoa, U. Juárez, E. Ríos, G. Benítez, and M. Nakano. *Revista Ingenierías Universidad de Medellín*, vol. 11, pp. 161-174, 2012.

Congresos internacionales:

- "Face Identification Based on Contrast Limited Adaptive Histogram Equalization (CLAHE)," G. Benitez-Garcia, J. Olivares-Mercado, G. Aguilar-Torres, G. Sanchez-Perez, and H. Perez-Meana. in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'11), WORLDCOMP*, vol. 11, pp. 363-369, 2011.
- "Comparison of Facial Expression Recognition Rates Depending on Combination of Different Facial Regions," G. Benitez-Garcia, G. Sanchez-Perez, H. Perez-Meana, K. Takahashi, and M. Kaneko. *2012年映像情報メディア学会冬季大会*, December 18-19, 2012.
- "Facial Expression Recognition Under Partial Occlusion Based on Facial Region Segmentation," G. Benitez-Garcia, G. Sanchez-Perez, H. Perez-Meana, K. Takahashi, and M. Kaneko. *映像情報メディア学会技術報告*, vol. 37, pp. 95-98, 2013.

Congresos institucionales:

- "Facial Expression Recognition under Partial Occlusion using Sub-Block Eigenphases," G. Benitez-Garcia and M. Kaneko. *Proceedings of the UEC International Mini-Conference for Exchange Students on Electro-Communications and Information Science No.28*, Abstract p.90, Poster p.91, August 9-10, 2012.
- "Facial Expression Recognition under Partial Occlusion Based on Modal Value Approach," G. Benitez-Garcia, G. Sanchez-Perez, H. Perez-Meana, K. Takahashi, and M. Kaneko. *Proceedings of the UEC International Mini-Conference for Exchange Students on Informatics & Engineering and Information Systems No.29*, pp. 55-62, March 5-6, 2013.