

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE CIENCIAS
MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA
MAESTRÍA EN CIENCIAS



**RECONOCIMIENTO Y SEGUIMIENTO FACIAL BASADO EN
CORRELACIÓN USANDO FILTROS COMPUESTOS NO LINEALES Y
NO RESTRINGIDOS**

TESIS PROFESIONAL
QUE PARA CUBRIR PARCIALMENTE LOS REQUISITOS PARA OBTENER
EL GRADO DE MAESTRO EN CIENCIAS

PRESENTA:
EVERARDO SANTIAGO RAMÍREZ

DIRECTOR DE TESIS:
DR. JOSÉ ÁNGEL GONZÁLEZ FRAGA

ENSENADA, BAJA CALIFORNIA, MÉXICO, ENERO DEL 2012.



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE CIENCIAS
Maestría y Doctorado en Ciencias e Ingeniería

ACTA DE REVISIÓN DE TESIS

TESIS DE GRADO

**RECONOCIMIENTO Y SEGUIMIENTO FACIAL BASADO EN
CORRELACIÓN USANDO FILTROS COMPUESTOS NO
LINEALES Y NO RESTRINGIDOS**

PRESENTADA POR

Everardo Santiago Ramírez

Y APROBADA POR EL SIGUIENTE COMITÉ

Dr. José Ángel González Fraga

Director del Comité

Dr. José Ignacio Ascencio López

Miembro del Comité

Dr. Juan Iván Nieto Hipólito

Miembro del Comité

RESUMEN de la tesis de Everardo Santiago Ramírez, presentada como requisito parcial para la obtención del grado de MAESTRO EN CIENCIAS. Ensenada, Baja California, México. Enero del 2012.

**RECONOCIMIENTO Y SEGUIMIENTO FACIAL BASADO EN CORRELACIÓN
USANDO FILTROS COMPUESTOS NO LINEALES Y NO RESTRINGIDOS**

Resumen aprobado por:



Dr. José Ángel González Fraga
Director de tesis

Reconocimiento facial, entendido como una tarea desafiante debido a los cambios que presenta la cara por la edad, expresiones, sombras, iluminación intensa, detalles faciales, oclusión, entre otros. Los actuales algoritmos de reconocimiento facial trabajan bajo entornos y personas controlables. Esta tesis aborda el problema de reconocimiento y seguimiento facial en donde no es posible tener control sobre la persona o el entorno.

Como resultado del análisis de filtros de correlación en la tarea de reconocimiento facial (verificación e identificación) con imágenes de prueba con variaciones en expresiones, sombras, iluminación intensa, detalles faciales, ligeros movimientos (pose y rotación) y ruido, así como la influencia de una operación de pre-procesamiento para la mejora de la calidad de las imágenes distorsionadas, se desarrollaron algoritmos de reconocimiento facial que emplean un filtro compuesto no lineal sin restricciones. Dicho filtro se construye con regiones faciales, alcanzando una efectividad aproximado de 95% y una alta capacidad de discriminación. El filtro trabaja bien con imágenes faciales parciales y cuando las imágenes contienen variaciones en sombras.

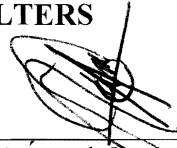
También, como consecuencia del análisis anterior, se desarrolló una estrategia de seguimiento facial basado en un filtro compuesto no lineal, sin restricciones y adaptativo, capaz de adecuarse a cambios ligeros que sufre una cara de una imagen a otra.

Para validar los algoritmos propuestos, se desarrolló un sistema de identificación y seguimiento facial. La base de datos (galería) usada contiene 39 plantillas biométricas de igual número de personas. El sistema fue probado con una secuencia de video en donde aparece la persona correspondiente al registro número 39 de la galería. El sistema identificó y localizó correctamente la cara de la personas en la mayoría de las imágenes de la secuencia de video.

ABSTRACT of the thesis of Everardo Santiago Ramírez, presented as partial requirement for obtaining the degree of Master of Science. Ensenada, Baja California, México. January, 2012.

**CORRELATION-BASED FACE RECOGNITION AND TRACKING USING
UNCONSTRAINED NONLINEAR COMPOSITE FILTERS**

Abstract approved by:



Dr. José Ángel González Fraga
Thesis advisor

Facial recognition, understood as a challenging task due to the changes introduced by the age, expressions, shadows, bright lighting, facial details, occlusion, among others. The current face recognition algorithms work under controlled environments and people. This thesis addresses the problem of facial recognition and tracking where it is not possible to have control over the person or the environment.

As a result of the analysis of correlation filters in the Face recognition task (verification and identification) with test images with variations in expressions, shadows, bright lighting, facial details, slight movements (pose and rotation) and noise, and the influence of a pre-processing operation to improve the quality of distorted images, face recognition algorithms using a unconstrained nonlinear composite filter were developed. The filter is constructed with facial regions, reaching an approximately 95% efficiency and a high capacity of discrimination. The filter works well with partial facial images and when the images contain variations in shadows.

Also, as result of the above analysis, a face tracking strategy based on an adaptive and unconstrained nonlinear composite filter was developed, able to adapt to slight changes that undergoes a face from one image to another.

To validate the proposed algorithms, a face identification and tracking system was developed. The database (gallery) contains 39 biometric templates, equal number of people. The system was tested with a video sequence where the person that appears in it, correspond to the record number 39 in the gallery. The system correctly identified and located the face of the people in most of the images from the video sequence.

Dedicatorias

A mis padres Lucia F. Ramírez Morales y Marciano Santiago López, los mentores de mi persona. A mis hermanas Victoria, Alejandra, Maura, Elizabeth, Edith Leticia y Siri Jazmín. A mis hermanos Javier y Héctor. Ustedes son la razón de todos mis sueños.

Agradecimiento

A dios, por darme la vida y permitirme hacer lo que me gusta. Por darme la fuerza para soportar el trabajo diario, por darme la sabiduría para resolver los problemas, pero sobre todo por ser la luz que ilumina mi sendero.

A mi director Dr. José Ángel González Fraga. Gracias por ser el pilar de apoyo para mis estudios. Cada vez que lo buscaba para presentar algún avance o con dudas, siempre salía con ideas nuevas y claras. Gracias sobre todo por la disponibilidad, paciencia, generosidad y por apoyarme en la divulgación de los resultados. Usted no sólo es un gran director, sino un gran amigo también.

A mis sinodales Dr. José Ignacio Ascencio López y Dr. Juan Iván Nieto Hipólito, por su apoyo en la revisión de este trabajo de tesis. Sus comentarios y observaciones fueron de gran utilidad para el enriquecimiento de este trabajo.

A mis padres, por enseñarme a que a base de trabajo, esfuerzo y respeto al prójimo se logra una vida mejor. Gracias madre por enseñarme el valor de la humildad, a ti padre por enseñarme el valor del trabajo.

A mis hermanas y hermanos por ser también mis amigos, gracias por soportarme y apoyarme. Victoria, Alejandra y Maura, gracias por su apoyo incondicional.

A mis sobrinos Ilich y Antonio, esos momentos de carcajadas fueron la forma más divertida de eliminar el estrés.

A mis amigos y compañeros de posgrado. Dolores, gracias por tus comentarios y apoyo. No solo logré una tesis si no una gran amistad. Francisco por los comentarios y apoyo en los ensayos para la presentación de avances. Alfredo, Claudia, y otros que por cuestiones de espacio no menciono pero que siempre recordaré.

A mis maestros, por compartir sus conocimientos y experiencias. A mi alma mater por las facilidades brindadas en la realización de mis estudios.

A CONACyT por la beca proporcionada para mis estudios de maestría.

Índice de contenido

1. Introducción	1
1.1 Justificación de la investigación	2
1.2 Planteamiento del problema	3
1.3 Objetivos	5
1.3.1 Objetivo general	5
1.3.2 Objetivos específicos	5
1.4 Organización de la tesis	6
2. Filtros de correlación	7
2.1 Filtro de acoplamiento	9
2.2 Funciones sintéticas discriminantes generalizadas	10
2.3 Filtro compuesto no lineal Ley k	12
2.4 Filtro del mínimo promedio de la energía de correlación	13
2.5 Filtro SDF de compromiso óptimo	14
2.6 Filtros de correlación no restringidos	15
2.6.1 Filtro MACE no restringido	16
2.6.2 OTSDF no restringido	16
2.7 Filtro del promedio de filtros sintéticos exactos	16
2.8 Filtro de salida mínima de la suma del error cuadrático	18
2.9 Medidas de desempeño para filtros de correlación	19
3. Introducción al procesamiento facial	22
3.1 Reconocimiento facial	23
3.1.1 Historia del reconocimiento facial	26
3.1.2 Principales enfoques	29
3.1.3 Reconocimiento facial usando filtros de correlación	32

3.1.4 Medidas de desempeño para algoritmos de reconocimiento facial	34
3.2 Seguimiento facial	36
3.2.1 Seguimiento de caras humanas mediante búsqueda logarítmica en rejilla	37
3.2.2 Seguimiento facial de objetos usando filtros de correlación adaptativos	37
4. Verificación facial usando filtros de correlación e imágenes faciales de prueba con distorsiones	
40	
4.1 Configuración del experimento	40
4.1.1 Umbral PSR de reconocimiento.	42
4.2 Resultados experimentales	43
4.2.1 Influencia de las expresiones faciales y variaciones en iluminación en la verificación facial	44
4.2.2 Influencia del ruido blanco Gaussiano aditivo en la verificación facial.....	57
4.2.3 Influencia de sombras en imágenes faciales	59
4.2.4 Desempeño de los filtros de correlación en verificación facial.....	61
5. Influencia del pre-procesado de las imágenes faciales con distorsiones en la verificación facial.....	64
5.1 Transformación logarítmica sobre imágenes faciales con variaciones en iluminación.....	65
5.2 Normalización de imágenes faciales.....	67
5.3 Ventana de coseno	68
5.4 Desempeño en la verificación facial de los filtros de correlación al usar operaciones de pre-procesamiento.....	69
6. Identificación facial usando filtros de correlación e imágenes con variaciones en sombras.....	75
6.1 Configuración del experimento	75
6.2 Resultados experimentales	76
7. Reconocimiento facial mediante un filtro compuesto no lineal sin restricciones construido con regiones faciales	83

7.1 Filtro compuesto no lineal sin restricciones	84
7.2 Regiones faciales	86
7.3 Reconocimiento facial mediante FCnLsR	88
7.4 Evaluación del método propuesto.....	92
8. Seguimiento facial mediante un filtro de compuesto no lineal, sin restricciones y adaptativo	101
8.1 Filtro compuesto no lineal, sin restricciones y adaptativo	101
8.2 Estrategia de seguimiento facial basado en correlación usando el filtro FCnLsR	102
8.3 Evaluación.....	105
8.3.1 Comparación con otro seguidor basado en filtros de correlación	108
9. Discusión de resultados y conclusiones.....	115
Bibliografía	119
Apéndice A. Sistema de identificación y seguimiento facial	124

Índice de figuras

Figura 2.1. Diseño de un filtro compuesto usando imágenes de entrenamiento.....	8
Figura 2.2. Salida 2D de correlación	10
Figura 2.3. Matriz que contiene las imágenes de entrenamiento en forma de vector columna.	11
Figura 2.4. Construcción de un filtro ASEF promediando filtros sintéticos exactos.	17
Figura 2.5. Plano de correlación y regiones indispensables para la obtención del PSR	20
Figura 2.6. Región de rechazo y región objetivo sobre un plano de correlación para la obtención de DC.....	21
Figura 3.1. Distorsiones faciales.	23
Figura 3.2. Cara humana.....	24
Figura 3.3. Proceso de verificación facial.....	25
Figura 3.4. Proceso de identificación facial.	26
Figura 3.5. Representación 3D de imágenes faciales: a) Nube de puntos en el espacio, b) mapa de profundidad, c) mapa de puntos más textura.	27
Figura 3.6. Diagrama de bloques del proceso de correlación en el reconocimiento facial.	33
Figura 3.7. Esquema general del proceso de seguimiento facial mediante búsqueda logarítmica. Tomado de (Vicente et al., 2002).....	37
Figura 3.8. Desempeño de MOSSE ante distorsiones faciales. A, B, E) Oclusión, C) Expresión, D) Pose y F) Pose y expresión. Tomado de (Bolme, y otros 2010).	39
Figura 4.1. Metodología para la evaluación de filtros de correlación.	42
Figura 4.2. Salida de correlación: A) Persona auténtica y B) Persona impostora.	44
Figura 4.3. Misma persona, diferentes expresiones faciales.....	45
Figura 4.4. Desempeño PSR del filtro K-Law para las 40 personas en la base de datos ORL usando el conjunto de prueba CP1.	46
Figura 4.5. Peor desempeño PSR del filtro K-Law con el conjunto de prueba CP1.....	47
Figura 4.6. Mejor desempeño PSR del filtro K-Law con el conjunto de prueba CP1.	48
Figura 4.7. Desempeño PSR del filtro MACE para las 40 personas de la base de datos ORL usando el conjunto de prueba CP1.	49
Figura 4.8. Peor desempeño PSR del filtro MACE (persona 36).	50

Figura 4.9. Mejor desempeño PSR del filtro MACE (persona 27).	50
Figura 4.10. Desempeño PSR del filtro UMACE para 40 personas y el conjunto de prueba CP1.	51
Figura 4.11. Peor desempeño PSR del filtro UMACE.	52
Figura 4.12. Mejor desempeño PSR del filtro UMACE.	53
Figura 4.13. Peor desempeño PSR del filtro OTSDF.	54
Figura 4.14. Mejor desempeño PSR del filtro OTSDF.	55
Figura 4.15. Peor desempeño PSR del filtro UOTSDF.	56
Figura 4.16. Mejor desempeño PSR del filtro UOTSDF.	56
Figura 4.17. A) Imagen facial, B) Ruido Gaussiano aditivo, C) Imagen facial con ruido blanco Gaussiano aditivo.	57
Figura 4.18. Correlación con una imagen con ruido Gaussiano aditivo y una imagen libre de ruido.	58
Figura 4.19. Misma imagen facial, diferentes varianzas en ruido; 5, 10, 15 y 20 pixeles, respectivamente.	58
Figura 4.20. Valores PSR de imágenes faciales de prueba con y sin ruido.	59
Figura 4.21. Diferentes imágenes faciales de la misma persona en diferentes condiciones de iluminación.	60
Figura 4.22. Valores PSR de imágenes faciales con y sin sombras usando el filtro K-Law.	61
Figura 5.1. Mapeado de intensidades de pixeles usando la transformación logarítmica base 10. Pixeles de intensidad 0 representan el negro y 255 el blanco.	65
Figura 5.2. Histograma de imágenes faciales sombreadas parcial o totalmente.	66
Figura 5.3. Resultado de la aplicación de la transformación logarítmica base 10 a las imágenes con regiones sombreadas.	67
Figura 5.4. Multiplicación de una imagen facial por una ventana de cosenos.	68
Figura 5.5. Desempeño de los filtros de correlación con operaciones de preprocesamiento sobre imágenes con variaciones en iluminación.	71
Figura 5.6. Desempeño de los filtros de correlación con el conjunto de prueba CP3, con y sin preprocesamiento.	73
Figura 5.7. Mejora en el desempeño del filtro K-Law con imágenes ruidosas del conjunto CP4, CP5 y CP6 con la operación de preprocesamiento D.	74

Figura 6.1. Imágenes faciales distintas de la misma persona.....	76
Figura 6.2. Desempeño PSR de un filtro construido con una sola imagen y correlacionada con 64 imágenes de la misma clase.	77
Figura 6.3. Muestras faciales de la misma clase para construir un filtro correlación.	78
Figura 6.4. Desempeño PSR del filtro construido con imágenes faciales sin variaciones en iluminación.....	78
Figura 6.5. Desempeño PSR del filtro construido con muestra que presenta una distorsion general.....	79
Figura 6.6. Desempeño PSR del filtro de correlación construido con dos muestras faciales distorsionadas.....	80
Figura 6.7. Desempeño del filtro construido con la estrategia 2 con tres imágenes faciales con distorsiones.....	80
Figura 6.8. Desempeño PSR del filtro de correlación construido con cuatro imágenes con variaciones en iluminación.	81
Figura 6.9. Muestras faciales con variaciones en iluminación para la construcción de un filtro. .	81
Figura 7.1. Principales regiones de una cara humana.	87
Figura 7.2. Esquema general del método de reconocimiento facial propuesto.	90
Figura 7.3. Salida de correlación del filtro FCnLsR para una imagen facial auténtica.	91
Figura 7.4. A) Imágenes parciales de prueba. B) Desempeño PSR del filtro UNCF con regiones faciales de prueba. Línea continua: filtro construido con regiones. Línea segmentada: filtro construido con imágenes completas.	93
Figura 7.5. Imágenes de entrenamiento. A) Sombra perfil derecho, B) sombra perfil izquierdo, C) sombra total y D) luz ambiental, sin sombras.	94
Figura 7.6. Desempeño PSR del filtro FCnLsR de la persona 1.....	95
Figura 7.7. Desempeño PSR del filtro de la persona 1 construido con 20 regiones faciales obtenido de cuatro imágenes faciales.	96
Figura 7.8. Mejor desempeño PSR del filtro FCnLsR.	98
Figura 7.9. Mejor desempeño PSR del filtro UMACE usando imágenes completas de entrenamiento.....	99
Figura 7.10. Mejor desempeño PSR del filtro UMACE usando imágenes completas de entrenamiento.....	99

Figura 8.1. Esquema general de la estrategia de seguimiento facial.	104
Figura 8.2. Inicialización del filtro de correlación para el seguimiento facial.	106
Figura 8.3. Desempeño del filtro compuesto no lineal, sin restricciones y adaptativo en imágenes donde caras presenta expresiones faciales.	107
Figura 8.4. Desempeño del filtro compuesto no lineal, sin restricciones y adaptativo en imágenes donde la cara presenta movimientos ligeros.	108
Figura 8.5. Salidas de localización de un seguidor facial.	109
Figura 8.6. Desempeño PSR de los filtros de correlación FCnLsRA y MOSSE en la secuencia de video Ever02.	111
Figura 8.7. Distancia entre la localización exacta y el pico de correlación producido por por los filtros FCnLsRA y MOSSE.	112
Figura 8.8. Muestras de las secuencias davidin300 (primera fila), dudek (segunda fila) y trellis70 (tercera fila).	113
Figura A.1. Esquema general del sistema de identificación y seguimiento facial.	124
Figura A.2. Selección del área en la imagen 1 de la secuencia de video la cara de la persona a observar.	125
Figura A.3. Persona en la secuencia de video identificada e iniciado a seguirlo en cada imagen de la secuencia.	126
Figura A.4. Resultados PSR de la correlación de la imagen facial de prueba con cada plantilla biométrica facial en galería.	127

Índice de tablas

Tabla 3.1. Aplicaciones del reconocimiento facial	24
Tabla 4.1. Distorsiones presentadas por las imágenes faciales de los conjuntos de prueba.	41
Tabla 4.2. Umbral PSR de reconocimiento general para cada filtro evaluado.	43
Tabla 4.3. Tasa de verificación de los filtros de correlación en la verificación facial.	62
Tabla 4.4. Tasa de falsa aceptación de los filtros de correlación en la verificación facial.	62
Tabla 5.1. Desempeño de los filtro usando el conjunto de prueba CP1 y pre-procesado de las imágenes.	70
Tabla 5.2. Porcentaje de imágenes faciales del conjunto de prueba CP1 confundidas por los filtros.....	70
Tabla 5.3. Tasa de verificación obtenida por los filtros al aplicar el pre-procesado a las imágenes faciales y usar el conjunto de prueba CP2.	71
Tabla 5.4. Tasa de falsa aceptación obtenida por los filtros al aplicar el pre-procesado a las imágenes faciales y usar el conjunto de prueba CP2.....	72
Tabla 5.5. Tasa de verificación obtenida por los filtros con el conjunto de prueba CP3 y operaciones de pre-procesamiento.	72
Tabla 6.1. Desempeño de los filtros de correlación en la tarea de identificación facial.	82
Tabla 7.1. Desempeño de filtros de correlación no restringidos.	97
Tabla 8.1. Desempeño de los filtros FCnLsRA y MOSSE en el seguimiento facial.	114

Capítulo 1

1. Introducción

El ser humano reconoce a una persona por medio de su cara con una gran exactitud y rapidez, pues es una habilidad que ha desarrollado a lo largo de su vida. Esta precisión en el reconocimiento facial se debe a que inconscientemente hace uso otras características, tales como la complexión física de la persona, cabello, color de piel, estatura, voz, entre otras características físicas detectables sensorialmente.

Con el avance de las capacidades computacionales de hoy en día, es posible realizar el reconocimiento automático de caras humanas, aunque con menos precisión con que la realiza el ser humano. Esta tarea resulta desafiante para los investigadores debido a los cambios que sufre la cara con la edad, iluminación, pose, rotación, detalles faciales (lentes, bello facial, maquillaje, etc.), cicatrices, oclusión por objetos, imágenes faciales parciales, entre otros.

El reconocimiento facial ha avanzado mucho en la última década pero solo en aplicaciones donde es posible tener control sobre las personas o el entorno. Sin embargo, como lo describe P. H. (2011) se requiere más investigación donde las imágenes faciales son capturadas en un entorno no controlado y/o las personas no cooperan o no están conscientes de que están siendo identificados. Por ejemplo, reconocer automáticamente a una persona en una secuencia de video de baja resolución tomado en exterior con fuertes sombras continua siendo un desafío.

Se han propuesto varios algoritmos con el objetivo de realizar el reconocimiento facial de forma automática, los cuales han evolucionado a través del tiempo a tal grado de convertir al reconocimiento facial en una disciplina científica con una amplia gama de aplicaciones. La National Science and Technology Council (2010) describe que el reconocimiento facial en sus inicios usó algoritmos que implementaban modelos geométricos simples, pero que hoy ese proceso de reconocimiento ha madurado en una ciencia de sofisticadas representaciones matemáticas y de procesos de acoplamiento.

A pesar de esta madurez, el desempeño de los algoritmos existentes se degrada cuando las imágenes faciales de entrada contienen variaciones en iluminación, expresiones faciales, oclusión parcial, pose y ruido. Esto se debe a que los algoritmos consideran distancias entre las

características faciales (ojos, nariz, fosas nasales, boca, mejilla, entre otro) o en espacios faciales, lo que no permite aprovechar la información que pudiera proporcionar la apariencia de la cara.

En este trabajo de investigación de reconocimiento y seguimiento facial, se aprovechan las premisas de los filtros de correlación, tales como: a) pueden usar toda la información del objeto, es decir, tanto su forma como su contenido, b) tienen un buen fundamento matemático, c) en su diseño se puede incluir invariancia a distorsiones geométricas y tolerancia a ciertos tipos de ruidos (aditivo, fondo, iluminación, entre otros) y d) son buenos candidatos para implementarse con algoritmos rápidos que utilizan la transformada de Fourier (TF).

Con esto se aprovecha tanto la forma geométrica de la cara como la información que proporciona su apariencia (textura, color, entre otros). Con la inclusión de un algoritmo de pre-procesamiento a las imágenes faciales se mejora el desempeño de los filtros, obteniendo una mayor precisión en el reconocimiento y seguimiento.

Se evaluó el desempeño de cinco filtros compuestos de correlación en la actividad de reconocimiento facial. Primero, el filtro compuesto no lineal K-Law con un factor de no linealidad $k = 0.3$. Segundo, el filtro del mínimo promedio de la energía de correlación (MACE, Minimum Average Correlation Energy). Este filtro produce picos de correlación altos para sujetos auténticos mientras que minimiza picos de sujetos impostores. Tercero, el filtro SDF de compromiso óptimo (OTSDF, Optimal Trade Synthetic Discriminant Function) que produce picos altos para sujetos auténticos y un pequeño grado de tolerancia al ruido. Los últimos dos filtros evaluados fueron los filtros no restringidos UMACE (Unconstrained MACE) y UOTSDF (Unconstrained OTSDF).

Los resultados de las evaluaciones anteriores se usaron para el desarrollo de algoritmos de reconocimiento y seguimiento facial basados en correlación. El desempeño de los algoritmos desarrollados fue comparado con el desempeño de otros filtros ya existentes.

1.1 Justificación de la investigación

En la literatura se han propuesto algoritmos de reconocimiento facial que primero realizan la localización y alineación de características faciales para su posterior comparación o determinación de las distancias con las características de una plantilla biométrica facial almacenada. El método de acoplamiento de mallas elásticas considera una cara como un grafo

conectado (Wiskott, 1996). Cada nodo del grafo se localiza en ciertas coordenadas espaciales. A cada nodo del grafo se le asocia un vector de características, llamado *jet*, que pueden ser los valores del brillo local de la región alrededor del nodo. Las técnicas de análisis de sub-espacios para reconocimiento facial se basan en el hecho de que una clase de patrones de interés, tales como la cara, residen en un sub-espacio de la imagen de entrada (Li y Jain, 2004).

Estos métodos generalmente no aprovechan toda la información de la cara, es decir, al basarse en distancias desaprovechan la información de la textura de la piel facial o la información que pudiera proporcionar la forma de las mejillas o la barbilla.

Los algoritmos biométricos faciales intentan acoplar una cara de prueba de una persona con una plantilla facial almacenada en una base de datos. Los filtros de correlación son excelentes candidatos para la precisión del acoplamiento en presencia de variabilidad (expresiones faciales, cambios de iluminación, oclusión parcial y desplazamiento), característica necesaria en el reconocimiento biométrico. Algunas de las ventajas de emplear filtros de correlación son: a) pueden utilizar toda la información del objeto, es decir, tanto la forma como su contenido (color e intensidad) (Viajaya Kumar et al., 2005); b) tienen un buen fundamento matemático (Vanderlugt, 1964); c) en su diseño se pueden incluir invariancias a distorsiones geométricas y tolerancias a ciertos tipos de ruido (aditivo, fondo, iluminación, etc.); d) son buenos candidatos para implementarse con algoritmos rápidos (Frigo y Johnson, 2010); y e) son capaces de tanto de reconocer como detectar (Vijaya Kumar, 1992). En la literatura revisada existen pocas propuestas que hagan uso de los filtros de correlación en el reconocimiento facial.

En este trabajo de investigación, se aprovecha el potencial de los filtros de correlación en el reconocimiento de objetos parcialmente ocultos (González Fraga, 2004), tolerancia a cierto tipos de ruidos y su capacidad de obtener el grado de similitud entre un objeto de referencia y un objeto de prueba (Viajaya Kumar *et al.*, 2005) en el reconocimiento y seguimiento de caras humanas.

1.2 Planteamiento del problema

El principal desafío para el reconocimiento facial, como cualquier tecnología biométrica no intrusiva, es el desarrollo de algoritmos que funcionen en entornos e individuos no

controlables y el uso de sensores de baja calidad, por ejemplo, sobre un video de baja resolución tomada en exterior con fuertes sombras (P. H., 2011).

Bonsor et al. (2008) describen un sistema de reconocimiento facial instalado en el año 2001 por el departamento de policías de Tampa, Estados Unidos de América. Se colocó en las calles cámaras equipadas con tecnologías de reconocimiento facial en un intento por reducir los crímenes en el área. El sistema falló y fue desechado en 2003 debido a su ineffectividad. Las personas usaban máscaras y/o hacían gestos obscenos, impidiendo a las cámaras obtener una imagen clara de la persona a identificar.

Rodriguez (2009) describe otro problema que enfrenta este tipo de sistemas, la cual radica en su inviabilidad para una aplicación masiva, por ejemplo, detectar sospechoso en un aeropuerto. Con una tasa de falsa aceptación (porcentaje de veces que el sistema declara iguales a dos individuos que no lo son) del 1% significaría para a una de cada 100 personas que pase por delante de una cámara y obligarlos a identificarse utilizando un método más fiable, por ejemplo: colocar su índice en un lector automático de huellas. Si en Barajas (el aeropuerto de Madrid) hay 150,000 pasajeros al día, tendríamos que parar a 1500 personas cada día, sin contar a aquellas personas que acuden solo a esperar a alguien.

En aplicaciones- tales como monitorización, supervisión, habitaciones inteligentes, entre otros- es necesario seguir una cara en una secuencia de video. Vicente et al. (2002) presentan un algoritmo de seguimiento facial en tiempo real mediante la técnica denominada búsqueda logarítmica en regilla, el cual trabaja bien solo en condiciones sin distorsiones. Bolme et al. (2010) presentan un filtro de correlación para el seguimiento de objetos en una secuencia de imágenes logrando resultados prometedores.

El reconocimiento facial es una tarea desafiante debido a que la cara sufre variaciones (distorsiones) debido a la edad, pequeñas oclusiones de la cara por objetos -tales como lentes oscuros, gorras, bufandas o cualquier otro objeto- que impiden al sensor de visión adquirir una imagen facial completa. A esto hay que agregarle distorsiones causadas por los siguientes factores: pose, ausencia o presencia de componentes estructurales, expresiones faciales, orientación en la imagen e iluminación. En el caso del seguimiento facial, la cara varía de una imagen a otra y es difícil controlar o predecir las condiciones del ambiente.

Este trabajo de investigación aborda algunas de estas variaciones evaluando el desempeño de filtros de correlación y desarrollando una nueva técnica de reconocimiento facial usando un

nuevo filtro de correlación invariante a expresiones, sombras o imágenes faciales incompletas. También se aborda el tema de seguimiento de una cara en una secuencia de imágenes desarrollando un nuevo filtro que se adapta rápidamente a los cambios.

1.3 Objetivos

1.3.1 Objetivo general

Proponer y diseñar nuevos algoritmos de reconocimiento y seguimiento facial basados en correlación, dichos algoritmos se espera sean robustos a variaciones en sombras, imágenes faciales parciales, expresiones faciales y una alta capacidad para distinguir una clase cara de otras clases diferentes.

1.3.2 Objetivos específicos

1. Entender los fundamentos teórico-matemáticos de los filtros compuestos de correlación.
2. Entender los filtros de correlación aplicados al reconocimiento facial, así como otros enfoques existentes.
3. Evaluar la capacidad de los filtros de correlación para distinguir una clase de cara de otras clases, considerando variaciones en expresión, iluminación, pose, oclusión parcial, rotación y ruido blanco aditivo de tipo Gaussiano.
4. Evaluar la influencia del pre-procesado de las imágenes faciales en la construcción de filtros de correlación para el reconocimiento facial, considerando las variaciones del punto anterior.
5. Desarrollar y evaluar un filtro de correlación no lineal y sin restricciones, que pueda sintetizarse con regiones faciales para el reconocimiento facial.
6. Desarrollar y evaluar un filtro de correlación no lineal, sin restricciones y adaptativo para el seguimiento facial.
7. Implementación de un sistema de identificación facial que utilice los filtros de los puntos 5 y 6.

1.4 Organización de la tesis

Este documento está estructurado de la siguiente manera. En el capítulo 2 se presenta el fundamento teórico de los filtros de correlación, así como algunos de los filtros más estudiados en la literatura con aplicaciones en reconocimiento, localización y seguimiento de objetos. En el capítulo 3 se introduce el procesamiento facial con algunos de los principales enfoques que han servido como punto de partida o comparación con otros métodos, y el enfoque de reconocimiento facial basado en correlación. También, en dicho capítulo se aborda el tema de seguimiento facial mediante filtros de correlación.

En el capítulo 4 se evalúa el desempeño de los filtros de correlación con imágenes de prueba distorsionadas basadas en la adición de ruido y sombras en las imágenes faciales. Esta evaluación permite observar la capacidad de los filtros para distinguir una clase de otras clases diferentes. En el capítulo 5 se presenta la influencia que tienen algunas operaciones de pre-procesamiento en las imágenes faciales y cómo se mejora (en algunos casos) el desempeño de filtros de correlación en el reconocimiento facial. En el capítulo 6 se evalúa el desempeño de los filtros de correlación en la tarea de identificación facial considerando imágenes de prueba con sombras.

En el capítulo 7 se presenta un filtro compuesto no lineal sin restricciones para el reconocimiento facial usando regiones de la cara en su entrenamiento. En el capítulo 8 se presenta una estrategia de seguimiento facial basada en un filtro de correlación compuesto no lineal, sin restricciones y adaptativo. Los capítulos 7 y 8 contienen las principales contribuciones de este trabajo.

Finalmente, en el capítulo 9 se discuten los principales resultados y conclusiones de este trabajo de investigación. Como evidencia de la validez de los métodos propuestos, en el Apéndice A se presenta brevemente un sistema de identificación y seguimiento facial.

Capítulo 2

2. Filtros de correlación

La correlación es una técnica de reconocimiento de patrones usada en el reconocimiento automático de objetivos (Vijaya et al., 2005), localización y reconocimiento de objetos (Campos et al., 1991), reconocimiento biométrico (Vijaya et al., 2004), seguimiento visual de objetos (Bolme et al., 2010), localización de objetos (Bolme et al., 2009), entre otros.

Esta técnica se basa en la selección o creación de una señal de referencia para luego determinar su grado de similitud con una señal de prueba. En este trabajo las señales corresponden a imágenes bidimensionales. La señal de referencia es un arreglo de imágenes (véase la Figura 2.1) en el dominio de las frecuencias usando la transformada de Fourier, lo que se le denomina como filtro de correlación.

Podemos definir dos tipos de filtros de correlación mediante la cantidad de imágenes que se usa para ser construido: el simple que usa una sola imagen y el compuesto que se sintetiza con una colección de imágenes que contienen las distorsiones esperadas para el objeto de interés. A la colección de imágenes que se usa para construir un filtro se le denomina conjunto de entrenamiento, estas imágenes representan las posibles distorsiones que se espera pudiera presentar el objeto objetivo.

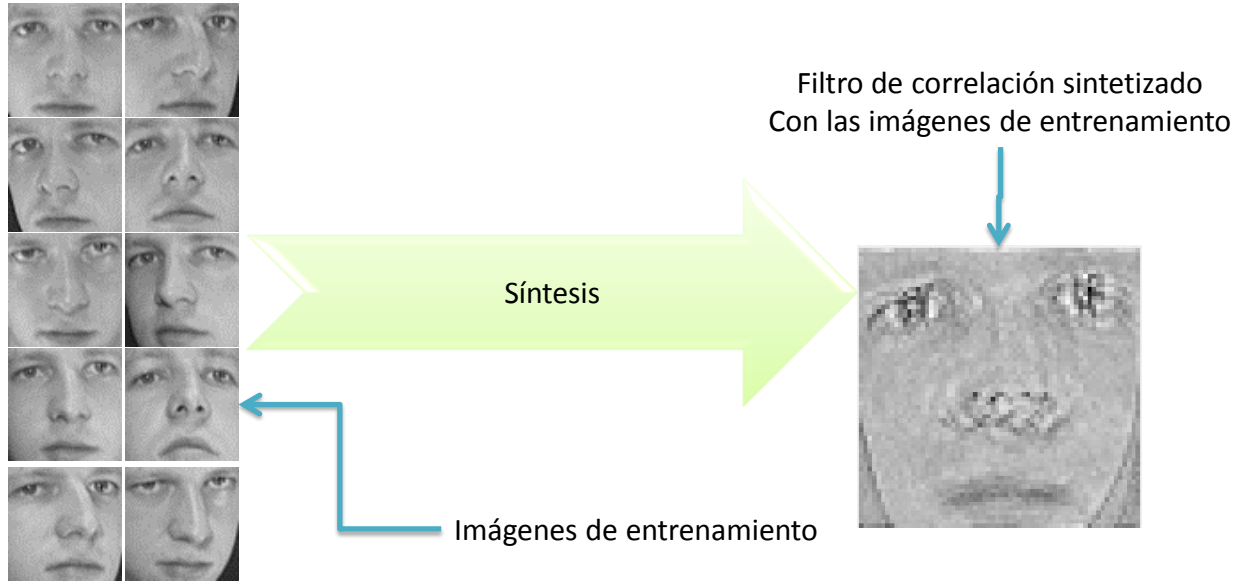


Figura 2.1. Diseño de un filtro compuesto usando imágenes de entrenamiento.

La definición estándar de la correlación espacial entre una señal de entrada 2-D $i(m, n)$ y un filtro 2-D $h(m, n)$, resultando un plano que se denomina como la salida de correlación 2-D $y(m, n)$, es de la siguiente manera:

$$y(m, n) = i(m, n) \circ h(m, n) \quad (1)$$

donde $\circ$ representa la correlación espacial. El uso de la transformada de Fourier (TF) en las imágenes permite obtener de forma rápida datos de frecuencias primarias los cuales son significativamente más discriminativos que el dato espacial del cual fue obtenido (Savvides et al., 2007). Mediante el teorema de convolución, la correlación en el dominio de las frecuencias entre dos señales se convierte en una multiplicación elemento a elemento de sus TF. Por lo tanto, la ecuación 1 en el dominio de Fourier queda de la siguiente manera:

$$Y(k, l) = F^{-1}\{I(k, l) \odot H^*(k, l)\} \quad (2)$$

donde $I(k, l)$ y $H(k, l)$ son la transformada de Fourier 2-D de $i(m, n)$ y $h(m, n)$ respectivamente. El símbolo F^{-1} , $\odot$, y $*$ representan la inversa de la transformada de Fourier (ITF), la multiplicación elemento a elemento de las dos señales 2-D (correlación en el dominio de las frecuencias) y la conjugación elemento a elemento, respectivamente.

Algunos filtros de correlación reportados en la literatura como el filtro del mínimo promedio de la energía de correlación (MACE, Minimum Average Correlation Energy), función discriminante sintética (SDF, Synthetic Discriminant Function) y el filtro de compromiso óptimo

(OTF, Optimal tradeoff filters). Estos filtros generalmente ofrecen buen desempeño en el reconocimiento de objetos. Aunque los filtros de correlación son capaces tanto de reconocer como localizar objetos en una escena, dentro de la literatura revisada solo se encontró el filtro del promedio de filtros sintéticos exactos (ASEF, Average of Synthetic Exact Filters) diseñado específicamente para la localización de objetos. En el caso de seguimiento de objetos mediante filtros de correlación se encontró el filtro de salida mínima de la suma del error cuadrático (MOSSE, Minimum Output Sum of Squared Error) que se ha aplicado en el seguimiento de objetos.

En este trabajo se analiza el desempeño de los filtros de correlación K-Law, MACE, UMACE (Unconstrained MACE), OTSDF y UOTSDF (Unconstrained OTSDF) en el reconocimiento facial. El interés en estos filtros se debe a que han sido ampliamente usados en el reconocimiento de objetos donde han obtenido un buen desempeño. Para el seguimiento facial se tiene interés en el filtro MOSSE, el cual sirve como comparación con la estrategia de seguimiento desarrollado en este trabajo. En esta sección se describen a detalle dichos filtros y algunos que permiten comprender la relevancia de la investigación y desarrollo de los filtros de correlación.

2.1 Filtro de acoplamiento

El filtro de correlación básico es el filtro de acoplamiento (MF, Matched Filter), el cual es óptimo para detectar una señal de referencia que ha sido afectada por ruido blanco aditivo. Sin embargo, el MF se desempeña pobremente cuando la imagen de referencia aparece con distorsiones (por ejemplo, iluminación, rotación y cambios de escala). El filtro MF está dado por la ecuación 3 (VanderLugt, 1964).

$$H(f) = S^*(f) \quad (3)$$

donde $S^*(f)$ es la conjugada elemento a elemento de la transformada de Fourier de la imagen de referencia f . Este filtro produce un plano de correlación con un pico (valor máximo en el plano de correlación) con muchos lóbulos laterales (valores grandes en el vecindario del pico, véase Figura 2.2). Un pico amplio no es adecuado en muchas aplicaciones debido a que puede afectar al reconocimiento, por lo que filtros más avanzados buscan producir un pico agudo y sin los lóbulos laterales.

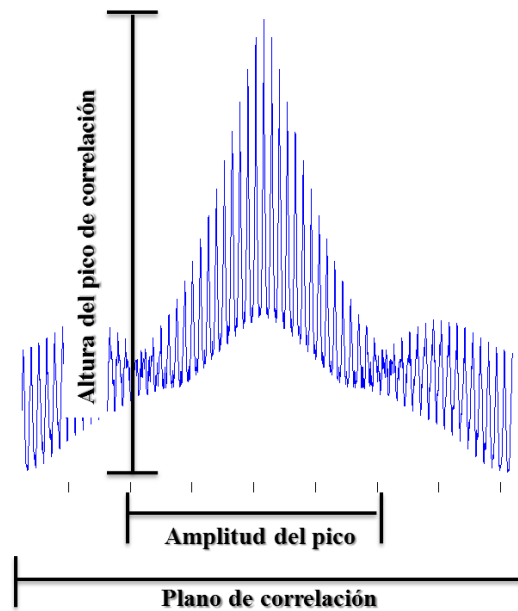


Figura 2.2. Salida 2D de correlación

2.2 Funciones sintéticas discriminantes generalizadas

La función sintética discriminante (SDF, Synthetic Discriminant Function) propuesto por Casasent y Weng (1986) fue el primer filtro de correlación enfocado a reconocer versiones distorsionadas de un objeto mediante una combinación lineal de imágenes de referencia, diseñado de tal forma que produzca un plano de correlación cuya salida en el origen produzca un valor pre-especificado. Dicha restricción sobre la salida de correlación permite una comparación normalizada y un grado de discriminación para el filtro. Por lo general, el plano de la salida de correlación se restringe (en el origen) a 1 para todos los datos de entrenamiento de la clase verdadera y 0 para la clase falsa.

Sea el vector columna $\mathbf{x}_n$ de dimensión d (cantidad de píxeles que contiene la imagen de entrenamiento) la j -ésima imagen de entrenamiento. Cada imagen de entrenamiento $i_j(x, y)$ se escanea de izquierda a derecha y de arriba abajo (véase la Figura 2.3) para producir el vector $\mathbf{x}_n$. Considere X como la matriz que contiene en sus columnas los vectores $\mathbf{x}_n$:

$$X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \quad (4)$$

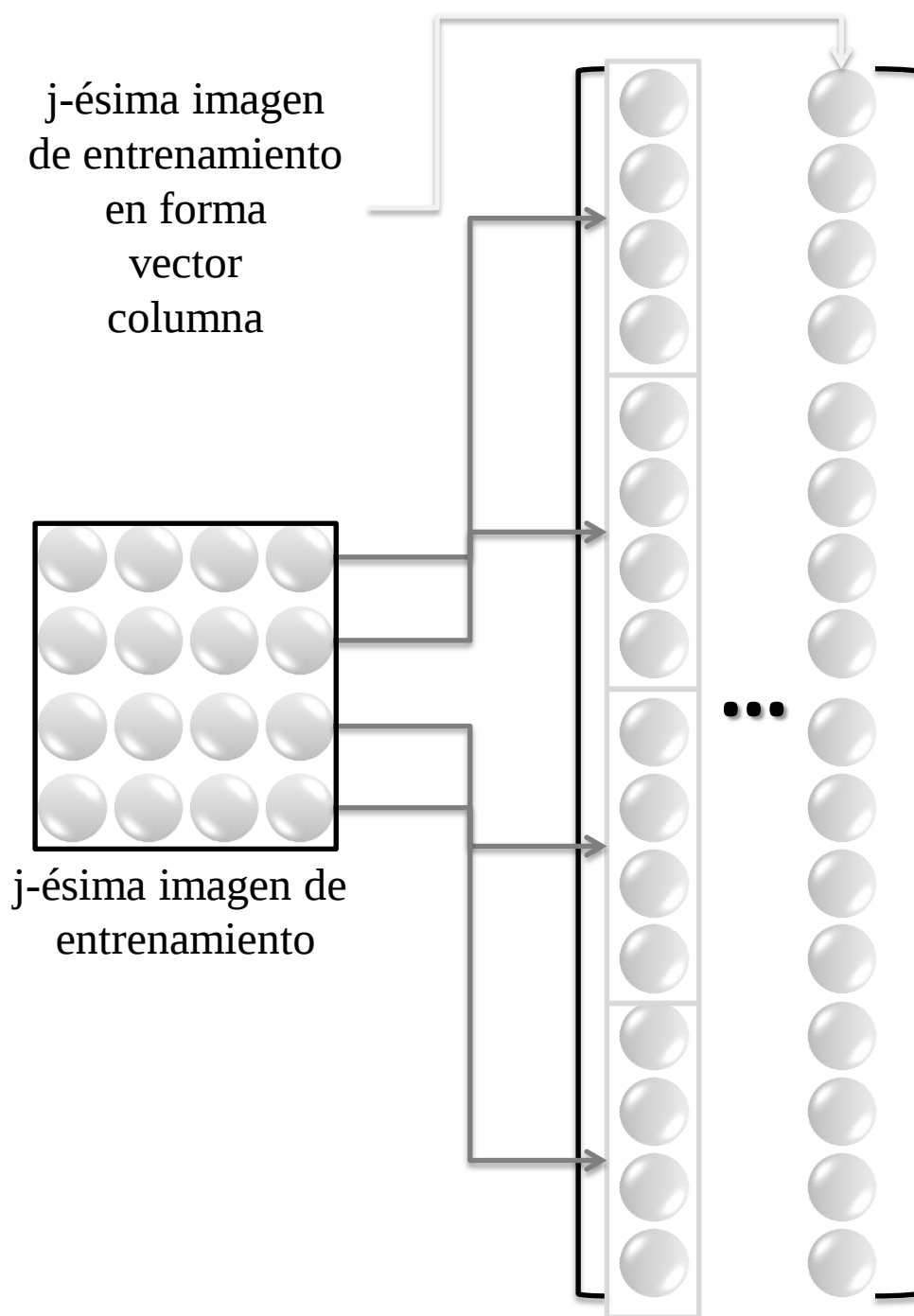


Figura 2.3. Matriz que contiene las imágenes de entrenamiento en forma de vector columna.

Considere a u como el vector columna que contiene los valores de salida esperada en el origen para cada imagen de entrenamiento. Entonces, la correlación con restricciones en el origen se puede expresar de la siguiente manera:

$$X^T \mathbf{h} = \mathbf{u} \quad (5)$$

donde T es la operación de transpuesta y $\mathbf{u} = [u_1, u_2, \dots, u_N]^T$ es un vector de tamaño $N \times 1$ que contiene los valores deseados para cada imagen de entrenamiento. Sin embargo, el número de imágenes de entrenamiento N por lo general es mucho menor que el número de frecuencias en el filtro. Para encontrar una única solución, suponemos que $\mathbf{h}$ es una combinación lineal de las imágenes de entrenamiento, es decir, $\mathbf{h} = X\mathbf{a}$ donde $\mathbf{a}$ es el vector de pesos para las combinaciones lineales de las columnas de la matriz de datos X . $X^T(X\mathbf{a}) = \mathbf{u} \Rightarrow \mathbf{a} = (X^T X)^{-1} \mathbf{u}$, sustituyendo esta solución en $\mathbf{h} = X\mathbf{a}$ obtenemos el filtro SDF:

$$\mathbf{h} = X(X^T X)^{-1} \mathbf{u} \quad (6)$$

donde $\mathbf{h}$ es el vector $d \times 1$ del filtro SDF en el dominio espacial y -1 es la inversa de la matriz. En este trabajo un filtro de correlación se expresa en su término vectorial con la letra $\mathbf{h}$ ya sea en el dominio espacial o de las frecuencias, excepto donde se especifique otra cosa. Un filtro 2D puede obtenerse fácilmente mediante una operación de re-ordenamiento sobre el vector $\mathbf{h}$.

2.3 Filtro compuesto no lineal Ley k

El filtro SDF a pesar de tener cierta tolerancia a distorsiones no considera el ruido y además produce un pico de correlación muy amplio. Para mejorar el desempeño de los filtros compuestos en términos de discriminación contra objetos similares al objetivo, agudeza del pico de correlación y robustez a ruido, se aplica la técnica de filtrado no lineal a la ecuación 6 tal y como lo presentan Javidi et al. (1997).

Primero se aplica la no linealidad con la Ley k (K-Law) a las TF de las imágenes de entrenamiento, tal y como se presenta en la ecuación 7. Cuando la no linealidad se aplica a una matriz, ésta se aplica a cada elemento.

$$I^k = |I|^k \exp[-i\varphi_I], \quad 0 \leq k \leq 1 \quad (7)$$

donde $|I|^k$ es la magnitud de I y φ es su fase. El valor k (factor de no linealidad) controla la fuerza de la no linealidad. Cuando el filtro es sintetizado con solo una imagen y se establece $k = 1$ se obtiene un filtro de acoplamiento clásico. Cuando $k = 0$, se establece la magnitud de todas las frecuencias a 1, obteniendo de esta forma un filtro de correlación de sólo fase. Sea el vector columna $\tilde{\mathbf{I}}_i^k$ la i -ésima imagen de entrenamiento con la ley k aplicada sobre su TF. Sea

$\tilde{X}^k = [\tilde{I}_1^k, \tilde{I}_2^k, \dots, \tilde{I}_i^k]$. Ahora, aplicando la TF y la no linealidad sobre la ecuación 6 para construir un filtro compuesto no lineal SDF no lineal ley k se obtiene:

$$h = \tilde{X}^k (\tilde{X}^{k+} \tilde{X}^k)^{-1} u \quad (8)$$

El factor de no linealidad eleva a la k-ésima potencia la magnitud del espectro de Fourier tanto de la imagen analizada como el filtro, mientras se mantiene intacta la información de la fase. Esta característica provoca que el filtro tenga una alta capacidad de discriminación.

Los filtros K-Law, con las imágenes de entrenamiento apropiadas pueden ser invariante a escala y rotación, distorsiones presentadas por las imágenes de prueba. También, pueden tener una mayor tolerancia al ruido. Para objetos de la clase verdadera produce un plano de correlación con un pico más agudo y fuerte, mientras que un filtro convencional SDF no.

2.4 Filtro del mínimo promedio de la energía de correlación

El filtro MACE (Mahalanobis et al., 1987) suprime los lóbulos laterales producidos por un filtro SDF convencional minimizando la energía en el plano de correlación. La energía de correlación promedio (ACE – average correlation energy) para N imágenes de entrenamiento se define de la siguiente forma:

$$ACE = \frac{1}{N} \sum_{i=1}^N \sum_m^{d_1} \sum_n^{d_2} |y_i(m, n)|^2 \quad (9)$$

donde y_i mantiene su definición anterior. Usando el teorema de Parseval, ACE se puede expresar en el dominio de la frecuencia como

$$ACE = \frac{1}{d \cdot N} \sum_{i=1}^N \sum_k^{d_1} \sum_l^{d_2} |Y_i(k, l)|^2 \quad (10)$$

donde $Y_i(k, l)$ es la transformada de Fourier de $y_i(m, n)$. De esta forma $Y_i(k, l) = H(k, l) X_i^*(k, l)$, ACE expresada en el dominio de las frecuencias se convierte en:

$$\begin{aligned} ACE &= \frac{1}{d \cdot N} \sum_{i=1}^N \sum_k^{d_1} \sum_l^{d_2} |H(k, l)|^2 |I_i(k, l)|^2 \\ &= \frac{1}{d \cdot N} \sum_{i=1}^N (\mathbf{h}^+ I_i) (I_i^* \mathbf{h}) \end{aligned} \quad (11)$$

donde $\mathbf{h}$ es un vector columna de longitud d cuyos elementos pertenecen $H(k,l)$ y X_i es una matriz diagonal de tamaño $d \times d$ cuyos elementos no ceros se toman de $I_i(k,l)$. Por simplicidad definamos la matriz diagonal $D = \frac{1}{d \cdot N} \sum_{i=1}^N I_i \cdot I_i^*$, cuyo diagonal contiene el promedio de la energía espectral de las imágenes de entrenamiento. Entonces la ecuación 11 puede re-escribirse de la siguiente manera:

$$ACE = \mathbf{h}^+ D \mathbf{h} \quad (12)$$

Las restricciones sobre el pico de correlación deben ser expresadas en el dominio de la frecuencia. Debido a que el producto interno en el dominio espacial es directamente proporcional a los productos internos en el dominio de la frecuencia, es posible re-escribir las restricciones sobre el pico expresado en la ecuación 5 de la siguiente forma:

$$\tilde{X}^+ \mathbf{h} = d \cdot \mathbf{u} \quad (13)$$

donde $\tilde{X}$ es una matriz cuyas columnas son las TF de las imágenes de entrenamiento en forma vector columna. De esta forma, el filtro $\mathbf{h}$ el cual minimiza la ecuación 12 mientras satisface las restricciones de la ecuación 5 es el filtro MACE. Una solución óptima se logra usando el método de los multiplicadores de Lagrange, el cual produce el filtro dado por la siguiente ecuación:

$$\mathbf{h} = D^{-1} X (X^+ D^{-1} X)^{-1} \mathbf{u} \quad (14)$$

Los filtros MACE eliminan el lóbulo lateral para asegurar un pico de correlación agudo y fácil de distinguir, reduciendo la probabilidad de errores. Sin embargo, tienen dos inconvenientes: no son inmunes a ruidos y son demasiado sensibles a cambios de variaciones.

2.5 Filtro SDF de compromiso óptimo

Una forma óptima de mantener el compromiso entre la tolerancia al ruido y la agudeza de los picos se logra mediante el uso del filtro OTSDF (Savvides, 2004), el cual está dado por la ecuación:

$$\mathbf{h} = T^{-1} X (X^+ T^{-1} X)^{-1} \mathbf{u} \quad (15)$$

donde $T = (\alpha D + \sqrt{1 - \alpha^2} C)$ y $1 \geq \alpha \geq 0$. La matriz D es la misma que la del filtro MACE. El parámetro α representa el compromiso entre la tolerancia al ruido y la discriminación. Variando α entre 0 y 1 se consigue una cantidad de tolerancia al ruido y agudeza del pico en el

filtro. Si $\alpha = 1$ entonces la ecuación 15 se reduce al filtro MACE, mientras que si $\alpha = 0$ el filtro se reduce al filtro MVSDF (Savvides et al., 2007) el cual es tolerante al ruido.

2.6 Filtros de correlación no restringidos

La mayoría de las investigaciones en el diseño de los filtros SDF asumen que los valores de correlación en el origen fueron especificados previamente. Sin embargo, no es necesario asumir tal restricción (Mahalanobis et al., 1994). Esta restricción es útil cuando el filtro SDF se sintetiza en el laboratorio óptico, sin embargo en la síntesis digital de filtros SDF en computadoras no es necesaria dicha restricción (Vijaya Kumar, 1992).

Otro hecho es que los valores preestablecidos están diseñados solo para imágenes de entrenamiento y no para imágenes de prueba. Los filtros de correlación son atractivos porque no solo reconocen un objetivo si no también lo localizan. Si el objetivo se desplaza una distancia conocida del origen, la salida de correlación también se desplaza la misma distancia. Pero el origen está restringido a 1 y los valores no restringidos del plano de correlación pueden ser mayores a 1 lo que hace imposible encontrar el desplazamiento del pico de correlación correspondiente al objeto de interés.

Por lo tanto, eliminando las restricciones se incrementa el número de posibles soluciones. También, sin la restricción se puede diseñar filtros que ofrezcan un buen desempeño en presencia de ruido y fondo estructurado mientras mantiene un pico de correlación agudo y fácil de localizar. En esta sección explicaremos detalladamente los filtros de correlación no restringidos UMACE (Unconstrained MACE) y UOTSDF (Unconstrained OTSDF) que son de nuestro interés.

Como se verá a continuación, la baja complejidad y bajo costo computacional de los filtros UMACE y UOTSDF posibilitan la implementación de dichos filtros en dispositivos con pocos recursos y de bajo poder de procesamiento. También es posible modificar estos filtros para que se actualicen de tal forma que siempre se consuma el mismo espacio de almacenamiento, tal y como lo presentan Savvides et al. (2003).

2.6.1 Filtro MACE no restringido

El filtro UMACE (Mahalanobis et al., 1993) es la versión no restringida del filtro MACE. El objetivo de UMACE es sólo maximizar el promedio de los valores de correlación en el origen, en lugar de restringirlo como se hace en la ecuación 5, mientras minimiza el promedio de la energía de correlación entre las imágenes de entrenamiento. El filtro UMACE está dado por la siguiente ecuación:

$$\mathbf{h} = D^{-1}\mathbf{m} \quad (16)$$

La matriz D mantiene su definición de la construcción del filtro MACE original, y el vector $\mathbf{m}$ es la TF de la imagen de entrenamiento promedio. Como se puede observar, UMACE no es tan costoso, computacionalmente hablando, ya que requiere solo de la inversión de una matriz. Otra ventaja del filtro UMACE es que puede producir diferentes valores en el origen para una misma imagen de entrenamiento.

2.6.2 OTSDF no restringido

La versión no restringida del filtro OTSDF, denominado UOTSDF (Kiat Ng, 2005) está dado por la siguiente ecuación:

$$\mathbf{h} = T^{-1}\mathbf{m} \quad (17)$$

donde T y $\mathbf{m}$ mantienen su definición de la construcción del filtro OTSDF original y UMACE, respectivamente.

2.7 Filtro del promedio de filtros sintéticos exactos

Los filtros presentados anteriormente generalmente se aplican en la actividad de reconocimiento, mientras que el filtro ASEF (Bolme et al., 2009) está diseñado especialmente para la localización de objetos. La primera diferencia entre los filtros discutidos anteriormente y los filtros ASEF es que los filtros ASEF están muy restringidos, es decir, los filtros SDF especifican un simple “valor en el origen” de la salida de correlación por cada imagen de entrenamiento, mientras que los filtros ASEF son entrenados usando imágenes que especifican una respuesta deseada en todo el plano de correlación. Esta respuesta es típicamente un pequeño brillo centrado sobre el objeto interés y pequeños valores alrededor. Esta imagen de respuesta esperada restringe todo el plano de correlación y no únicamente un valor.

Una consecuencia de especificar completamente la salida de correlación es un perfecto balance entre las restricciones y los grados de libertad para cada imagen de entrenamiento, y así se determina un filtro exacto para cada imagen de entrenamiento. Para cada imagen de entrenamiento i_j , especificamos la salida deseada g_i .

Note que los pares de entrenamiento i_j, g_i consisten de una imagen de entrenamiento asociada a una salida de correlación deseada. La salida de correlación g_i es sintéticamente generada con un pico brillante en el centro del objetivo y valores pequeños en el resto del plano. Específicamente, en el trabajo original g_i es un Gaussiano 2-D centrado en la posición (x_i, y_i) y con radio σ :

$$y_i(x, y) = e^{-\frac{(x-x_i)^2 + (y-y_i)^2}{\sigma^2}} \quad (18)$$

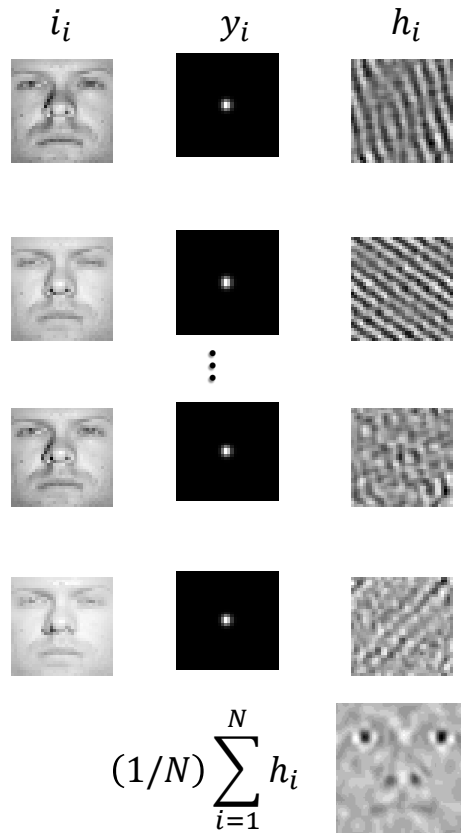


Figura 2.4. Construcción de un filtro ASEF promediando filtros sintéticos exactos.

La Figura 2.4 muestra un filtro ASEF en el dominio espacial construido con 2432 imágenes faciales. Porque la transformada de Fourier es una transformación lineal, el promedio se puede calcular ya sea en el dominio de Fourier o espacial. Para construir el filtro ASEF en el

dominio de Fourier, se construye un filtro exacto para cada imagen de entrenamiento con su salida esperada como sigue:

$$H_i^*(w, v) = \frac{G_i(w, v)}{I_i(w, v)} \quad (19)$$

donde la división es elemento a elemento, $G_i(w, v)$ es la TF de la salida esperada g_i y $I_i(w, v)$ es la TF de la imagen de entrenamiento i_i . El promedio de los filtros exactos se calcula de la siguiente manera:

$$H_\mu^*(w, v) = \frac{1}{N} \sum_{i=1}^N H_i^*(w, v) \quad (20)$$

donde H_μ es el filtro de correlación ASEF en el dominio de la frecuencia. Como se puede observar en la Figura 2.4, el filtro h_i parecen no tener una estructura con la cual se pueda acoplar una cara desconocida a reconocer. Para resolver este problema, se promedian múltiples filtros sintéticos exactos con el objetivo de generalizar el conjunto de entrenamiento. El promediado enfatiza las características comunes entre las imágenes de entrenamiento, tal y como se puede observar en el filtro ASEF final de la Figura 2.4.

2.8 Filtro de salida mínima de la suma del error cuadrático

El filtro MOSSE (Bolme et al., 2010) se genera como un filtro ASEF pero usando pocas imágenes de entrenamiento. Para iniciar, el filtro necesita un conjunto de imágenes de entrenamiento i_i y una salida esperada g_i . La salida deseada g_i mantiene su definición del filtro ASEF. El entrenamiento se realiza en el dominio de Fourier para tomar ventaja de la simple relación elemento a elemento de la entrada y la salida.

Para encontrar un filtro que mapea una entrada de entrenamiento a la salida de entrenamiento deseada, MOSSE encuentra un filtro H que minimiza la suma del error cuadrático entre la salida actual de la convolución y la salida deseada de la convolución. Este problema de minimización toma la forma:

$$\min_{H^*} \sum_i |I_i \odot H^* - G_i|^2 \quad (21)$$

En el problema de seguimiento, el objetivo no está siempre centrado y el pico en G_i se mueve para seguir el objetivo en i_i . En un caso más general, g_i puede tener cualquier forma.

Resolver este problema de optimización no es particularmente difícil pero requiere un tipo de cuidado especial porque la función a optimizar es una función de valores reales de una variable compleja. Primero, cada elemento de H (indexado por w y v) se pueden resolver independientemente porque todas las operaciones en el dominio de Fourier se realizan elemento por elemento. Esto involucra re-escribir la función en términos tanto de H_{wv} y H_{wv}^* .

$$0 = \frac{\partial}{\partial H_{wv}^*} \sum_i |I_{i w v} H_{w v}^* - G_{i w v}|^2 \quad (22)$$

Resolviendo para H^* en la ecuación 22 se encuentra una expresión para el filtro MOSSE:

$$H^* = \frac{\sum_i G_i \odot I_i^*}{\sum_i I_i \odot I_i^*} \quad (23)$$

El numerador es la correlación entre la entrada y la salida deseada y el denominador es el espectro de energía de la entrada.

2.9 Medidas de desempeño para filtros de correlación

Cuando se considera el uso de correlación en el reconocimiento de patrones, en este caso para reconocimiento de rostros humanos, es necesario definir una métrica que cuantifique la bondad o exactitud de una correlación. Una forma simple pero a veces efectiva es seleccionar el valor más alto en el plano de correlación y usar un umbral para determinar una coincidencia. Este enfoque trabaja bien cuando hay variaciones pequeñas en el dato tal que la varianza en el valor del pico de correlación sea pequeño. Sin embargo, cuando la imagen de prueba está muy distorsionada, el valor del pico puede ser muy grande o muy pequeño, siendo difícil predecir ese valor.

En el reconocimiento facial, se espera que un filtro de correlación produzca un pico agudo y alto para imágenes faciales de un sujeto auténtico. Algunas medidas de desempeño para un filtro de correlación están basadas en la capacidad del filtro para reconocer a un objeto afectado por ruido blanco aditivo, otros en la capacidad de filtro para rechazar objetos falsos, entre otros. Pero lo que se busca en común es caracterizar la agudeza del pico de correlación. A continuación se mencionan algunas de esas métricas de desempeño.

Relación pico de correlación a lóbulos laterales (PSR, Peak to Sidelobe Ratio): es la razón del pico de correlación entre la desviación estándar de una región o vecindad alrededor pero no incluyendo el pico (Véase Figura 2.5).

$$PSR = \frac{(pico - media\ de\ la\ región\ del\ lóbulo)}{desviación\ estandar\ de\ la\ región\ del\ lóbulo} \quad (24)$$

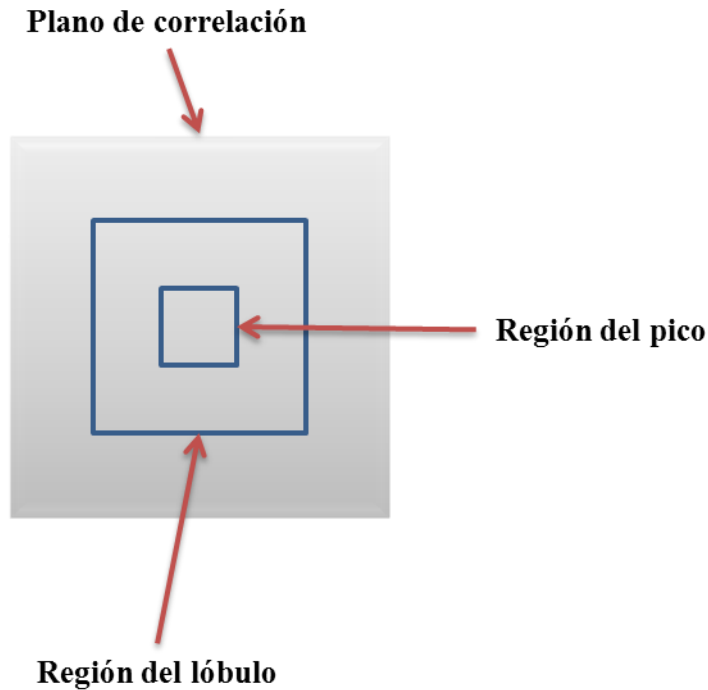


Figura 2.5. Plano de correlación y regiones indispensables para la obtención del PSR

Relación pico a energía de correlación (PCE, Peak to Correlation Energy): cuando se desea medir la magnitud de la diferencia entre el pico y el resto del plano de correlación se puede usar PCE, el cual se define de la siguiente manera:

$$PCE = \frac{(pico - media\ del\ plano)}{desviación\ estandar} \quad (25)$$

Capacidad de discriminación (DC, Discriminant Capacity): La capacidad de discriminación es la habilidad que posee un filtro de correlación para distinguir a un objeto de otros diferentes.

$$DC = 1 - \left(\frac{C^R(0,0)^2}{C^O(0,0)^2} \right) \quad (26)$$

donde C^R es el valor máximo del plano de correlación sobre la región a rechazar y C^O es valor máximo del plano de correlación sobre la región donde se encuentra el objeto patrón (Véase Figura 2.6). Si DC es un valor muy cercano a uno, indica que el filtro tiene un buen desempeño, en caso contrario si es un valor que tiende a cero o valores negativos indica que el filtro tiene una mala capacidad de discriminación.

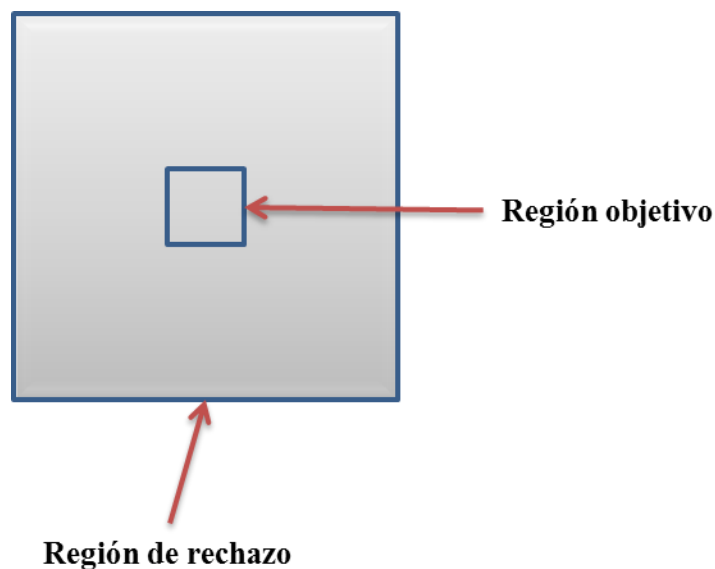


Figura 2.6. Región de rechazo y región objetivo sobre un plano de correlación para la obtención de DC

Estas tres métricas pueden ser usadas con cualquier filtro de correlación, pero la métrica óptima depende de la aplicación. En situaciones donde se espera que el objeto de interés aparezca varias veces sobre una escena se puede usar PSR, ya que es una métrica que se basa en regiones del plano de correlación. En situaciones donde se espera que el plano de correlación solo deba de producir un pico en el plano de salida, basta con usar PCE o DC.

Capítulo 3

3. Introducción al procesamiento facial

A todas aquellas técnicas computacionales que trabajan de forma sistemática con la cara humana para lograr un objetivo se denomina procesamiento facial. Esto incluye la detección, localización, detección de características faciales, seguimiento, reconocimiento y reconocimiento de expresiones. En los siguientes párrafos se describe brevemente alguna de las tareas de procesamiento facial (Yang et al., 2002).

Dada una imagen arbitraria, el objetivo de la detección facial es determinar si existe o no una cara en la imagen. En caso de que la imagen contenga caras, se devuelve la ubicación y la dimensión de cada una de ellas. En cualquier sistema de procesamiento facial, la detección facial es el primer paso para realizar todas las demás actividades.

La localización facial es una versión simplificada de la detección facial. En esta actividad se tiene como objetivo encontrar las coordenadas de la ubicación de una imagen facial, se asume que la imagen contiene una imagen facial. Los métodos de seguimiento facial estiman la localización y posible orientación de una cara en una secuencia de imágenes en tiempo real.

El objetivo de la detección de características faciales es determinar la presencia y ubicación de características tales como los ojos, nariz, fosas nasales, cejas, boca, labios, entre otros, asumiendo que existe solo una cara en la imagen de entrada.

El objetivo del reconocimiento facial es obtener la identidad de una persona mediante una imagen de su rostro. Para que esto sea posible es necesario haber registrado previamente a la persona. Los métodos de reconocimiento de expresiones faciales tienen como objetivo identificar el estado afectivo (feliz, triste, disgustado, entre otros) de una persona a través del análisis de la cara y sus características faciales.

El gran crecimiento en el procesamiento facial se debe a que la información acerca de la identidad del usuario, su estado de ánimo y su intención en un determinado escenario puede ser extraído de imágenes y las computadoras pueden reaccionar de acuerdo a ellos. Por ejemplo, reconociendo las expresiones faciales de un niño con autismo podemos determinar su estado anímico y tomar las medidas necesarias.

El procesamiento facial es una disciplina muy amplia y describirla detalladamente está fuera de los alcances de este trabajo. Solo nos enfocaremos a detallar el reconocimiento y el seguimiento facial que son las áreas donde en este trabajo aplicamos filtros de correlación.

3.1 Reconocimiento facial

El reconocimiento facial está motivado no únicamente por las distorsiones (véase la Figura 3.1) propias de la cara, tales como pose, rotación, iluminación, oclusión, expresión, edad, entre otros, sino por numerosas aplicaciones prácticas donde es necesario el reconocimiento del usuario a través de su cara. Algunos ejemplos presentados por Zhao et al. (2003) se muestran en la Tabla 3.1.

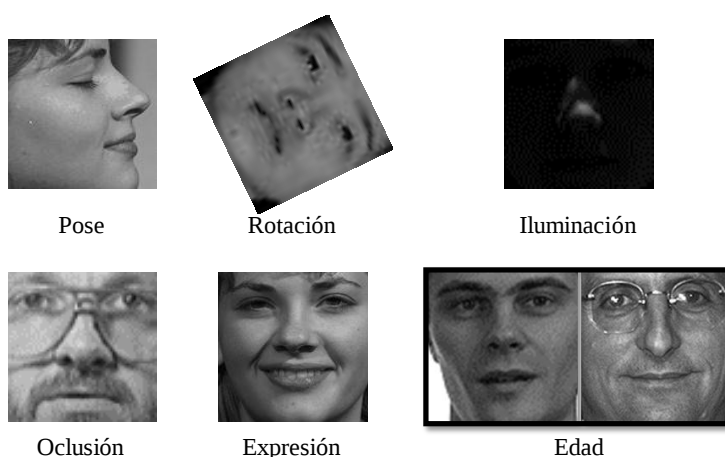


Figura 3.1. Distorsiones faciales.

Usar la cara como un rasgo biológico para reconocer a una persona tiene varias ventajas (véase Figura 3.2). Primero, la posibilidad de hacer un reconocimiento no intrusivo, es decir, la persona no tiene que interactuar de forma directa con algún dispositivo e inclusive ni estar consciente de que está siendo identificada. Segundo, su naturalidad. Al ser la cara un rasgo biológico inseparable del ser humano, éste no tiene que cargar con alguna tarjeta o algún otro objeto que sirva como identificador. Finalmente, su facilidad de uso. Al usar un sistema de reconocimiento facial, el mayor esfuerzo que el usuario hará y solo en caso de ser necesario es dirigir la mirada hacia un sensor de visión. No tiene que memorizar alguna clave o aprender a usar el sistema.

El reconocimiento facial es un problema de reconocimiento de patrones visuales. Aunque tiene varias décadas de investigación –desde 1960 según Li et al. (2004)–, no ha existido un gran avance. Además de las distorsiones propias de la cara, esta tarea también se vuelve difícil cuando se quiere aplicar en ambientes donde no es posible controlar, por ejemplo la iluminación, expresión, oclusión y punto de vista con el sensor de visión.

Tabla 3.1. Aplicaciones del reconocimiento facial

Áreas	Especificación de aplicaciones
Entretenimiento	Video juegos, realidad virtual, programas de entrenamiento, interacción humano-robot, interacción humano-computadora.
Tarjetas inteligentes	Licencias de manejo, inmigración, identificación nacional, pasaportes, registro de votantes, acceso a servicios de salud
Aseguramiento de la información	Control parental de la televisión, autenticación en dispositivos personales, autenticación en computadoras de escritorio, aseguramiento de aplicaciones, aseguramiento de bases de datos, encriptación de archivos, seguridad en intranets, acceso a internet, registros médicos
Reforzamiento de la ley y vigilancia	Vigilancia en video, control de portales web, análisis post-eventos, robo en tiendas, seguimiento de sospechosos he investigación



Figura 3.2. Cara humana.

El reconocimiento facial se clasifica en dos tareas principales, verificación e identificación. En la tarea de verificación facial (véase Figura 3.3), una persona presenta su muestra biométrica facial al sistema y declara una identidad presumiblemente de una persona registrada en la base de datos. La muestra facial presentada se le denomina prueba. Entonces el sistema compara la prueba con la plantilla biométrica facial presumiblemente de la persona almacenada en la base de datos. La comparación produce un grado de similitud. El sistema acepta la identidad declarada si el grado de similitud es mayor a un umbral de reconocimiento establecido previamente.

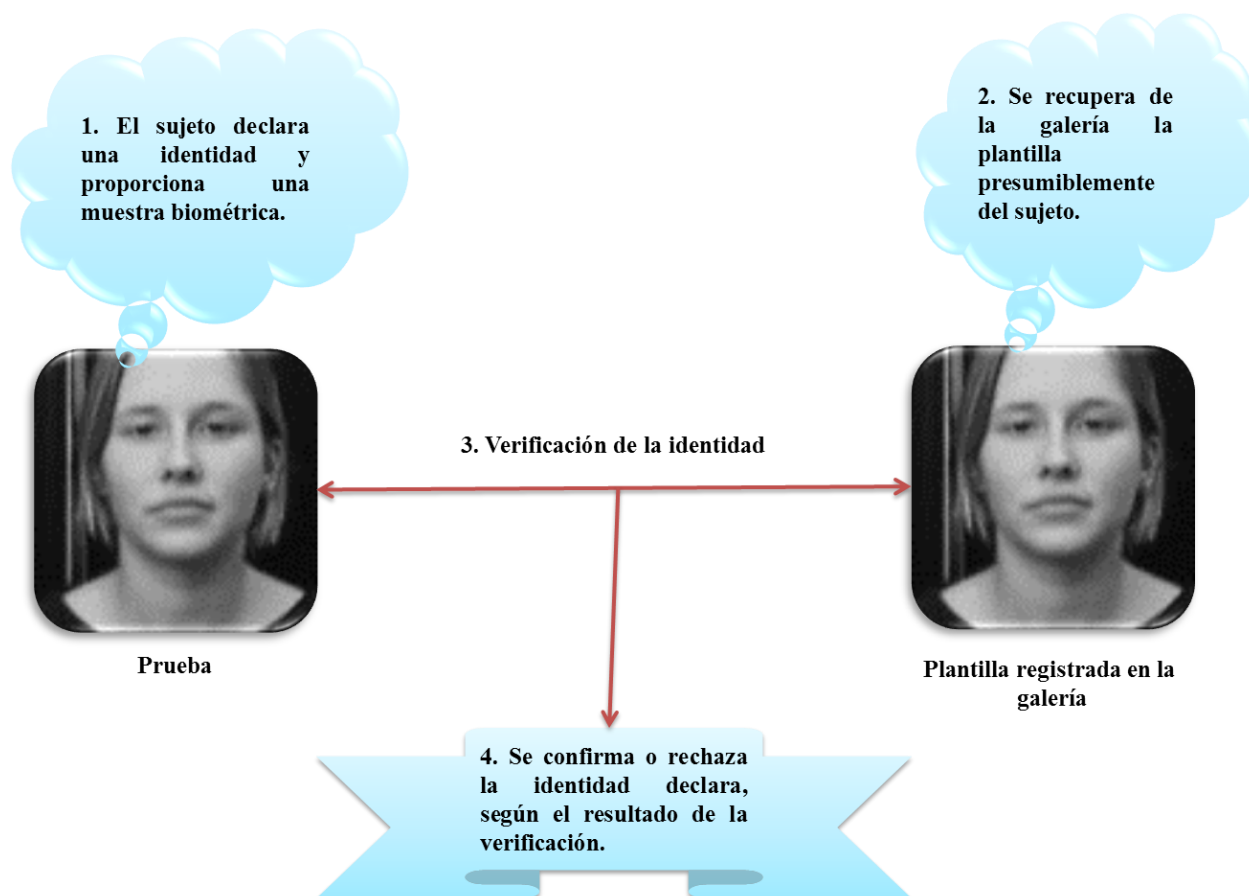


Figura 3.3. Proceso de verificación facial¹

La verificación facial se usa en aplicaciones donde el sujeto está consciente de que tiene que proporcionar una muestra biométrica y declara una identidad. Esto tiene aplicaciones para el acceso a un edificio gubernamental, acceso a una aplicación computacional, entre otros. Por lo general, este tipo de sistemas se instala en ambientes controlados.

En la tarea de identificación facial (véase figura 3.4), una prueba es comparada con cada una de las plantillas biométricas faciales almacenadas en una base de datos con el fin de determinar si existe o no una correspondencia. Si existe una correspondencia con un fuerte grado de similitud, entonces se devuelve la identidad de la persona. En otro caso se indica que la persona no se encuentra entre los conocidos por el sistema.

¹ Esta figura contiene imágenes faciales de Weyrauch et al. (2004)

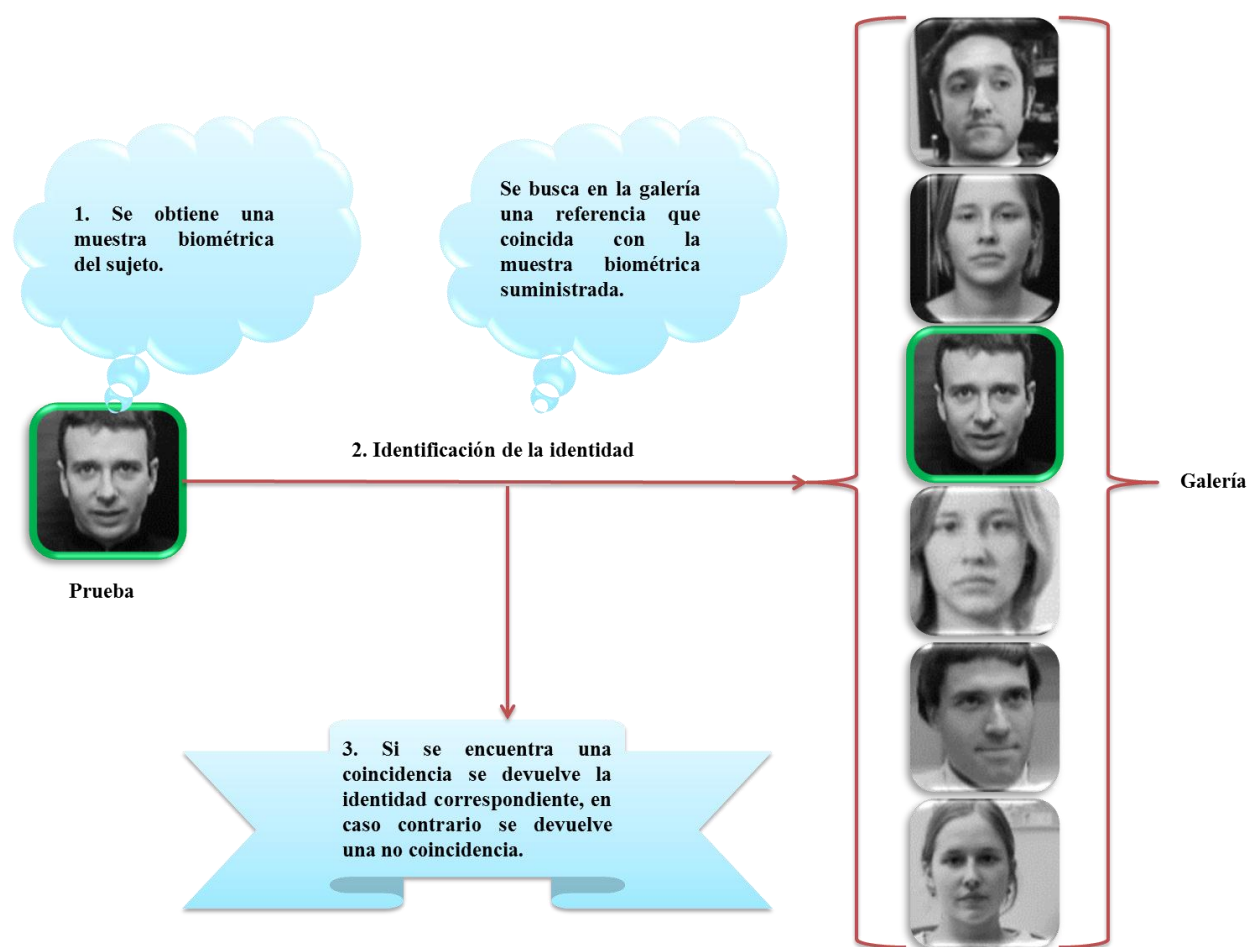


Figura 3.4. Proceso de identificación facial².

Los sistemas de identificación facial se pueden implantar con el objetivo de, por ejemplo, buscar criminales, buscar niños perdidos, entre otros. En este tipo de sistemas los sujetos pueden no estar conscientes de la existencia de dicho sistemas ni de que están siendo identificados.

La base de datos que almacena las plantillas biométricas se denomina técnicamente como galería. Las plantillas biométricas son registros faciales de una persona que será reconocida. Este registro puede consistir de imágenes faciales completas o simplemente de características.

3.1.1 Historia del reconocimiento facial

El primer sistema semi-automatizado para el reconocimiento facial fue desarrollado en 1960. Este sistema requería que una persona localizara las características (ojos, nariz, boca, etc.)

² Esta figura contiene imágenes faciales de Weyrauch et al. (2004).

sobre una imagen facial antes de calcular sus distancias y proporciones para un punto de referencia en común. En los años setentas, se establecieron 21 marcadores específicos, tales como el color del cabello y la distancia de los ojos. El problema con ambas soluciones fue el cálculo manual de las medidas y localizaciones de las características.

Sirovich et al. (1987) aplicaron un método conocido como análisis de componentes principales el cual se basa en representar las caras como una combinación lineal de ciertas caras básicas denominadas Autocaras. Posteriormente, Turk y Pentland (1991) descubrieron que mientras se usaba la técnica de Autocaras, el error residual puede ser usado para detectar caras en imágenes. Este fue un descubrimiento que posibilitó el desarrollo de sistemas de reconocimiento facial automatizado en tiempo real.

La gran mayoría de los sistemas de reconocimiento facial, tanto comerciales como de investigación, usan imágenes bidimensionales (2D) a color o de intensidad (escala de grises). Otro enfoque que ha surgido es el que utiliza las imágenes faciales en tercera dimensión (3D). Este enfoque representa la información tridimensional de diferentes formas (véase Figura 3.5): a) como una nube de puntos en el espacio, los cuales se pueden representar en forma de malla, donde cada punto es un vértice y a su vez los vértices unidos por aristas, b) mapas de profundidad o imágenes de rango. Estas imágenes son en escala de grises, donde el nivel de intensidad de cada píxel está relacionado con la coordenada de profundidad z de un punto tridimensional, y c) la información tridimensional registrada con la imagen de textura.

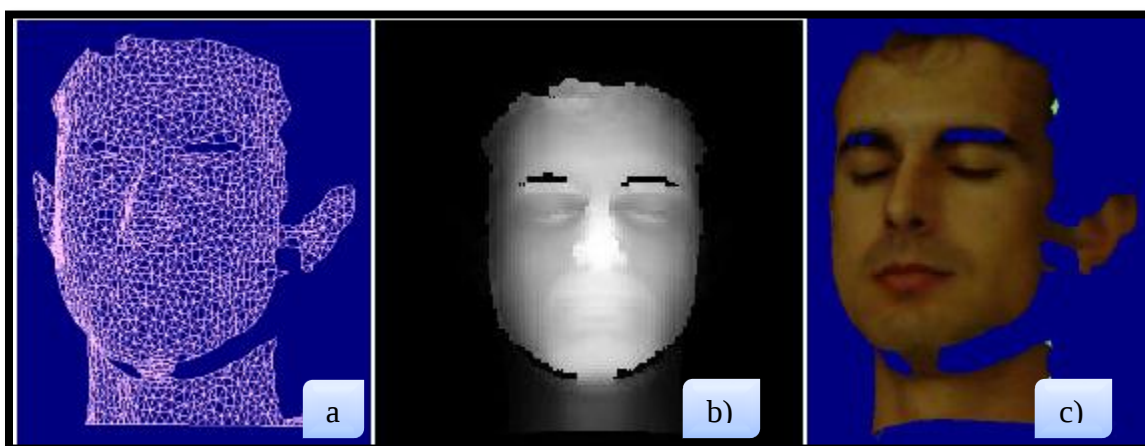


Figura 3.5. Representación 3D de imágenes faciales: a) Nube de puntos en el espacio, b) mapa de profundidad, c) mapa de puntos más textura.

El enfoque de reconocimiento facial 3D inicio al principio de la década de los noventas, específicamente en 1991, que es a partir de entonces cuando se tienen trabajos reportados. Esto fue gracias al abaratamiento de los sensores de adquisición 3D y el aumento en su capacidad de procesamiento. Este enfoque aunque ya tiene dos décadas, aún se encuentra en desarrollo pues existe menos información comparado con los enfoques basados en 2D.

El reconocimiento facial mediante filtros de correlación se basa en la capacidad de los filtros para ofrecer invariancia a distorsiones. Vijaya et al. (2004) aplicaron filtros de correlación en el reconocimiento biométrico. También, Savvides et al. (2002) presentan un trabajo de verificación facial usando un filtro MACE que fue capaz de producir mejores resultados que algunos de los algoritmos más estudiados en la literatura. Otro trabajo en donde usan la correlación para el reconocimiento facial es el que se presenta Kiat Ng (2005), en este trabajo se usó el filtro UOTSDF para desarrollar un sistema de reconocimiento facial en tiempo real sobre una PDA. Estos trabajos de reconocimiento facial mediante filtros de correlación están limitados a imágenes faciales estáticas y frontales, considerando solo variaciones en iluminación y expresiones faciales.

Desde el año 2005 a la fecha se han reportado en la literatura diversos trabajos sobre reconocimiento facial con nuevos enfoques, como el marco de trabajo que presenta (Nair et al., 2009) para la detección y el reconocimiento facial, localización de características y un buen registro de mallas faciales mediante el ajustamiento de un modelo facial basado en un modelo de distribución de puntos en tercera dimensión 3D. Gorodnichy (2005) presenta un marco de trabajo basado en video para el reconocimiento facial, el cual se basa en el principio de la neuro-asociación empleado por el cerebro humano para lidiar con la baja resolución y la baja calidad de los cuadros individuales de imágenes del video, de esta forma, tanto la memorización y el reconocimiento de datos se basan en un flujo de cuadros más que de un solo cuadro. IDNoticias (2010) presenta un sistema de reconocimiento facial para el control de acceso a recintos donde se realicen eventos públicos. Este proyecto combina el análisis biométrico de rostros con el empleo de la tecnología de identificación por radiofrecuencia (RFID).

Actualmente, tenemos aplicaciones de reconocimiento facial en teléfonos inteligentes, visas para acceso a países y en investigación criminal y militar. Los gigantes como Google, Apple y Facebook ahora incluyen reconocimiento facial en sus productos y el desarrollo

comercial de sensores faciales de bajo costo ha sido de gran apoyo para el progreso en este campo.

Aunque el reconocimiento facial ya tiene muchas décadas de desarrollo, aún se encuentra lejos el día en que existan algoritmos que ofrezcan desempeños óptimos, es decir, que tengan la misma precisión de reconocimiento que un humano. Por tal motivo el reconocimiento facial es una disciplina de gran auge en la actualidad que merece la atención de la comunidad científica.

3.1.2 Principales enfoques

Hay dos enfoques predominantes en el reconocimiento facial: geométrico (basado en características) y fotométricas (basado en vistas). A través del tiempo se han desarrollado diferentes algoritmos, tres de los más estudiados en la literatura son: Análisis de componentes principales (PCA, Principal Component Analysis, Sirovich y Kirby (1987)), análisis discriminante lineal (LDA, Linear Discriminant Analysis, Lu et al. (2003)) y acoplamiento de grafos de mallas elásticas (EBGM, Elastic Bunch Graph Matching, Wiskott et al. (1999)). A continuación se presenta una descripción detallada PCA y LDA, además se describe también trabajos donde aplican filtros de correlación en el reconocimiento facial.

3.1.2.1 Análisis de componentes principales

PCA extrae la información relevante en una imagen facial, lo codifica eficientemente como sea posible y lo compara con una cara codificada de forma similar almacenada en una base de datos de modelos faciales. En términos matemáticos, se desea encontrar los componentes principales de la distribución de caras o los eigenvectores de la matriz de covarianza del conjunto de imágenes faciales, tratando una imagen como un punto (o vector) en un espacio dimensional muy grande.

Con PCA las imágenes de prueba y las imágenes de la galería deben ser del mismo tamaño y deben ser normalizadas previamente para alinear los ojos y la boca de los sujetos en las imágenes. La aproximación de PCA es luego utilizado para reducir la dimensión de los datos por medio de fundamentos de compresión de datos y revela la más efectiva estructura de baja dimensión de los patrones faciales.

Esta reducción en las dimensiones quita información que no es útil y descompone de manera precisa la estructura facial en componentes ortogonales conocidos como Autocaras. Cada imagen facial puede ser representada como una suma ponderada (vector de rasgo) de los Autocaras, las cuales son almacenadas en un conjunto 1D.

Una imagen de prueba es comparada con una gallería de imágenes midiendo la distancia entre sus respectivos vectores de rasgos. La aproximación PCA típicamente requiere la cara completa y de frente para producir buenos resultados; de otra forma la imagen dará un resultado de bajo rendimiento del algoritmo.

A continuación se resume el algoritmo para reconocimiento facial mediante los componentes principales:

1. Que el conjunto de entrenamiento de imágenes faciales sea $\Gamma_1, \Gamma_2, \dots, \Gamma_M$.
2. La cara promedio del conjunto se define por $\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n$.
3. Cada cara difiere del promedio por el vector $\Phi_i = \Gamma_i - \Psi$.
4. Este conjunto de vectores muy grandes es entonces sujeto a análisis de componentes principales, el cual busca un conjunto de M vectores orto-normales, u_n , el cual describe de la mejor la distribución del dato. El k –ésimo vector, u_k , se selecciona tal que: $\lambda_k = \frac{1}{M} \sum_{n=1}^M (u_k^T \Phi_n)^2$, el cual es un máximo, sujeto a $u_l^T u_k = \delta_{lk} = \begin{cases} 1, & \text{si } l = k \\ 0, & \text{en otro caso} \end{cases}$
5. Los vectores u_k y los escalares λ_k son los eigenvectores y eigenvalores, respectivamente, de la matriz de covarianza C.
6. Obtenemos la matriz de covarianza $C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T$, donde $A = [\Phi_1, \Phi_2, \dots, \Phi_M]$.
7. Siguiendo este análisis, construimos la matriz $L = A^T A$, donde $L_{mn} = \Phi_m^T \Phi_n$, y encontramos los M eigenvectores, v_l , de L.
8. Una nueva cara (Γ), se transforma en sus componentes Autocaras (proyectado dentro del espacio facial) mediante una simple operación: $w_k = u_k^T (\Gamma - \Psi)$, donde $k = 1, 2, \dots, M'$. Los pesos forman un vector $\Omega^T = [w_1, w_2, \dots, w_{M'}]$ que describen la contribución de cada eigencara en la representación de la imagen facial de entrada. Por lo tanto, este vector puede usarse, junto con un algoritmo de reconocimiento de patrones para encontrar en un número predefinido de clases el que mejor describe la cara.

9. El método más simple para determinar qué clase de cara provee la mejor descripción de una imagen facial de entrada es encontrar la clase facial k que minimice la distancia euclidiana: $\epsilon_k^2 = ||\Omega - \Omega_k||^2$, donde Ω_k es un vector que describe la k –ésima clase facial.

Con los experimentos que se han hecho en la literatura, se ha mostrado un buen desempeño de este enfoque en entornos controlados y con pocos usuarios. Turk y Pentland (1991) presentan un marco de trabajo usando este algoritmo para reconocimiento facial, el cual muestra buen desempeño bajo escenarios controlados y capaz de registrar de forma “on-line” a nuevos sujetos.

3.1.2.2 Análisis discriminante lineal Fisher

LDA es un enfoque estadístico para clasificar muestras de clases desconocidas en base a muestras de entrenamiento con clases conocidas. Este algoritmo intenta encontrar el conjunto óptimo de vectores w_i que maximicen la separación entre clases mientras que de forma simultánea minimicen la separación de los elementos dentro de las clases. Esto se logra maximizando la ecuación:

$$J_{FLDA}(W) = \frac{|W^T S_B W|}{|W^T S_W W|} \quad (27)$$

donde S_B y S_W se definen como:

$$S_B = \sum_{i=1}^c (\mu_i - \mu)(\mu_i - \mu)^T \quad (28)$$

$$S_W = \sum_{i=1}^c \sum_{j=1}^{N_i} (X_j^i - \mu_i)(X_j^i - \mu_i)^T \quad (29)$$

donde N_i , μ_i y μ con el número de imágenes de entrenamiento para la i –ésima clase, la media de la i –ésima clase y la media global de todas las clases respectivamente. Cuando se aplica FLDA al reconocimiento facial, la dimensión del dato d es típicamente mayor que el número de datos muestra N . Esta situación crea un problema de deficiencia en S_W . Para evitar este problema, se puede ejecutar PCA sobre los datos para reducir sus dimensiones a $N - c$ y luego ejecutar FLDA. Finalmente, FLDA está dado por la siguiente ecuación:

$$W_{FLDA} = \arg \max_W \frac{|W^T W_{PCA}^T S_B W_{PCA} W|}{|W^T W_{PCA}^T S_W W_{PCA} W|} \quad (30)$$

$$W_{Fisherface}^T = W_{FLDA}^T W_{PCA}^T \quad (31)$$

3.1.3 Reconocimiento facial usando filtros de correlación

Los filtros de correlación ofrecen varias ventajas para el reconocimiento facial. La más importante de ellas es la invariancia al desplazamiento en la operación de filtrado, ya que si la imagen de entrada se traslada una cierta cantidad de píxeles, entonces la salida del filtro se desplaza la misma cantidad de píxeles. Cuando se usan los filtros de correlación, el desplazamiento se estima fácilmente mediante la localización de los picos de correlación (Vijaya et al., 2004).

En la Figura 3.6 se muestra el proceso de la correlación en el reconocimiento facial, dicha operación involucra el uso de un conjunto de entrenamiento para el diseño de un filtro y de una imagen de prueba. Tanto a la imagen de prueba como al conjunto de imágenes de entrenamiento se les aplica la transformada de Fourier (TF), para posteriormente realizar una multiplicación elemento a elemento, con la inversa de la transformada rápida de Fourier (ITF) de este producto se obtiene el plano de correlación.

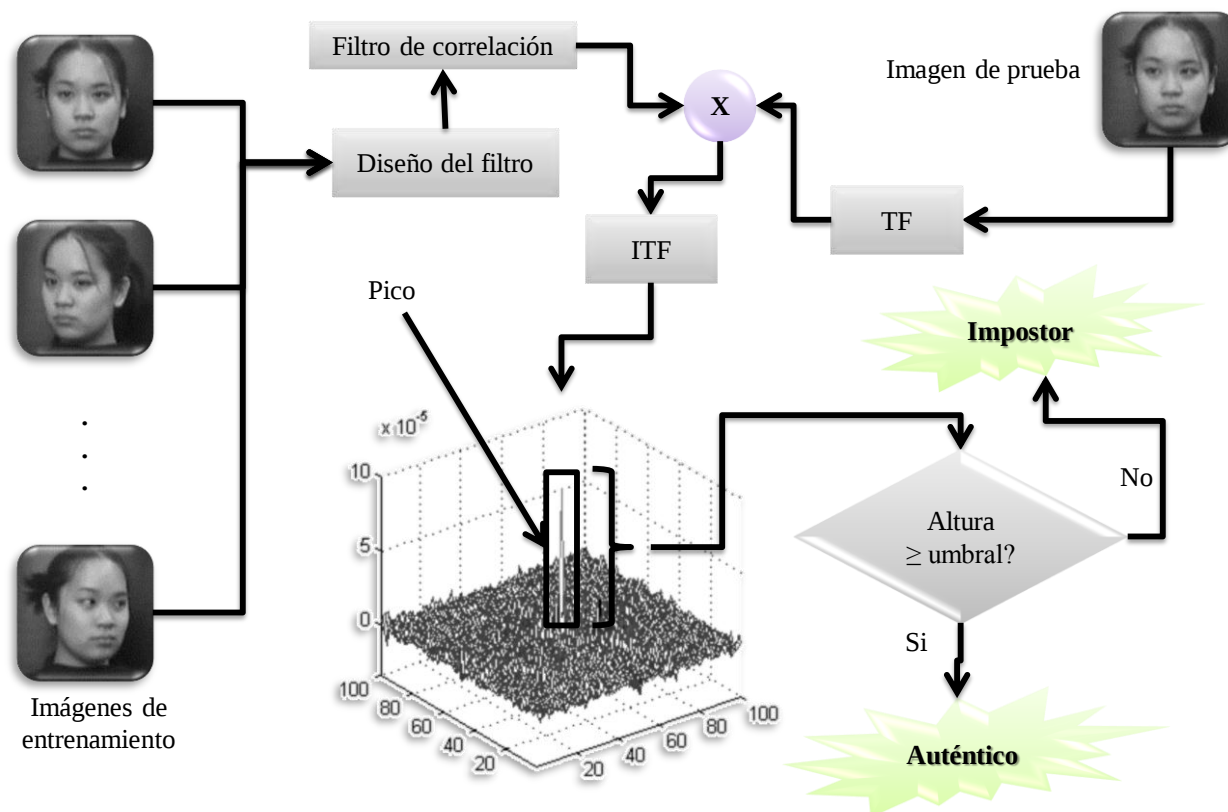


Figura 3.6. Diagrama de bloques del proceso de correlación en el reconocimiento facial.

En el proceso de análisis de la salida de correlación, se buscan los picos para determinar sus alturas mediante el uso de alguna de las métricas de desempeño para determinar si un objeto de entrada pertenece a una clase verdadera o a una clase falsa.

Savvides et al. (2002) presentan un trabajo de verificación facial mediante filtros de correlación. La simulación mostró resultados muy prometedores, demostrando que los filtros de correlación pueden ser una solución alternativa para realizar la tarea de verificación facial. Las ventajas que se mostraron son la invariancia al desplazamiento y la habilidad para suprimir imágenes faciales impostoras usando un umbral universal.

Otro trabajo en donde se demuestra el uso de filtros de correlación para verificación e identificación facial se presenta en Kiat Ng (2005). En este trabajo se desarrolló un sistema de reconocimiento facial en tiempo real sobre una PDA. Con dicho trabajo se demuestra que reconocimiento facial mediante filtros de correlación sobre dispositivos con poder computacional y recursos de memoria limitados.

3.1.4 Medidas de desempeño para algoritmos de reconocimiento facial

Una forma de conocer la efectividad de esos sistemas es midiendo el desempeño de sus algoritmos de forma estadística (NSTC: Subcommittee on Biometrics and Identity Management, 2008). Existen varios protocolos de evaluación de algoritmos de reconocimiento facial, dos de las más importantes son el FRVT (Phillips et al., 2007) y los protocolos de evaluación FERET (Phillips et al., 1999). Los protocolos establecen cómo realizar una evaluación y cómo se calculan los resultados.

Las evaluaciones de los algoritmos de reconocimiento facial no solo proporcionan una imagen en el tiempo de las capacidades del reconocimiento facial, sino también dirigen el continuo avance de la tecnología de reconocimiento facial en el mundo (NSTC: Subcommittee on Biometrics and Identity Management, 2008).

El desempeño de los algoritmos de reconocimiento facial se reporta sobre las tareas de verificación e identificación. Cada tarea tiene sus propias medidas de desempeño. Calcular el desempeño de un algoritmo de reconocimiento facial requiere de tres conjuntos de imágenes faciales. El primero es la galería G (base de datos) el cual contiene plantillas biométricas faciales de personas conocida por el sistema. Los otros dos son conjuntos de pruebas. Una prueba es una muestra biométrica facial que se presenta al sistema para reconocimiento. El primer conjunto de prueba se denomina P_G el cual contiene muestras biométricas de personas registradas en la galería (estas muestra son diferentes a las que están en la galería). El otro conjunto de prueba se denomina P_N , el cual contiene muestras biométricas de personas que no están en la galería.

Una galería G consiste de un conjunto de plantillas biométricas faciales $\{g_1, g_2, \dots, g_{|G|}\}$, con una plantilla por persona. Cuando una prueba P_j se presenta a un sistema, ésta se compara con toda la galería (identificación) o únicamente con alguna plantilla de la galería en específico (verificación). La comparación entre una prueba P_j y una plantilla biométrica facial g_i produce un grado de similitud s_{ij} . Si P_j es una muestra biométrica de una persona que está en la galería, entonces considere g^* como el único acoplamiento en la galería. El grado de similitud entre P_j y g^* se denota como s_{*j} .

3.1.4.1 Identificación

El desempeño de un algoritmo de identificación facial se caracteriza mediante dos estadísticas de desempeño: la tasa de identificación y detección (TID) y la tasa de falsa alarma (TFA).

En la identificación los grados de similitud entre P_j y la galería se examinan y ordenan. Una prueba P_j tiene un rango n si s_{*j} es la n – ésima medida de similitud más alta. Esto se denota por $rank(P_j) = n$. Se considera $rank = 1$ al acoplamiento más fuerte.

Consideremos el caso donde la identidad de una prueba es de alguien en la galería (es decir, $P_j \in P_G$). Una prueba es detectada e identificada si la prueba es identificada correctamente y el grado de similitud está por encima de un umbral τ . La tasa de identificación y detección está dada por:

$$TID = \frac{|\{p_j: rank(p_j) = 1 \text{ y } s_{*j} \geq \tau\}|}{|p_G|} \quad (32)$$

La segunda estadística de desempeño es la tasa de falsa alarma. La tasa de falsa alarma provee el desempeño cuando una prueba es de alguien que no está en la galería (es decir, $p_j \in p_N$). Este tipo de prueba también se le conoce como un impostor. Una falsa alarma ocurre cuando una medida de similitud de un impostor está arriba del umbral τ .

La tasa de falsa alarma es la porción de pruebas $p_j \in p_N$ que son falsas alarmas. Esto se calcula por:

$$TFA = \frac{|\{p_j: \max_i s_{ij} \geq \tau\}|}{|p_N|} \quad (33)$$

Lo ideal sería tener una tasa de detección e identificación de 1.0 y una tasa de falsa alarma de 0; todas las personas en la galería son detectadas e identificadas y no hay falsas alarmas.

3.1.4.2 Verificación

Los resultados obtenidos en la verificación facial se obtienen siguiendo el protocolo Round Robin. En el protocolo Round Robin, los conjuntos de prueba p_G y p_N son el mismo conjunto y son referidos como P . Se calculan todos los grados de similitudes entre la galería y las muestras del conjunto de prueba. Los resultados que sean mayores al umbral de reconocimiento y

la imagen de prueba y la plantilla en la galería pertenecen a la misma clase son usados para calcular la tasa de verificación. Los resultados que son mayores al umbral de reconocimiento y que la imagen de prueba y la plantilla no pertenecen a la misma clase se usan para calcular la tasa de falsa aceptación. El desempeño de un algoritmo de verificación facial se caracteriza por las estadísticas: tasa de verificación (TV) y tasa de falsa aceptación ($TFAc$).

La tasa de verificación, para el método Round Robin, está dada por:

$$TV = \frac{|\{p_j: s_{ij} \geq \tau, id(g_i) = id(p_j)\}|}{|P|} \quad (34)$$

Donde $id(g_i)$, $id(p_j)$ representan la identidad de la plantilla biométrica almacenada en la galería y la muestra facial de prueba, respectivamente. La tasa de falsa aceptación está dada por:

$$TFAc = \frac{|\{s_{ij}: s_{ij} \geq \tau \text{ y } id(g_i) \neq id(p_j)\}|}{(|P| - 1)|G|} \quad (35)$$

3.2 Seguimiento facial

En algunas aplicaciones computacionales, es necesario identificar y seguir a una persona en una secuencia de video. Esto hace posible inferir automáticamente la actividad que realiza una persona en un escenario, así como su identidad.

Una secuencia de video es una colección de imágenes estáticas y una de sus propiedades más importantes es la continuidad en el tiempo (Chellappa et al., 2005). De esta forma, podemos usar algoritmos de reconocimiento facial basados en imágenes estáticas para identificar a una persona que aparece en dicha secuencia, mientras que un algoritmo de seguimiento se encarga de localizar la ubicación de la cara identificada en cada imagen de la secuencia de video.

El seguimiento facial ha cobrado gran interés tanto en la comunidad científica como en desarrolladores de tecnologías biométricas y de entretenimiento. Existen varias estrategias de seguimiento, como el basado en partículas, seguimiento visual incremental (Ross et al., 2008), seguimiento basado en fragmentos (Adam et al., 2006), aprendizaje discriminativo basado en grafos (Zhang et al., 2007) y la estrategia de múltiples instancias (Babenko et al., 2009).

En esta sección describimos brevemente dos estrategias de seguimiento facial reportados en la literatura para introducir al lector al tema.

3.2.1 Seguimiento de caras humanas mediante búsqueda logarítmica en rejilla

La técnica de búsqueda logarítmica en rejilla (Vicente et al., 2002) lleva acabo una reducción logarítmica del espacio de búsqueda de los objetos de interés con el objetivo de ejecutar en tiempo real los procesos de localización y seguimiento facial en secuencias de video a color. Este método se apoya en el color de la piel humana para delimitar la region facial y en la Figura 3.7 se muestra el esquema general del procesos de seguimiento facial.

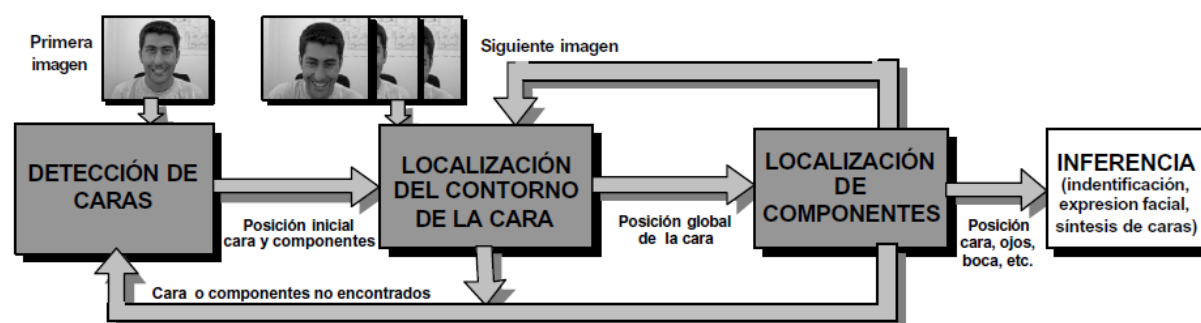


Figura 3.7. Esquema general del proceso de seguimiento facial mediante búsqueda logarítmica. Tomado de (Vicente et al., 2002).

Esta técnica asume una detección de la cara y un modelo HSI para color de la piel. Luego de la detección, se procede a la localización del contorno de la cara en la siguiete imagen de la secuencia mediante la técnica de búsqueda logarítmica. La búsqueda logarítmica consiste básicamente de dos fases: primero se calcula el rectángulo que contiene la cara y posteriormente see obtiene el contorno. Una vez obtenida la región de la cara, se calculan sus ejes de simetría, que permiten la localización de características faciales.

Al ser una técnica basada en color, la tarea de la delimitación de la región facial puede ser muy costosa computacionalmente. Además de que no se considera los diferentes colores de la piel humana y los cambios de iluminación que pudieran afectar la apariencia de la cara.

3.2.2 Seguimiento facial de objetos usando filtros de correlación adaptativos

Al igual que el método anterior, este método asume una detección. Una vez detectado el objetivo, éste es seleccionado en base a una ventana de seguimiento centrado en el objeto objetivo. Durante el seguimiento, un objetivo puede cambiar su apariencia mediante el alteraciones de rotación, escala, pose y moviéndose a través de diferentes condiciones de iluminación o incluso sometido a deformaciones rígidas. Por lo tanto, los filtros necesitan

adaptarse rápidamente a con el fin de seguir los objetos. Para lograr este objetivo, en este método el entrenamiento y el seguimiento trabajan juntos y se realizan en el dominio de Fourier.

Con el objeto detectado en la primera imagen de la secuencia de video, se generan ocho perturbaciones con las cuales se construye un filtro MOSSE (véase sección 2.8) que técnicamente es la ventana de seguimiento. El objetivo es seguido correlacionando el filtro sobre una ventana de búsqueda en la siguiente imagen. La ubicación correspondiente al valor máximo en la salida de correlación, indica la nueva posición del objetivo. Entonces se realiza una actualización del filtro basado en la nueva posición.

Para detectar fallas, por ejemplo oclusión y/o desaparición del objeto en la imagen, los autores usan la medida de desempeño PSR. Un objeto auténtico produce valores PSR en el intervalo de 20 a 60. Los valores PSR alrededor de 7 indican oclusión o que el seguimiento falló.

Los autores de este método emplean las siguientes operaciones de pre-procesamiento sobre las imágenes. Primero, los valores de los pixeles se transforman usando una función log que ayuda con las situaciones de bajo contraste de iluminación. Los valores de pixeles se normalizan para tener un valor medio de 0.0 y una norma de 1. Finalmente, la imagen es multiplicada por una ventana de coseno la cual reduce gradualmente los valores del pixel cercanos a los bordes a cero.

Una virtud de esta técnica es la rapidez con que se realiza el seguimiento, 669 imágenes por segundo, y con la capacidad de recuperarse en caso de una falla. En la Figura 3.8 se muestran algunos resultados de esta estrategia donde la cara presenta variaciones.

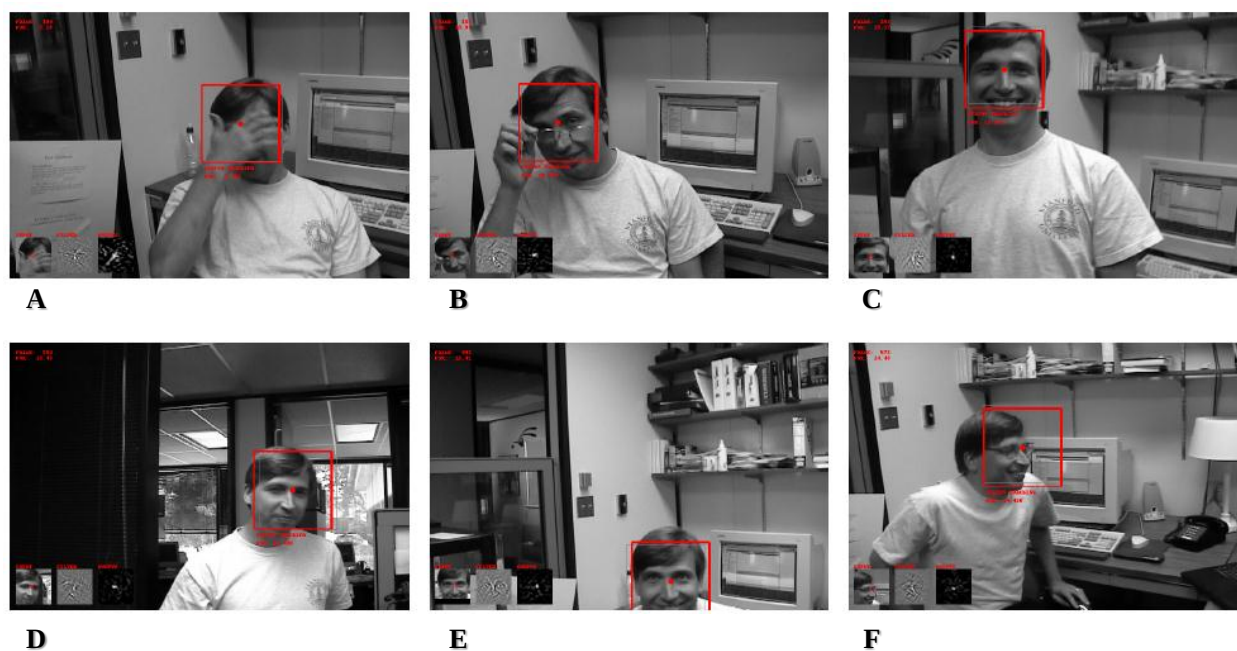


Figura 3.8. Desempeño de MOSSE ante distorsiones faciales. A, B, E) Oclusión, C) Expresión, D) Pose y F) Pose y expresión. Tomado de Bolme et al. (2010).

Capítulo 4

4. Verificación facial usando filtros de correlación e imágenes faciales de prueba con distorsiones

La apariencia de la cara de una persona puede diferir considerablemente en diferentes entornos. Este cambio de apariencia puede ser causado por variaciones en iluminación debido diferentes fuentes de luz, ruido causado por el sensor de visión al momento de capturar la imagen facial o expresiones (cuando la persona habla, canta, ríe, llora, hace gestos, entre otros). A estas variaciones se les denomina distorsiones.

Cuando una persona se registra en sistema de reconocimiento facial, generalmente se hace en un escenario controlado. De tal forma que las muestras faciales capturadas para la construcción de la plantilla biométrica contienen distorsiones mínimas. Sin embargo, en la etapa de reconocimiento en un escenario donde no exista un control, las imágenes faciales de prueba pueden aparecer con distorsiones mayores y difíciles de predecir o calcular. Este problema disminuye el desempeño de los algoritmos de reconocimiento facial.

Un buen algoritmo de reconocimiento facial es capaz de reconocer aquellas imágenes faciales que pertenecen a la misma clase que con las que fue entrenado y de rechazar todas aquellas imágenes faciales de otras clases. Sin embargo, ésta es una tarea desafiante cuando las imágenes faciales de prueba presentan alguna de las distorsiones descritas anteriormente.

Esta sección tiene como objetivo mostrar el desempeño de los filtros de correlación K-Law, MACE, UMACE, OTSDF y UOTSDF en el reconocimiento de imágenes faciales de prueba con distorsiones. Se consideran variaciones en expresiones, detalles faciales, iluminación y ruido blanco Gaussiano aditivo.

4.1 Configuración del experimento

En este experimento se usaron imágenes faciales de la base de datos ORL (Olivetti Research Laboratory, Samaria et al. (1994)) la cual contiene 10 imágenes faciales diferentes de 40 personas. Algunas de las imágenes fueron tomadas en tiempos diferentes, ligeras variaciones en iluminación, expresiones faciales (ojos abiertos/cerrados, sonrisa/no sonrisa) y detalles

faciales (lentes/no lentes). La posición de las personas en las imágenes faciales son frontales (con algunos ligeros movimientos laterales) y con los ojos arriba. Cada imagen facial fue recortada y escalada manualmente a una resolución de 64×64 píxeles. Todas las imágenes son en tonos de gris con un rango dinámico de 0-255.

Con el objetivo de llevar a cabo la evaluación usando las estadísticas para la verificación facial reportadas en la literatura por organismos especializados (vease sección 2.9) se creó una galería de plantillas biométricas faciales y seis conjuntos de prueba. La galería se compone de 40 filtros de correlación, una para cada persona en la base de datos ORL. Cada filtro de correlación se construyó con las 10 imágenes faciales diferentes de cada persona. Cada conjunto de prueba contiene $40 \times 10 = 400$ imágenes faciales de personas desconocidas a las cuales hay que verificar su identidad.

El primer conjunto de prueba (denominado CP1) contiene las imágenes faciales originales de la base de datos ORL. El segundo conjunto de prueba (CP2) fue creado agregando aleatoriamente sombras a los datos del conjunto CP1. Al agregar ruido blanco Gaussiano aditivo al conjunto CP1, fueron creados los conjuntos de prueba CP3, CP4, CP5 y CP6 con media 0 y varianzas de 5, 10, 15 y 20 píxeles, respectivamente. Esta información se resume en la Tabla 4.1.

Tabla 4.1. Distorsiones presentadas por las imágenes faciales de los conjuntos de prueba.

Conjuntos de prueba	Variaciones presentadas por las imágenes faciales							
	Expresiones	Detalles (Oclusión parcial)	Iluminación	Sombras	Ruido blanco Gaussiano aditivo			
					5	10	15	20
CP1	×	×	×					
CP2	×	×	×	×				
CP3	×	×	×		×			
CP4	×	×	×			×		
CP5	×	×	×				×	
CP6	×	×	×					×

Cada filtro se evaluó siguiendo la metodología representada en la Figura 4.1. Para cada persona se construyó un filtro de correlación usando 10 imágenes faciales, el cual es la plantilla biométrica. Las estadísticas de desempeño se calcularon correlacionando el filtro de cada sujeto con el conjunto de consulta. De este paso, por cada filtro se producen 400 salidas de correlación de los cuales 390 corresponden a impostores. Por lo tanto, $400 \times 40 = 16000$ salidas de

correlación y solo 400 son auténticas para cada conjunto de prueba. Se usa un umbral de reconocimiento general calculado como se describe en la sección 4.1.1.

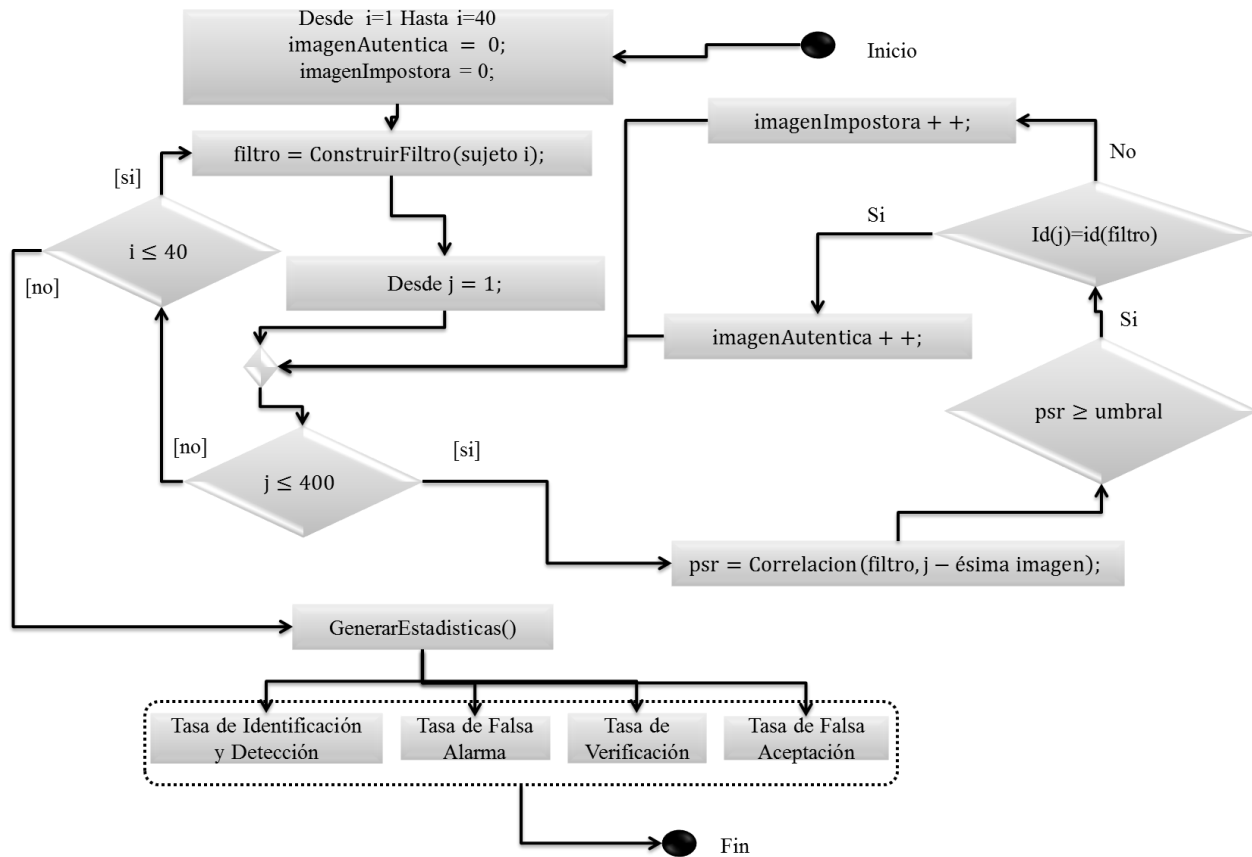


Figura 4.1. Metodología para la evaluación de filtros de correlación.

4.1.1 Umbral PSR de reconocimiento.

En la correlación de una plantilla biométrica facial con una imagen facial de prueba, un umbral PSR de reconocimiento τ_{psr} permite determinar si existe o no un acoplamiento. De esta forma, si el resultado de la comparación es mayor o igual que el umbral τ_{psr} entonces existe un acoplamiento y la imagen facial de prueba se determina como auténtica, en caso contrario se determina el no acoplamiento y la imagen facial de prueba se declara como impostora.

En este trabajo se considera un umbral de reconocimiento general que pueda ser usado para comparar el resultado PSR de cualquier correlación entre una imagen facial de prueba con una plantilla biométrica facial. Sea p_i una imagen facial de prueba de la misma clase que la plantilla g_i , y psr_j el grado de similitud (agudeza del pico de correlación calculado con la

métrica PSR, expuesto en la sección 2.9) entre p_i y g_i . Sea R el conjunto de todos los valores PSR de imágenes faciales auténticas, el cual está dado por la ecuación:

$$R = \{psr_j: psr_j = PSR(p_i \otimes g_i)\} \quad (36)$$

donde $\otimes$ representa la operación de correlación. Sea $\bar{x}$ y σ la media y la desviación estándar de R , respectivamente. El umbral PSR general está dado por:

$$\tau_{psr} = \bar{x} - \sigma \quad (37)$$

Se calculó el umbral PSR de reconocimiento general para cada filtro evaluado en este trabajo. La Tabla 4.2 muestra los umbrales calculados con 400 valores psr auténticos, estos valores fueron redondeados al entero más próximo.

Al emplear tres desviaciones estándar nos aseguramos de que al menos el 88%, según el teorema de Chebyshev, de los valores psr 's de los sujetos auténticos estarán dentro del umbral

Tabla 4.2. Umbral PSR de reconocimiento general para cada filtro evaluado.

Filtro de correlación	$\bar{x}$	σ	$\bar{x} - 3\sigma$
K-Law	19.63	3.0792	10
MACE	20.00	4.15	16
UMACE	29.43	6.89	9
OTSDF	21.44	3.61	11
UOTSDF	22.37	5.60	6

Los umbrales mostrados en la Tabla 4.2 se calcularon usando tres desviaciones estándar. Mientras mayor es el umbral de reconocimiento, la probabilidad de rechazar imágenes faciales impostoras es mayor. El método propuesto para el cálculo del umbral de reconocimiento es flexible para usar ya sea uno, dos o tres desviaciones estándar dependiendo de precisión con que se desea el reconocimiento.

4.2 Resultados experimentales

Cuando se correlaciona un filtro correspondiente a una persona con una imagen facial de prueba de la misma persona, la salida de correlación que se produce debe de exhibir un pico alto y agudo. Esto indica que la imagen facial es auténtica, es decir, la persona es realmente quien dice ser. Pero si el filtro y la imagen facial de prueba corresponden a personas diferentes, la salida de correlación debe de exhibir un pico bajo, o en el mejor de los caso no exhibir pico alguno. La

Figura 4.2 muestra una salida de correlación para una imagen facial auténtica y una salida de correlación para una imagen facial impostora.

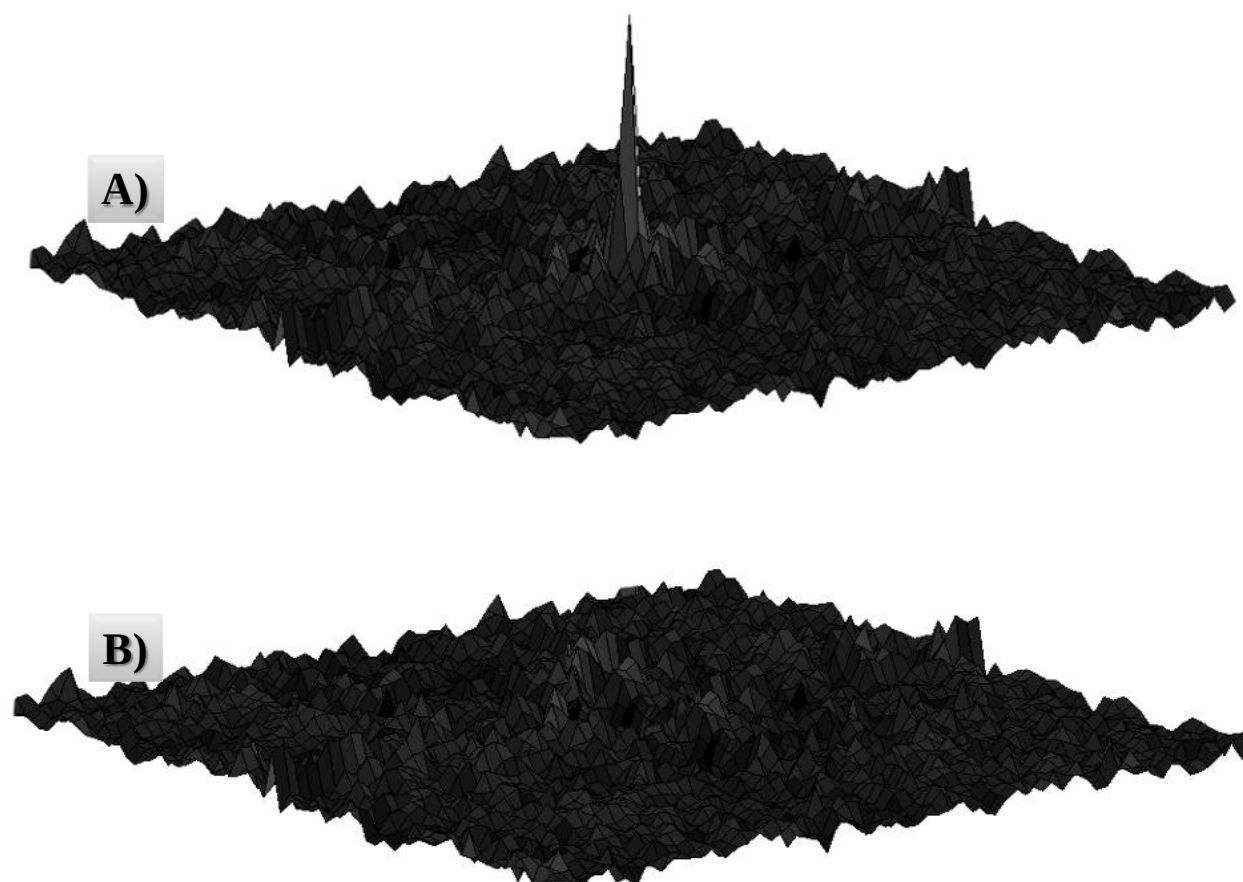


Figura 4.2. Salida de correlación: A) Persona auténtica y B) Persona impostora.

Cuando una imagen facial de prueba de una persona que no corresponda al filtro produce valores PSR iguales o mayores al umbral de reconocimiento PSR se dice que existe un falso reconocimiento. Es decir, una imagen facial impostora ha sido incorrectamente declarada como auténtica. En caso contrario, cuando una imagen facial auténtica de prueba produce un valor PSR menor al umbral de reconocimiento se dice que existe un falso rechazo.

4.2.1 Influencia de las expresiones faciales y variaciones en iluminación en la verificación facial

Cuando una persona habla, canta, murmura, ríe, hace algún gesto con la cara, o alguna otra acción que implique el movimiento de los músculos faciales, éstos afectan la apariencia de la cara. Estas acciones faciales son conocidas como expresiones, algunos ejemplos se presentan en

la Figura 4.3. La mayoría de estas expresiones muestran el estado de ánimo de la persona, lo que ha sido aprovechado por investigadores para desarrollar algoritmos que reconozcan el estado anímico (feliz, triste, desanimado, entre otros) o el grado de satisfacción del cliente al consumir un producto mediante el reconocimiento de sus expresiones faciales.



Figura 4.3. Misma persona, diferentes expresiones faciales

Si consideramos la cara como un objeto compuesto de características faciales, que éstas a su vez pueden ser otros objetos, entonces una expresión modifica la forma de dicha características. Por ejemplo, la boca tiene diferente forma al estar abierta y otra forma al estar cerrada. Dicho comportamiento de la cara hace que la mayoría de los algoritmos funcionen solo en ambientes restringidos y/o personas restringidas.

En esta sección se describe el desempeño de filtros de correlación con imágenes de prueba con expresiones, detalles faciales y variaciones en iluminación, tal y como es el conjunto de prueba CP1. Se presenta el mejor y el peor desempeño PSR para cada filtro.

Todas las imágenes de faciales correspondientes a una persona en la cual estamos interesados en verificar se les denomina clase auténtica, mientras que las imágenes faciales de otras personas se les considera como clases impostoras. En la Figura 4.4 se muestra el diagrama de cajas para el desempeño PSR de las 40 personas (clases auténticas) con el conjunto de prueba CP1. Cada caja representa valores PSR para una clase de cara y se obtuvo correlacionando el filtro de una persona con sus diferentes imágenes faciales.

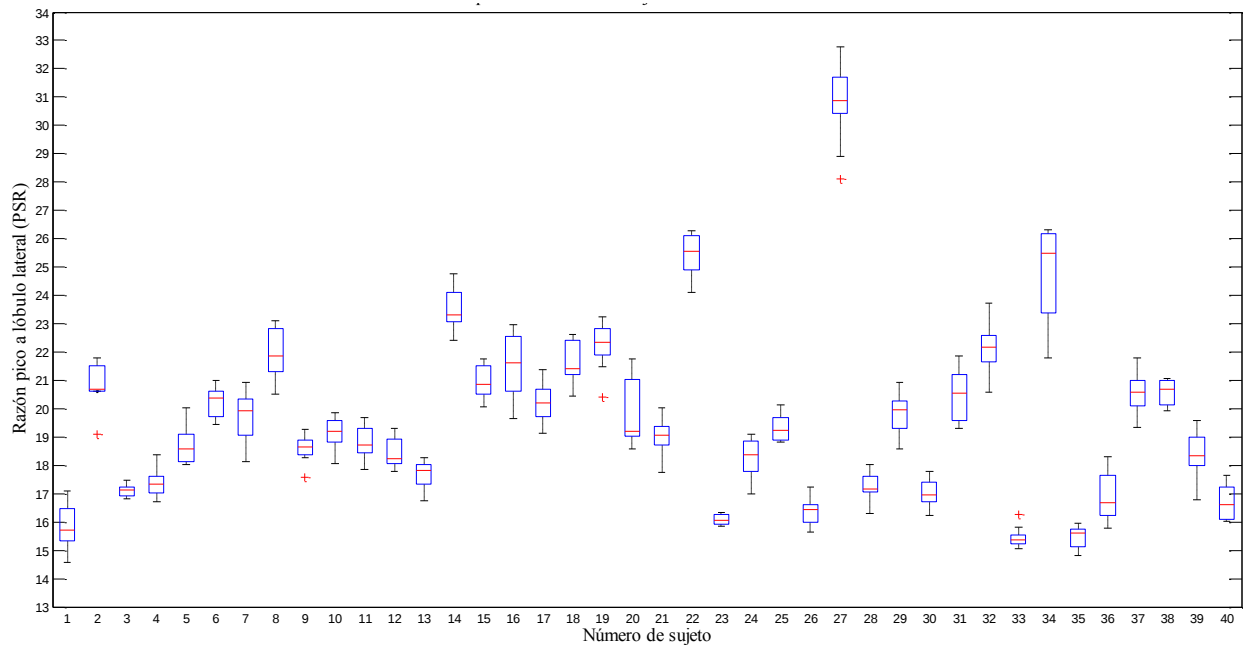


Figura 4.4. Desempeño PSR del filtro K-Law para las 40 personas en la base de datos ORL usando el conjunto de prueba CP1.

Como se puede observar, mientras más bajo sea el valor PSR que produce una imagen facial auténtica, mayor es el riesgo de que ésta sea rechazada. Podemos deducir fácilmente de la Figura 4.4 que el peor y el mejor desempeño para el filtro de correlación K-Law usando el conjunto de prueba CP1 corresponden a la persona 1 y a la persona 27, respectivamente.

De los diez valores PSR producidos por la persona 1, el más pequeño fue de 14.56 con la imagen de prueba 10. Esta imagen facial contiene un ligero pose, desplazamiento hacia la derecha y una pequeña rotación. Los valores PSR bajos se obtienen con imágenes faciales con variaciones en pose, tal es el caso de las imágenes faciales 2, 4, 5, 9 y 10. En la Figura 4.5 muestra el diagrama de cajas de la correlación del filtro de la persona 1 con las imágenes faciales de los 40 individuos, incluyendo a la persona 1.

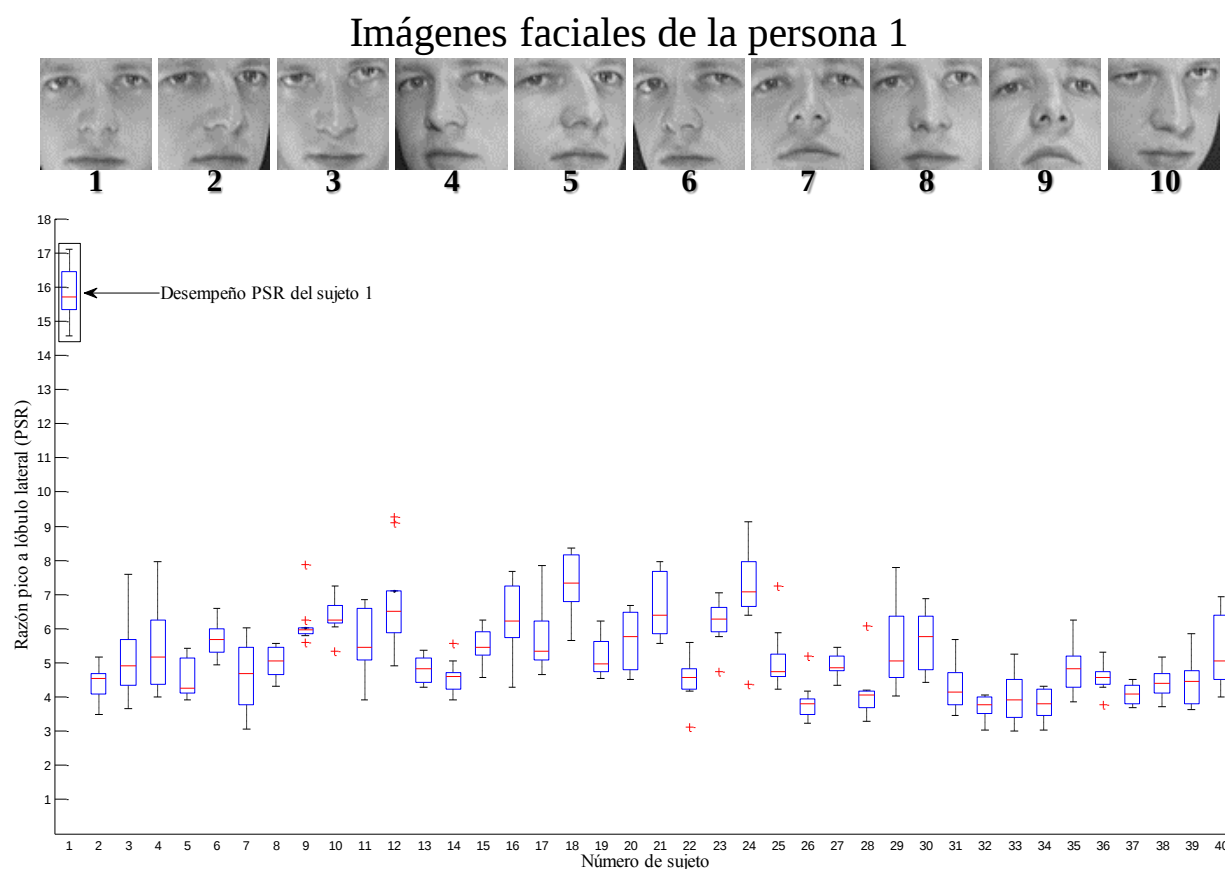


Figura 4.5. Peor desempeño PSR del filtro K-Law con el conjunto de prueba CP1.

Las imágenes faciales de la persona 27 (el mejor caso para el filtro K-Law) contiene variaciones menores. Éstas consisten en expresiones (ojos cerrados/abiertos y sonrisas/no sonrisas) lo cual no fueron difíciles de ser reconocidos lo que explica los altos valores PSR obtenidos. Se obtienen valores PSR en el intervalo de 28.12 a 32.79. Como se puede observar en la Figura 4.6, la separación que existe entre la clase auténtica y las clases impostoras es mayor que la presentada por el sujeto 1.

El filtro K-Law es capaz de producir valores PSR altos para sujetos auténticos y valores PSR bajos para impostores, esto se traduce en la separación existente entre estas clases aun en el peor de los casos (persona 1) del conjunto de pruebas CP1.

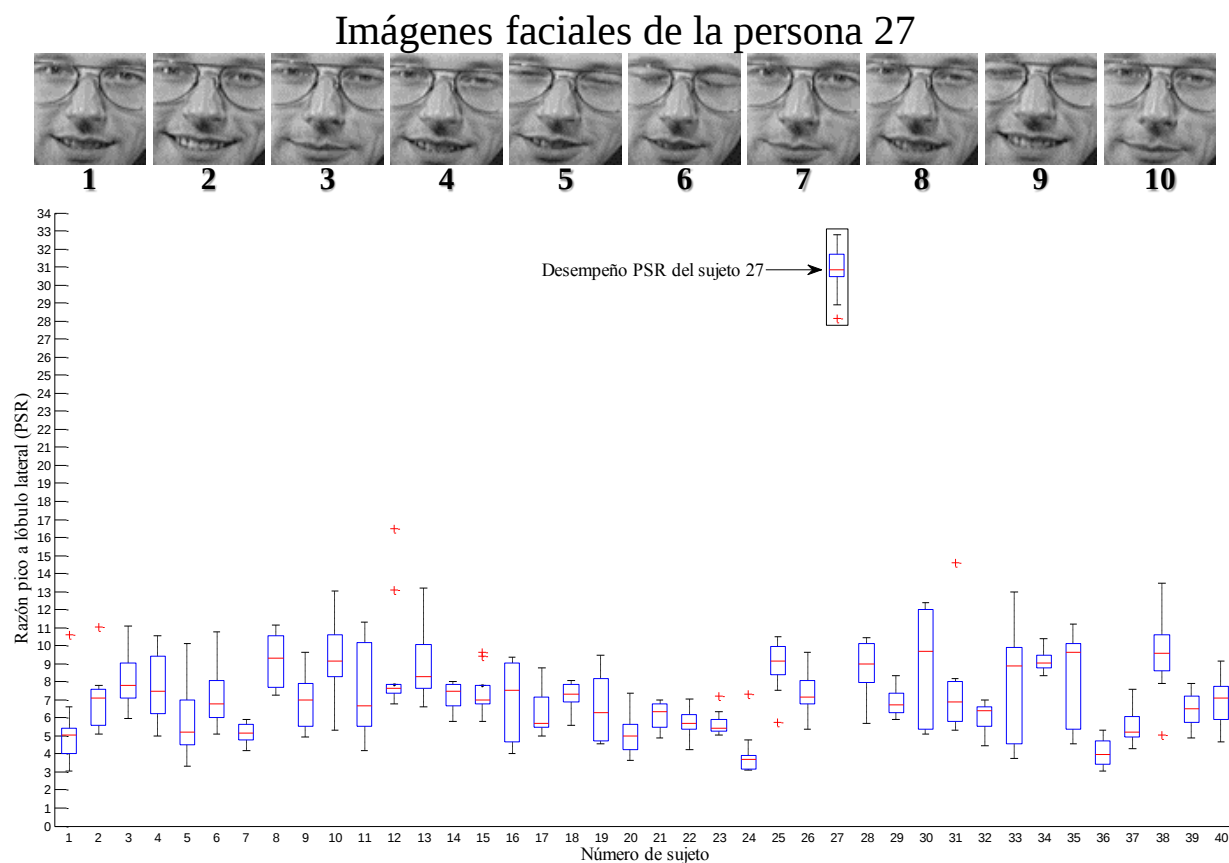


Figura 4.6. Mejor desempeño PSR del filtro K-Law con el conjunto de prueba CP1.

El filtro MACE produce picos de correlación más agudos, esto se infiere de los altos valores PSR que se obtiene para los sujetos auténticos presentados en el diagrama de cajas de la Figura 4.7. El peor desempeño lo obtuvo la persona número 36, mientras que al igual que en el filtro K-Law, el mejor desempeño fue para la persona 27.

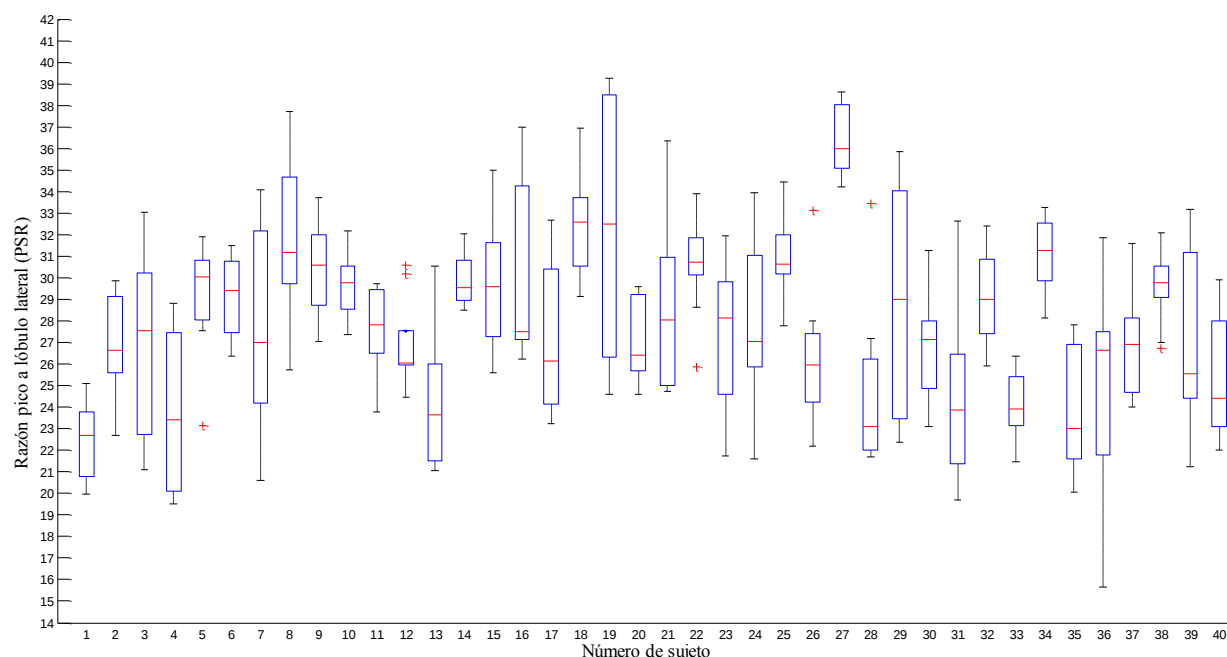


Figura 4.7. Desempeño PSR del filtro MACE para las 40 personas de la base de datos ORL usando el conjunto de prueba CP1.

Las imágenes faciales de la persona 36 contienen variaciones en iluminación y detalles faciales, tal y como se puede observar en la Figura 4.8. Los valores PSR más bajos se obtienen con las imágenes faciales con detalles, en este caso lentes. Esto se debe a que la mayoría de las imágenes de entrenamiento no contienen este detalle facial, por lo que el peso de las imágenes faciales 9 y 10 es menor en el filtro de correlación. En la Figura 4.9 se presenta el mejor caso para el filtro MACE. Como se puede observar, este sujeto produce valores PSR muy altos para el sujeto auténtico y los valores PSR no están muy dispersos con respecto a la media.

Imágenes faciales de la persona 36

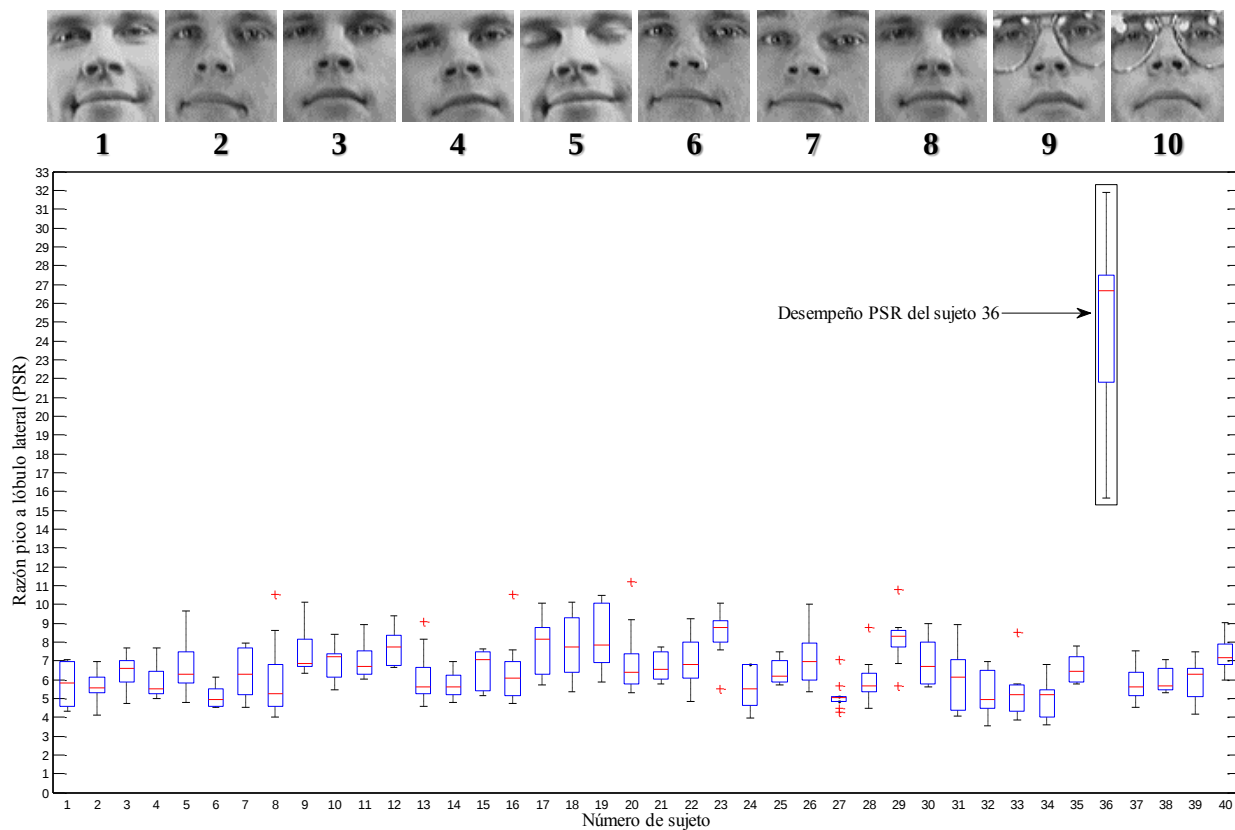


Figura 4.8. Peor desempeño PSR del filtro MACE (persona 36).

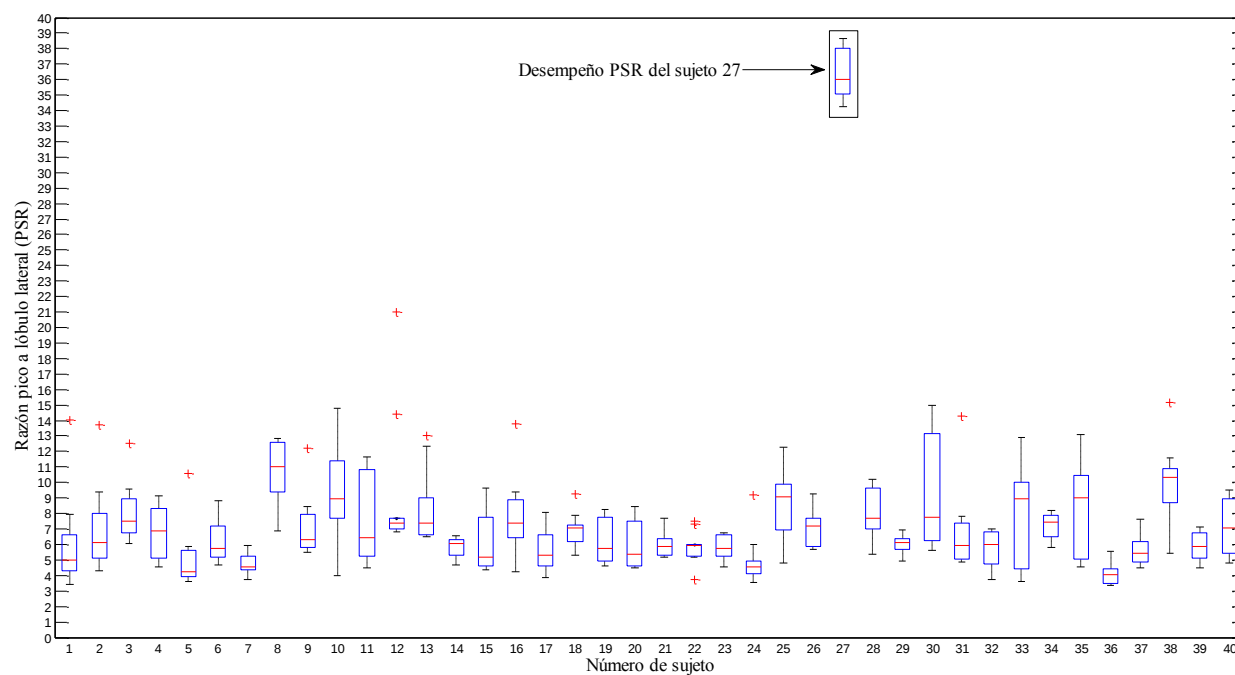


Figura 4.9. Mejor desempeño PSR del filtro MACE (persona 27).

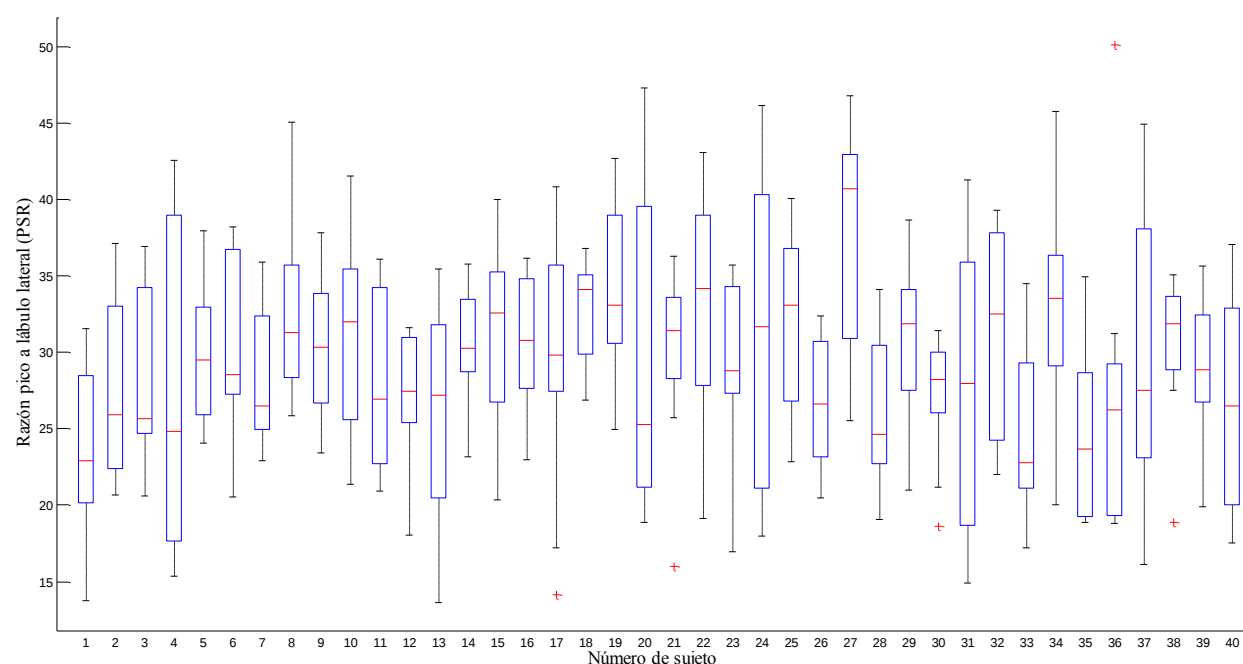


Figura 4.10. Desempeño PSR del filtro UMACE para 40 personas y el conjunto de prueba CP1.

El filtro UMACE produce valores PSR altos (véase Figura 4.10) al igual que el filtro MACE. La media de los valores PSR para imágenes faciales de prueba auténticas es mayor a 20, sin embargo los valores PSR son más dispersos con respecto a la media. El peor desempeño lo obtuvo la persona 13 (Figura 4.11) y el mejor desempeño lo obtuvo la persona 38 (Figura 4.12).

La persona 13 contiene detalles faciales y variaciones en pose. Esta variación provoca que se obtengan los bajos valores y que no se reconozca con la suficiente certeza a las mismas imágenes de entrenamiento. Las imágenes faciales de la persona 38 únicamente contienen variaciones pose, dicha variación es menor por lo que el filtro no tiene ningún problema en reconocer sus imágenes de entrenamiento.

Imágenes faciales de la persona 13

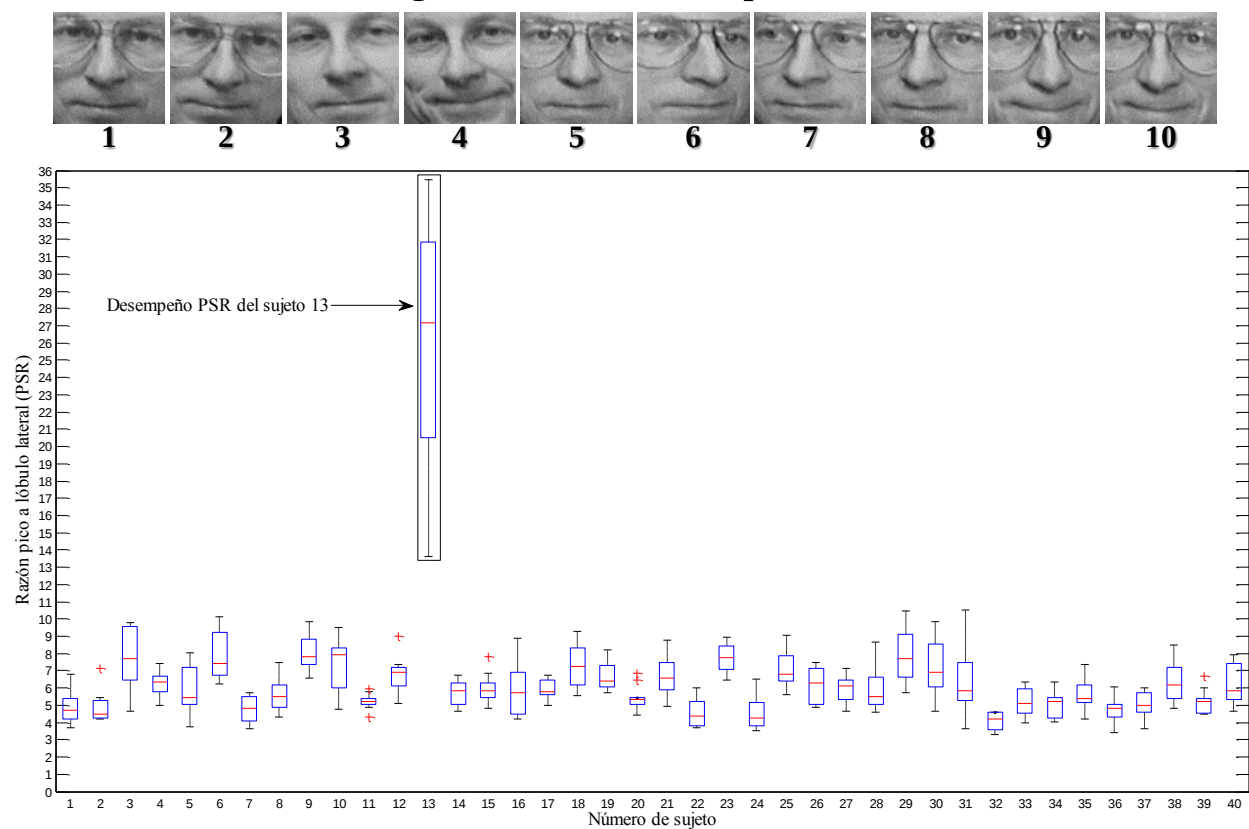


Figura 4.11. Peor desempeño PSR del filtro UFACE.

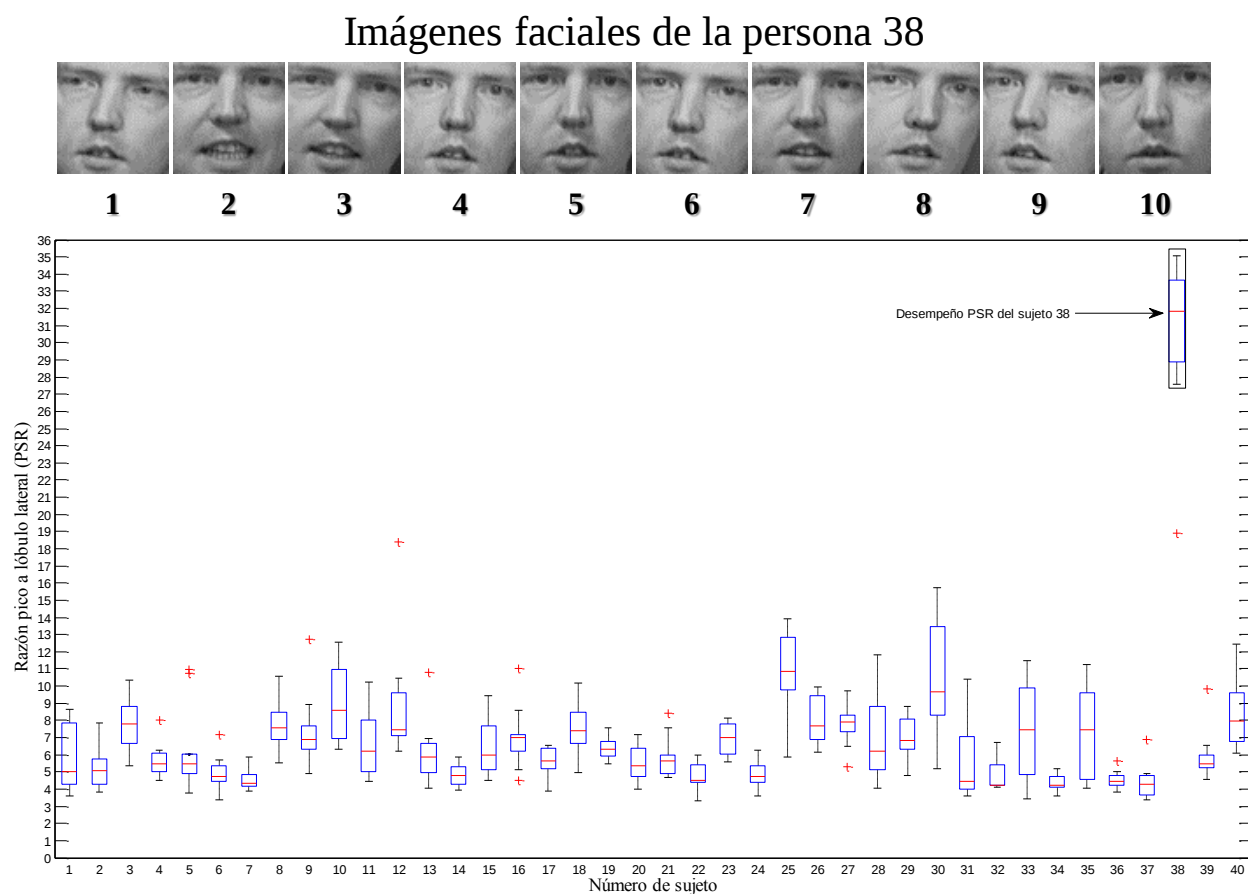


Figura 4.12. Mejor desempeño PSR del filtro UMACE.

El filtro OTSDF mantiene un compromiso entre el ruido y la agudeza del pico. Es decir, es tolerante a pequeños grados de ruido mientras produce un pico agudo. Esto es inversamente proporcional, ya que mientras mayor sea la tolerancia al ruido menor es la agudeza del pico producido. Con el conjunto de prueba CP1, el peor desempeño lo obtuvo la persona 36. Como se puede observar en la Figura 4.13, existe muy poco espacio entre el sujeto auténtico y los sujetos impostores. El mejor caso del filtro OSTDF fue para el sujeto 9 (Figura 4.14) cuyas imágenes faciales contienen no varían mucho entre ellas. Solo presentan pequeñas variaciones en pose.

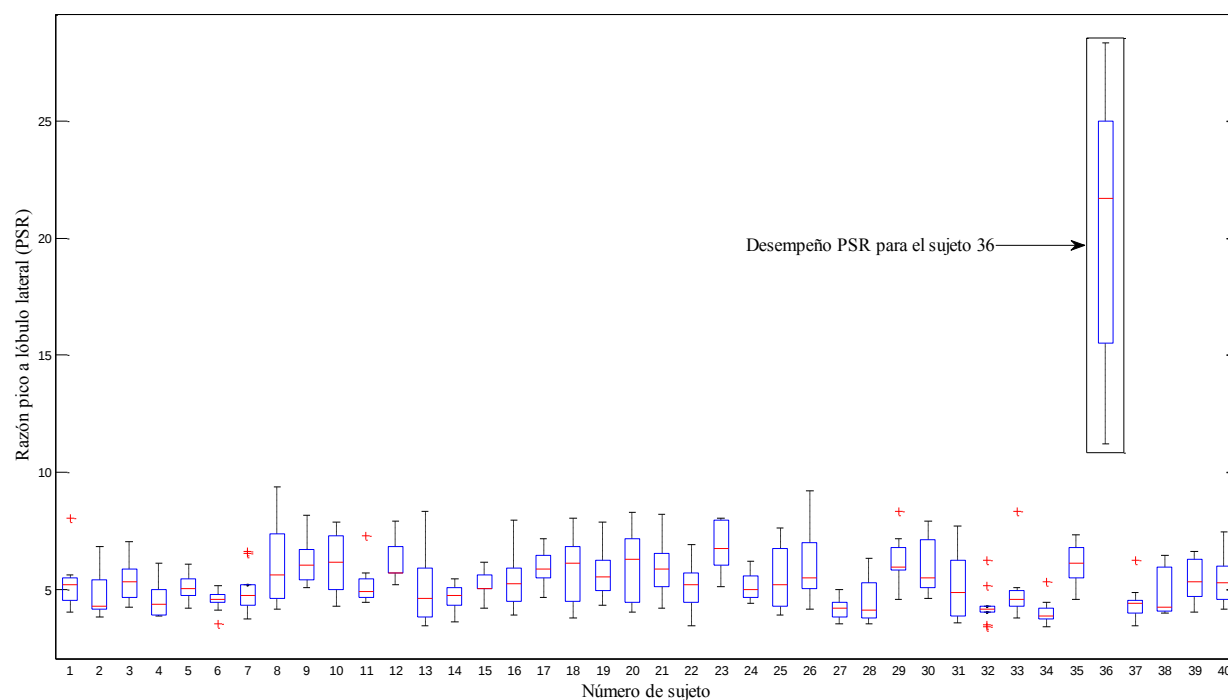


Figura 4.13. Peor desempeño PSR del filtro OTSDF.

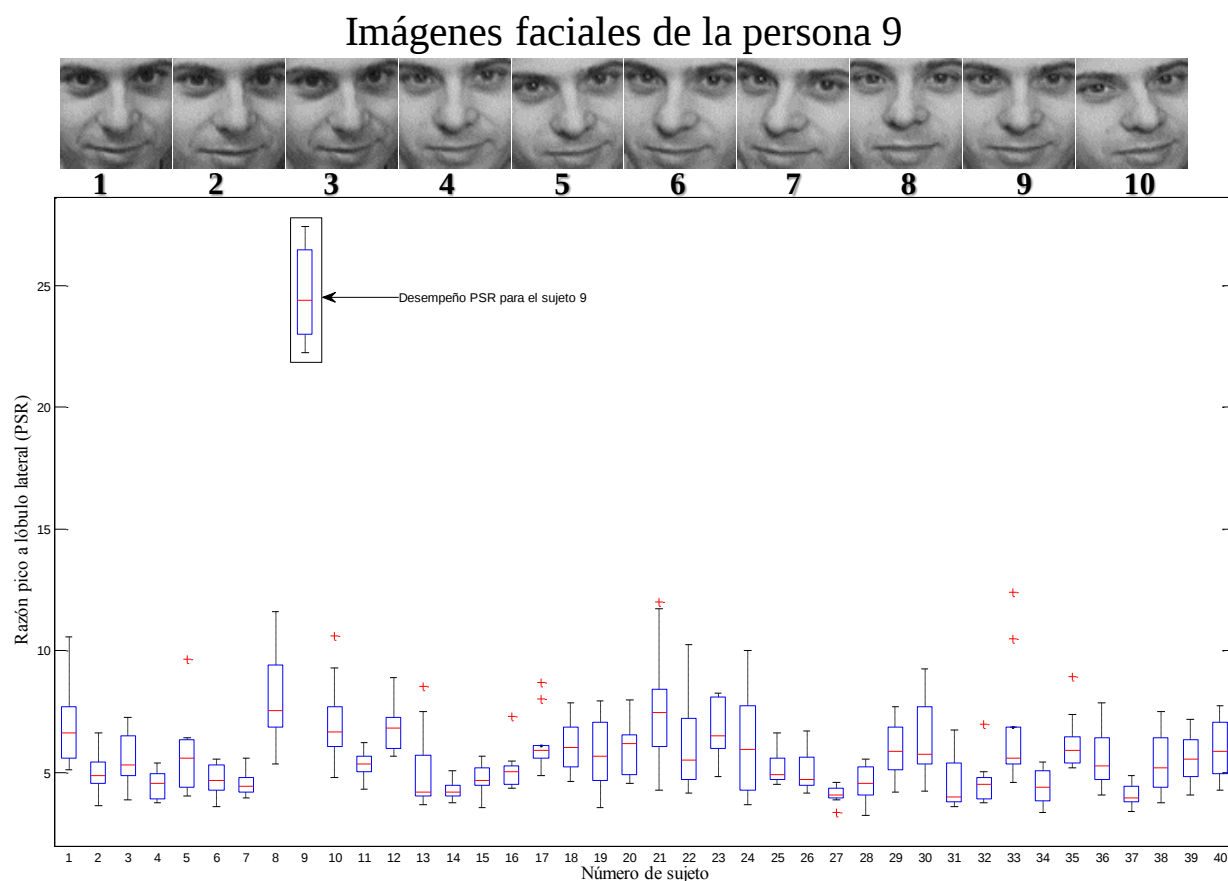


Figura 4.14. Mejor desempeño PSR del filtro OTSDF.

El filtro UOTSDF produce picos de correlación similares a su versión restringida. El peor desempeño para este filtro fue para la persona 13 (Figura 4.15) mientras que el mejor desempeño lo obtuvo la persona 18 (Figura 4.16). Una virtud que se observó de este filtro es que produce valores PSR muy bajos para sujetos impostores, los cuales son menores que 10.

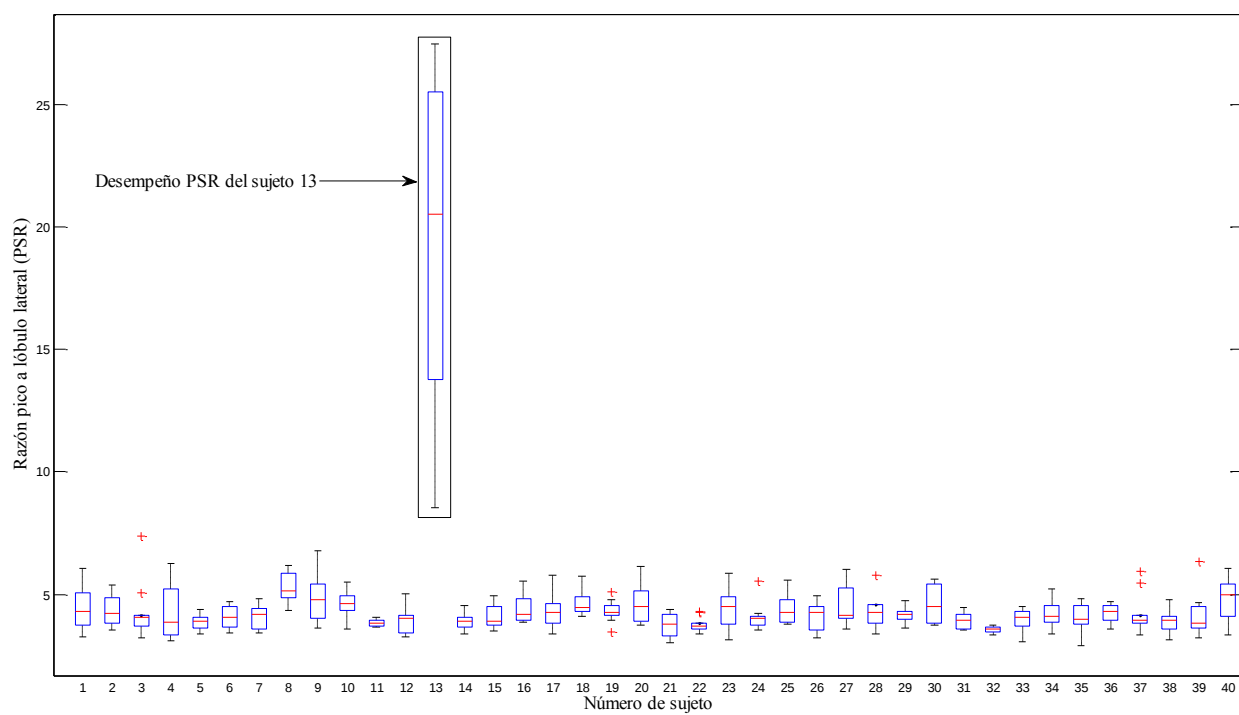


Figura 4.15. Peor desempeño PSR del filtro UOTSDF.

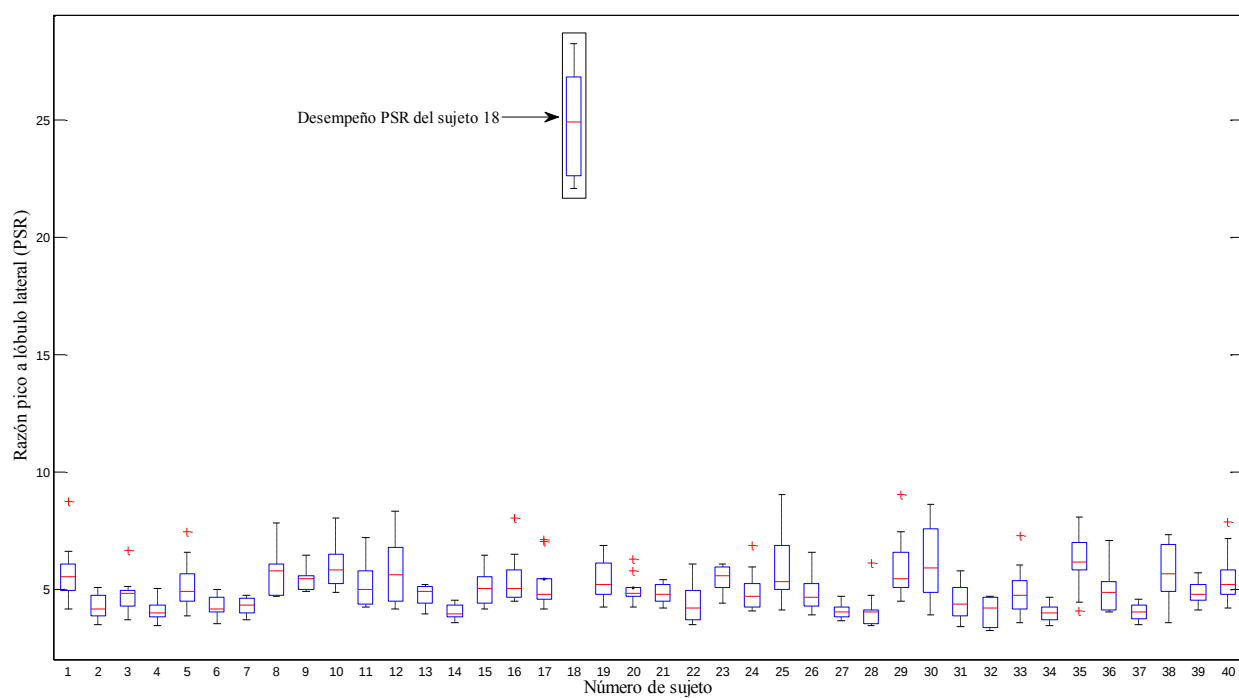


Figura 4.16. Mejor desempeño PSR del filtro UOTSDF.

4.2.2 Influencia del ruido blanco Gaussiano aditivo en la verificación facial

El ruido en la imagen es una variación aleatoria de la información de color o brillo (Wikimedia Foundation, 2011), y es un componente no deseado (Boncelet, 2005). El ruido más común en muchas señales es el ruido Gaussiano. Sea $i(x,y)$ una imagen facial que se puede descomponer en, $k(x,y)$ el componente deseado y $r(x,y)$ el componente ruido. Si sumamos los dos componentes obtenemos la imagen con ruido Gaussiano, es decir, tenemos una descomposición aditiva (de aquí el término ruido Gaussiano aditivo):

$$i(x,y) = k(x,y) + r(x,y) \quad (38)$$

En el modelo aditivo el ruido es independiente de $i(x,y)$. Algunos ejemplos de este modelo son ruido termal, ruido fotográfico y ruido de cuantización. En la Figura 4.17 se muestra un ejemplo de imagen facial a la cual se le añade ruido Gaussiano aditivo.

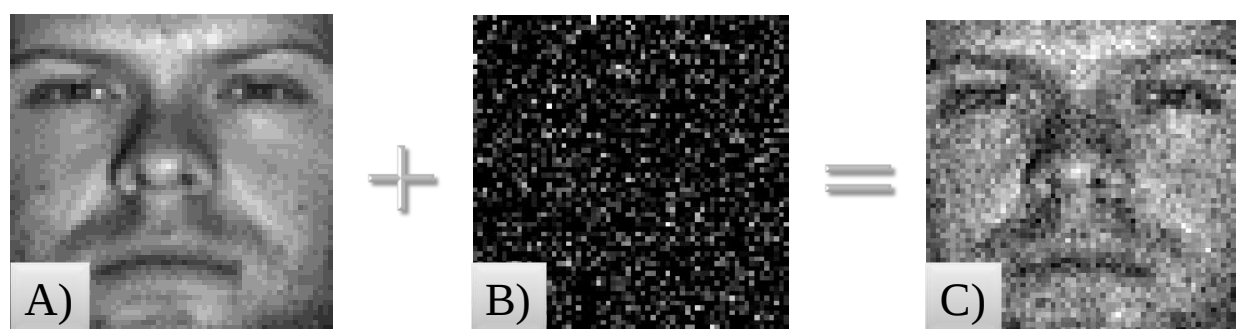


Figura 4.17. A) Imagen facial, B) Ruido Gaussiano aditivo, C) Imagen facial con ruido blanco Gaussiano aditivo.

Como se puede observar en la Figura 4.17-C, el ruido sobre la imagen facial afecta sobre todo la información de bordes que pudieran ser de gran utilidad en el reconocimiento. Los ojos ya no son distinguibles. Esta distorsión dificulta el proceso de reconocimiento. En la Figura 4.18 se presenta el resultado obtenido al correlacionar un filtro con dos imágenes faciales: Una con ruido Gaussiano aditivo y otra libre de ruido. La imagen con ruido produce un plano de correlación con un pico de correlación muy amplio y de menor altura. La imagen libre de ruido produce un pico de correlación alto y agudo, con un plano de correlación menos ruidosa.

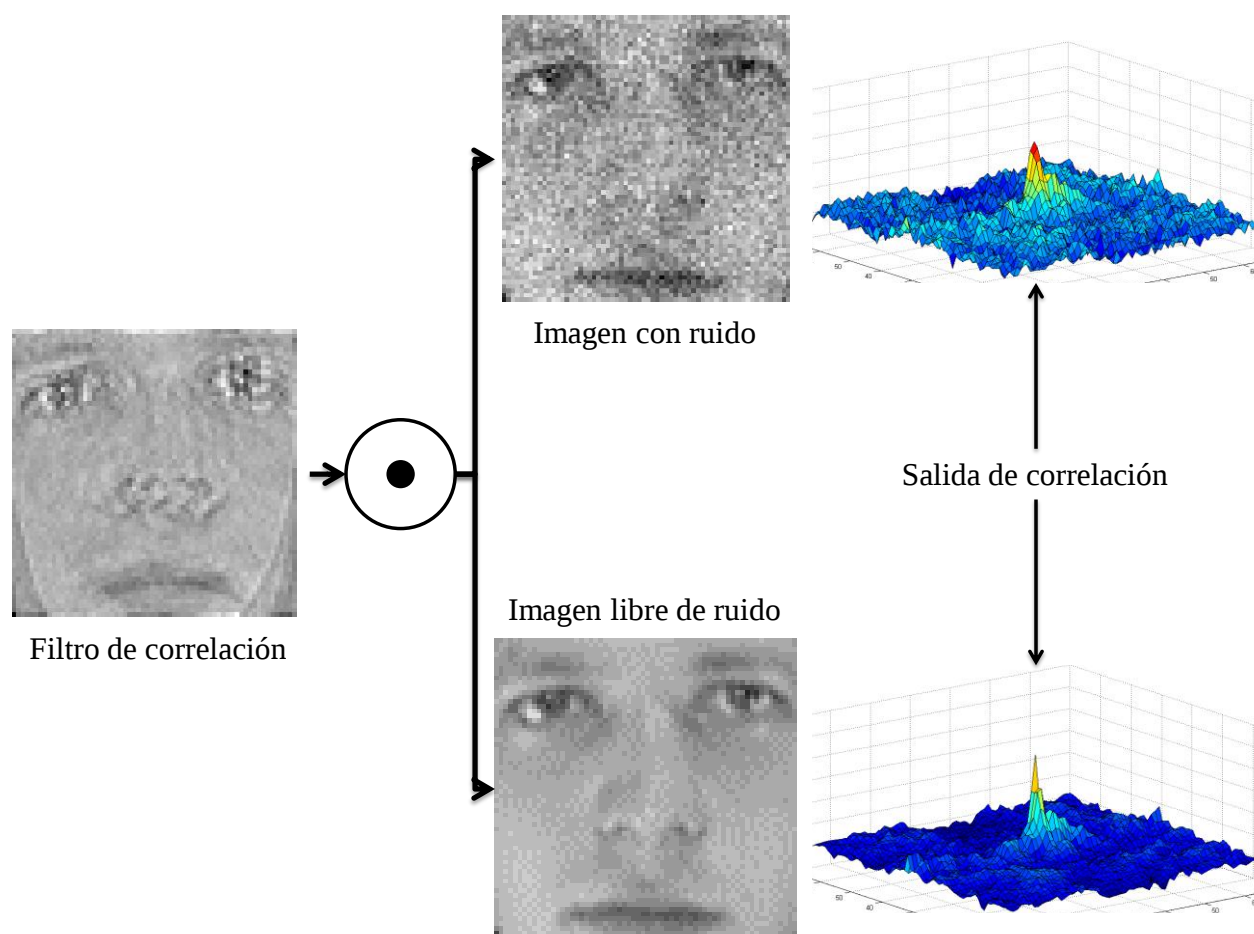


Figura 4.18. Correlación con una imagen con ruido Gaussiano aditivo y una imagen libre de ruido.

En la figura 4.19 se presenta una muestra de cuatro imágenes faciales con ruido. Cada imagen facial pertenece, de izquierda a derecha, a los conjuntos de prueba CP3, CP4, CP5 y CP6, respectivamente.

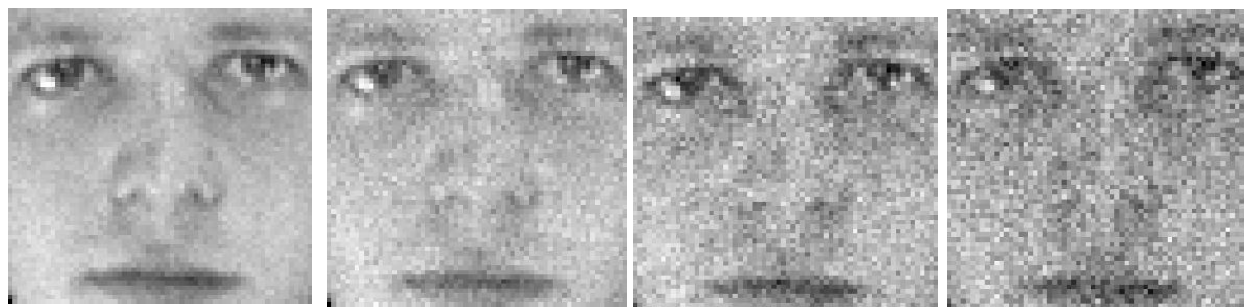


Figura 4.19. Misma imagen facial, diferentes varianzas en ruido; 5, 10, 15 y 20 píxeles, respectivamente.

Los filtros de correlación de cada persona se construyeron con imágenes faciales libres de ruido. Cuando este filtro es correlacionado con imágenes faciales de prueba afectadas por ruido blanco Gaussiano aditivo, su desempeño disminuye considerablemente. En la figura 4.20 se muestran los valores PSR producidos por un filtro de correlación K-Law e imágenes faciales de prueba afectadas por ruido blanco Gaussiano aditivo con media 0 y varianza de 20 pixeles. Las imágenes faciales de prueba con ruido producen valores PSR menores que aquellas que están libres de ruido.

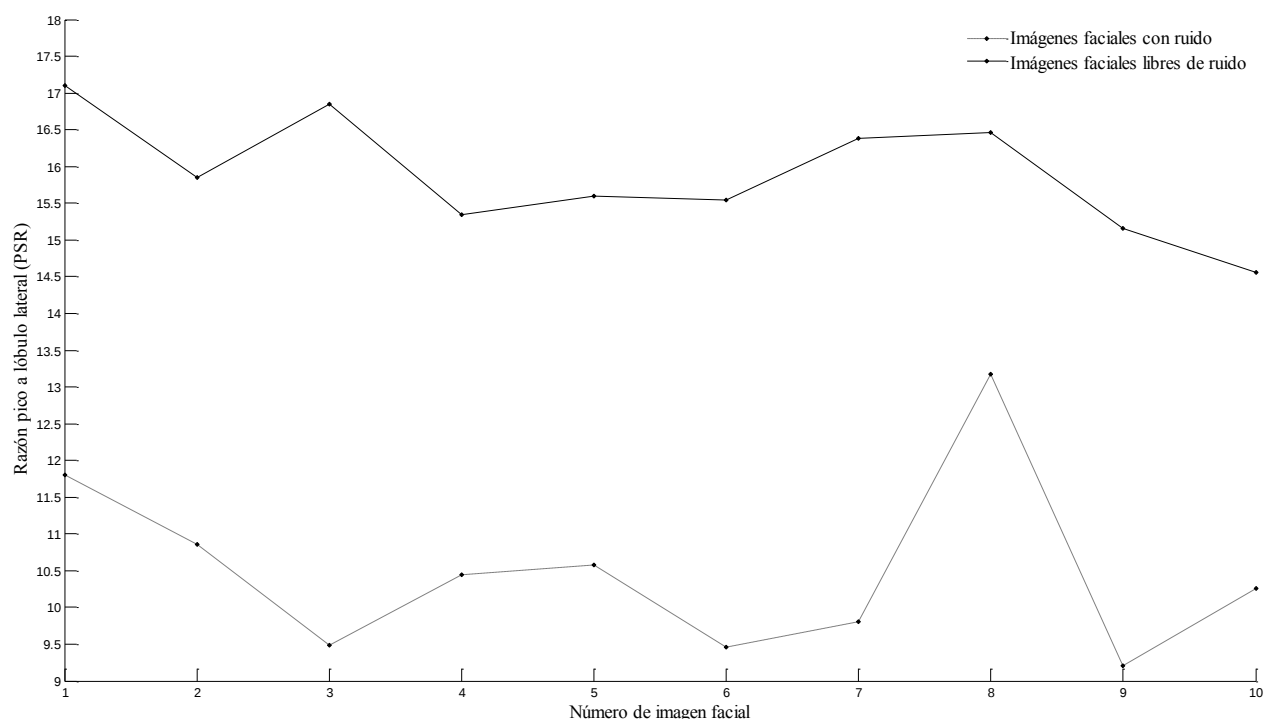


Figura 4.20. Valores PSR de imágenes faciales de prueba con y sin ruido.

4.2.3 Influencia de sombras en imágenes faciales

Las diferentes fuentes de luz existentes en un escenario no controlado pueden producir grandes variaciones en la apariencia de una cara, tal y como se muestra en la Figura 4.21. Esta variación se manifiesta en la cara con regiones más iluminadas que otras o con regiones sombreadas en el caso de que no llegue la luz de alguna fuente de iluminación.



Figura 4.21. Diferentes imágenes faciales de la misma persona en diferentes condiciones de iluminación.

Las imágenes faciales usadas en este experimento fueron sombreadas aleatoriamente. Tal y como se muestra en la Figura 4.21, se sombreó aleatoriamente el lado derecho, izquierdo, centro, región superior, región inferior o toda la cara. Es claro que la diferencia existente entre una región sombreada y otra con mucha iluminación, pero en este trabajo ambas variancias se consideran como problemas de iluminación. Esto debido a que una región con sombra es consecuencia de la ausencia una fuente de luz que la ilumine o en otros caso la obstrucción entre la superficie facial y la fuente de luz por un objeto y como consecuencia la sombra del objeto se refleja sobre la cara. En la Figura 4.22 se muestran los valores PSR de imágenes faciales de prueba pertenecientes a la persona 1 en la base de datos ORL y los valores PSR producidos por las mismas imágenes pero con sombra aleatoria.

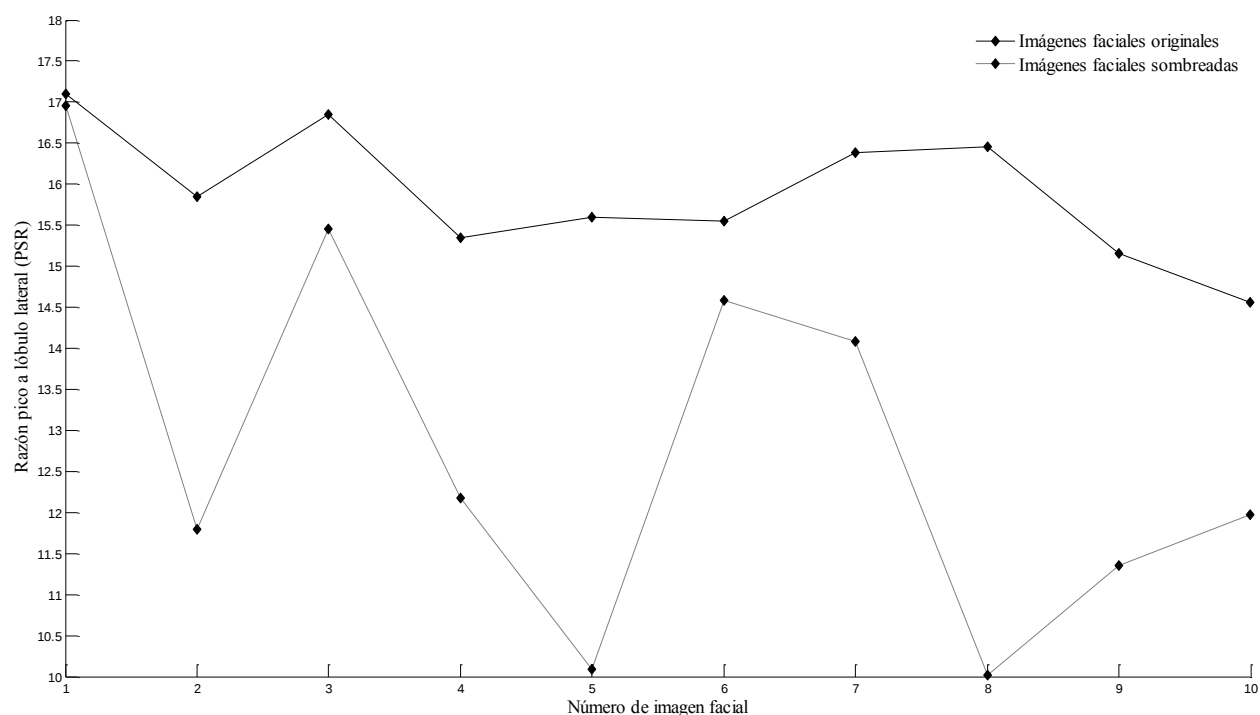


Figura 4.22. Valores PSR de imágenes faciales con y sin sombras usando el filtro K-Law.

4.2.4 Desempeño de los filtros de correlación en verificación facial

Con la configuración presentada en este experimento, ningún filtro tuvo mayor dificultad en reconocer imágenes faciales con las que fue construido. Como se puede observar en la Tabla 4.3, todos los filtros K-Law, UMACE, OTSDF y UOTSDF obtuvieron una tasa de verificación de 100% con el conjunto de prueba CP1, mientras que el filtro MACE obtuvo una tasa de verificación del 99.75%. Sin embargo, cuando las imágenes faciales de prueba presentan sombras o ruido el desempeño de los filtros cae considerablemente en algunos casos. Con el conjunto de prueba CP2 el cual contiene imágenes faciales sombreadas parcialmente el mejor desempeño lo obtuvo el filtro UOTSDF con 99.50%. Este resultado se obtuvo con $\alpha = 0.9999$, se obtiene una pequeña tolerancia al ruido mientras que el pico producido es agudo.

Cuando las imágenes faciales presentan variaciones en ruido, el desempeño de los filtros desciende conforme la cantidad de ruido aumenta. El mejor desempeño lo obtiene el filtro K-Law con los conjuntos CP3, CP4, CP5 Y CP6 CON 100%, 100%, 99.50% y 97% de tasa de verificación, respectivamente. Estos resultados fueron obtenidos con un factor de no linealidad $k = 3$ con el cual se obtiene una mayor tolerancia a distorsiones, el cual se ve reflejado en el

desempeño obtenido. El filtro OTSDF obtuvo un desempeño muy pobre pues no reconoció ninguna imagen facial afectada por ruido.

Tabla 4.3. Tasa de verificación de los filtros de correlación en la verificación facial.

Filtro	Conjunto de prueba					
	CP1	CP2	CP3	CP4	CP5	CP6
K-Law	100.00	92.50	100.00	100.00	99.50	97
MACE	99.75	80.25	5.50	0.00	0.00	0.00
UMACE	100	95.00	58.50	11.25	3.50	0.00
OTSDF	100	97.25	0.00	0.00	0.00	0.00
UOTSDF	100	99.50	40.75	3.5	0.00	0.00

El porcentaje de imágenes faciales impostoras que fueron declaradas como auténticas se presenta en la Tabla 4.4. Durante la simulación se observó que mientras mayor sea la distorsión presentada por una imagen facial de prueba impostora, la probabilidad de que ésta sea declarada como auténtica es muy bajo. Por el ejemplo, los conjuntos de prueba CP4, CP5 y CP6 contiene grandes cantidades ruido y los filtros MACE, UMACE, OTSDF y UOTSDF obtuvieron resultados óptimos con dichos conjuntos. Sin embargo, el conjunto de prueba CP2 que solo contiene sombras en regiones pequeñas las tasas de falsa alarma que se obtuvieron son altas, el filtro UOTSDF fue el filtro que obtuvo el peor desempeño con este conjunto declarando el 15.53% de 15600 imágenes impostoras como auténticas.

Tabla 4.4. Tasa de falsa aceptación de los filtros de correlación en la verificación facial.

Filtro	Conjunto de prueba					
	CP1	CP2	CP3	CP4	CP5	CP6
K-Law	3.58	2.24	2.56	1.43	0.90	0.51
MACE	0.06	0.01	0.00	0.00	0.00	0.00
UMACE	8.83	9.94	0.01	0.00	0.00	0.00
OTSDF	0.28	0.19	0.00	0.00	0.00	0.00
UOTSDF	16.52	15.53	0.04	0.00	0.00	0.00

Mientras en la estadística de tasa de verificación se desea un valor lo más cercano al 100, en la estadística de tasa de falsa aceptación se busca un valor lo más pequeño posible. Un 0% en tasa de falsa alarma es el resultado óptimo deseado. Estas métricas indican la efectividad de un filtro algoritmo de reconocimiento facial para reconocer correctamente una imagen facial o rechazar aquellos que intentan usurpar alguna identidad ajena.

En esta sección se ha presentado el desempeño de los filtros de correlación para reconocer imágenes faciales con la que fue construido. El mejor desempeño lo obtiene el filtro K-Law con imágenes faciales afectadas por ruido. Mientras que para aquellas que contienen variaciones en iluminación el filtro UOTSDF obtiene el mejor desempeño. También se mostró cómo las variaciones en iluminación y ruido afectan el desempeño de los filtros.

Capítulo 5

5. Influencia del pre-procesado de las imágenes faciales con distorsiones en la verificación facial

Como se vio en el capítulo 4, cuando se le presenta al algoritmo de reconocimiento facial una imagen facial auténtica que presenta distorsiones, esta tiene una probabilidad muy alta de ser declarada como impostora incorrectamente. Existen dos maneras de añadir invariancias a los filtros, i) analíticamente, si la distorsión se puede modelar o ii) empleando imágenes de entrenamiento que representen de forma general la distorsión que se espera presente el objeto de interés. En este trabajo se usan muestras y algunas operaciones de pre-procesado que mejora la calidad de las imágenes.

En algunas variaciones es difícil usar muestras, por ejemplo, imágenes faciales con variaciones en iluminación. En un ambiente no controlado las variaciones en iluminación (sombra o iluminación intensa) cambian bruscamente de un instante de tiempo a otro, y de un escenario a otro. Este cambio tan brusco hace difícil predecir qué variaciones presentará la imagen facial. En el caso del ruido, como se verá más adelante algunos filtros son invariantes a pequeñas cantidades.

Una solución que han adoptado muchos investigadores es el uso de una operación de pre-procesado a las imágenes faciales, tanto las de la galería como las de conjunto de prueba, que mejore la calidad de las imágenes antes de procesarse. Por ejemplo, una operación de pre-procesado que elimine el ruido contenido en una imagen o que aclare las regiones sombreadas o de intensa iluminación. Cabe recalcar que tanto las imágenes de entrenamiento como las de prueba tienen que ser pre-procesadas de la misma manera.

En esta sección se presenta la influencia que tienen las operaciones de pre-procesado sobre imágenes faciales y su influencia en la mejora del desempeño de los algoritmos de verificación facial. Las operaciones que se consideran son: A) Transformación logarítmica, B) Normalización, C) Ventana de coseno, D) Transformación logarítmica y normalización, E) transformación logarítmica y ventana de coseno, F) Normalización y ventana de coseno y G) Transformación logarítmica, normalización y ventana de cosenos. De ahora en adelante, en cada

tabla donde se presente resultados de desempeño las operaciones de pre-procesamiento será identificado con las letras A-G como se presentó anteriormente. Las secciones 5.2, 5.2 y 5.3 describen brevemente las operaciones de pre-procesamiento A, B y C, respectivamente. Las operaciones de pre-procesamiento D-G son combinaciones de las operaciones A, B y C.

5.1 Transformación logarítmica sobre imágenes faciales con variaciones en iluminación

Una imagen facial puede contener regiones con sombras, iluminación intensa o iluminación ambiental como resultado de las diferentes fuentes de luz que pudieran existir en el ambiente donde fue capturada. Considere una imagen facial en escala de grises con valores de intensidad en el intervalo 0-255, el logaritmo base 10 mapeará esos valores a la escala de 0-2.4082 (véase Figura 5.1).

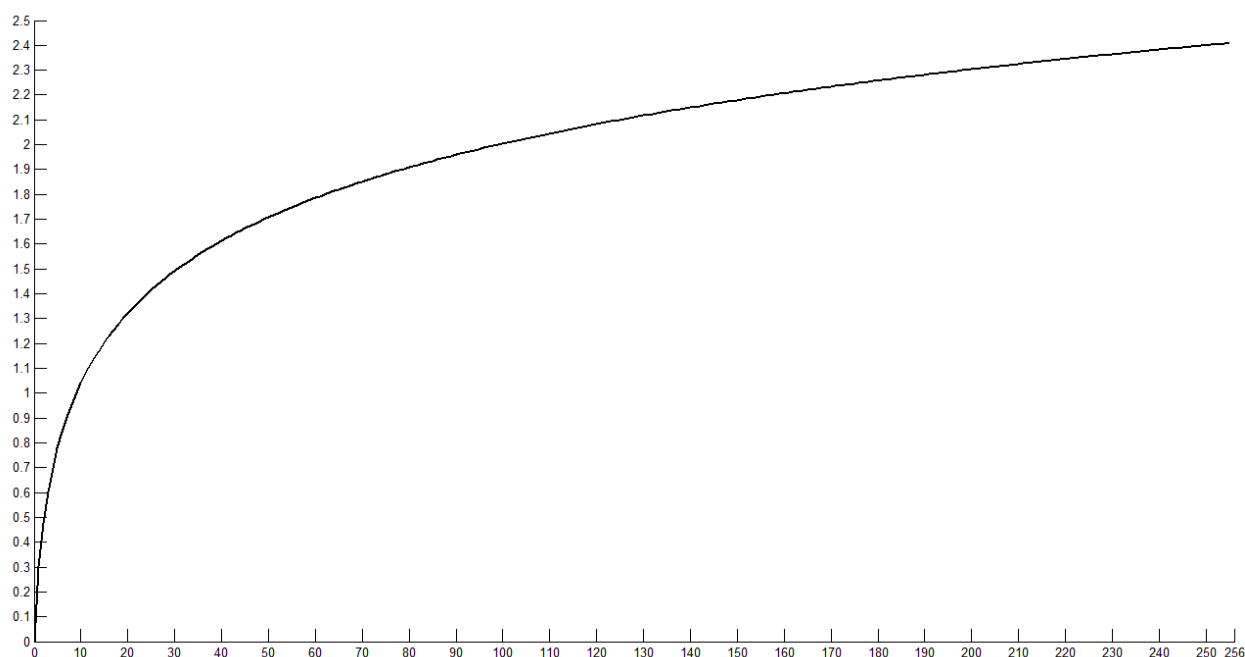


Figura 5.1. Mapeado de intensidades de píxeles usando la transformación logarítmica base 10. Píxeles de intensidad 0 representan el negro y 255 el blanco.

En imágenes sombreadas parcial o totalmente, la mayor parte de sus píxeles son de intensidades bajas. En aquellos donde la iluminación es intensa los píxeles son de intensidades altas. En la Figura 5.2 se muestran imágenes faciales sombreadas parcial y totalmente, así como una imagen facial sin sombras. Como se puede observar en los histogramas de las imágenes

faciales sombreadas, éstos contienen un sesgo positivo que indica que la mayoría de los píxeles son de regiones oscuras. El histograma de una imagen facial sin sombras (columna 3 de la figura 5.2) tiene una distribución casi simétrica.

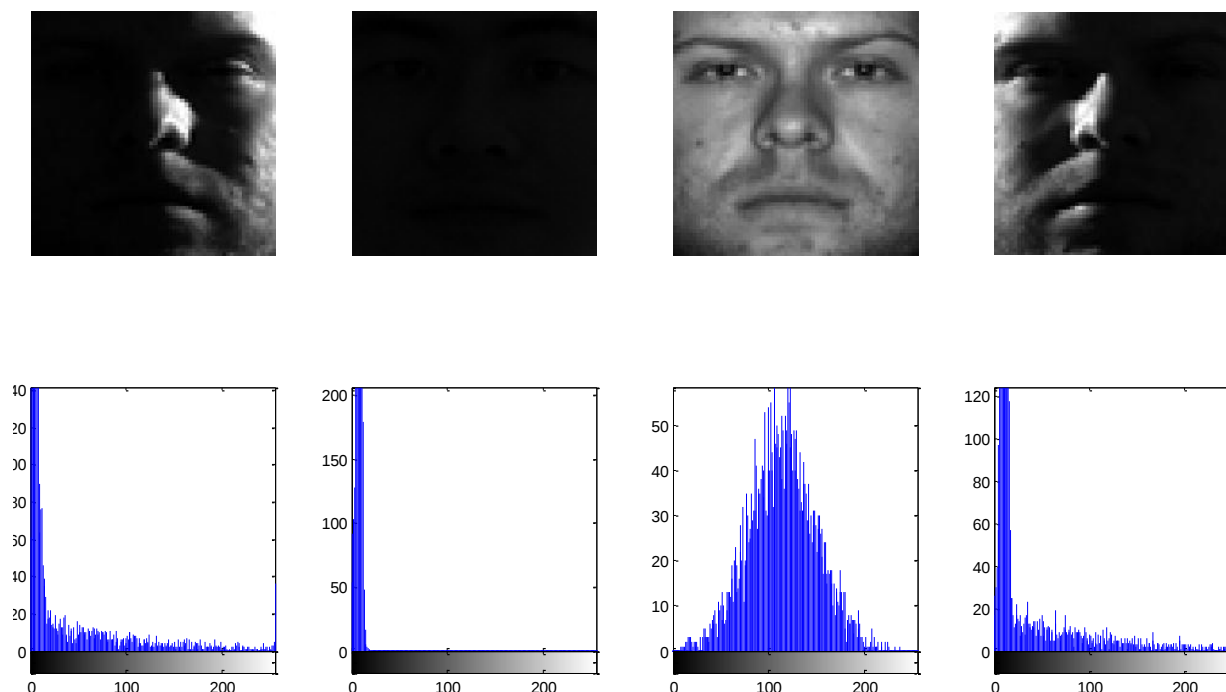


Figura 5.2. Histograma de imágenes faciales sombreadas parcial o totalmente.

Las imágenes faciales sombreadas parcial o totalmente contienen píxeles de intensidad 0, para el cual no existe logaritmo. Para evitar problemas, a cada píxel se le suma la unidad, es decir $\log_{10}(\text{valor del píxel} + 1)$. Así, el valor más pequeño para el cual hay que calcular el logaritmo es la unidad. En la Figura 5.3 se muestran las imágenes faciales resultantes de la aplicación de la transformación logarítmica. Como se puede observar, los píxeles tienen una distribución casi simétrica. Se reducen los efectos de las sombras y la iluminación, obteniendo con ello imágenes faciales cuyas características son fáciles de distinguir.

La transformación logarítmica no iguala las intensidades de las regiones sombreadas con las regiones iluminadas. Como se puede observar en la primera y cuarta columna (donde es más notorio el efecto de la transformación) de la Figura 5.3, esta transformación realiza un escalamiento, elevando un poco las intensidades de los píxeles de las regiones sombreadas pero mantiene una proporcionalidad con las intensidades de los píxeles de las regiones iluminadas. Esta transformación produce una imagen facial más clara y fácil de distinguir sus características sin afectar la información de los bordes.

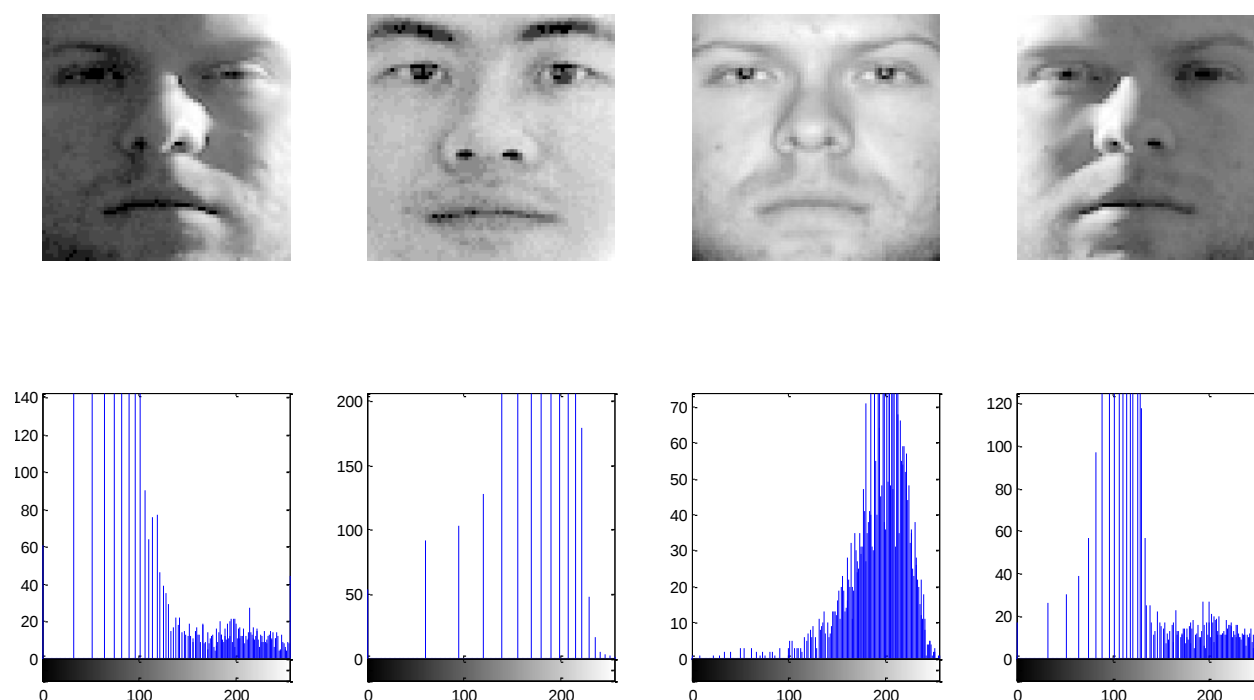


Figura 5.3. Resultado de la aplicación de la transformación logarítmica base 10 a las imágenes con regiones sombreadas.

Esta transformación ha sido usada de forma exitosa en el reconocimiento facial usando los filtros MACE y UMACE, tal y como los describe Savvides y Vijaya (2003). Bolme (2009) y Bolme (2010) usa esta operación de preprocesado como primero de tres operaciones para mejorar la calidad de objetos con sombra e iluminación intensa.

5.2 Normalización de imágenes faciales

La normalización es un proceso que cambia el intervalo de los valores de intensidad de los píxeles en una imagen. Esta técnica intenta mejorar el contraste de una imagen extendiendo el intervalo de sus valores de intensidades a un intervalo de valores deseados. En este trabajo, la normalización se realiza en la energía de la imagen con el objetivo de obtener independencia de las intensidades de las fuentes de luz (Belhumeur et al., 1997).

Sea $i(x,y)$ la imagen en escala de grises a normalizar. Primero, se calcula la media de imagen:

$$\bar{x} = \left(\frac{1}{d}\right) \sum_{x=1}^d \sum_{y=1}^d i(x, y) \quad (39)$$

A la imagen de entrada se resta la media calculada en la ecuación 42:

$$i_2(x, y) = i(x, y) - \bar{x} \quad (40)$$

Se calcula la energía de la imagen $i_2(x, y)$:

$$Energía = \sqrt{\sum_{x=1}^d \sum_{y=1}^d i_2(x, y)^2} \quad (41)$$

Finalmente, la imagen normalizada está dada por la ecuación:

$$i_{normalizado}(x, y) = i_2(x, y) / Energía \quad (42)$$

Con esta operación sobre una imagen se tiene una media aproximado de 0.0 y una suma del cuadrado de los valores de intensidad igual a 1, para obtener imágenes con una intensidad consistente.

5.3 Ventana de coseno

Aplicar una ventana de coseno sobre una imagen se obtiene una imagen resultante con valores de intensidades altos en el centro. Esta operación sobre una imagen facial se traduce en una imagen con la región central resaltada y valores bajos en las esquinas y lados, lo cual reduce los efectos de la frecuencia de los bordes de la imagen cuando es transformada por Fourier.

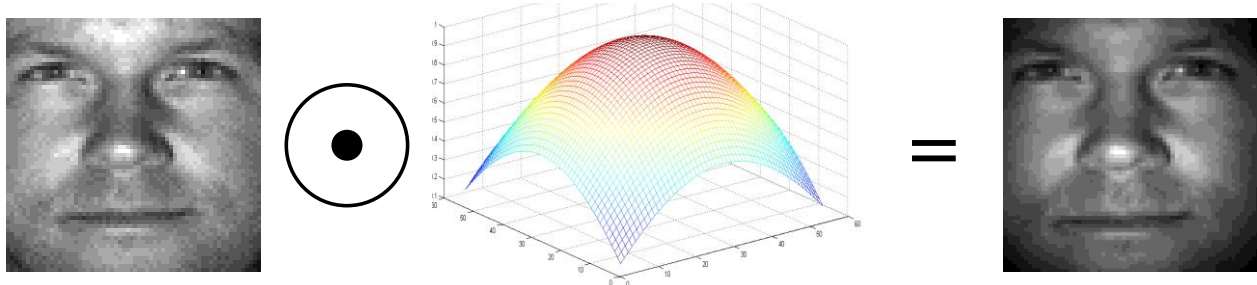


Figura 5.4. Multiplicación de una imagen facial por una ventana de cosenos.

Sean fr_1 y fr_2 vectores de frecuencias bidimensionales para una matriz $m \times n$. Se generan matrices F_{r1} y F_{r2} con los vectores fr_1 y fr_2 para poder evaluar una función de dos

variables, específicamente una imagen facial bidimensional. Las filas de la matriz F_1 son copias del vector fr_1 y las columnas de Fr_2 son copias del vector fr_2 . Sea $d = \sqrt{Fr_1^2 + Fr_2^2}$ que contiene la raíz cuadrada de la suma de los cuadrados de cada valor de Fr_1 y Fr_2 . Obtenemos la ventana de coseno con la siguiente ecuación:

$$w_{cos} = \cos(d) \quad (43)$$

La imagen facial multiplicada por la ventana de coseno (véase Figura 5.4) se obtiene $I_2 = I \odot w_{cos}$, donde la multiplicación es elemento a elemento. Esta operación es usada por Bolme et al. (2010) después de aplicar la transformación logarítmica y la normalización a una imagen. Esta operación difumina los bordes de la imagen facial que son esenciales para el algoritmo de reconocimiento lo que produce un bajo desempeño, tal y como se verá más adelante.

5.4 Desempeño en la verificación facial de los filtros de correlación al usar operaciones de pre-procesamiento

El desempeño de los filtros de correlación usando operaciones de pre-procesamiento se calculó siguiendo la configuración presentada en la sección 4.1. La diferencia radica en que a las imágenes faciales tanto de entrenamiento del filtro como las de prueba fueron codificadas de la misma forma con alguna de las operaciones de pre-procesamiento mencionadas al principio de esta sección.

La Tabla 5.1 muestra las tasas de verificación obtenidas por los filtros con el conjunto de prueba CP1 aplicando operaciones de pre-procesamiento a las imágenes. Los filtros UMACE, OTSDF y UOTSDF obtuvieron un desempeño óptimo con cualquiera de las operaciones de pre-procesamiento. El filtro K-Law, con la operación de la transformada logarítmica y normalización (preprocesamiento E) se obtiene una tasa de verificación de 87.25%. Con la normalización de la energía de las imágenes (operación de preprocesamiento B) se obtiene el desempeño óptimo para todos los filtros de correlación, y el desempeño del filtro MACE pasa de 99.75% (véase Tabla 4.3) a 100% de probabilidad de reconocimiento.

Tabla 5.1. Desempeño de los filtro usando el conjunto de prueba CP1 y pre-procesado de las imágenes.

Tasa de Verificación					
Preprocesamiento	Filtro de correlación				
	K-Law	MACE	UMACE	OTSDF	UOTSDF
A	100.00	99.75	100.00	100.00	100.00
B	100.00	100.00	100.00	100.00	100.00
C	100.00	99.75	100.00	100.00	100.00
D	100.00	100.00	100.00	100.00	100.00
E	87.25	99.75	100.00	100.00	100.00
F	100.00	99.75	100.00	100.00	100.00
G	100.00	99.50	100.00	100.00	100.00

La Tabla 5.2 muestra la tasa de falsa aceptación de cada filtro. Como se puede observar, el filtro UOTSDF obtuvo el peor desempeño donde el mejor de los casos obtuvo el 10.61% con la aplicación de la ventana de coseno sobre la imagen facial.

Tabla 5.2. Porcentaje de imágenes faciales del conjunto de prueba CP1 confundidas por los filtros.

Tasa de Falsa Aceptación					
Preprocesamiento	Filtro de correlación				
	K-Law	MACE	UMACE	OTSDF	UOTSDF
A	3.04	0.22	12.00	2.68	53.45
B	3.77	0.06	8.82	1.84	49.39
C	0.47	0.00	4.23	0.00	10.61
D	3.70	0.22	12.14	2.76	53.93
E	0.00	0.03	6.41	0.92	44.05
F	2.26	0.00	4.74	0.70	38.25
G	2.00	0.03	5.68	0.86	41.78

Con el conjunto de imágenes de prueba con variaciones en iluminación, ya sea sombra o iluminación intensa, los filtro K-Law y OTSDF obtuvieron los mismos resultados que con el conjunto CP1, esta información se presenta en la Tabla 5.3. En la Figura 5.5 se muestra el

desempeño de los filtros evaluados con y sin preprocesamiento donde el filtro K-Law, UMACE y UOTSDF mejoraron su desempeño con la transformación logarítmica y la ventana de coseno sobre las imágenes faciales. Los filtros MACE Y OTSDF no mejoraron sus desempeños. En la Tabla 5.4 se presenta la tasa de falsa aceptación obtenida por cada filtro donde se evidencia la mala capacidad del filtro UOTSDF para distinguir clases de caras.

Tabla 5.3. Tasa de verificación obtenida por los filtros al aplicar el pre-procesado a las imágenes faciales y usar el conjunto de prueba CP2.

Preprocesamiento	Filtro de correlación				
	K-Law	MACE	UMACE	OTSDF	UOTSDF
A	100.00	70.75	94.00	88.50	100.00
B	100.00	79.25	95.00	90.50	100.00
C	100.00	77.25	96.00	89.50	100.00
D	100.00	71.25	93.75	87.00	100.00
E	87.75	79.50	99.25	95.00	100.00
F	100.00	74.75	95.50	91.00	100.00
G	100.00	74.25	96.50	91.00	100.00

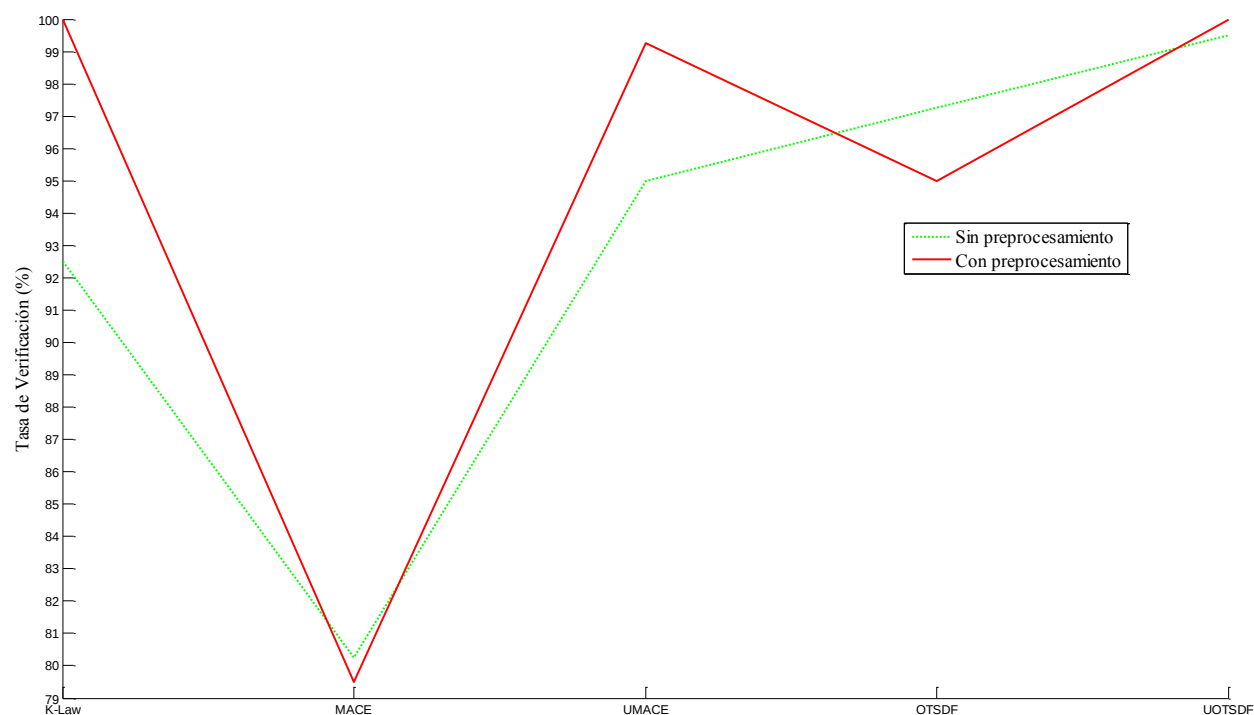


Figura 5.5. Desempeño de los filtros de correlación con operaciones de preprocesamiento sobre imágenes con variaciones en iluminación.

Tabla 5.4. Tasa de falsa aceptación obtenida por los filtros al aplicar el pre-procesado a las imágenes faciales y usar el conjunto de prueba CP2.

Preprocesamiento	Filtro de correlación				
	K-Law	MACE	UMACE	OTSDF	UOTSDF
A	3.18	0.26	12.80	3.19	53.45
B	3.88	0.02	9.91	1.78	49.39
C	0.50	0.00	9.59	0.07	10.61
D	3.79	0.29	12.90	3.46	53.93
E	0.00	0.01	12.16	1.63	44.05
F	2.33	0.00	10.19	2.21	38.25
G	2.05	0.01	7.25	1.15	41.78

El ruido fue la distorsión que más afectó el desempeño de los filtros, pues mientras mayor fuera este menor era el desempeño del filtro. En la Tabla 5.5 se muestra el desempeño de los filtros con el conjunto CP3. El filtro MACE fue el más sensible al ruido. Con este conjunto de prueba, todos los filtros mejoraron su desempeño (véase Figura 5.6) con la operación de preprocesamiento D, la cual consiste en la aplicación de la transformada logarítmica y la normalización de la energía de las imágenes faciales.

Tabla 5.5. Tasa de verificación obtenida por los filtros con el conjunto de prueba CP3 y operaciones de pre-procesamiento.

Preprocesamiento	Filtro de correlación				
	K-Law	MACE	UMACE	OTSDF	UOTSDF
A	100.00	31.00	82.50	75.50	100.00
B	100.00	5.75	62.25	32.00	100.00
C	99.25	3.25	46.00	1.25	100.00
D	100.00	33.50	85.50	78.00	100.00
E	62.75	26.50	79.75	65.75	100.00
F	100.00	2.75	43.25	21.75	100.00
G	100.00	24.25	79.00	63.50	100.00

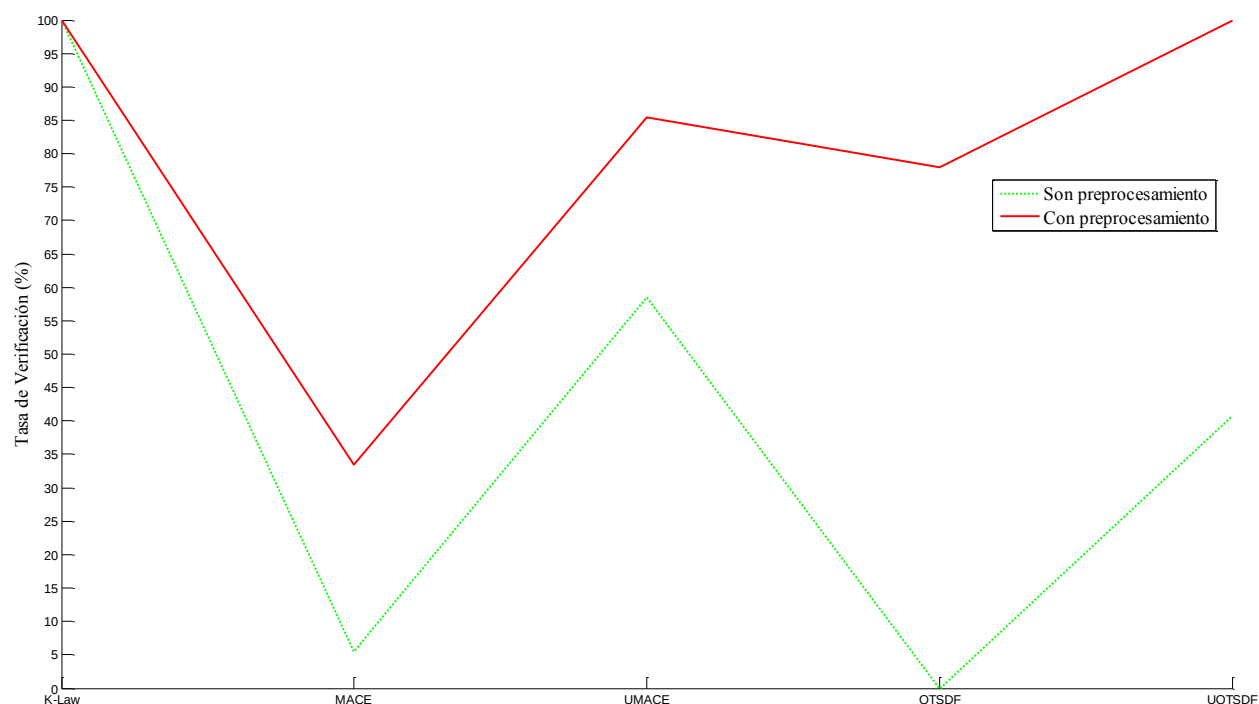


Figura 5.6. Desempeño de los filtros de correlación con el conjunto de prueba CP3, con y sin preprocesamiento.

Con los conjuntos CP4, CP5 y CP6 el desempeño de los filtros disminuyó considerablemente, en donde el filtro K-Law fue el obtuvo los mejores desempeños con 100%, 99.75 y 98.25% de efectividad usando la transformación logarítmica y la ventana de coseno (al igual que en el conjunto de prueba CP2) como operación de preprocesado a las imágenes (véase Figura 5.7). El resto de los filtro obtuvieron un desempeño mejor que sin preprocesado pero menor al 75% en la verificación con los conjuntos de prueba CP4, CP5 y CP5.

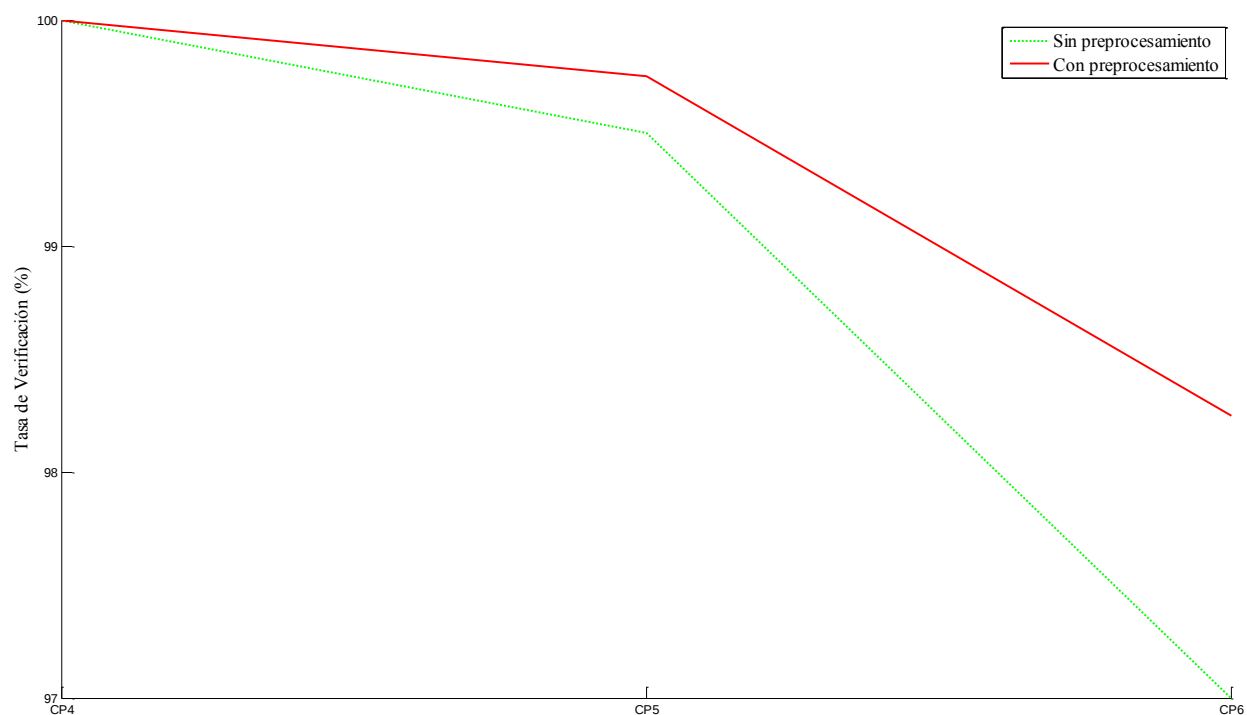


Figura 5.7. Mejora en el desempeño del filtro K-Law con imágenes ruidosas del conjunto CP4, CP5 y CP6 con la operación de preprocesamiento D.

En esta sección se ha presentado el desempeño de los filtros de correlación usando operaciones de preprocesado para mejorar la calidad de las imágenes faciales. El filtro K-Law obtuvo el mejor desempeño con los 6 conjuntos de prueba usando las operación A) transformación logarítmica y D) Transformación logarítmica y normalización.

Capítulo 6

6. Identificación facial usando filtros de correlación e imágenes con variaciones en sombras

En los capítulos 4 y 5 se evaluó el desempeño de los filtros de correlación en la tarea de verificación facial. El objetivo de dichos capítulos fue mostrar que los filtros de correlación son capaces de distinguir una clase de cara de otras clases distintas y que el bajo desempeño de los filtros debido a distorsiones mejora (en algunos casos) con el pre-procesado de las imágenes. Sin embargo, en dichos experimentos no se consideró el reconocimiento de imágenes de la misma clase que el filtro pero que no estaban dentro del conjunto de entrenamiento.

En esta sección se evalúa el desempeño de los filtros de correlación contruidos con la mínima cantidad posible de imágenes de entrenamiento para reconocer imágenes de la misma clase que no están en el conjunto de entrenamiento. Se presenta dos estrategias de selección de las imágenes de entrenamiento para la construcción del filtro de correlación y se muestra el desempeño de cada estrategia. Los experimentos se realizaron en la tarea de identificación facial y se aborda el problema de la variación en iluminación. En este trabajo se decidió usar la transformación logarítmica como operación de pre-procesado a las imágenes faciales, ya que ésta ha sido usada en la literatura para tal propósito (Savvides y Vijaya Kumar, 2003) y en el capítulo 5 fue una de las operaciones con que mejor resultados se obtuvo en el filtro K-law con las imágenes de prueba CP2.

6.1 Configuración del experimento

Todos los experimentos usaron imágenes faciales de las bases de datos Yale B (Samaria y Harter, 1994) y ORL (presentado en el capítulo 4). La base de datos YALE B contiene 2432 imágenes faciales diferentes de 38 personas. Cada persona tiene 64 imágenes faciales las cuales algunas contienen regiones sombreadas e iluminación ambiental. Cada imagen facial, tanto de la base de datos YALE B y ORL, fueron recortados y escalados manualmente a una resolución de 64×64 pixeles y en un rango dinámico de 0-255.

Para fines de evaluación siguiendo lo descrito en la sección 3.1.4.1, se crearon tres conjunto de datos faciales. El conjunto G que contiene las plantillas biométricas fáciles contruidos con imágenes faciales de la base de datos Yale B. El segundo conjunto P_G contiene imágenes faciales de prueba de personas registradas en la galería, este conjunto contiene las 2432 imágenes faciales de la base de datos YALE B y con ellos se calcula la tasa de identificación y detección. El tercer conjunto P_N contiene imágenes faciales de prueba de personas que no están en la base de datos, en este conjunto están las 400 imágenes faciales de la base de datos ORL con los cuales se calcula la tasa de falsa alarma. Se usaron los umbrales de reconocimiento determinados en la sección 4.1.1 para cada filtro.

6.2 Resultados experimentales

En la Figura 6.1 se muestra el conjunto de 64 imágenes faciales diferentes de la misma clase. Este conjunto de imágenes se caracteriza por contener variaciones en iluminación, el cual se manifiesta sombras en regiones de la cara o en su totalidad como la imagen facial *img_facial29* de la Figura 6.1. El objetivo es la selección de la cantidad mínima de imágenes faciales representativas para la construcción de un filtro de correlación que pueda ser usado para reconocer las 64 imágenes.

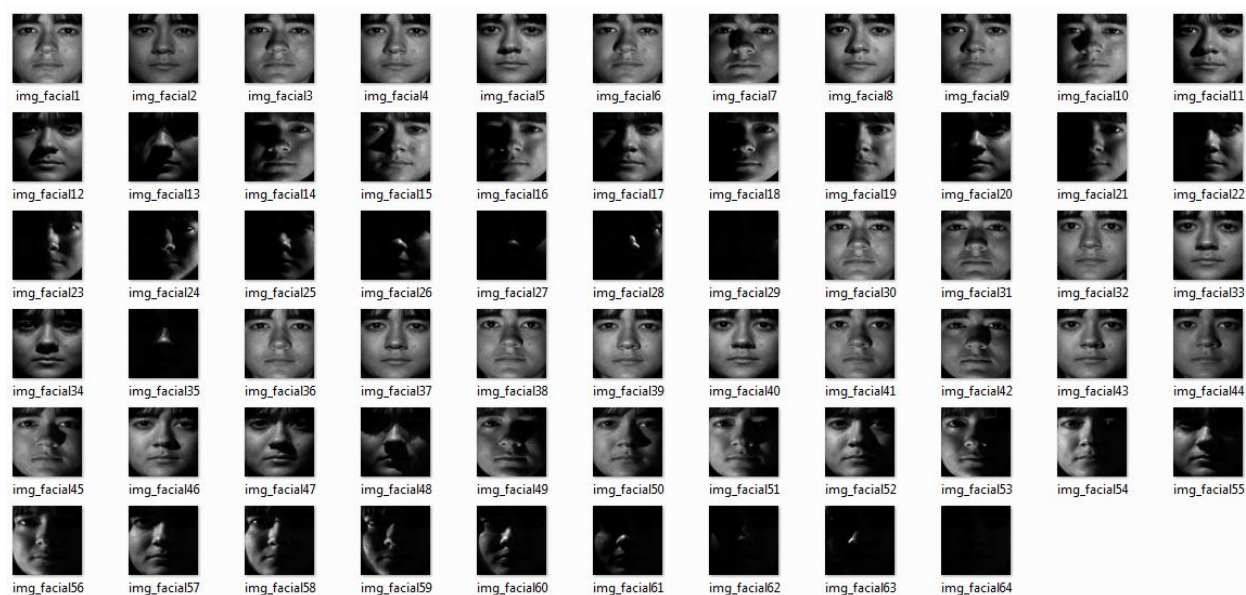


Figura 6.1. Imágenes faciales distintas de la misma persona.

Para lograr dicho propósito se analizaron dos estrategias de construcción del filtro considerado la variación presentada por el conjunto de datos. La primera estrategia consiste en construir el filtro con muestras faciales sin variaciones en iluminación. Considere a la imagen *img_facial32* del Figura 6.1 para construir el filtro. La correlación del filtro con las 64 imágenes faciales de forma consecutiva como están en la Figura 6.1 produce los valores PSR que se muestran en la Figura 6.2. Como se puede observar, este filtro solo produce valores PSR altos para imágenes similares a él, es decir, aquellos que no contienen sombras.

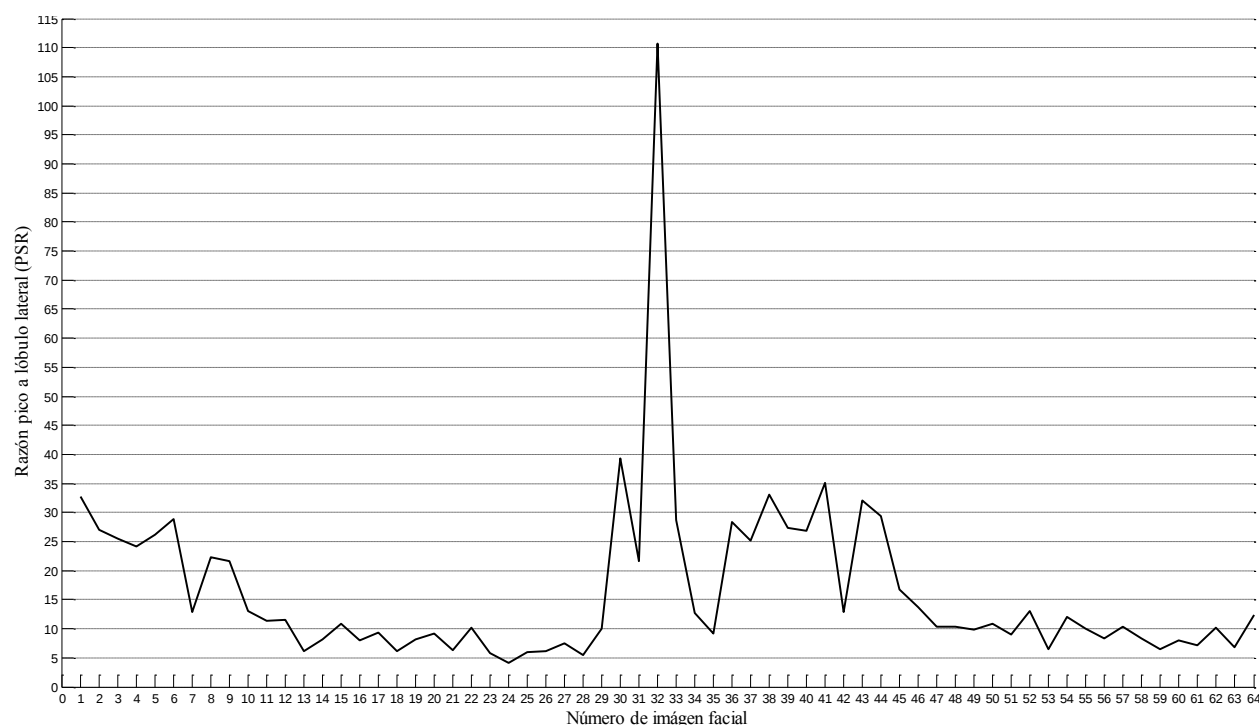


Figura 6.2. Desempeño PSR de un filtro construido con una sola imagen y correlacionada con 64 imágenes de la misma clase.

Con el fin de reconocer todas las imágenes, se consideraron más imágenes de entrenamiento y se determinó a base de experimentos y observaciones que con un mínimo de 4 imágenes se obtiene un desempeño aceptable. En la Figura 6.3 se muestran las cuatro imágenes faciales seleccionadas para la construcción del filtro y en la Figura 6.4 se muestra el desempeño PSR de este filtro con imágenes faciales auténticas (línea negra continua superior) e imágenes faciales impostoras (líneas segmentadas en la parte inferior de la gráfica). Aunque los valores PSR auténticos son mayores que el de los impostores, el uso del umbral de reconocimiento causa que se rechacen muchas imágenes auténticas sobre todo las imágenes con sombras.



Figura 6.3. Muestras faciales de la misma clase para construir un filtro correlación.

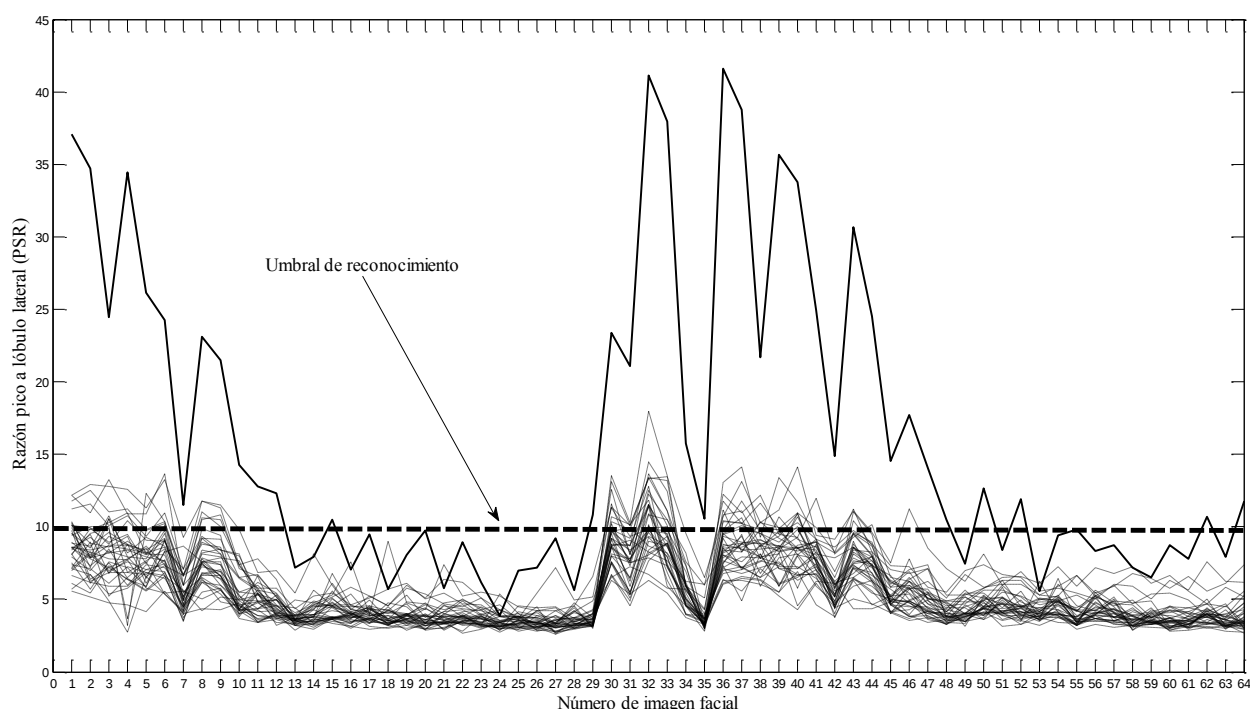


Figura 6.4. Desempeño PSR del filtro construido con imágenes faciales sin variaciones en iluminación.

La segunda estrategia construye un filtro de correlación con muestras que contiene las distorsiones más generales del conjunto de datos faciales. Un de esas distorsiones es la sombra en la mitad derecha de la cara. Un buen candidato para representar esta distorsión es la imagen `img_facial22` de la Figura 6.1. El resultado obtenido al crear el filtro con esta imagen se muestra en la Figura 6.5. La contraparte de esta distorsión es la región mitad izquierda sombreada, por lo que lo integramos al conjunto de entrenamiento y se obtiene el resultado que se presenta en la Figura 6.6. Con estas dos imágenes se producen valores PSR altos para las imágenes similares al filtro. Se agrega al conjunto de entrenamiento la imagen `img_facial35` de la Figura 6.1 la cual está completamente sombreada y el resultado con estas tres imágenes se muestra en la Figura 6.7. Con una cuarta imagen (`img_facial36`) libre de sombras se producen valores PSR altos con aquellas imágenes libres de sombras. El desempeño PSR del filtro con las cuatro imágenes

faciales se muestra en la Figura 6.8 donde todos los valores PSR de imágenes auténticas son mayores al umbral de reconocimiento.

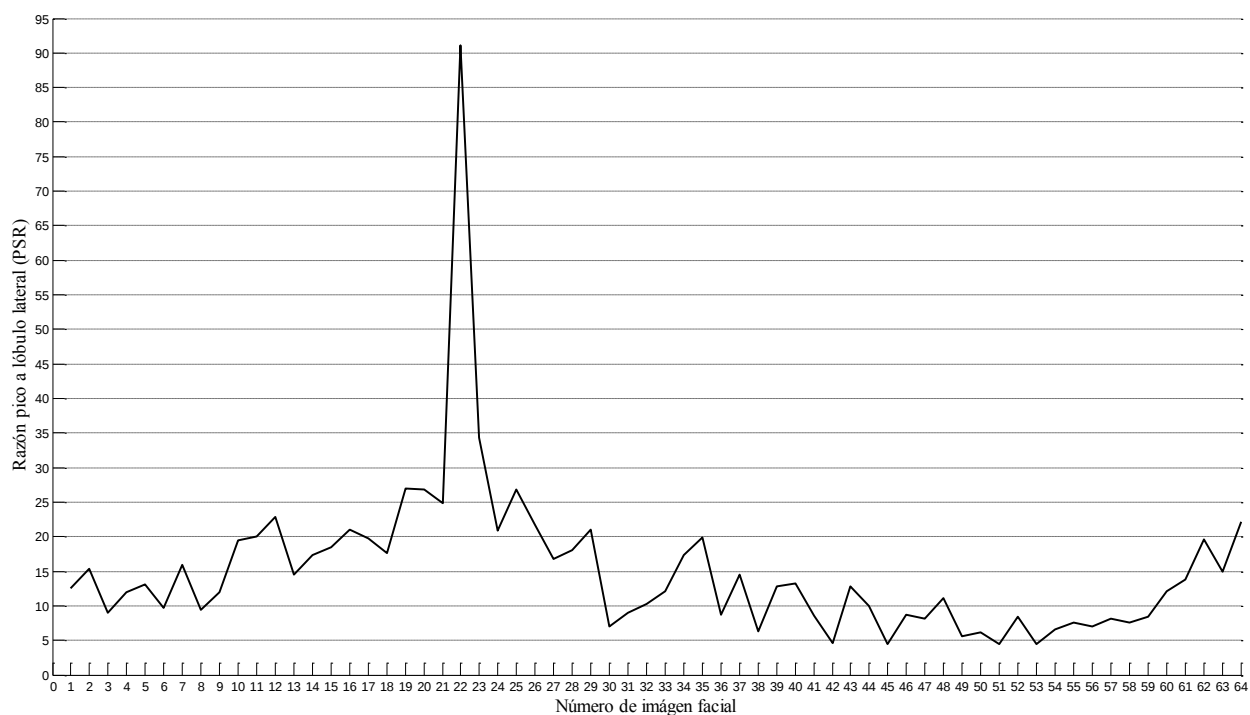


Figura 6.5. Desempeño PSR del filtro construido con muestra que presenta una distorsion general.

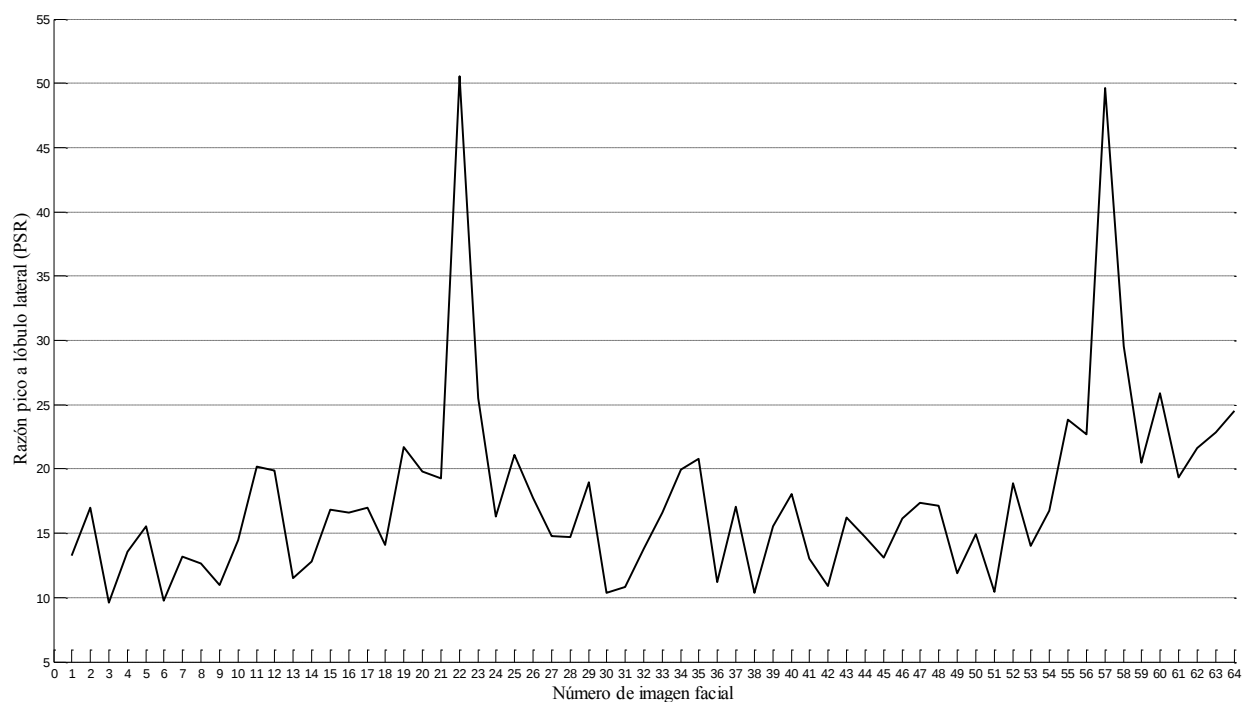


Figura 6.6. Desempeño PSR del filtro de correlación construido con dos muestras faciales distorsionadas.

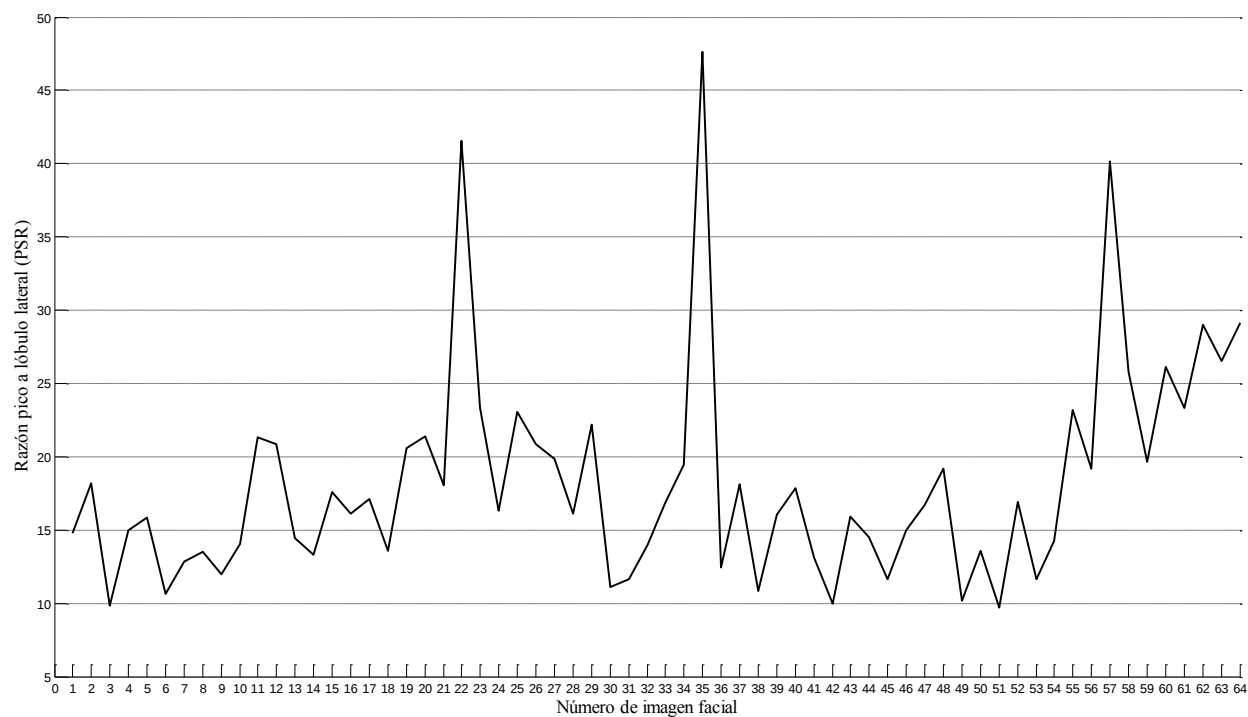


Figura 6.7. Desempeño del filtro construido con la estrategia 2 con tres imágenes faciales con distorsiones.

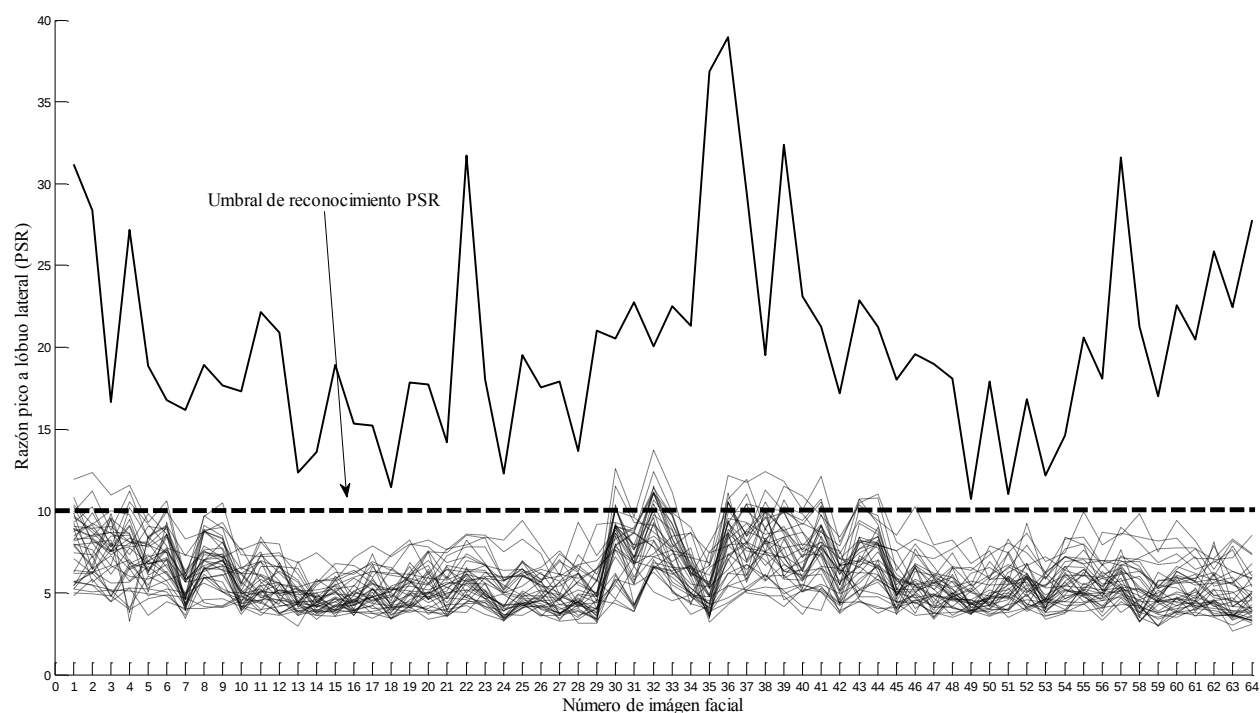


Figura 6.8. Desempeño PSR del filtro de correlación construido con cuatro imágenes con variaciones en iluminación.

La Figura 6.8 muestra el desempeño PSR de un filtro construido con la segunda estrategia usando las imágenes faciales mostradas en la Figura 6.9. Todos los resultados PSR de la correlación de un conjunto de imágenes de la misma clase con un filtro son alto y mayores al umbral de reconocimiento lo que aumenta la probabilidad de ser reconocidos correctamente.



Figura 6.9. Muestras faciales con variaciones en iluminación para la construcción de un filtro.

Se evaluó el desempeño de los filtros K-LAW, MACE, UMACE, OTSDF y UOTSDF en la tarea de identificación facial con las dos estrategias de selección de imágenes de entrenamiento. El resultado obtenido se muestra en la Tabla 6.1. El filtro K-Law usando la estrategia 2 fue el que obtuvo el mejor desempeño en términos de la tasa de identificación y detección (TID), el cual se obtuvo aplicando la transformación logarítmica sobre las imágenes faciales. El filtro MACE fue el que obtuvo el mejor desempeño en lo referente a la métrica de tasa de falsa alarma (TFA).

Tabla 6.1. Desempeño de los filtros de correlación en la tarea de identificación facial.

Filtro de correlación	TID		TFA	
	Estrategia 1	Estrategia 2	Estrategia 1	Estrategia 2
K-Law	80.14	99.09	25.75	25.25
MACE	45.19	75.69	2.75	3.50
UMACE	72.45	97.82	40.25	26.00
OTSDF	64.02	92.85	17.50	17.00
UOTSDF	80.02	98.31	86.00	89.50

En esta sección se ha mostrado que los filtros de correlación ofrecen un buen desempeño en la tarea de identificación facial, como en el caso de del filtro K-Law que solo rechazó incorrectamente el 0.91% de las imágenes faciales auténticas. El filtro que obtuvo el peor desempeño en esta actividad fue el filtro UOTSDF con la tasa de falsa alarma más alta que es el 98.31% con la segunda estrategia de selección de imágenes de entrenamiento. Esto muestra que este filtro no es capaz de distinguir una clase de otra.

Capítulo 7

7. Reconocimiento facial mediante un filtro compuesto no lineal sin restricciones construido con regiones faciales

Alcanzar un buen desempeño con los algoritmos de reconocimiento facial es una tarea desafiante, ya que muchos de los algoritmos fueron diseñados para optimizar cierta distorsión o porque su naturaleza no permite usar toda la información facial. El enfoque PCA necesita que las imágenes de prueba y de la galería sean del mismo tamaño y con las características faciales alineadas, esto implica que para que PCA proporcione un buen desempeño, las imágenes faciales tienen que ser frontales y alineadas. El enfoque LDA trata de maximizar la varianza entre las clases de caras y minimizar la varianza dentro de las clases, para lograr tal propósito se necesitan imágenes faciales de entrenamiento con pocas variaciones entre ellos. Cuando se trata con datos faciales de dimensiones grandes este enfoque necesita de un número grande de imágenes de entrenamiento para poder reconocer dichos datos faciales, en caso contrario surgen problemas. El método EBGm requiere de una localización exacta de los puntos de interés. Jun et al. (2010) proponen un método que descompone una imagen facial de entrada en varias versiones que exhiben diferentes condiciones de iluminación. Cada versión de la imagen se representa en un histograma, para posteriormente realizar el reconocimiento mediante un acoplamiento de histogramas. Una desventaja de este método es lo difícil que es representar todas las condiciones de iluminación debido a la gran variabilidad y a diferentes fuentes de luz que pudieran existir en un ambiente no controlado.

La mayoría de los métodos anteriores e incluso los filtros de correlación evaluados en los capítulos cuatro, cinco y seis requieren de una cara frontal completa y con las características faciales alineadas. Muchos de los algoritmos de reconocimiento facial requieren de un proceso de alineación de características faciales antes de realizar el reconocimiento. Algunos no requieren de la alineación de características, como los enfoques basados en correlación, sin embargo, requieren de una imagen facial completa y sin oclusiones.

Otra cuestión es que la mayoría de los algoritmos existentes tratan una imagen facial como un solo objeto y no consideran el hecho de que la cara sea un objeto compuesto de

características faciales con sus propios comportamientos, tal que a su vez pueden tratarse como otros objetos. Es decir, la persona puede abrir y cerrar sus ojos con libertad o abrir un ojo mientras se cierra el otro.

Así, esta sección propone una técnica de reconocimiento facial usando filtros de correlación compuesto no lineal sin restricciones. Se hace uso de regiones faciales que encierran una o más características faciales, considerando que cada región contribuye en la salida de correlación y que la cara completa se forma con $f = \bigcup_{i=1}^N r_i$, donde f es la cara completa, r_i es i -ésima región facial. Por lo que se espera que una imagen facial (completa o parcial) de prueba sea reconocida con al menos una de las regiones.

Esta propuesta se basa en los resultados obtenidos en los capítulos cuatro, cinco y seis para el diseño del filtro de correlación. Principalmente se aprovecha la característica y el buen desempeño evidenciado por el filtro K-Law, así como algunas características de las versiones no restringidas (UMACE y UOTSDF) de los filtros restringido en el origen MACE y UMACE.

7.1 Filtro compuesto no lineal sin restricciones

Considere $T = \{t_1(x, y), t_2(x, y), \dots, t_N(x, y)\}$ el conjunto de N imágenes de entrenamiento que representan las posibles distorsiones que se espera presente la imagen de referencia $r(x, y)$. El filtro SDF $h(x, y)$ es una combinación lineal de las imágenes de entrenamiento:

$$h(x, y) = \sum_{i=1}^N a_i t_i(x, y) \quad (44)$$

donde el coeficiente a_i de la combinación lineal se escoge para satisfacer la restricción:

$$\mathbf{h}^+ t_i(x, y) = c_i \quad (45)$$

con c_i constante pre-especificada, que para imágenes de entrenamiento de la clase que se quiere reconocer se le asigna el valor 1. El símbolo $+$ representa la transpuesta conjugada. El objetivo es que el filtro $\mathbf{h}$ produzca el valor constante c_i en el origen del plano de salida cuando se usa la i – ésima imagen de entrenamiento como entrada. Cuando la entrada es una imagen que no pertenece al conjunto de entrenamiento pero pertenece a la misma clase, se espera que la salida de correlación en el origen sea lo más cercano a la constante c_i .

Este filtro SDF convencional presenta muchos problemas. No considera el ruido aleatorio en la entrada, lo que dificulta un acoplamiento perfecto, es decir, que no se produzca en el origen la constante c_i , incluso si la prueba es una imagen de entrenamiento. Como la restricción es únicamente en el origen, éste puede no corresponder al pico de correlación. Ya que no se restringen los demás valores los cuales pueden ser mayores que c_i . Como se mencionó anteriormente, los filtros de correlación son invariantes al desplazamiento, entonces cuando la imagen de referencia se desplaza cierta cantidad de pixeles en el plano se producirá un valor alto en una posición diferente al origen que es difícil de detectar al únicamente estar restringiendo el origen. Por lo tanto, para que un filtro SDF produzca un pico agudo en el origen se requiere de una imagen de entrada centrada. La restricción está pensada solo para imágenes de entrenamiento pero no considera las imágenes de prueba (Mahalanobis, 1994), además de que fue diseñado para que los filtros fueran sintetizados en un laboratorio usando múltiples técnicas de exposición (Vijaya, 1994).

Entonces, nuestra hipótesis es que eliminando la restricción sobre el origen impuesto por un filtro SDF y aplicando el filtrado no lineal se incrementa la capacidad de los filtros para discriminar una clase de caras de otras clases.

No linealidad. Sea X_i la transformada de Fourier discreta de la i -ésima imagen de entrenamiento. Aplicamos el filtrado no lineal (Javidi et al., 1997) sobre la transformada de Fourier de cada imagen de entrenamiento X_i como sigue:

$$X_i^k = |X_i|^k \exp(-i\varphi_{X_i}) \quad (46)$$

donde k es el factor de no linealidad, $|X_i|$ es el módulo de X_i y φ es su fase. El factor de no linealidad k eleva a la k -ésima potencia la magnitud de la transformada de Fourier mientras se mantiene intacta la información de la fase.

Sin restricciones. Sustituyendo la ecuación 46 en la ecuación 44 obtenemos:

$$h(x, y) = a_1 X_1^k + a_2 X_2^k, \dots, a_N X_N^k$$

$$= a_i \sum_{i=1}^N X_i^k \quad (47)$$

con $a_i = \frac{1}{N}$, el filtro compuesto no lineal no restringido (FCnLsR) en dominio de Fourier está dado por la ecuación:

$$H^* = \left(\frac{1}{N}\right) \sum_{i=1}^N X_i^k \quad (48)$$

El coeficiente a_i no se selecciona para satisfacer la ecuación 45, en su lugar se hace el promediado, el cual enfatiza las características comunes de las imágenes de entrenamiento, mientras que elimina aquellas características no comunes. Con esto se consigue un que el filtro no esté restringido en el origen.

7.2 Regiones faciales

La cara humana es un objeto tridimensional el cual se está intentando reconocer en este trabajo mediante su imagen digital bidimensional. La mayoría de los métodos de reconocimiento facial consideran la cara como un solo objeto, y no consideran que las características faciales como el ojo izquierdo, el ojo derecho, la nariz, la boca y las mejillas tengan su propio comportamiento. Este comportamiento influye en el reconocimiento de la cara completa. Cuando una persona habla la boca cambia de forma. Este cambio de forma modifica la estructura no únicamente de la característica sino de una región que rodea dicha característica.

Considere una cara completa $f = \bigcup_{i=1}^N r_i$, donde r_i es i-ésima región facial. Cada región facial puede ser considerada como un clasificador débil que puede ser capaz de reconocer una cara completa o parcial. La unión de muchos clasificadores débiles produce un clasificador fuerte más discriminativo y robusto para reconocer imágenes faciales parciales. Esto si se considera que si la imagen de entrada está incompleta, la probabilidad de que ésta se parezca a una de las regiones de entrenamiento es alta y por ende sea reconocido con una probabilidad alta. También, el efecto que sufre la cara del cambio de forma de una de las características faciales influye

menos ya que es solo una región la afectada. Una región facial es una área de la imagen facial que contiene una o más características faciales, tal y como se muestra en la Figura 7.1.

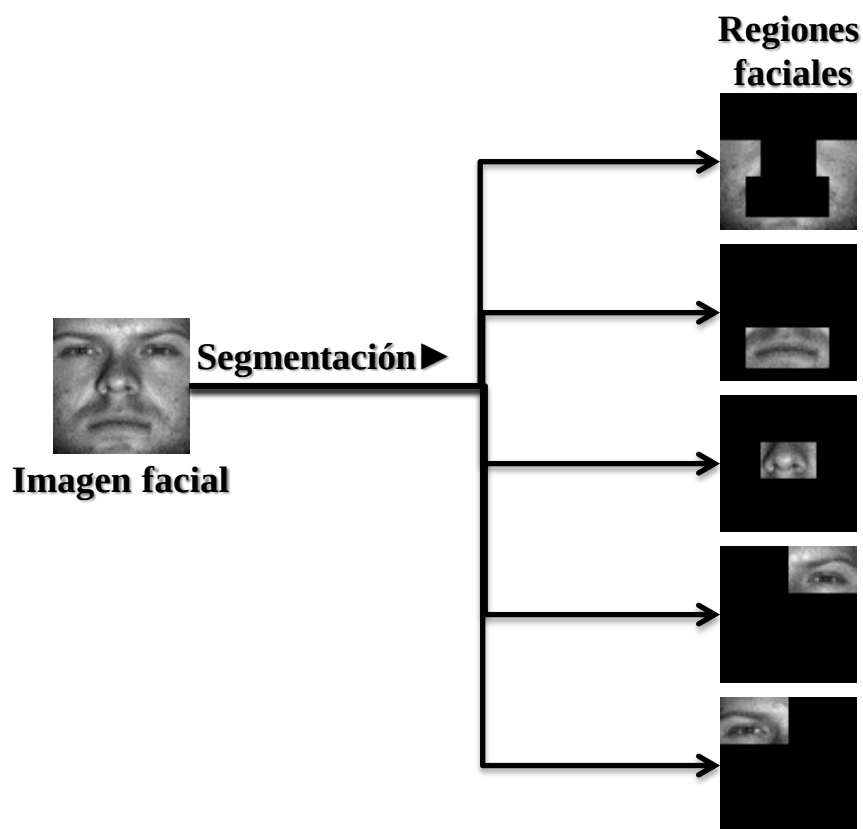


Figura 7.1. Principales regiones de una cara humana.

El procedimiento para obtener regiones faciales como se muestra en la Figura 7.1 puede ser de diferentes formas, siempre y cuando se mantenga la esencia que son las regiones. En este trabajo se usan máscaras binarias, de tal forma que una imagen facial multiplicada elemento a elemento con una máscara produce la región deseada. El procedimiento completo para la obtención de regiones se presenta a continuación.

Algoritmo 1. Obtener regiones faciales

Entrada:

- Imagen facial $i(x, y)$
- Un conjunto de máscaras binarias $M = \{m_1, m_2, \dots, m_N\}$

Salida:

- Un conjunto de regiones faciales $R = \{r_1, r_2, \dots, r_N\}$

Procedimiento:

- (1) Inicializar conjunto $R = \{\}$
- (2) Multiplicar elemento a elemento la máscara m_i con la imagen facial de entrada:
 - a. $r_i = i(x, y) \odot m_i$, donde $\odot$ representa la multiplicación elemento a elemento.
- (3) Agregar al conjunto R la región r_i
- (4) Repetir los pasos (2) y (3) $|M|$ veces

Cada región facial permanece en su posición original, con el objetivo de que la salida de correlación coincida en el origen para cada región facial. Una de las ventajas de usar regiones faciales para el entrenamiento del filtro es que si una imagen de entrada solo contiene una parte de la cara, ésta se parezca a una de las regiones de entrenamiento y se produzca un valor alto de correlación. Con esto se pretende reconocer la cara fragmentada con alto grado de certeza. Otra ventaja es que regiones como la de las mejillas y barbillas permanecen sin muchos cambios cuando una persona habla, por lo que podría utilizarse para seguimiento. La región de los ojos también cambia con menos frecuencia.

7.3 Reconocimiento facial mediante FCnLsR

El esquema de reconocimiento facial propuesto en esta sección está compuesto generalmente de dos módulos. El primero, el de registro de un nuevo usuario al sistema en donde éste proporciona N imágenes faciales de muestra para construir su plantilla biométrica. El procedimiento es como sigue.

Algoritmo 2. Construcción de la plantilla biométrica facial

Entrada:

-Conjunto de imágenes faciales de la misma clase

$$I = \{i(x, y)_1, i(x, y)_2, \dots, i(x, y)_N\}$$

Salida:

-Filtro de correlación FCnLsC (plantilla biométrica facial)

Procedimiento:

- (1) Obtener en el conjunto R todas regiones faciales de cada imagen de entrada $i_j(x, y)$ con el Algoritmo 1, donde $i_j(x, y) \in I$
- (2) Inicializar matriz de sumas X con sus elementos a 0

(3) Para cada región facial contenido en R:

- a. Se obtiene la transformada de Fourier de la región r_i , donde $r_i \in R$
- b. Se aplica la no linealidad a la transformada de Fourier de r_i como se expresa en la ecuación 46.
- c. Se le suma a la matriz X la matriz obtenida en el paso (3).b

(4) Promediar la matriz X para obtener el filtro FCnLSR como en la ecuación 48.

El segundo módulo es de reconocimiento y entra en funcionamiento cuando hay una persona a reconocer. El procedimiento está dado por el siguiente algoritmo.

Algoritmo 3. Reconocimiento facial mediante correlación

Entrada:

- Imagen facial de prueba $i(x, y)$
- Plantilla biométrica facial $g_i(x, y)$
- Umbral PSR de reconocimiento τ_{psr}

Salida:

- Confirmación de la autenticidad o falsedad de la imagen de entrada
- Valor PSR resultante de la comparación

Procedimiento:

- (1) Obtener la transformada de Fourier de la imagen facial de entrada $i(x, y)$. Sea F el resultado de este paso
- (2) Aplicar la no linealidad a la transformada de Fourier obtenida en el paso 1 como en la ecuación 46. Sea F^k el resultado de este paso.
- (3). Sea H^* el complejo conjugado del filtro $g_i(x, y)$. Multiplicarlo elemento a elemento con la matriz F^k para producir el plano de salida de correlación: $C = F^k \cdot H^*$
- (4) Examinar la salida de correlación C calcular la agudeza del pico de correlación con la ecuación 24.
- (5) Si el valor PSR producido por el paso (4) es mayor a τ_{psr} entonces se determina que la imagen facial es auténtica, en caso contrario se determina que es falsa.

En la Figura 7.2 se presenta el esquema general del método propuesto y en la Figura 7.3 se muestra el proceso de correlación de un filtro FCnLsR con una imagen facial completa de prueba. Se produce un plano de correlación con poco ruido y un pico agudo y fácil de distinguir y localizar.

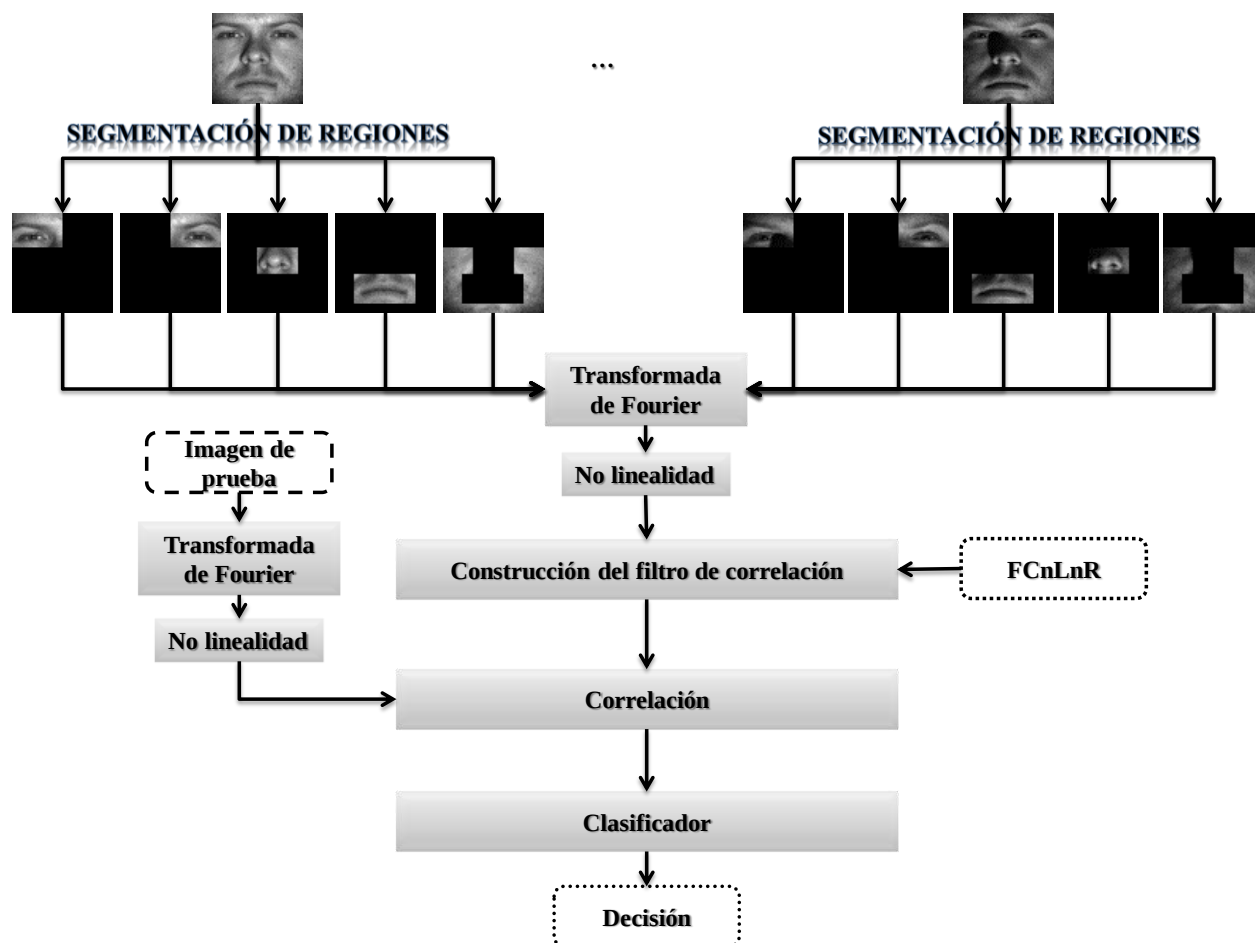
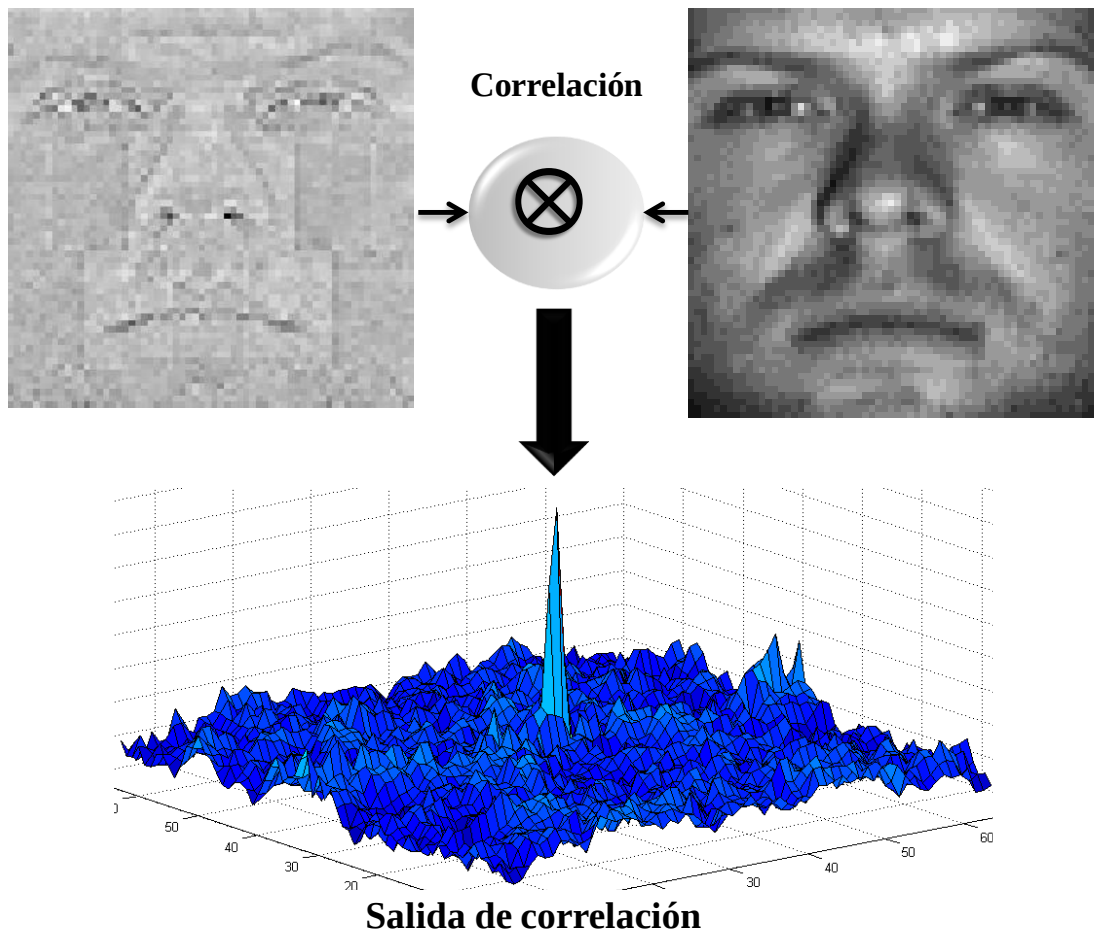


Figura 7.2. Esquema general del método de reconocimiento facial propuesto.

Filtro de correlación FCnLsR**Imagen facial de prueba****Figura 7.3.** Salida de correlación del filtro FCnLsR para una imagen facial auténtica.

La simplicidad del método propuesto evidencia la posibilidad de implementarse en un sistema de reconocimiento facial automático en tiempo real o en dispositivos con recursos computacionales limitados, con los procedimientos descritos en Algoritmo 4 y Algoritmo 5, respectivamente.

Algoritmo 4. Verificación facial mediante el filtro FCnLsR

Entrada:

- Imagen facial de prueba $i(x, y)$
- Presunta identidad id
- Umbral de reconocimiento τ_{psr}

Salida:

-Confirmación o negación de la identidad id declarad

Procedimiento:

- (1) Se recupera de la galería G la plantilla $g_i(x, y)$, donde $id(g_i(x, y))$ es presumiblemente igual a $id(i(x, y))$.
- (2) Se verifica la autenticidad de la imagen facial de entrada mediante el Algoritmo 3 con $i(x, y)$, $g_i(x, y)$ y τ_{psr} como parámetro.
- (3) Si en el paso (2) se confirma la autenticidad de la imagen facial de entrada, entonces se confirma que $id(g_i(x, y)) = id(i(x, y))$. En caso contrario se niega la identidad.

Algoritmo 5. Identificación facial mediante el filtro FCnLsR

Entrada:

- Imagen facial de prueba $i(x, y)$
- Umbral de reconocimiento τ_{psr}

Salida:

- Id de la persona en caso de que esté registrada en la galería

Procedimiento:

- (1) Inicializar el vector de resultados PSR $\mathbf{v}$
- (2) Se recupera la plantilla $g_i(x, y)$ de la galería G , donde $i = 1, 2, \dots, |G|$.
- (3) Se verifica la autenticidad de la imagen facial de entrada mediante el Algoritmo 3 con $i(x, y)$, $g_i(x, y)$ y τ_{psr} como parámetro.
- (4) Si en el paso tres se determina que la imagen facial es auténtica, entonces se registra en el vector $\mathbf{v}$ el valor PSR producido
- (5) Devolver la identidad del usuario $id(\max(\mathbf{v}))$. Si $\mathbf{v}$ está vacío se declara como sujeto no conocido

7.4 Evaluación del método propuesto

Para evaluar el método propuesto con imágenes faciales parciales, se construyeron dos filtros FCnLsR. El primer filtro se construyó con cinco regiones faciales extraídas de una imagen facial completa de entrenamiento. El segundo filtro se construyó con una imagen facial completa de entrenamiento. Se generaron 6 versiones de una imagen facial de prueba, con 17.26%,

17.29%, 10.89%, 18.19%, 81.47% y 100% de información facial, respectivamente. Estas imágenes faciales parciales fueron usadas como pruebas a reconocer. La Figura 7.4 muestra el desempeño PSR obtenido de cada filtro ante las imágenes parciales de prueba. Los valores PSR resultantes del filtro construido con regiones faciales son más altos cuando la de prueba contiene poca información, tal es el caso de las parciales 1, 2, 3 y 4. Para las imágenes parciales 5 y 6, con 81.47% y 100% de información, respectivamente, el filtro FCnLsR construido con imágenes completas obtiene valores PSR más altos.

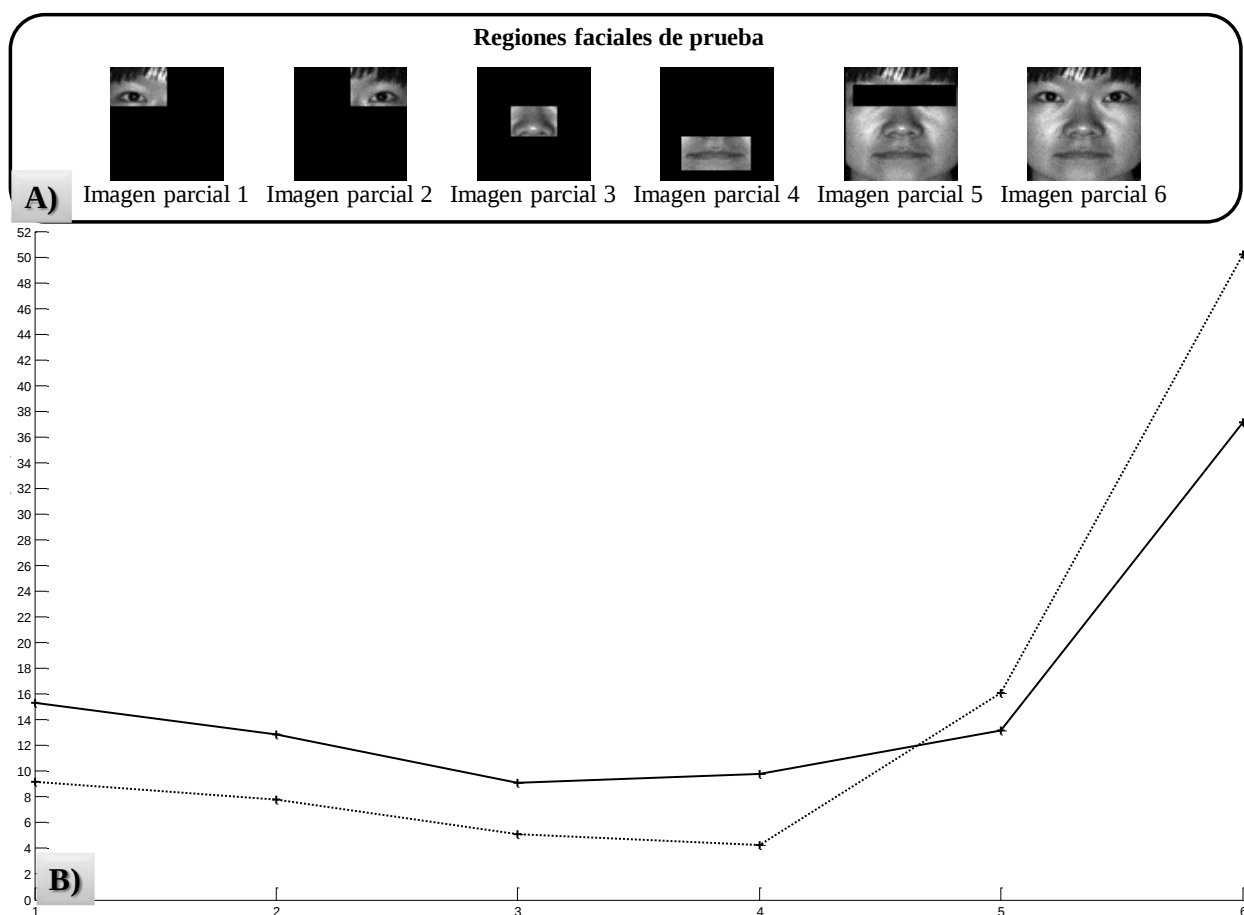


Figura 7.4. A) Imágenes parciales de prueba. B) Desempeño PSR del filtro FCnLsR con regiones faciales de prueba. Línea continua: filtro construido con regiones. Línea segmentada: filtro construido con imágenes completas.

El método propuesto fue evaluado usando las bases de datos de imágenes faciales YALE B y ORL en la tarea de identificación y verificación facial. Cada imagen facial, de ambas bases de datos, fue recortado y escalado manualmente a una resolución de 64×64 píxeles. Todas las imágenes son en escala de grises con un rango dinámico de 0-255.

Para calcular el desempeño del método propuesto se crearon tres conjuntos de imágenes faciales. El primero es la galería $G = \{g_1, g_2, \dots, g_{|G|}\}$ el cual contiene la plantilla biométrica facial de cada persona conocida por el sistema creado con el Algoritmo 2. Los otros dos son conjuntos de pruebas, el primer conjunto P_G de prueba que contiene muestras faciales de personas en la galería. El conjunto P_G contiene $38 \times 64 = 2432$ imágenes faciales del conjunto YALE B. El otro conjunto de prueba se denomina P_N , el cual contiene muestras biométricas de personas que no están en la galería. El conjunto P_N contiene $40 \times 10 = 400$ de la base de datos ORL.

Los filtros de la galería fueron creados con regiones faciales de 4 imágenes faciales seleccionadas de tal forma que se pudiera reconocer 64 imágenes faciales (4 de entrenamiento y 60 restantes). Las imágenes de entrenamiento fueron seleccionadas siguiendo la estrategia 2 presentada en la sección 6.2, la Figura 7.5 muestra un ejemplo de imágenes de entrenamiento para una clase. Éstas representan de forma general las distorsiones del conjunto de datos YALE B de cada persona.

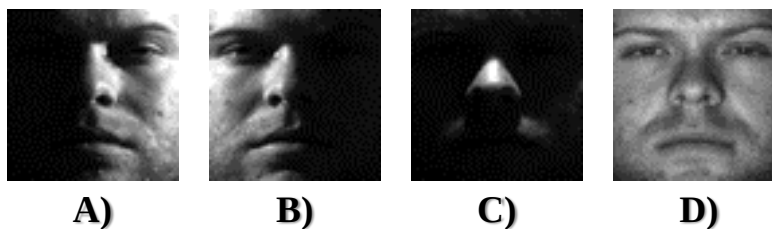


Figura 7.5. Imágenes de entrenamiento. A) Sombra perfil derecho, B) sombra perfil izquierdo, C) sombra total y D) luz ambiental, sin sombras.

Cada imagen facial se segmentó en cinco regiones faciales. Por lo tanto cada filtro $g_i(x, y)$ en la galería G se construyó con $4 \times 5 = 20$ regiones faciales. Antes de la segmentación se aplica la transformación logarítmica a las imágenes para mejorar su calidad en caso de que presenten sombras o iluminación intensa. La agudeza del pico producido por la correlación se mide usando la métrica PSR, el cual a su vez representa el grado de similitud entre una imagen facial de prueba y plantilla biométrica facial.

A base de muchos experimentos se determinó usar 4 imágenes de entrenamiento para obtener el mínimo desempeño aceptable. Sin embargo, mientras más imágenes se usen para construir el filtro el desempeño mejor. Considere la imagen facial D de la Figura 7.5 para construir un filtro FCnLsR de la persona 1. Al correlacionar dicho filtro con las 64 imágenes de la misma clase, se obtiene el desempeño que se muestra en la Figura 7.6. Los valores PSR de la

línea continua que caen dentro del rectángulo negro sesgado de la Figura 7.6 corresponden a imágenes sombreadas, ya sea total o parcialmente y su bajo valor se debe a que no existe un acoplamiento perfecto con el filtro.

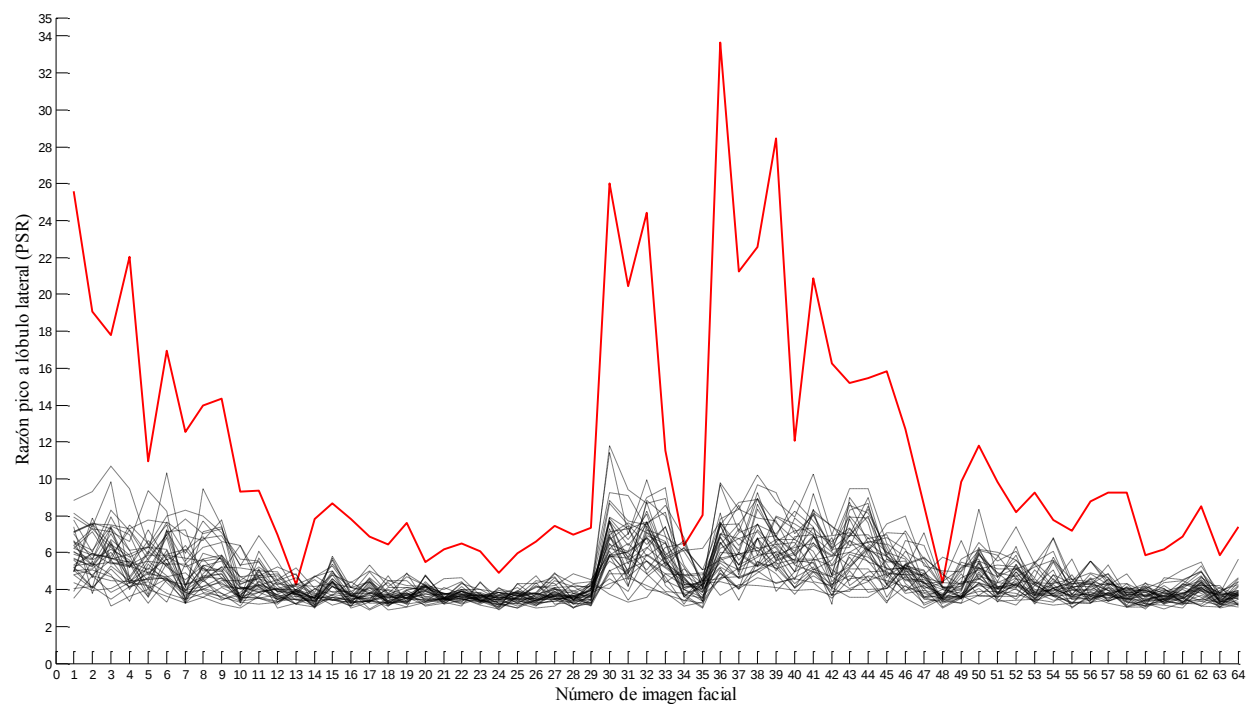


Figura 7.6. Desempeño PSR del filtro FCnLsR de la persona 1.

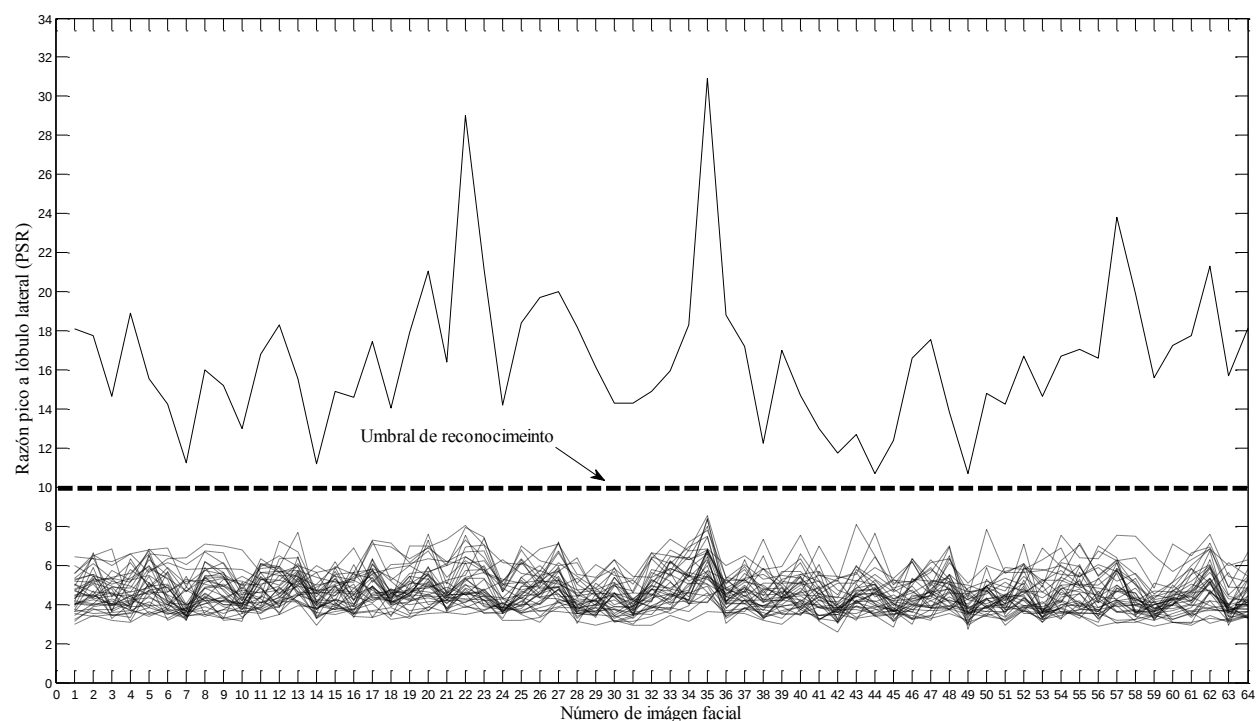


Figura 7.7. Desempeño PSR del filtro de la persona 1 construido con 20 regiones faciales obtenido de cuatro imágenes faciales.

En la Figura 7.7 se muestra el desempeño obtenido con cuatro imágenes faciales para la construcción del filtro. Se reconoce correctamente aquellas que están en el conjunto de entrenamiento e incluso la mayoría de las que pertenecen a la misma clase pero no están en el conjunto de entrenamiento. La cantidad de imágenes faciales auténticas rechazadas incorrectamente son muy pocas. Usando el umbral de reconocimiento $PSR = 10$ se rechazan correctamente todas las imágenes faciales impostoras.

Con la configuración descrita anteriormente, se calculó el desempeño del método propuesto a las medidas de desempeño presentados en la sección 3.1.4, las cuales son: Tasa de identificación y detección (TID), tasa de falsa alarma (TFA), tasa de verificación (TV) y tasa de falsa aceptación (TFAC). Los resultados obtenidos con estas estadísticas se muestran en la Tabla 7.1 en porcentajes.

Tabla 7.1. Desempeño de filtros de correlación no restringidos en el reconocimiento facial.

Filtro	Datos de entrenamiento	TID	TFA	Precisión	TV	TFAC	Precisión
UNCF	Regiones	95.60	1.25	99.78	95.72	0.25	90.87
UMACE	Regiones	96.42	19.75	96.74	97.57	2.18	54.07
	Completas	97.82	26	95.81	98.43	2.42	51.75

El filtro FCnLsR obtuvo un buen desempeño al reconocer una persona de forma correcta con una probabilidad de 95.60% y 96.72% para la identificación y verificación facial, respectivamente. En lo referente a las tasas de falsa alarma y falsa aceptación se obtuvieron valores muy bajos. En el caso de la identificación, la probabilidad de que el método propuesto declare como auténtico una imagen facial impostora cuando éste ni siquiera es de una persona que está en la galería es muy baja. De 400 imágenes faciales impostoras, solo 5 fueron declaradas como auténticas. En la verificación, la probabilidad de que el método declare como iguales a dos imágenes faciales que no lo son es de solo el 0.25%. Es decir, de las 89994 imágenes faciales impostoras, solo 225 se declararon incorrectamente como auténticas. Una falsa aceptación sucede cuando el algoritmo confunde dos imágenes de prueba de diferentes clases.

El filtro UMACE fue evaluado también con regiones faciales como conjunto de entrenamiento. El desempeño en términos de TID y TV obtenido por este filtro (entrenado con regiones e imágenes completas) es mayor al obtenido por el filtro propuesto en esta sección. Sin embargo la TFA y TFAC son muy altos, esto indica el filtro UMACE tiene una alta probabilidad de confundir dos clases diferentes de imágenes faciales. En este sentido el filtro FCnLsR ofrece el mejor desempeño. Para evidenciar estadísticamente la superioridad del filtro FCnLsR con respecto al filtro UMACE se midió la precisión con que cada filtro realiza la clasificación de la siguiente manera:

$$Pr = \frac{\text{No. de reconocimientos correctos}}{\text{No. de reconocimientos correctos} + \text{No. de reconocimientos incorrectos}} \quad (49)$$

donde Pr es la precisión. Un reconocimiento correcto sucede cuando una imagen facial de prueba $i(x, y)$ es comparado con una plantilla $g_i(x, y)$ y resulta un valor PSR mayor al umbral de reconocimiento τ_{psr} y $id(i(xy)) = id(g_i(x, y))$. Un reconocimiento incorrecto sucede cuando una imagen facial de prueba $i(x, y)$ es comparado con una plantilla $g_i(x, y)$ y resulta un valor

PSR mayor al umbral de reconocimiento τ_{psr} y $id(i(xy)) \neq id(g_i(x,y))$. $id(\cdot)$ representa la identidad.

Como se puede ver en la Tabla 7.1, el filtro FCnLsR tiene el mejor desempeño con 99.78% y 90.87%, en la tarea de identificación y verificación respectivamente. Siendo el mejor desempeño de los filtros evaluados en cada tarea.

En la Figura 7.8 se muestra el mejor desempeño PSR obtenido con el filtro FCnLsR. Como se puede observar, el espacio que separa la clase auténtica de las clases impostoras es mayor que el obtenido con el mejor caso de filtro UMACE construido con imágenes faciales completas. Este espacio de separación se traduce como una mayor capacidad de discriminación del filtro FCnLsR. En las Figuras 7.9 y 7.10 se muestra los mejores casos del filtro UMACE entrenado con imágenes faciales completas y regiones, respectivamente. La separación existente entre los valores PSR de la clase auténtica con respecto de las clases impostoras en ambas gráficas es muy poco, lo que evidencia una pobre capacidad de discriminación.

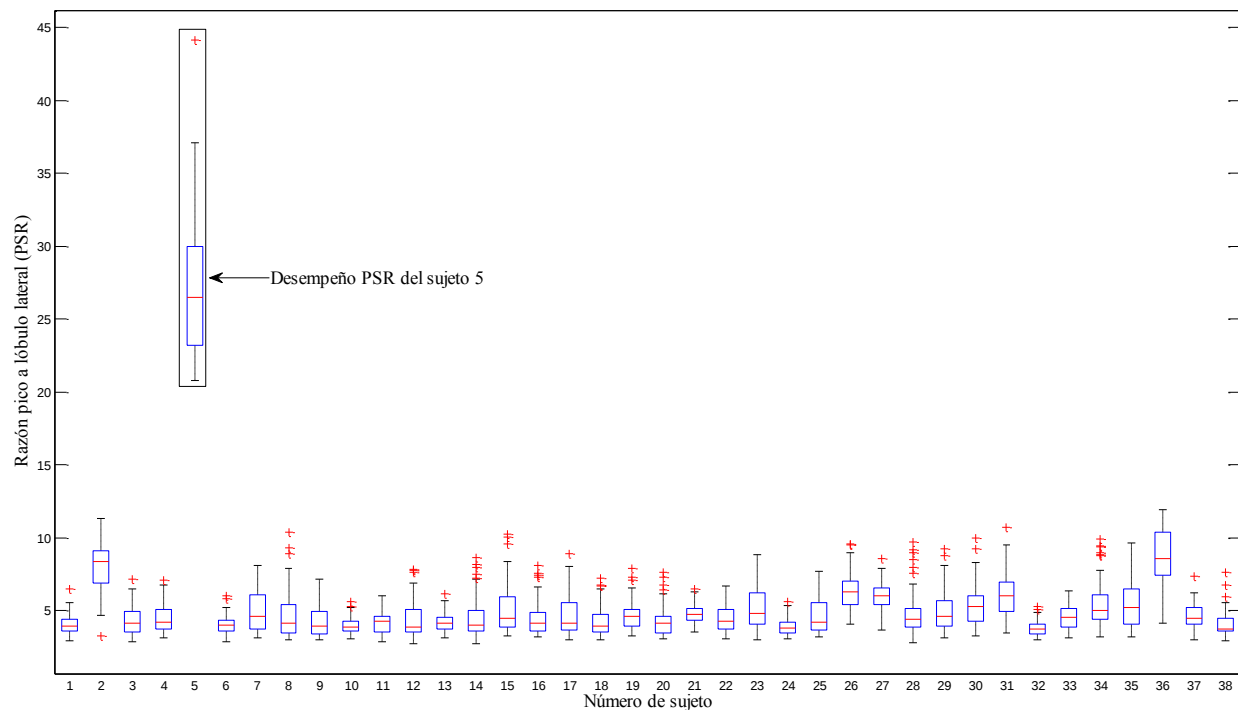


Figura 7.8. Mejor desempeño PSR del filtro FCnLsR.

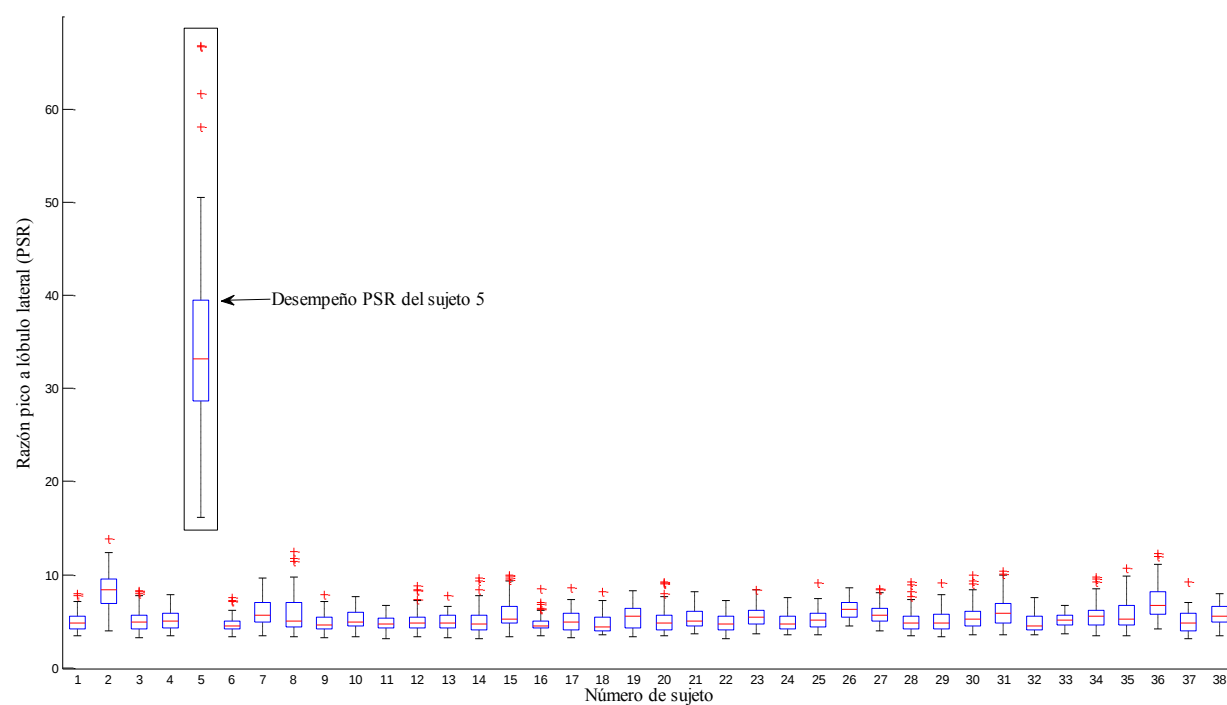


Figura 7.9. Mejor desempeño PSR del filtro UMACe usando imágenes completas de entrenamiento.

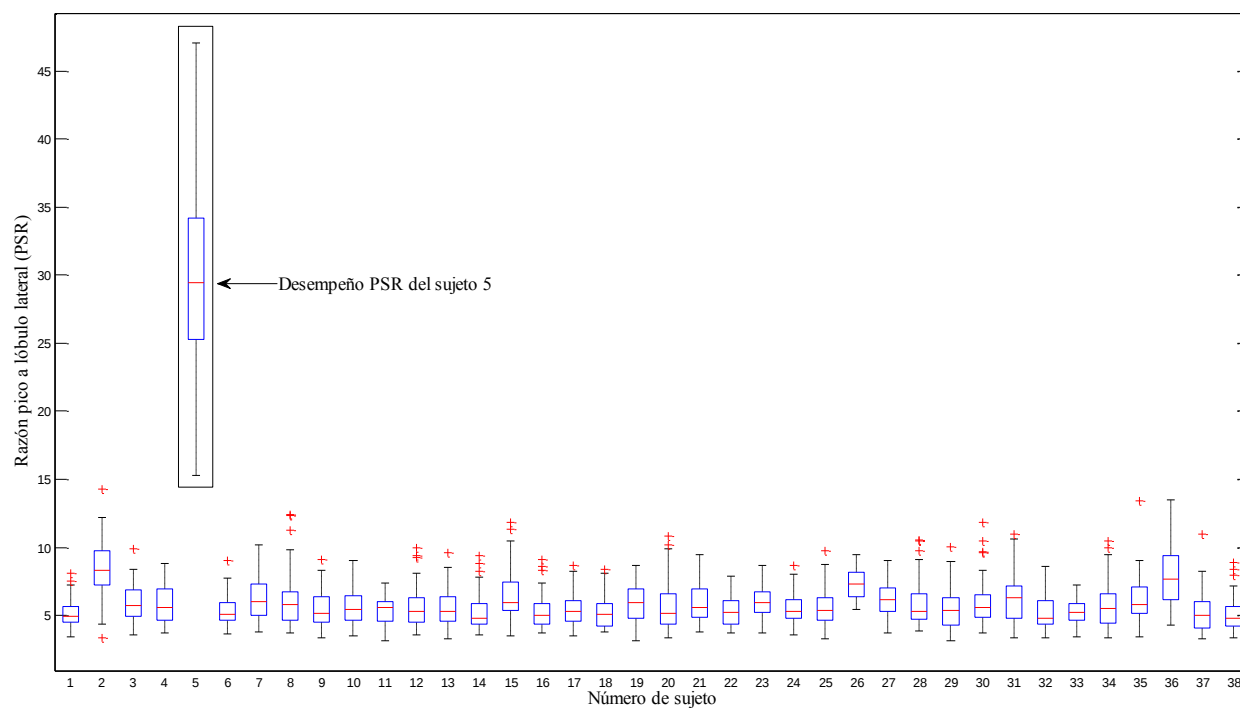


Figura 7.10. Mejor desempeño PSR del filtro UMACe usando imágenes completas de entrenamiento.

En esta sección se propuso un método de reconocimiento facial que usa un filtro de correlación SDF compuesto no lineal sin restricciones. En el proceso de la construcción del filtro, cada imagen de entrenamiento es segmentado en cinco regiones faciales que con contiene una o más características faciales. Este proceso provoca que un filtro de correlación sea construido con veinte imágenes (regiones) y la probabilidad de error disminuye. El filtro de correlación se construye promediando la transformada de Fourier con la Ley k de las imágenes de entrenamiento. Cuando se tiene una nueva imagen de prueba, se le aplica la no linealidad a su transformada de Fourier y se le correlaciona con el filtro. El plano de salida producido se examina en busca de un pico. El pico, mediante la métrica PSR se mide su agudeza y se compara con un umbral de reconocimiento.

Los resultados experimentales usando dos bases de datos evidencian la efectividad del método propuesto. Se alcanza una tasa de identificación y detección de 95.69% y 95.72%, en las tareas de detección y verificación facial, respectivamente. El método propuesto funciona bien con imágenes faciales parciales y completas. El uso de la transformación logarítmica sobre las imágenes faciales se logra la invariancia a la iluminación.

Capítulo 8

8. Seguimiento facial mediante un filtro de compuesto no lineal, sin restricciones y adaptativo

En algunas aplicaciones donde se trabaja con la cara de una persona existe la necesidad de seguir dicha cara dentro un escenario por distintas razones. Por ejemplo, para inferir la actividad que realiza dicha persona.

Existen muchas estrategias de seguimiento de objetos que han sido aplicados exitosamente al seguimiento facial como el filtrado de partículas y el método de búsqueda logarítmica en rejillas. Estos métodos funcionan bien bajo ambientes con poca o nula variabilidad ambiental. El filtro de correlación conocido como MOSSE ha sido aplicado al seguimiento de objetos obteniendo resultados prometedores, tales como la velocidad de procesamiento y la capacidad de adaptarse rápidamente a los cambios que sufre la cara debido a la iluminación, pose, rotación, escala, oclusión, entre otros.

En esta sección se propone una estrategia de seguimiento facial basado en un filtro de correlación compuesto no lineal sin restricciones que se actualiza de forma incremental para adaptarse a cambios de la cara en un ambiente natural y en secuencias de imágenes, tales como las expresiones.

La hipótesis de esta propuesta es que un filtro no lineal, sin restricciones y que sea capaz de adaptarse a los cambios que sufre la cara a través de las secuencias de imágenes es más robusto en cuanto capacidad para recuperarse de fallas y mantener el seguimiento en presencia de distorsiones.

8.1 Filtro compuesto no lineal, sin restricciones y adaptativo

En el seguimiento facial es necesario adaptar al filtro los cambios que sufra la región facial a través de la secuencia de imágenes. Por lo tanto es necesario que el filtro de correlación construido en una etapa inicial pueda ser actualizado en cada imagen de la secuencia en donde el objeto aparece. Esta actualización debe ser de tal forma que el filtro no crezca en tamaño para

que no ocupe mucho espacio en memoria y no disminuya la velocidad del algoritmo. La estrategia de actualización del filtro se describe a continuación.

No linealidad y no restricción. Sea X la matriz que contiene la suma de la transformada de Fourier con la ley k en la imagen $N - 1$ de la secuencia:

$$X = \sum_{i=1}^{N-1} X_i^k \quad (50)$$

donde X_i^k se obtiene como en la ecuación 45.

Adaptabilidad. Considere $f_j(x,y)$ la región de la imagen N de la secuencia donde se localizó la cara. El filtro compuesto no lineal adaptativo (FCnLsRA) con actualización incremental está dado por:

$$H_i^* = \left(\frac{N-1}{N}\right)X + \left(\frac{1}{N}\right)F_j^k \quad (51)$$

donde F_j^k es la transformada de Fourier con la ley k de $f_j(x,y)$ obtenido con la ecuación 46. Como se puede ver en la ecuación 51, este filtro de correlación siempre mantendrá el mismo tamaño, esto permite mantener una velocidad constante en el seguimiento. Esta estrategia de actualización incremental del filtro de correlación no es nueva, pues fue introducido por Savvides et al. (2003) para los filtros de correlación UMACE y UOTSDF, en este trabajo lo aplicamos a un filtro compuesto no lineal adaptativo y sin restricciones.

8.2 Estrategia de seguimiento facial basado en correlación usando el filtro FCnLsR

La estrategia de seguimiento facial que se propone en esta sección está compuesta por un conjunto de procedimientos. El primer procedimiento inicializa el filtro de correlación generando versiones de la cara inicial de entrada, el cual se describe a continuación:

Algoritmo 6. Creación del filtro no lineal incremental inicial

Entrada:

- Imagen facial $f(x,y)$
- N, Cantidad de versiones a generar

Salida:

- X matriz que contiene la suma de las transformadas de Fourier con la ley k
- Filtro de correlación inicial H^*

Procedimiento:

(1) Inicializar X

(2) Desde $n=1$ hasta N

- a. Generar la distorsión $f'_n(x, y)$ aplicando una transformación espacial a $f(x, y)$
- b. Obtener la transformada de Fourier de $f'_n(x, y)$
- c. Aplicar la no linealidad de la ecuación 46 a la transformada de Fourier obtenida en el paso (2).b.
- d. Añadir a X la matriz F_n^k obtenida en el paso (2).c.

(3) $H^* = X$

La estrategia de seguimiento que se propone en esta sección está descrita en el siguiente procedimiento:

Algoritmo 7. Seguimiento facial mediante filtros de correlación**Entrada:**

- Secuencia de video
- Imagen facial inicial $f(x, y)$
- Umbral de reconocimiento τ_{psr}

Salida:

- localización visual de la cara en la secuencia de video

Procedimiento:

(1) Inicializar filtro H^* con el Algoritmo 6 y pasándole como parámetro $f(x, y)$

(2) Para cada imagen $s_i(x, y)$ de la secuencia de video de entrada

- a. Obtener la transformada de Fourier de $s_i(x, y)$
- b. Aplicar la no linealidad a $s_i(x, y)$ según la ecuación 46, este paso produce $S_i^k(x, y)$.
- c. Correlacionar el filtro H^* con $S_i^k(x, y)$

- d. Localizar y calcular la agudeza del pico de correlación de la salida producida en el paso (2).c según la ecuación 24.
- e. Si el valor PSR resultante del paso (2).d es mayor o igual al umbral τ_{psr}
 - i. Mostrar visualmente al usuario la localización de la cara.
 - ii. Segmentar la región donde se localizó el pico
 - iii. Actualizar el filtro H^* según la ecuación 51 con la región segmentada en el paso (2).e.ii.

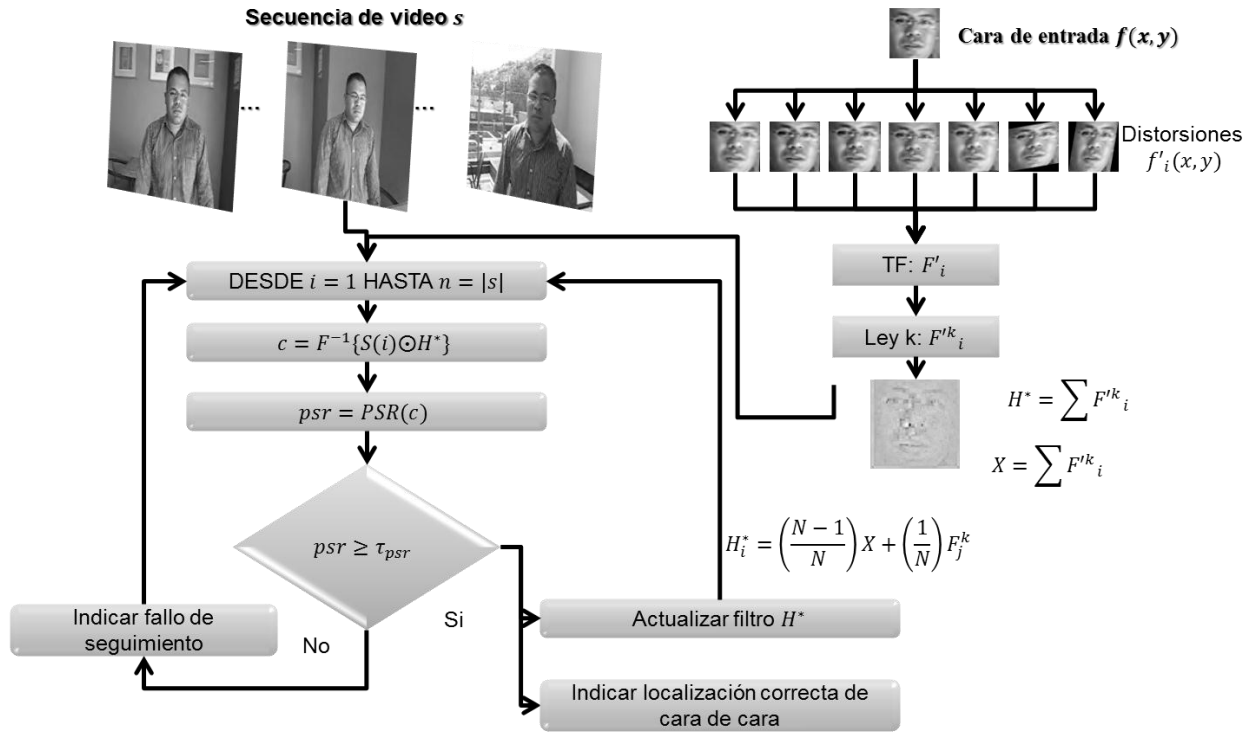


Figura 8.1. Esquema general de la estrategia de seguimiento facial.

En la Figura 8.1 se presenta el esquema general de la estrategia de seguimiento facial. Se recibe como parámetro una cara detectada ya sea manual o automáticamente en un paso anterior al iniciar el seguimiento y una secuencia de video. La detección facial automática está fuera del objetivo de este trabajo por lo que no se aborda. Un algoritmo de detección facial que ha sido ampliamente usado y que puede usarse es el algoritmo presentado por Viola y Jones (2003) o también puede entrenarse el filtro ASEF presentado en la sección 2.7 para que detecte la cara inicial en la secuencia de video.

8.3 Evaluación

La estrategia de seguimiento facial propuesto en este trabajo fue evaluado con una secuencia de video capturado con una cámara fotográfica Canon. El video consiste de 120 imágenes en donde aparece una persona con el rostro visible, dicha secuencia se denomina Ever03. Las imágenes de la secuencia de video fueron capturadas en escala de grises y con una resolución de 240×320 píxeles. Esta secuencia de video fue capturada bajo condiciones de luz ambiental. En la secuencia de video se observa a una persona moverse en el escenario teniendo como fondo una pared, haciendo algunas expresiones faciales, con lentes/sin lentes, ligeros movimientos hacia los lados y acercamiento hacia el sensor.

El área donde se localiza la cara en la primera imagen de la secuencia fue seleccionada de forma manual. Esta área tiene una resolución de 54×54 píxeles (conocida como ventana de seguimiento) y encierra correctamente la cara humana. Siguiendo el algoritmo 6 se generaron 8 versiones del área inicial, con diferentes grados de rotación, ruido y otras transformaciones espaciales, para construir el filtro de correlación inicial. Este proceso se representa en la Figura 8.2.

Cada versión generada fue procesada de la siguiente manera. Cada valor de pixel se transformó usando una función logarítmica la cual ayuda a mejorar la calidad de las imágenes de bajo contraste. Posteriormente se normalizó la energía de la imagen para tener una media de 0 y norma 1. Luego se multiplicó por una ventana de coseno para resaltar la región central de la imagen y disminuir el efecto de los bordes. Finalmente se multiplicó la imagen resultante de los pasos anterior por una ventana Gaussiana 2D para lograr una consistencia uniforme en las esquinas de la imagen pero con la parte central de la imagen con las características distinguibles. El objetivo de aplicar el conjunto de operaciones de pre-procesamiento anterior, es resaltar la región central de la imagen de referencia mientras se difuminan las esquinas, obteniendo un filtro de correlación como el que se muestra en la Figura 8.2.

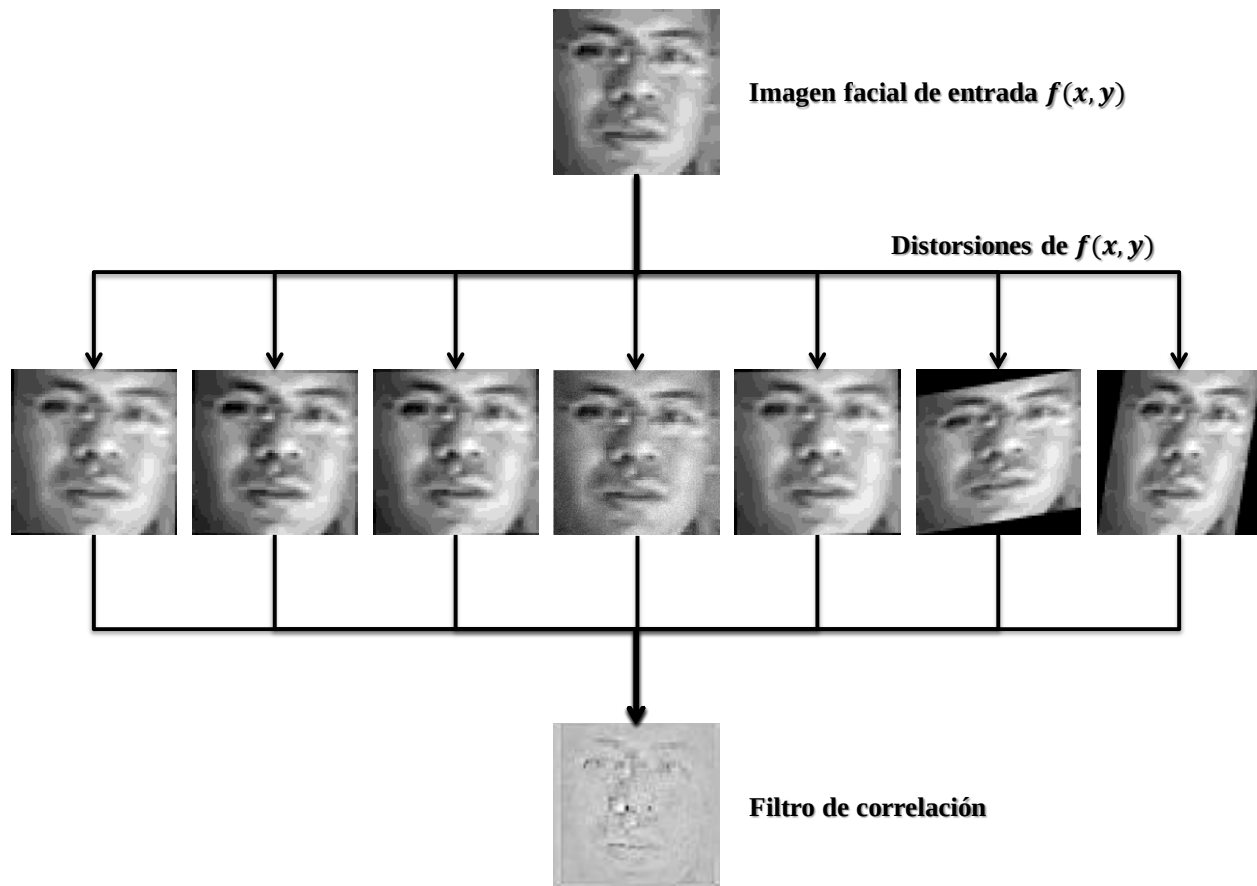


Figura 8.2. Inicialización del filtro de correlación para el seguimiento facial.

Una vez inicializado el filtro de correlación, éste se correlaciona con la primera imagen de la secuencia de video. La salida de correlación es examinada en busca del pico de correlación. Una vez localizado se mide su agudeza con la métrica PSR, si este valor es mayor o igual que el umbral τ_G entonces se determinan un reconocimiento y localización correctos y se actualiza el filtro para adaptarse a los cambios.

Un valor PSR menor que τ_G indica una localización y reconocimiento fallido. El enfoque propuesto y al igual que otro enfoque basado en filtro de correlación tiene la capacidad de recuperarse y la velocidad de procesamiento es buena.

La estrategia de seguimiento propuesto ofrece tolerancia a expresiones faciales, por ejemplo, abrir la boca, cerrar los ojos y muecas con la boca, tal y como se muestra en la Figura 8.3. También el filtro funciona bien cuando la cara presenta ligeros giros hacia los lados, por ejemplo los presentados en la Figura 8.4. En esta misma figura se muestra una imagen facial con iluminación intensa en la mejilla izquierda y reconocida y localizada de forma exitosa.



Figura 8.3. Desempeño del filtro compuesto no lineal, sin restricciones y adaptativo en imágenes donde caras presenta expresiones faciales.



Figura 8.4. Desempeño del filtro compuesto no lineal, sin restricciones y adaptativo en imágenes donde la cara presenta movimientos ligeros.

8.3.1 Comparación con otro seguidor basado en filtros de correlación

En esta sección se evalúa el desempeño del filtro de correlación FCnLsRA y el filtro MOSSE. Ambos filtros fueron evaluados con cuatro secuencias de videos en cuatro escenarios diferentes. La primera secuencia se llama Ever03, captura en un ambiente con ligeros cambios de

iluminación en el interior de un edificio, cuenta con 578 imágenes en escala de grises donde aparece una persona (con rostro visible) en movimiento y haciendo expresiones.

Las otras tres secuencias se obtuvieron de <http://www.cs.toronto.edu/~dross/ivt/>. davidin300, dudek y trellis70, con 462, 573 y 502 imágenes, respectivamente. Estos videos de prueba son en escalas de grises e incluyen fuertes variaciones en iluminación, pose y apariencia. La cámara se mueve en las cuatro secuencias de video evaluados.

El desempeño de un algoritmo de seguimiento facial se puede caracterizar por cuatro métricas que se basan en la exactitud de la localización y reconocimiento de la cara bajo observación.

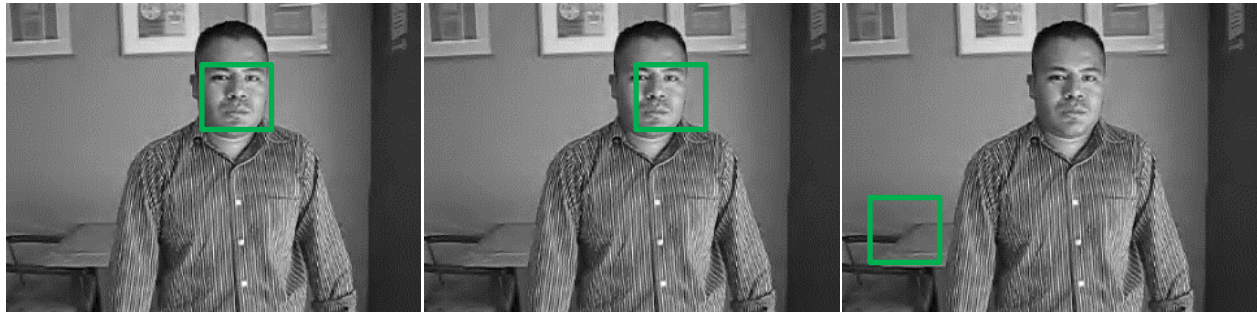


Figura 8.5. Salidas de localización de un seguidor facial.

Considere la Figura 8.5 de izquierda a derecha. La primera imagen muestra una localización y reconocimiento correcto de la cara bajo observación en una imagen de la secuencia. La segunda imagen muestra una localización desviada del centro de la cara en dicha imagen. Si se usa algoritmo de reconocimiento facial (como los presentados en la sección 7) si recibe dicha región de entrada debería de reconocer a la persona. En la tercera imagen se muestra una falsa localización. El algoritmo determinó que dicha región corresponde a una cara cuando en realidad no lo es. En estos tres casos no se considera cuando el algoritmo no localiza la cara en una imagen pues lo que se busca es caracterizar la efectividad de localización y reconocimiento.

Sea s una secuencia de video y s_i la i -ésima imagen de dicha secuencia. Una localización y reconocimiento sucede cuando se produce un pico de correlación en la posición correcta (o lo más próximo a ella) y la agudeza del pico es mayor o igual a un umbral de reconocimiento. El porcentaje de veces que un algoritmo de seguimiento localiza y reconoce correctamente una cara en una secuencia de video está dado por:

$$TLR = \frac{|\{s_i: d_{s_i} \leq d'_L\}|}{|s|} \quad (52)$$

donde $d_{s_i} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ es la distancia existente entre la localización correcta y la localización del pico de correlación producido por el algoritmo al correlacionar el filtro de seguimiento con la imagen s_i . Esta distancia representa el error de localización. La localización correcta está dada por el punto $p(x_1, y_1)$ y la localización del pico de correlación está dado por $p(x_2, y_2)$. La distancia mínima que debe existir entre la posición correcta y la posición del pico está representado por d'_L , $|s|$ es la cantidad de imágenes que contiene la secuencia.

Cuando el pico de correlación se localiza en una posición mayor que d'_L pero menor que la distancia que hay del centro de la cara hacia una de sus esquinas y la agudeza del pico de correlación es mayor a un umbral de reconocimiento, la ventana de seguimiento encierra parte de la cara (que puede ser usado para reconocimiento de la persona) sucede una localización y reconocimiento desviado. El porcentaje de localización y reconocimiento desviado está dado por la ecuación 52. La ecuación 53 produce el porcentaje de veces que un algoritmo de seguimiento produce un pico con una agudeza mayor al umbral de reconocimiento en un lugar donde no hay una cara.

$$TLRD = \frac{|\{s_i: d_{s_i} > d'_L \text{ y } d_{s_i} \leq d_f\}|}{|s|} \quad (53)$$

$$TFLR = \frac{|\{s_i: d_{s_i} > d_f\}|}{|s|} \quad (54)$$

donde d_f es la distancia máxima que debe existir entre la posición correcta de la cara y la determinada por el algoritmo.

Cuando una salida de correlación contiene un pico en la posición correcta pero la agudeza de ésta es menor al umbral de reconocimiento, entonces sucede un falso rechazo pues el algoritmo no cuenta con la suficiente certeza de que sea el objeto correcto y puede ser un impostor. Esta estadística está caracterizada por la ecuación:

$$TFR = \frac{|\{s_i: (d_{s_i} \leq d'_L) \wedge (PSR_{s_i} < \tau_{psr})\}|}{|s|} \quad (55)$$

Como puede observar en las ecuaciones 52, 53, 54 y 55, el valor máximo que se puede producir es 1. Éste valor es multiplicado por 100 para convertirlo a porcentajes.

La Tabla 8.1 muestra el desempeño de los filtros de correlación con la secuencia de video Ever02. La estrategia propuesta por este trabajo tiene un desempeño (con un factor de no linealidad $k = 0.3$) muy similar al filtro MOSSE, siendo mejor el filtro FCnLsRA al tener una tasa de falsa localización y falso reconocimiento nulo. Se observó que ambos filtros tiene la capacidad de recuperarse después de una falla. Un cambios brusco de la cara en una imagen a otra afecta más al desempeño del filtro MOSSE que al filtro FCnLsRA.

En las Figura 8.6 se muestran el desempeño PSR de los filtros de correlación el seguimiento facial al procesar la secuencia Ever02. Ambos filtros después de cierta cantidad de imágenes el valor PSR que se produce disminuye, pero el seguimiento se mantiene. Estos se debe a que surge el efecto conocido como sobre ajuste lo que dificulta que el acoplamiento sea perfecto.

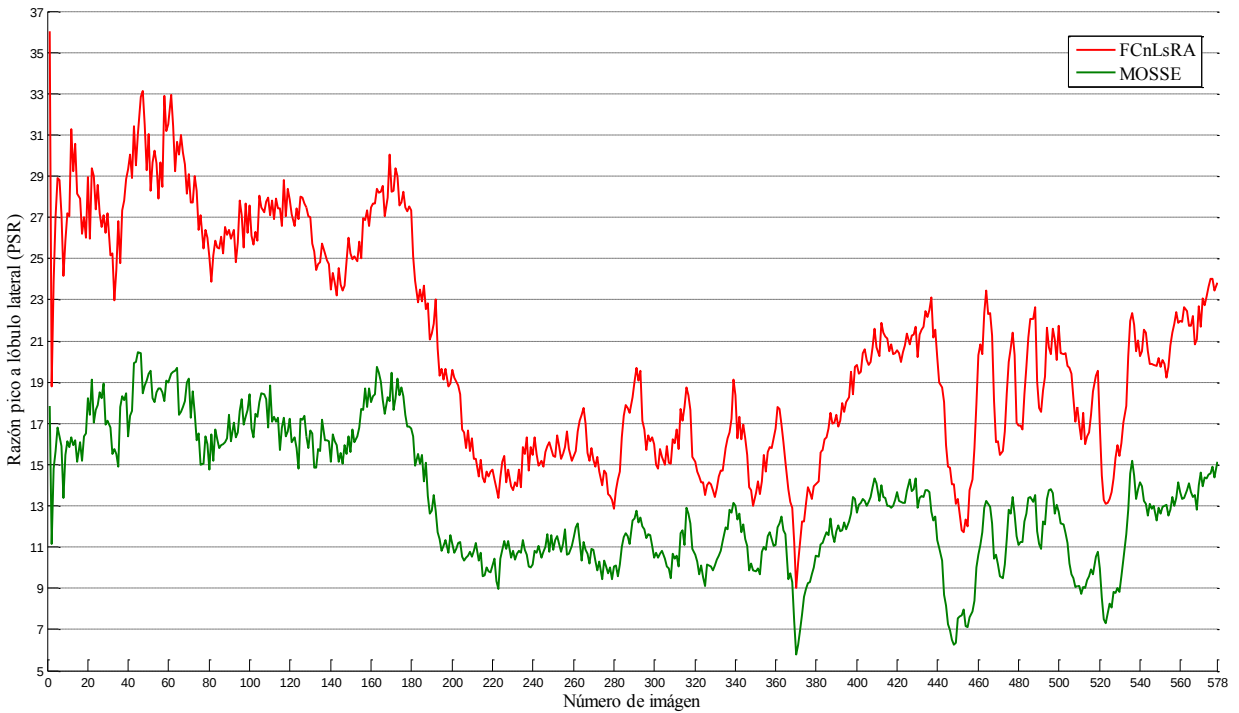


Figura 8.6. Desempeño PSR de los filtros de correlación FCnLsRA y MOSSE en la secuencia de video Ever02.

La distancia producida por los dos filtros fue menor a 11 píxeles cuando la localización era correcta. Cuando el centro de la ventana de seguimiento se desviaba mucho del centro de la cara pero se la ventana de seguimiento contiene parte de la cara que posibilita un reconocimiento de la persona, la distancia está entre 11 y 35. El umbral de reconocimiento fue $psr = 5$. En la Figura 8.7 se muestra los errores de localización de los filtros. El filtro MOSSE únicamente en una imagen de la secuencia Ever02 determinó una localización incorrecta de la cara en seguimiento, mientras que el filtro FCnLsRA ninguno. El filtro FCnLsRA la ventana de seguimiento tuvo un desvío mayor a la distancia de tolerancia (11 píxeles) el 0.35% de las veces.

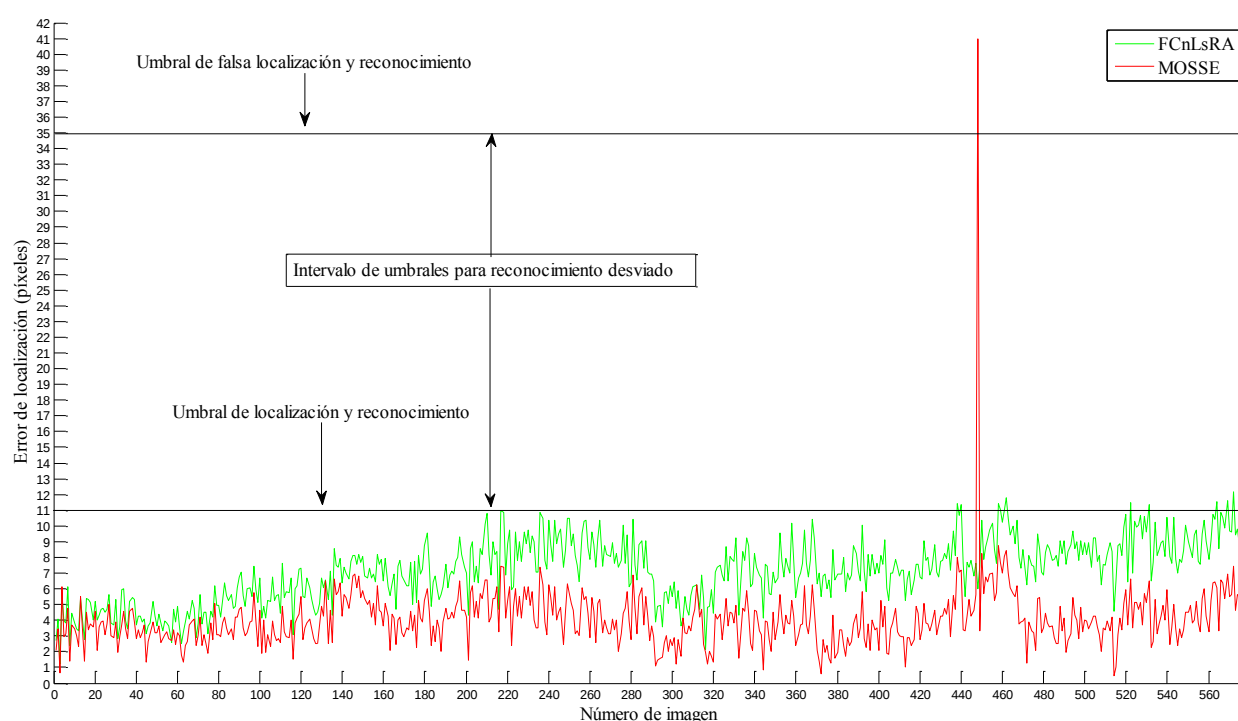


Figura 8.7. Distancia entre la localización exacta y el pico de correlación producido por los filtros FCnLsRA y MOSSE.



Figura 8.8. Muestras de las secuencias davidin300 (primera fila), dudek (segunda fila) y trellis70 (tercera fila).

Las secuencias davidin300, dudek y trellis70 presentan imágenes con variaciones muy fuerte. Ambos filtros tuvieron un pobre desempeño (véase tabla 8.1), en algunos caso no fue posible localizar ni reconocer una cara en la imagen porque presentaban poses laterales, iluminación intensa y algunas imágenes no tenían las características faciales distinguibles. En la figura 8.8 se muestran algunos ejemplos de estas distorsiones.

Tabla 8.1. Desempeño de los filtros FCnLsRA y MOSSE en el seguimiento facial.

	Ever02		Davidin300		dudek		trellis70	
	FCnLsRA	MOSSE	FCnLsRA	MOSSE	FCnLsRA	MOSSE	FCnLsRA	MOSSE
TLR	99,65	99,82	22,68	12,12	53,6	53,6	49,56	79,43
TLRD	0,346	0	0,21	0	1,23	2	5,41	7,79
TFLR	0	0,17	77,52	24,02	7,42	7,42	3,24	12,33
TFR	0	0	0	1,29	11,54	10	5,19	0

En esta sección se ha presentado una estrategia de seguimiento facial basada en un filtro de correlación compuesto no lineal, sin restricciones y adaptativo, cuya evaluación de desempeño evidenció resultados superiores a un filtro ya reportado en la literatura en escenarios donde existe pocas variaciones en sombras, iluminación intensa, pose y las caras en las imágenes tienen las características distinguibles o que pueden ser mejoradas mediante alguna operación de pre-procesamiento.

Capítulo 9

9. Discusión de resultados y conclusiones

El reconocimiento facial automático es una herramienta importante en la obtención de la identidad de una persona de forma no intrusiva, natural y fácil de usar. Además cuenta con una amplia gama de aplicaciones tales como en seguridad, industria, entretenimiento, entre otros. La cara sufre constantes cambios de apariencia debido la edad, diferentes fuentes de luz, detalles faciales, oclusión parcial, etcétera. Los actuales algoritmos de reconocimiento facial solo funcionan restringiendo el entorno y las personas procesan una sola imagen. Así, es necesario el desarrollo de algoritmos que identifiquen y sigan a una persona en una secuencia de video a través de su cara.

En este contexto, los resultados obtenidos en esta tesis se pueden dividir en cinco partes:

1. El análisis del desempeño y capacidad de discriminación de algunos filtros de correlación en la verificación facial con imágenes de prueba distorsionadas.
2. El análisis de la influencia de algunas operaciones de pre-procesamiento en las imágenes faciales para la mejora (en algunos casos) del desempeño de los filtros de correlación en el reconocimiento facial.
3. El análisis del desempeño de los filtros de correlación en la tarea de identificación facial.
4. Desarrollo y análisis de un algoritmo de reconocimiento facial basado en correlación usando un filtro compuesto no lineal sin restricciones.
5. Desarrollo y análisis de una estrategia de seguimiento facial basado en correlación usando un filtro compuesto no lineal, sin restricciones y adaptativo.

Las primeras tres partes evalúan el desempeño de los filtros de correlación en las tareas de verificación e identificación facial. Los resultados de esta evaluación mostraron las características de dichos filtros y de las cuales se consideraron algunos para el desarrollo de los algoritmos de reconocimiento y seguimiento facial propuestos en este trabajo, capaces de funcionar en una secuencia de video en tiempo real. A continuación se presentan las principales contribuciones hechas en cada parte.

1. El análisis de los filtros de correlación en la verificación facial arrojó los siguientes resultados:
 - a. El mejor desempeño en la tarea de verificación facial lo obtiene el filtro K-Law con imágenes faciales afectadas por ruido. Este filtro con un factor de no linealidad $k = 3$ muestra una buena capacidad de discriminación al distinguir una clase de caras de otras diferentes incluso en presencia de ruido.
 - b. Para imágenes faciales con sombras parciales o totales el filtro UOTSDF obtiene el mejor desempeño.
 - c. El desempeño de los filtros de correlación MACE, UMACE y UOTSDF es pobre cuando las imágenes faciales presentan ruido.
 - d. Al utilizar los filtros UMACE y UOTSDF existe una alta probabilidad de confundir elementos de dos diferentes clases.
2. El análisis de la influencia de las operaciones de pre-procesamiento: a) transformación logarítmica, b) normalización, c) ventana de coseno, d) transformación logarítmica y normalización, e) transformación logarítmica y ventana de coseno, f) normalización y ventana de coseno y g) transformación logarítmica, normalización y ventana de cosenos, en las imágenes faciales:
 - a. La transformación logarítmica mejora el contraste de una imagen facial parcial o totalmente sombreada, sin pérdida de información importante como lo son los bordes o las características faciales. Esto debido a que realiza un escalamiento de los datos, incrementando el valor de los píxeles cercanos a cero y disminuyendo el valor de los píxeles altos.
 - b. Para imágenes de prueba con sombras, el filtro K-Law ofrece un buen desempeño con la mayoría de las operaciones de pre-procesamiento excepto con la operación de transformación logarítmica y ventana de coseno.
 - c. Para imágenes de prueba con ruido, la aplicación de la transformación logarítmica y la ventana de cosenos mejora el desempeño de los filtros MACE, UMACE y UOTSDF.
 - d. El filtro UOTSDF obtiene el desempeño óptimo, en términos de tasa de verificación, con cualquiera de las operaciones de pre-procesamiento.

- e. El filtro K-Law obtiene el mejor desempeño al procesar imágenes afectadas por ruido, cuando éstas presentan 5, 10, 15 y 20 píxeles de variancia, con 100%, 100%, 99.75% y 98.25%, respectivamente.
3. Análisis de los filtros de correlación en la identificación facial con imágenes con variaciones en iluminación:
 - a. El filtro de correlación K-Law tiene un buen desempeño alcanzado 99.09% en la tasa de identificación y detección al ser entrenado con muestras faciales representativas de variaciones en iluminación.
 - b. El filtro de correlación K-Law es capaz de reconocer con una buena precisión aquellas imágenes faciales que no están en el conjunto de entrenamiento pero que pertenecen a la misma clase. También rechaza con una probabilidad alta aquellas imágenes faciales impostoras.
 4. Las conclusiones con el algoritmo de reconocimiento facial desarrollado:
 - a. Eliminando las restricciones en el origen de la salida de correlación y aplicando el filtrado no lineal de una imagen facial permite aprovechar mejor la fase de la transformada de Fourier de la imagen.
 - b. Entrenar un filtro compuesto no lineal sin restricciones con regiones faciales incrementa la probabilidad de reconocer imágenes parciales.
 - c. Con el filtro compuesto no lineal sin restricciones se alcanza una probabilidad alta de éxito en el reconocimiento facial, mientras se mantiene una baja tasa de falso reconocimiento. Esto se traduce en una muy buena capacidad de discriminación.
 5. Con la estrategia de seguimiento facial:
 - a. Inicializar el filtro de correlación con versiones distorsionadas de la imagen facial inicial incrementa la probabilidad de reconocer/localizar correctamente la cara bajo observación en las siguientes imágenes de una secuencia de video.
 - b. El enfoque adaptativo incremental del filtro FCnLsRA permite adaptarse rápidamente a los cambios que sufre una cara de una imagen a otra.
 - c. La simplicidad de la actualización incremental del filtro FCnLsRA permite actualizar el filtro sin incrementar demasiado el tamaño del filtro. Esto conlleva a mantener constante la velocidad del filtro en el seguimiento.

- d. El filtro FCnLsRA tiene un desempeño similar a un filtro propuesto en la literatura para fines de seguimiento.

El algoritmo de reconocimiento facial basado en filtros de correlación tiene como principal limitación la invariancia a rotación y pose, distorsiones que no fueron considerados para su diseño. Estas limitaciones abren una oportunidad de investigación para añadir invariancia a estas distorsiones.

La estrategia de seguimiento facial propuesta en este trabajo y que usa nuestro filtro FCnLsRA, tiene como principal limitante el problema de altos grados de escala (es tolerante a pequeños grados de escalamiento), altos grados de rotación (es tolerante a pequeños grados de rotación) e iluminación intensa. Estas limitantes también abren una oportunidad de investigación para aumentar invariancia a estas distorsiones. Debido a que el filtro FCnLsRA se actualiza sin incrementar mucho su tamaño, abre la posibilidad de implementarlo en dispositivos con pocos recursos computacionales, tales como teléfonos inteligentes.

En ambos algoritmos desarrollados, queda como trabajo futuro determinar formalmente la cantidad y tamaño de las imágenes faciales para la construcción del conjunto de entrenamiento. Que los algoritmos sean capaces de localizar y reconocer una cara que presenta alguna pose es otra área de trabajo que queda a futuro.

Este trabajo de investigación es un pequeño paso hacia el reconocimiento facial en entornos y personas no controlados, problema que aún no ha sido resueltos en su totalidad debido a la complejidad del área. Además, se considera que se ha dado un paso hacia la re-autenticación de personas, por ejemplo para identificar y seguir a una persona en una secuencia de imágenes provenientes de diferentes sensores de visión ubicadas de diferentes posiciones geográficas.

Bibliografía

- Adam, A., Rivlin, E., y Shimshoni, I. (2006). Robust fragments-based tracking using the integral histogram. CVPR.
- Babenko, B., Yang, M.-H., y Belongie, S. (2009). Visual tracking with online multiple instance learning. CVPR.
- Belhumeur, P. N., Hespanha, J. P., y Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis And Machine Intelligence, 711-720.
- Bolme, D. S., Beveridge, R. J., Draper, B. A., y Lui, Y. M. (2010). Visual object tracking using adaptive correlation Filters. IEEE Conference on Computer Vision and Pattern Recognition, 2544-2550.
- Bolme, D. S., Draper, B. A., y Beveridge, J. R. (2009). Average of synthetic exact filters. IEEE Conference on Computer Vision and Pattern Recognition, 2105-2112.
- Bonchelet, C. (2005). Handbook of image and video processing, (págs. 397-409). United State of America: Elsevier Academic Press.
- Bonsor, K., y Johnson, R. (2008). How facial recognition systems work. Recuperado el 10 de diciembre de 2009, de: www.HowStuffWorks.com
- Campos, J., Turon, F., Yaroslavsky, L. P., y Yzuel, M. J. (1991). Some filters for reliable recognition and localization of objects by optical correlators: a comparison. International Journal of Optical Computing, 341-365.
- Casasent, D., y Weng Thong, C. (1986). Correlation synthetic discriminant functions. Applied Optics, 2343-2350.
- Chellappa, R., y Zhou, S. K. (2005). Face tracking and recognition from video. En S. Z. Li, & A. K. Jain, Handbook of face recognition (págs. 169-192). New York: Springer.
- Choi, Y. J., Ro, M. Y., y Plataniotis, K. N. (2009). Color face recognition for degraded face images. IEEE Transactions on Systems, Man, and Cybernetics, 1217-1230.
- De Dios, J. J. (2007). Skin color and feature-based segmentation for face localization. Optical Engineering, (037007-1)-(037007-6).

- Dornaika, F., y Raducanu, B. (2009). Three-dimensional face pose detection and tracking using monocular videos: tool and application. *IEEE Transactions on Systems, Man, and Cybernetics*, 935-944.
- Frigo, M., y Johnson, S. G. (2010). FFTW. Recuperado el 02 de January de 2011, de <http://www.fftw.org/>
- González Fraga, J. Á. (2004). Reonocimiento de objetos parcialmente ocultos utilizando filtros correlacionadores invariantes en posición. Ensenada, Baja California: Centro de Investigación Científica y de Educación Superior de Ensenada.
- Gorodnichy, D. (2005). Video-based framework for face recognition in video. *Proceedings*, 1-10.
- IDNoticias. (2010). Tecnología de identidad y seguridad, tarjetas inteligentes e identificación. Recuperado el 26 de Mayo de 2010, de: <http://www.idnoticias.com/2010/01/26/sistema-de-reconocimiento-facial-para-sistemas-de-control-de-acceso-a-eventos>
- Javidi, B., Wang, W., y Zhang, G. (1997). Composite Fourier-plane nonlinear filter for distortion-invariant pattern recognition. *Society of Photo-Optical Instrumentation Engineers*, 2690–2696.
- Jun, B., Lee, J. y Kim, D. (2010). A novel illumination-robust face recognition using statistical and non-statistical method. *Pattern Recognition Letters*, 329-336.
- Kiat Ng, C. (2005). PDA face recognition system using advanced correlation filters. Carnegie Mellon University.
- Li, S. Z., & Jain, A. K. (2004). *Handbook of face recognition*. Springer.
- Lu, J., Plataniotis, K. N., y Vanetsanopoulos, A. N. (2003). Regularized discriminant analysis for the small samples size problem in face recognition. *Pattern Recognition Letters*, 3079-3087.
- Mahalanobis, A., Vijaya Kumar, B. V., y Casasent, D. (1987). Minimum average correlation energy filters. *Optical Society of America*, 3633-3640.
- Mahalanobis, A., Vijaya Kumar, B. V., Song, S., Sims, S. R., y Epperson, J. F. (1993). Unconstrained correlation filters. *Optical Society of America*, 3751-3759.
- Mahalanobis, A., Vijaya, K., Song, S., Sims, S. R., y Epperson, J. F. (1994). Unconstrained correlation filters. *Applied Optics*, 3751-3759.
- Nair, P., y Cavallaro, A. (2009). 3-D Face detection, landmark localization, and registration using a point distribution model. *IEEE Transactions on Multimedia*, 611-623.

- NSTC Subcommittee on Biometrics. (2007). Biometrics. Recuperado el 6 de Marzo de 2010, de Face Verification using Correlation Filters: www.biometrics.gov
- NSTC: Subcommittee on Biometrics and Identity Management. (2008). Biometrics in Government Post 9/11: Advancing science, enhancing operations. NSTC.
- P. H., J. (2011). The national biometrics challenge. Washington, D. C.: Executive Office of the President.
- Phillips, P. J., Grother, P., y Micheals, R. (2005). Handbook of face recognition. Michigan: Springer.
- Phillips, P. J., Moon, H., Rizvi, S. A., y Rauss, P. J. (1999). The FERET evaluation methodology for face-recognition algorithms. United State of America: Face Recognition Technology.
- Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L. (2007). FRVT 2006 and ICE 2006 Large-Scale Results. NSTC.
- Rodriguez, F. M. (2009). Reconocimiento biométrico a través de la cara. Occams Razor, 54-60.
- Ross, D., Lim, J., y Yang, M.-H. (2004). Adaptative probabilistic visual tracking whit incremental subspace update. Proc. Eighth European Conference in Computer Vision.
- Ross, D., Lim, J., Lin, R., y Yang, M. (2008). Incremental learning for robust visual tracking. IJCV, 125-141.
- Samaria , F., y Harter , A. (1994). Parameterisation of a stochastic model for human face identification. 2nd IEEE Workshop on Applications of Computer Vision.
- Santiago Ramírez, E. (2010). Reconocimiento de patrones por correlación aplicado a un sistema de visión artificial para la supervisión del armado de un tangrama. Ensenada, México: Facultad de Ciencias, Universidad Autónoma de Baja California.
- Savvides , M., Bhagavatula, R., Li, Y.-h., y Abiantun, R. (2007). Frequency domain face recognition. Department of Electrical and Computer Engineering, Carnegie Mellon University, 495-527.
- Savvides, M. (2004). Reduced complexity face recognition using advanced correlation filters and fourier subspace methods for biometric applications. Carnegie Mellon University.
- Savvides, M., y Vijaya Kumar, B. V. (2003). Illumination normalization using logarithm transforms for face authentication. Dept. of Electrical and Computer Engineering, Carnegie Mellon University, 549-556.

- Savvides, M., Venkataramani, K., y Vijaya Kumar, B. V. (2003). Incremental updating of advanced correlation filters for biometric authentication systems. ICME, 229-232.
- Savvides, M., Vijaya Kumar, B. V., y Khosla, P. (2002). Face verification using correlation filters. Carnegie Melon University, 1-6.
- Sirovich, L., y Kirby, M. (1987). A low dimensional procedure for the characterization of human faces. J. Optical Soc. Am., 519-524.
- Turk, M. A., y Pentland, A. P. (1991). Face recognition using eigenfaces. Proc. IEEE, 586-591.
- Turk, M. A., y Pentland, A. P. (1991). Eigenfaces for Recognition. Journal of Cognitive Neuroscience. 71-86.
- VanderLugt, A. B. (1964). Signal detection by complex spatial filtering. IEEE Trans. on Inf. Theory, 139-145.
- Viajaya Kumar, B., Mahalanobis, H., y Juday, R. (2005). Correlation pattern recognition. New York, U. S. A.: Cambridge University Press.
- Vicente, C., García, M., y García, A. (202). Seguimiento de caras humanas mediante búsqueda logarítmica en rejilla. Cartagena: Univ. Politécnica de Cartagena.
- Vijaya Kumar, B. V. (1992). Tutorial survey of composite filter designs for optical correlators. Applied Optics, 4773-4800.
- Vijaya Kumar, B. V., Savvides, M., Xie, C., Venkataramani, K., Thornton, J., y Mahalanobis, A. (2004). Biometric verification with correlation filters. Applied Optics, 391-402.
- Vijaya, K. (1992). Tutorial survey of composite filter design for optical correlators. Applied Optics, 4773-4800.
- Viola, P., y Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Computer Vision And Pattern Recognition, 1-9.
- Weyrauch, B., Huang, J., Heisele, B., y Blanz, V. (2004). Component-based face recognition with 3D morphable models. First IEEE Workshop on Face Processing in Video.
- Wikimedia Foundation. (2011). Image noise. Recuperado el 07 de Octubre de 2011, de Wikipedia: The Free Enciclopedia: http://en.wikipedia.org/wiki/Image_noise
- Wiskott, L., Fellous, J. M., Kruger, N., y Malsburg, C. v. (1999). Face recognition by elastic bunch graph matching. Intelligent Biometrics Techniques in Fingerprint and Face Recognition, 355-396.

- Yang, M.-H., Kriegman, D., y Ahuja, N. (2002). Detecting faces in images: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34-58.
- Zhang, X., Hu, W., Maybank, S., & Li, X. (2007). Graph based discriminative tracking. ICCV.
- Zhao, W., Chellapa, R., Phillips, P., y Rosenfeld, A. (2003). Face recognition: A literature survey. ACM Computing Surveys, 399-458.

Apéndice A. Sistema de identificación y seguimiento facial

En algunos escenarios donde interactúan personas es necesario identificarlas a través de sus caras y seguirlas en dicho escenario. Esto posibilita inferir la actividad realizada por cada persona observada y etiquetar dicha actividad con la identidad de la persona que la realiza.

Como caso de uso se implementó un sistema que dado una secuencia de video de entrada y una región facial seleccionada de forma manual obtiene la identidad de la persona y localiza en cada imagen de la secuencia donde aparece la cara de la persona. El esquema general del sistema de muestra en la Figura A.1, la ejecución del sistema genera muchos datos que pueden ser usados para diferentes propósitos.

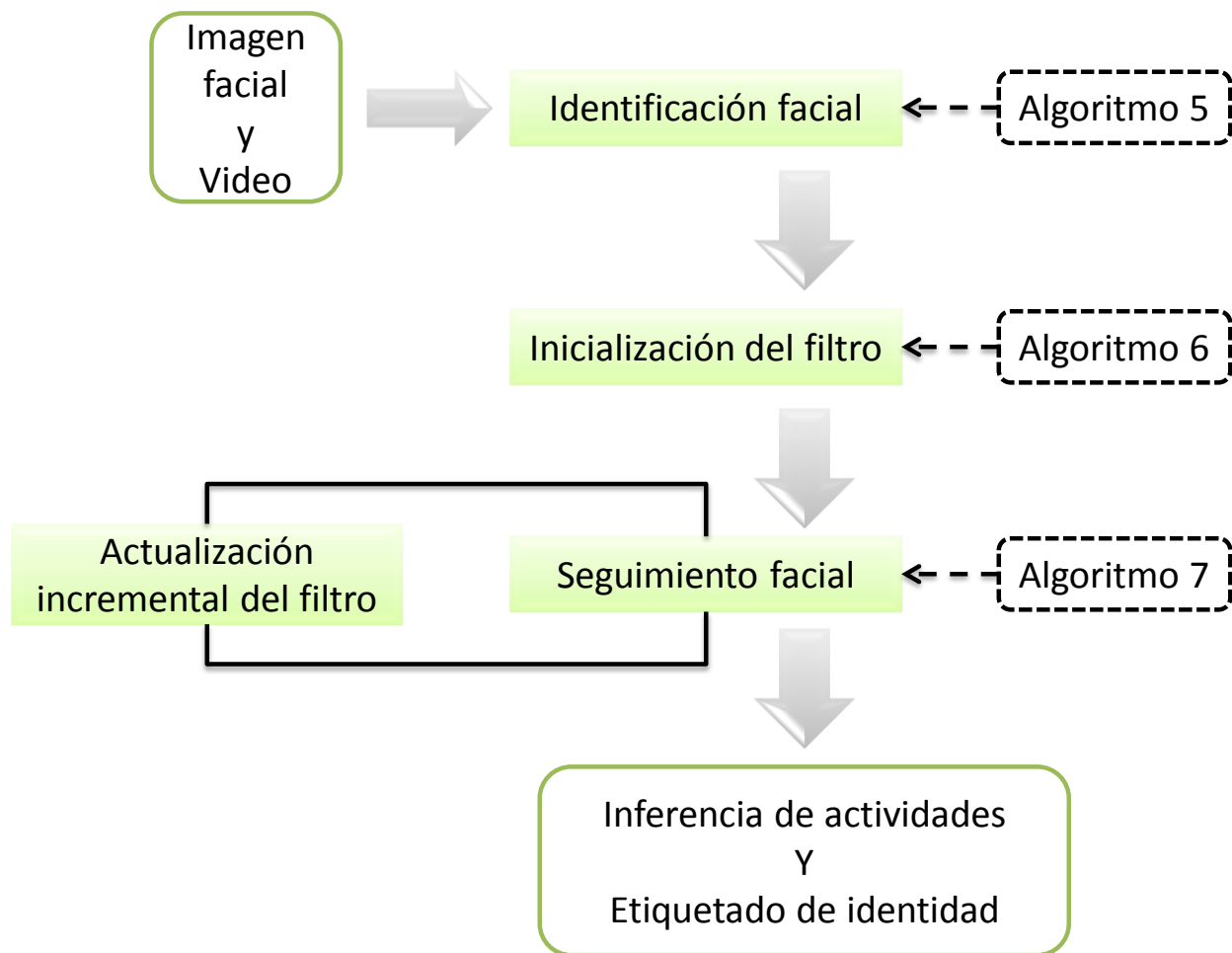


Figura A.1. Esquema general del sistema de identificación y seguimiento facial.

Se generó una galería con 39 personas $G = \{g_1, g_2, \dots, g_{39}\}$, de las cuales 38 corresponden a las 38 personas de la base de datos YALE B y la otra persona corresponde a quien aparece en la secuencia de video Ever02 presentado en la sección 8. Para cada persona se construyó un filtro de correlación FCnLsRA usando la estrategia 2 de selección de imágenes de entrenamiento presentado en la sección 6.2.

El procedimiento para usar el sistema es el siguiente. Primero, se selecciona el área donde aparece la cara en la primera imagen de la secuencia (véase Figura A.2). Segundo, ejecutar el sistema. El sistema identifica automáticamente la imagen facial de entrada tal y como se muestra en la Figura A.3.



Figura A.2. Selección del área en la imagen 1 de la secuencia de video la cara de la persona a observar.



Figura A.3. Persona en la secuencia de video identificada e iniciado a seguirlo en cada imagen de la secuencia.

En la Figura A.4 se muestra el resultado PSR del proceso de identificación facial. Como se puede observar el valor PSR que produce el sujeto auténtico, que corresponde a la plantilla biométrica facial 39 tiene el valor PSR más alto y mayor que el umbral de reconocimiento. El resto de las plantillas biométricas produjeron valores PSR por debajo del umbral, lo que muestra una buena capacidad de discriminación del filtro FCnLsRA.

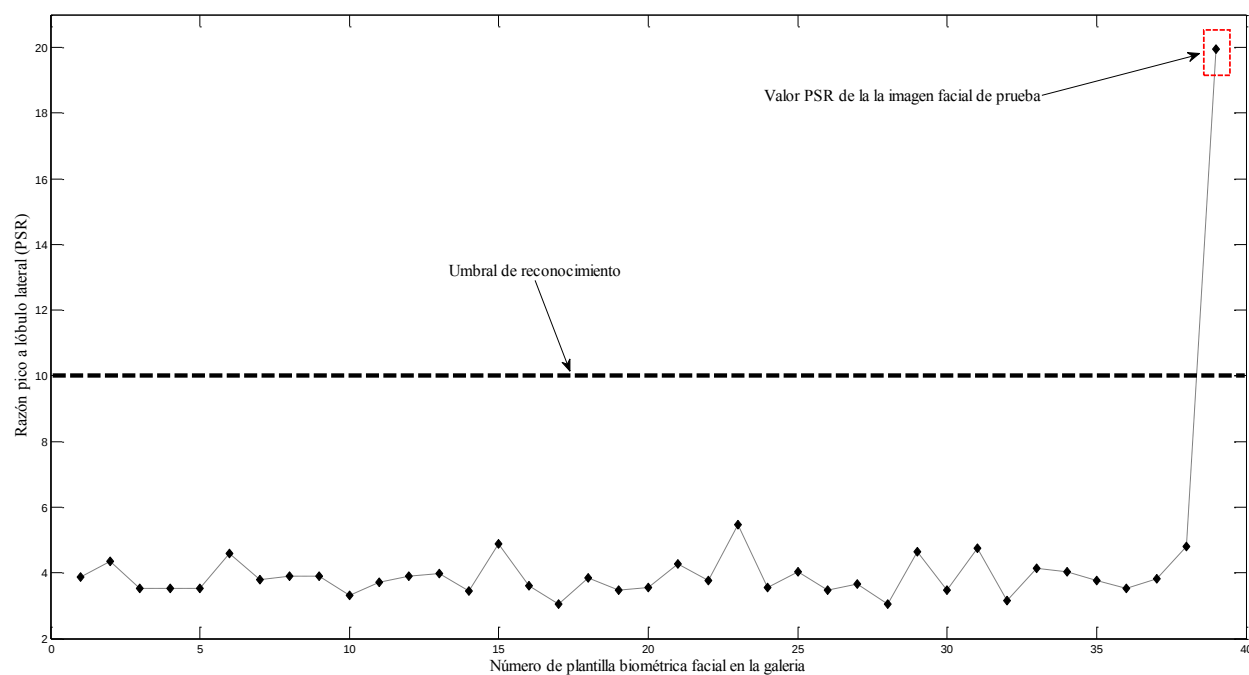


Figura A.4. Resultados PSR de la correlación de la imagen facial de prueba con cada plantilla biométrica facial en galería.

Este sistema fue desarrollado en el lenguaje de programación Matlab[®] (Matwork, 2009) como caso de uso y evidencia de que los algoritmos presentados en este trabajo de tesis funcionan en un sistema real. Al ser Matlab un lenguaje técnico que no fue hecho para desarrollar sistemas en tiempo real es preferible considerar otros lenguajes adecuados para sistemas en tiempo real para optimizar la ejecución de los algoritmos. Este documento de tesis viene acompañado de un pequeño video que muestra el funcionamiento del sistema.