RNPS No. 2142
ISSN 2072-6287
Versión Digital

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

REPORTE TÉCNICO
Reconocimiento de Patrones

SERIE AZUL

Reconocimiento de rostros en video

Yoanna Martínez Díaz,
Heydi Méndez Vázquez, y
Edel García Reyes

RT_045 **enero 2012**

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

RNPS No. 2142
ISSN 2072-6287
Versión Digital

SERIE AZUL

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

REPORTE TÉCNICO
Reconocimiento de Patrones

Reconocimiento de rostros en video

Yoanna Martínez Díaz,
Heydi Méndez Vázquez, y
Edel García Reyes

RT_045 **enero 2012**

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

Tabla de contenido

Lista de figuras

Lista de tablas

Reconocimiento de rostros en video

Yoanna Martínez Díaz[1], Heydi Méndez Vázquez[1], y Edel García Reyes[2]

[1] Dpto. Ingeniería en Sistemas, Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV),
La Habana, Cuba
{ymartinez, hmendez}@cenatav.co.cu
[2] Dpto. Reconocimiento de Patrones, Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV),
La Habana, Cuba
egarcia@cenatav.co.cu

RT_045, Serie Azul, CENATAV
Aceptado: 31 de octubre de 2011

Resumen. El reconocimiento automático de rostros se ha convertido en una de las áreas más investigadas en la biometría. En entornos no controlados esta tarea sigue siendo un reto para la mayoría de las aplicaciones prácticas. Recientemente, el centro de atención de la investigación se ha desplazado hacia los enfoques basados en video, con el objetivo de aprovechar la abundancia de información en estos para un mayor rendimiento en tiempo real. Además, problemas como las variaciones de iluminación, la baja resolución y la oclusión, que tanto afectan el rendimiento de los enfoques tradicionales basados en imágenes fijas, pueden ser tratados con una mayor eficacia. La detección y el seguimiento automático de rostros son dos tareas muy importantes durante el proceso de reconocimiento de rostros. Por esta razón en este reporte se revisan las principales investigaciones desarrolladas para el reconocimiento de rostros en video, haciéndose un mayor énfasis en los métodos de detección y seguimiento de rostros existentes. Por último, se realiza un análisis de las posibles brechas en las que se puedan basar investigaciones futuras.

Palabras clave: biometría, reconocimiento de rostros en video, detección de rostros, seguimiento de rostros.

Abstract. Automatic face recognition has become one of the most active research areas in biometrics. For uncontrolled environments, this task remains a challenge in most practical applications. Recently, the focus of research has shifted towards video-based approaches, in order to harness the wealth of data for improved performance in real time. In addition, problems such as lighting variations, low resolution and occlusion, which affect the performance of conventional approaches based on still images, can be treated with greater effectiveness. Automatic face detection and tracking are very important tasks in the face recognition process. Therefore, in this report we make a review of the main investigations carried out for face recognition in video, emphasizing on available methods for face detection and tracking. Finally, we discuss the open problems on which future research can be based.

Keywords: biometric, video face recognition, face detection, face tracking.

1 Introducción

En la actualidad, debido a los avances tecnológicos y la creciente demanda de sistemas de seguridad y video vigilancia, las aplicaciones biométricas han tomado un gran auge [11]. La biometría es la disciplina que permite identificar a los individuos basándose en sus características físicas y/o en su comportamiento. Entre las principales características físicas se pueden mencionar las huellas dactilares, el iris, la retina, la

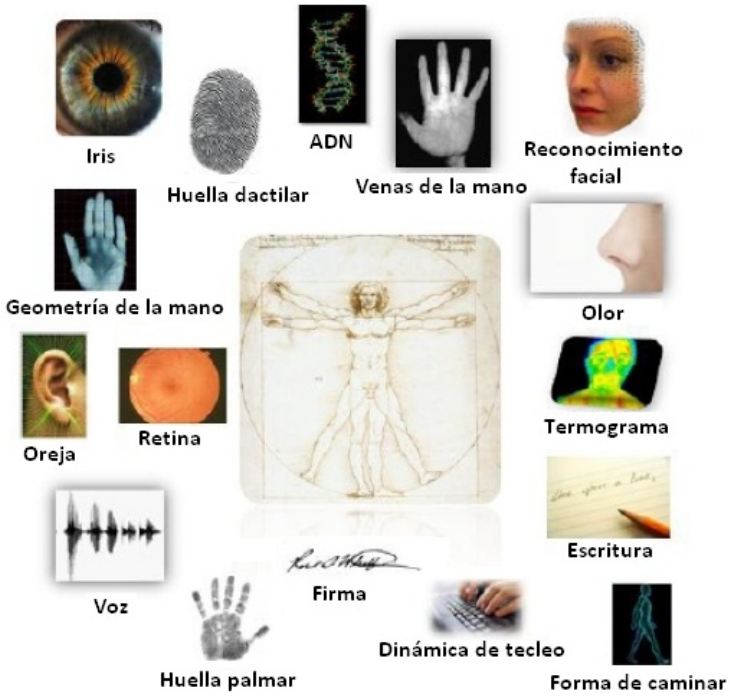


Fig. 1. Ejemplos de rasgos biométricos.

geometría de la palma de la mano y los rasgos faciales, entre otras. Por otra parte, dentro de las características del comportamiento más comunes se encuentran la firma, la forma de caminar, la voz y la pulsación en teclado. En la Figura 1 se muestran algunas de las características mencionadas anteriormente.

Los sistemas biométricos son usados en dos tareas fundamentales: la verificación y la identificación. La verificación es un proceso de reconocimiento de uno contra uno (1 : 1), basado en la comparación de dos muestras biométricas. Es una forma explícita de autenticación, tal como se conoce generalmente en la informática, donde se trata de comprobar si una persona es quien dice ser o no. La identificación, por su parte, responde al reconocimiento de una muestra biométrica en un conjunto de n muestras previamente almacenadas ($1 : n$). Esta tarea tiene como propósito identificar a una persona entre un grupo de individuos. En la Tabla 1 se muestran algunos de los principales sectores donde se utilizan sistemas biométricos [12].

Tabla 1. Aplicaciones de la biometría.

Forénsicas	**Gubernamentales**	**Comerciales**
Identificación de Cadáveres	Documentos de Identificación	Cajeros Automáticos
Investigación Criminal	Registros Electorales	Sistemas de Control de Accesos
Determinación de Paternidad	Pagos de Asistencia Social	Autenticación de usuarios en PC
Niños Extraviados	Chequeo de Inmigración en fronteras	Sistemas Bancarios

Los sistemas biométricos incluyen un dispositivo de captura y un *software* que interpreta la muestra física y la transforma en una secuencia numérica. Un sistema biométrico ideal debe ser fácil de usar, tener una elevada precisión en las mediciones, una alta velocidad de respuesta, un mínimo contacto con

el usuario y una alta aceptación; todo al menor costo posible. Cada rasgo biométrico, ya sea físico o de comportamiento, es diferente y existen varios factores que determinan su idoneidad para ser empleado en un sistema biométrico, como son [13]:

- **Universalidad:** La característica biométrica debe estar presente y bien definida en la mayor cantidad de individuos posible.
- **Capacidad distintiva:** La característica biométrica a utilizar debe ser lo suficientemente diferenciable entre los individuos sobre los que se aplica, brindando una alta precisión en la identificación.
- **Estabilidad de los datos:** La característica biométrica debe ser lo suficientemente estable para no cambiar significativamente con el tiempo o en distintos medios.
- **Facilidad de uso:** La señal biométrica debe ser posible de adquirir y digitalizar utilizando el *hardware* adecuado sin causar inconveniencias al individuo.
- **Nivel de aceptación:** Indica el grado de tolerancia de la sociedad.
- **Seguridad ante ataques:** Se refiere a la dificultad para burlar el sistema.

En la Tabla 2 se muestra una comparación del comportamiento de algunos de estos factores en las diferentes técnicas biométricas [12].

Tabla 2. Comparación de varias técnicas biométricas.

	Huella Dactilar	**Iris**	**Voz**	**Geometría de la mano**	**Rostro**
Universalidad	Media	Alta	Media	Media	Alta
Seguridad ante ataques	Media	Muy Alta	Media	Alta	Media
Nivel de aceptación	Media	Baja	Muy alta	Media	Muy alta
Estabilidad de los datos	Alta	Alta	Media	Alta	Media
Facilidad de uso	Alto	Bajo	Alto	Medio	Alto
Capacidad distintiva	Muy Alto	Muy Alto	Media	Alto	Alto

La mayoría de las técnicas biométricas requieren una acción voluntaria y de cierto modo invasiva por parte del usuario. Por ejemplo, la persona debe colocar su mano sobre un dispositivo para capturar la huella dactilar y la geometría de la mano o tiene que estar en una posición fija y bien cerca, frente a una cámara de alta resolución, para la captura del iris o la retina. La adquisición de datos en general es otra tarea llena de complejidades, por ejemplo, técnicas que se basan en las manos y los dedos pueden ser inútiles si el tejido de la epidermis está dañado de alguna manera (es decir, golpeado o rajado).

Sin embargo, el uso de las características faciales es una técnica no invasiva, que se puede realizar pasivamente sin ningún tipo de acción explícita o participación por parte del usuario. Esto permite que pueda ser utilizado en operaciones encubiertas, ya que las imágenes son capturadas por cámaras de las que el individuo no siempre tiene conocimiento. Además, el rostro es la característica más natural a los seres humanos. Se puede decir que es fácil de usar ya que el reconocimiento se realiza a partir de imágenes y no implica necesariamente el uso de dispositivos (cámaras) de altos costos. Por estas razones, como se observa en la Tabla 2, las aplicaciones biométricas basadas en el rostro tienen una mayor aceptación [12].

El reconocimiento a partir de las imágenes de rostros es un problema complejo, pero de gran interés, ya que el ámbito de aplicación es muy amplio. El rostro puede ser empleado como rasgo biométrico en la mayoría de las aplicaciones que se reflejan en la Tabla 1. En el proceso de reconocimiento automático de rostros se espera poder llevar a cabo la identificación o la verificación de los rostros presentes en una imagen o un video. Para esto es necesario ejecutar una serie de pasos intermedios. El primer paso es la

detección del rostro dentro de la imagen o cuadro del video. Una vez que se tiene la localización del rostro, esta región es alineada o normalizada con el objetivo de estandarizar las diferentes imágenes de rostros. Luego, se extraen los rasgos o características que representan el rostro y que serán usadas en el último paso donde se realiza la comparación de manera automática de la información extraída contra uno o más modelos de rostros, dando lugar a la clasificación del rostro que se está analizando [14].

En el reconocimiento automático de rostros existen cuatro escenarios diferentes de acuerdo a los diferentes tipos de entrada y al origen de los datos que se utilicen como base para la comparación (galería). Como se aprecia en la Tabla 3, el reconocimiento se puede realizar utilizando imágenes fijas o una secuencia de estas, que formen un video. El reconocimiento automático de rostros en video es uno de los escenarios que ha ganado mayor atención en los últimos años, debido principalmente al desarrollo de una gran cantidad de sistemas de video-vigilancia. El reconocimiento de rostros en video se originó a partir de las técnicas basadas en imágenes fijas [15]. No obstante, en una secuencia de video hay más información disponible que en una imagen fija. Además de la información fisiológica del rostro presente en las imágenes, se suma la información temporal u otras señales que se pueden obtener como la voz, el movimiento o los gestos, lo que hace posible que el reconocimiento se logre de manera más sólida y estable, siempre y cuando toda la información se aproveche e integre correctamente.

Tabla 3. Escenarios del reconocimiento de rostros.

Galería / **Entrada**	**Imagen Fija**	**Video**
Imagen Fija	Imagen-Imagen	Imagen-Video
Video	Video-Imagen	Video-Video

En este reporte, se estudian los principales métodos utilizados para el reconocimiento de rostros en video. En la sección 2 se describe de manera general cómo se lleva a cabo el proceso de reconocimiento de rostros en video y los principales enfoques desarrollados para este propósito. Al analizar los diferentes métodos se observa que para poder realizar este proceso, los pasos de detección y seguimiento del rostro son de gran importancia, ya que garantizan obtener la información adecuada para el reconocimiento en cada uno de los cuadros del video y mantener la relación entre estos. Por esta razón, en las secciones 3 y 4 se hace énfasis en los métodos de detección y seguimiento de rostros existentes, buscando los algoritmos que garanticen mayor eficiencia y eficacia a la hora de realizar la clasificación. Finalmente, en las conclusiones se resumen los principales problemas abiertos que constituyen motivo de investigación en esta temática.

2 Reconocimiento de rostros en video

El reconocimiento de rostros en video es la técnica de establecer la identidad de una o varias personas presentes en el video. Un video no es más que un conjunto de imágenes o cuadros consecutivos capturados durante un tiempo determinado. Recientemente, el reconocimiento de rostros en video ha recibido una mayor atención por parte de la comunidad científica [16]; debido principalmente a la creciente demanda de aplicaciones como video-vigilancia, seguridad, monitoreo y control de accesos, entre otras. Como se puede apreciar en la Figura 2, el proceso de reconocimiento de rostros en video parte de la entrada al sistema de la secuencia de imágenes o cuadros que forman el video que contiene el rostro de la persona que se quiere analizar. A partir de esto, se procede a realizar la detección y luego el seguimiento del

rostro durante toda la secuencia del video para finalmente extraer los rasgos y realizar la clasificación. Nótese que tanto la detección como el seguimiento resultan tareas de gran importancia en el proceso de reconocimiento, pues la calidad de sus resultados influyen directamente en la respuesta de la clasificación del rostro. Existen métodos en la literatura en los que algunos de los pasos mencionados se realizan de manera simultánea [17], [18], [19].

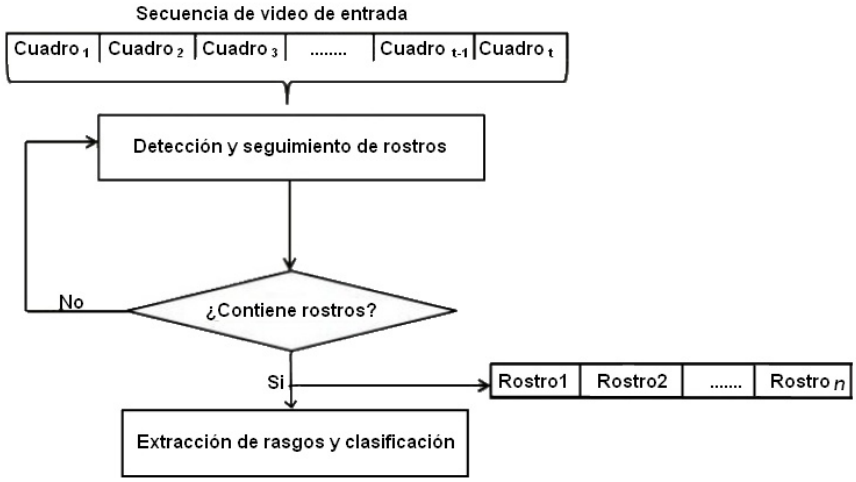


Fig. 2. Proceso de reconocimiento de rostros en video.

El proceso de reconocimiento de rostros en video, al igual que en imágenes fijas, se ve afectado por problemas como variaciones de iluminación, cambios de pose o expresiones faciales, oclusiones parciales y envejecimiento; aunque, el problema de la resolución se ha identificado como el reto principal en el reconocimiento de rostros en video [15]. Esto se debe principalmente a que en la mayoría de las aplicaciones mencionadas anteriormente la distancia entre la cámara y los sujetos es muy larga, y/o la resolución espacial de los dispositivos (cámara) es baja, ya que dispositivos de mayores prestaciones tienen elevados precios y frenarían el despliegue de aplicaciones reales. Por esta razón, en los videos las regiones de interés a menudo se encuentran empobrecidas o borrosas y las imágenes de los rostros son pequeñas, lo que afecta significativamente la eficacia del proceso de reconocimiento. En la Figura 3 se muestran ejemplos de rostros afectados por cada uno de estos problemas en cuadros de secuencias de videos.

Por lo mencionado anteriormente, el reconocimiento (verificación/identificación) de rostros de baja resolución ha ganado mucha atención en los últimos años. En la literatura se han propuesto técnicas de súper-resolución (SR) para la reconstrucción de una imagen o una secuencia de alta resolución mediante la combinación de la información de múltiples imágenes de baja resolución [20], [21]. Estas técnicas son utilizadas como un paso de preprocesamiento en la clasificación de las imágenes de rostros.

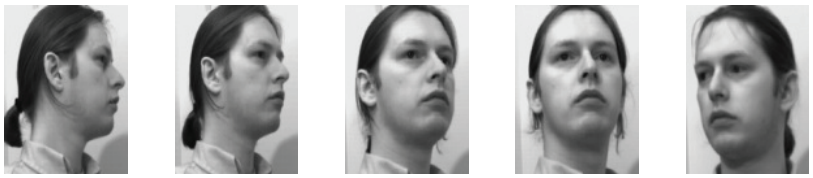
La mayoría de los métodos de SR se basan en tres pasos fundamentales: el registro de la imagen, interpolación y restauración. Estos pasos pueden ser implementados de manera independiente o simultánea, en dependencia del algoritmo SR a utilizar. De manera general, los métodos SR existentes para la reconstrucción pueden ser divididos en tres categorías fundamentales: métodos en el dominio de frecuencia, métodos en el dominio espacial y métodos basados en aprendizaje [22]. Estas técnicas de SR asumen que existen pequeñas diferencias entre las imágenes de entrada, por lo que funcionan de manera efectiva cuando varias imágenes de baja resolución contienen perspectivas ligeramente diferentes del mismo objeto. De esta manera la información total sobre el objeto excede la información de cualquier cuadro por si solo. El

(a) Envejecimiento.

(b) Diferentes expresiones faciales.

(c) Variaciones en las condiciones de iluminación.

(d) Oclusión.



(e) Variaciones en la pose.

(f) Resolución.

Fig. 3. Problemas a enfrentar en el proceso de reconocimiento de rostros.

mejor caso es cuando un objeto se mueve en el video. Situaciones no ideales se dan cuando un objeto no se mueve en lo absoluto y es idéntico en todos los cuadros o cuando se mueve o se transforma demasiado rápido, lo que hace que se vea muy diferente en distintos cuadros. En estos casos es difícil obtener la información adicional para la reconstrucción de la imagen [22].

Otra técnica empleada para el reconocimiento de rostros en imágenes o video de baja resolución es la conocida como *downsample*, la cual consiste en disminuir la resolución de todas las imágenes y luego realizar la correspondencia en el dominio de baja resolución. Sin embargo, en el empleo de esta técnica la pérdida de información es inevitable.

A pesar de la abundante bibliografía existente, factores como el manejo de caras coincidentes con oclusiones, la robustez, la eficiencia en el cálculo y uso de la memoria, la consideración del color y la selección automática de los parámetros en los métodos de súper-resolución aún requieren de un mayor estudio por parte de la comunidad científica.

Una ventaja del reconocimiento de rostros en video sobre el reconocimiento de rostros en imágenes fijas es la posibilidad de analizar no solo la apariencia facial sino también los movimientos y la trayectoria del rostro, así como los cambios en las expresiones faciales. Al contar con más de una imagen por persona se pueden generar representaciones del rostro más eficaces como los modelos 3D [12]. Además, se puede llevar a cabo el aprendizaje y la actualización de modelos de individuos sobre el tiempo. Las ventajas mencionadas hasta el momento pueden ser combinadas y aprovechadas para enfrentar los problemas antes descritos [23], [24] y obtener una mayor eficacia en el proceso de reconocimiento de rostros.

2.1 Principales enfoques del reconocimiento de rostros en video

Existe un conjunto de métodos desarrollados hasta el momento enfocados en el reconocimiento de rostros en video. Estos métodos han sido motivados por las dos características distintivas disponibles en una secuencia de video: los múltiples cuadros para un mismo sujeto y la información temporal existente. A continuación se realiza un resumen de estos métodos agrupados en tres categorías fundamentales: los métodos basados en técnicas para imágenes fijas, los basados en la información temporal y los basados en señales híbridas. Algunos de los principales trabajos que evidencian el uso de estos enfoques se resumen en la Tabla 4. Para consultar otros ejemplos ver [16], [25], [26].

2.1.1 Métodos basados en técnicas para imágenes fijas

Los métodos que pertenecen a esta categoría no tienen en cuenta la información temporal disponible en las secuencias de un video, pues consideran el problema como el problema del reconocimiento de rostros en imágenes fijas. Estos métodos aplican las técnicas de reconocimiento para imágenes fijas de dos maneras diferentes: en cuadros-claves o mediante el cotejo de conjuntos de imágenes.

En la primera variante cada video es considerado como una colección de imágenes, de la cual se selecciona una o un conjunto de ellas (cuadros-claves) mediante el empleo de heurísticas. Luego se realiza la clasificación de cada uno de los cuadros-claves seleccionados, aplicando técnicas de reconocimiento basadas en imágenes fijas [15], [33]. En el caso que se seleccione más de un cuadro-clave, el resultado final se obtiene utilizando un método de combinación de clasificadores [34].

La segunda variante utiliza todas las imágenes del rostro y formula el reconocimiento como la correspondencia entre un conjunto de imágenes de prueba contra un conjunto de imágenes de la galería, donde cada uno representa un sujeto. Una posible dificultad o desventaja de estos métodos es la obtención de varias imágenes de rostro por cada individuo para construir la galería. La mayoría de estos métodos utilizan

Tabla 4. Ejemplos de trabajos representativos del reconocimiento de rostros en video.

Enfoques	Trabajos	Breve descripción
Métodos basados en técnicas para imágenes fijas	Gorodnichy, 2002 [27]	El sistema propuesto para el reconocimiento facial se basa en la búsqueda y selección de un cuadro-clave en una secuencia de video dada. Para esto, utilizan las posiciones de tres puntos correspondientes a la nariz y los ojos del rostro. Si la ubicación de estos tres puntos forma un triángulo equilátero, se realiza el reconocimiento en el cuadro actual, de lo contrario se continúa con la búsqueda hasta encontrar un ”buen”marco. Este método aprovecha sólo la abundancia de los cuadros de la secuencia de vídeo y no la dinámica facial.
	Shakhnarovich, 2002 [28]	Proponen un enfoque para el reconocimiento de rostros utilizando conjuntos de imágenes, en el que se comparan directamente los modelos de distribuciones de probabilidad de los rostros observados y el modelo. Para esto desarrollan un método basado en un modelo Gaussiano multivariado de la distribución de la apariencia y una medida de similitud entre las distribuciones de los conjuntos usando la divergencia *Kullback-Leibler* (KL).
Métodos que utilizan la información temporal	Liu, 2003 [29]	Proponen el uso de una adaptación de los modelos ocultos de Markov (HMM, por sus siglas en inglés) para el reconocimiento de rostros basado en video. Para esto en la etapa de entrenamiento crean un HMM para cada individuo con el objetivo de aprender las estadísticas y dinámicas temporales. Luego, durante el proceso de reconocimiento la secuencia de prueba es analizada sobre el tiempo a través del HMM correspondiente a cada sujeto. La identidad es determinada mediante el modelo que proporcione la probabilidad más alta.
	Hadid, 2009 [30]	Proponen un enfoque efectivo para la reconocimiento de rostros en los videos basado en la combinación de la apariencia y el movimiento facial. Para esto mediante el uso del conjunto de características conocido como patrones binarios locales de volumen extendido (EVLBP, por sus siglas en inglés) y un esquema *boosting*, seleccionan sólo la información relacionada con la identidad, tratando de descartar la información relacionada con la expresión facial y las emociones.
Métodos basados en señales híbridas	Zhou, 2006 [31]	Proponen un sistema innovador de fusión para el reconocimiento de individuos no cooperativos que se encuentran distantes en un escenario de una sola cámara. Informaciones de dos fuentes de datos biométricos: el rostro y la forma de caminar, son utilizadas e integradas a través de diferentes estrategias de fusión con el objetivo de mejorar el rendimiento del reconocimiento. Para la combinación de los resultados de los clasificadores del rostro y de la forma de caminar se aplicaron varias reglas: la suma, el producto y el máximo; mostrando la efectividad del enfoque propuesto.
	Micheloni, 2009 [32]	Proponen un sistema que emplea e integra técnicas de reconocimiento de rostros y reconocimiento del locutor con el objetivo de lograr un mayor rendimiento en video. En el módulo de reconocimiento de rostros, se utilizan métodos como la normalización de histograma, el *boosting* y el análisis discriminante lineal (LDA) para resolver problemas como la variación en la iluminación, la oclusión y poses no frontales. Para la reducción de ruido en el habla emplean el filtro de *Kalman* extendido (EKF, del inglés *Entended Kalman Filter*). Cada clasificador (rostro y locutor) obtiene la probabilidad de pertenencia a una clase. Bajo el supuesto de independencia de estos procesos, el producto de estas probabilidades define una nueva probabilidad de pertenencia a una clase; mediante la cual se integran los resultados de ambos métodos.

modelos estadísticos, representando cada conjunto de imágenes mediante una distribución paramétrica y hallando la similitud entre dos distribuciones [35], [36].

2.1.2 Métodos que utilizan la información temporal

Los métodos que pertenecen a este enfoque hacen uso de todas las imágenes disponibles en una secuencia de video teniendo en cuenta su orden temporal, modelando y explotando tanto la información espacial como las dinámicas faciales existentes. Estudios realizados han demostrado que la información dinámica, o sea, las características de los movimientos y los gestos de los individuos, es muy importante en el reconocimiento de rostros [37]. Algunas desventajas de este enfoque es que estos métodos asumen coherencia temporal entre imágenes consecutivas, cuando pudiera ocurrir que las imágenes coleccionadas se hayan obtenido desde varias vistas y sobre largos períodos de tiempo, por lo que estarían desordenadas. Además, a veces se les dan iguales pesos a las características espacio-temporales cuando en realidad algunas de ellas contribuyen más que otras en el reconocimiento [16]. Por otra parte, en la mayoría de los métodos existentes basados en representaciones espacio-temporales la información local de los rostros no es explotada [16], cuando se ha demostrado su importancia para el reconocimiento [38]. A pesar de las desventajas mencionadas, lograr métodos que exploten al máximo la información redundante y la información local de las secuencias de videos, es una vía posible para alcanzar resultados óptimos.

2.1.3 Métodos basados en señales híbridas

Debido a la abundancia de información presente en una secuencia de video, los métodos pertenecientes a esta categoría no solo se basan en el rostro para realizar la tarea del reconocimiento sino que emplean conjuntamente otras señales como el movimiento, la manera de caminar, la voz, entre otras. Estos sistemas son más complejos pues integran el procesamiento de varias rasgos biométricos en un solo sistema. Algunos trabajos que emplean este enfoque se describen en [16].

En resumen, en esta sección se revisaron los principales enfoques y métodos propuestos en la literatura para el reconocimiento de rostros en video, así como sus principales ventajas y desventajas. Después de esto se puede concluir que el mejor camino para llevar a cabo esta tarea es explotar toda la información disponible en una secuencia de video. Dado que solo estamos contando con el rostro como rasgo biométrico, los métodos basados en la información temporal o en las relaciones espacio-temporales son los más apropiados para realizar el reconocimiento de rostros en video. Estos métodos evidencian la principal diferencia existente entre el reconocimiento de rostros en imágenes fijas y el reconocimiento de rostros en videos: la abundancia de información disponible. Sin embargo, para realizar el proceso de reconocimiento de rostros de manera eficiente y eficaz, es imprescindible contar con algoritmos robustos de detección y seguimiento.

3 Detección de rostros en video

La detección de rostros es el primer paso en toda tarea de un sistema de procesamiento de rostros dígase, la alineación, modelación, identificación, verificación, seguimiento, u otros. Por lo tanto, mientras más preciso sea este paso, más preciso será cualquier procesamiento posterior. El objetivo de la detección de rostros es dada una imagen o secuencia de video, determinar si existen o no rostros en la misma y, si hay, retornar la ubicación y tamaño de cada uno de ellos. Factores como la pose, las expresiones faciales, la oclusión, la orientación, entre otros, adicionan complejidad al proceso y convierten la detección de rostros en uno de los tópicos más estudiados por la comunidad científica [39], [40], [41].

En video, la detección de rostros se puede realizar de dos formas diferentes: en cada cuadro o integrada con el seguimiento. La primera forma no es más que aplicar el algoritmo de detección en cada cuadro de la secuencia del video, sin tener en cuenta la información temporal disponible, ni las relaciones existentes entre los cuadros. Cuando se realiza la detección integrada con el seguimiento se puede detectar el rostro en el primer cuadro y después seguirlo a través de toda la secuencia, o se puede realizar una especie de predicción y actualización del seguimiento, donde la detección se realiza cada cierto tiempo para actualizar así el seguimiento, con el objetivo de detectar múltiples rostros y evitar la pérdida de otros durante la secuencia [16]. Si embargo, en cualquiera de los casos mencionados los enfoques para la localización de los rostros en un video están basados en los métodos de detección en imágenes fijas, que después pueden ser integrados con cualquier método de seguimiento. A continuación se presenta un estudio de los principales enfoques propuestos en la literatura para la detección de rostros en imágenes y se presta una mayor atención a los métodos basados en la apariencia por ser los que mejores resultados han mostrado [41].

3.1 Métodos de detección de rostros en imágenes

La detección de rostros se puede realizar utilizando diferentes señales extraídas de las imágenes como el color de la piel, la forma facial y de la cabeza, la apariencia facial o cualquier combinación de estos parámetros. Varios trabajos han hecho una revisión de los métodos propuestos para detectar rostros en imágenes [39], [40], [41]. Tradicionalmente, los métodos existentes pueden ser agrupados en cuatro categorías: los métodos basados en el conocimiento, los métodos basados en características invariantes, los métodos basados en correspondencia de plantillas y los métodos basados en la apariencia [40]. Recientemente, ha aparecido una nueva tendencia de trabajos que utilizan la información contextual para la detección de rostros. A partir de lo mencionado anteriormente, en la Figura 4 se muestra una taxonomía que organiza los métodos existentes para llevar a cabo la detección de rostros en imágenes.

A continuación se describen los métodos de cada categoría, haciendo énfasis en los métodos basados en la apariencia ya que actualmente dominan los avances en la investigación de la detección de rostros [41].

3.1.1 Métodos basados en conocimiento

Los métodos basados en conocimiento utilizan ciertas reglas basadas en el conocimiento humano sobre las características que describen un rostro como son la forma, el tamaño, la textura y otros rasgos faciales como los ojos, la nariz, la barbilla, las cejas; así como las relaciones entre ellos (posiciones relativas y distancias) [42], [43]. La principal desventaja de estos métodos es encontrar una forma exitosa de traducir este conocimiento en reglas significativas y bien definidas, pues si las reglas son demasiado restrictivas muchos rostros podrían ser descartados (falsos negativos), mientras que, si las reglas son demasiado generales patrones que no pertenecen a la clase rostro serán incluidos en ella (falsos positivos). Otra desventaja es que estos métodos no son robustos ante variaciones en la pose u orientaciones de la cabeza, pues resulta difícil definir reglas que tengan esto en cuenta.

3.1.2 Métodos basados en características invariantes

Los métodos basados en características invariantes tienen como objetivo principal encontrar rasgos que no desaparezcan del rostro bajo ninguna condición. Luego, usando las características extraídas se construye un modelo estadístico para describir sus relaciones y verificar la existencia de un rostro. Varios métodos han propuesto detectar características faciales como los ojos, la boca, la nariz y las cejas [44]. Estas

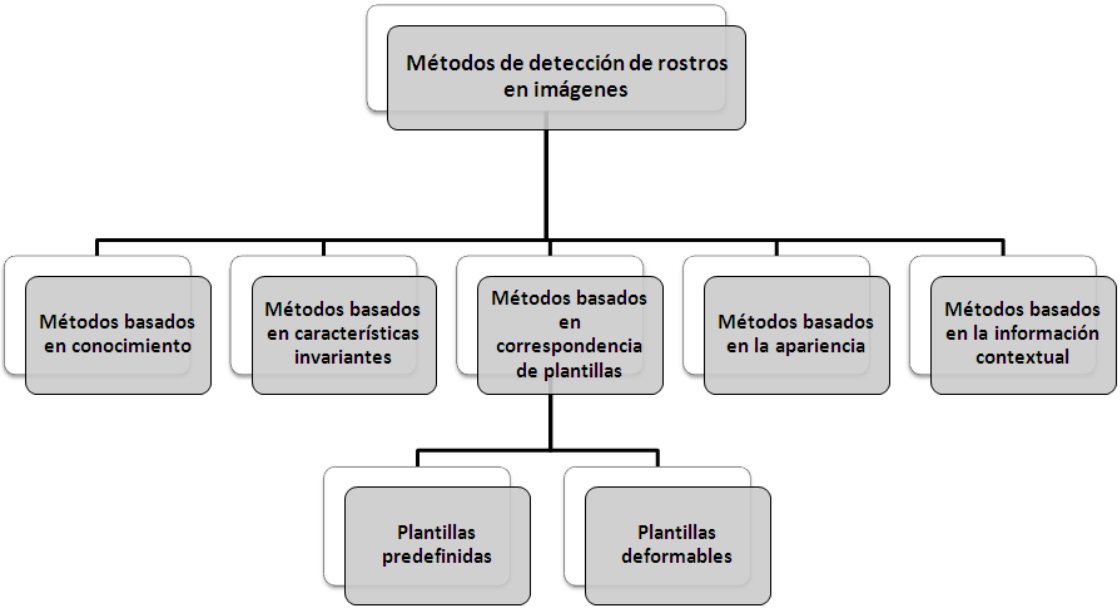


Fig. 4. Taxonomía propuesta para los métodos de detección de rostros en imágenes.

características comúnmente se extraen utilizando detectores de bordes. Otros enfoques más recientes se basan en la combinación de características [45]. La mayoría de estos métodos utilizan características globales como el color de la piel, la textura y la forma para encontrar los rostros candidatos y después verifican estos usando las características faciales. Las principales desventajas de los métodos basados en características invariantes es precisamente la detección de los rostros candidatos y de los rasgos dentro de estos, en fondos complejos no homogéneos y ante la presencia en la imagen de otras zonas similares a la piel. Además, la localización de características faciales falla cuando existe la influencia de factores como la iluminación, el ruido, y la oclusión.

3.1.3 Métodos basados en la correspondencia de plantillas

Los métodos basados en la correspondencia de plantillas comparan la imagen de entrada con una plantilla previamente almacenada utilizando métodos de correlación para localizar los rostros. Estas plantillas pueden ser predefinidas (basadas en bordes o regiones) o deformables (basadas en el contorno facial). Las plantillas predefinidas son aquellas que se construyen a partir de un patrón facial estándar predefinido de forma manual [46]. Los métodos que se basan en este tipo de plantillas no pueden tratar de manera efectiva las variaciones en escala, pose y forma del rostro. Para dar solución a este problema, surgieron las plantillas deformables, las cuales son construidas mediante una función con parámetros [47]. Estas plantillas son lo suficientemente flexibles como para poder cambiar su tamaño, y otros parámetros, para ajustarse a los datos. Sin embargo, una dificultad de los métodos basados en plantillas deformables es que estas plantillas deben ser inicializadas cerca del objeto de interés, en este caso el rostro. De manera general, aunque los métodos basados en la correspondencia de plantillas (predefinidas o deformables) son

fáciles de implementar, es difícil enumerar las plantillas para las diferentes poses y variaciones existentes. Además, estos métodos computacionalmente son muy costosos.

3.1.4 Métodos basados en la apariencia

Los métodos basados en la apariencia aprenden modelos de rostros a partir del entrenamiento sobre la base de un conjunto representativo de imágenes etiquetadas. Para ello se basan en técnicas de análisis estadístico y de aprendizaje automático con el objetivo de encontrar las características más relevantes de las imágenes y luego clasificar los rostros candidatos en una nueva imagen. Varias estrategias han sido propuestas para obtener una descripción del contenido de las imágenes; tales como el uso de características globales, de desplazamiento de ventanas, de segmentos de imagen, de rejillas fijas o de regiones aleatorias de la imagen [48].

En el contexto de la detección de objetos, específicamente de rostros, el desplazamiento de ventanas ha sido uno de los enfoque más utilizados por los métodos basados en la apariencia [49], [48]. La idea principal de estos métodos es escanear la imagen mediante una ventana deslizante que selecciona en cada desplazamiento diferentes subregiones de la imagen. En cada una de estas regiones se evalúa una función de clasificación y se seleccionan como regiones candidatas aquellas donde se obtuvo las tasas máximas de clasificación. Formalmente esto puede ser definido como:

$$R_{obj} = \arg\max_{R \subseteq I} f(R), \tag{1}$$

donde R son todos los rangos de las regiones rectangulares en la imagen I y f la función de clasificación. Al centrarse en las subregiones de la imagen, estos métodos son capaces de detectar objetos a pesar de los fondos cambiantes, e incluso si el objeto sólo cubre un pequeño porcentaje del área total de la imagen. Sin embargo, debido a que el número de rectángulos en una imagen $n \times m$ es de orden n^2m^2, no se puede comprobar de forma exhaustiva todas las subregiones posibles de la imagen.

Varias heurísticas han sido propuestas para acelerar la búsqueda de estos métodos mediante la reducción del número de evaluaciones necesarias. Una primera variante es realizar la búsqueda sólo con rectángulos de ciertos tamaños fijos como candidatos, a diferentes escalas [50]. Aunque se han aplicado métodos de imágenes integrales [51] e histogramas integrales [52] con el objetivo de acelerar los cálculos, este enfoque continua siendo computacionalmente muy costoso; pues estas técnicas solo pueden ser empleadas para calcular algunos tipos específicos de características. Además, debido a que el desplazamiento de ventanas prueba múltiples sub-ventanas que se solapan significativamente con el rostro verdadero, como resultado se obtienen multiples detecciones sobre este, por lo que después estas detecciones deben ser procesadas.

Otra alternativa es el uso de métodos de optimización local, primero identificando regiones candidatas en la imagen y después maximizando f mediante un procedimiento discreto de ascenso de gradiente para perfeccionar la detección [53]. Sin embargo, estos métodos en ocasiones fallan ante la presencia de mínimos o máximos locales.

Recientemente fue propuesto un método, basado en una técnica de optimización global, conocido como búsqueda de sub-ventana eficiente (ESS, del inglés *efficient subwindow search*) para predecir la mejor ubicación de un objeto en una imagen [54], [55]. Este método utiliza el esquema de optimización *branch-and-bound* [56] para encontrar el óptimo global de una función de calidad sobre todas las posibles sub-imágenes. Dicha función de calidad se construye sobre la base de un algoritmo de aprendizaje (clasificador) entrenado. El método ESS resulta muy rápido ya que no realiza una búsqueda exhaustiva. Al mismo tiempo, requiere muchas menos evaluaciones del clasificador que regiones candidatas existentes en la imagen, lo que permite el uso de clasificadores más complejos. Este enfoque es actualmente uno de los

más eficaces para encontrar la localización óptima de objetos arbitrarios en las imágenes, con un resultado equivalente al que se obtiene con una búsqueda exhaustiva usando desplazamiento de ventanas.

Existen varios problemas a la hora de aplicar este enfoque. El primero es encontrar una buena función acotadora para otros clasificadores y características. Otro aspecto lo constituye la detección de múltiples instancias de un objeto, ya que con el uso del esquema *branch-and-bound* solo se obtiene un óptimo global, es decir, la mejor ubicación del objeto en la imagen. Una variante para enfrentar esta última limitación consiste en aplicar el método ESS repetidamente, eliminando de la imagen, en cada iteración, la localización óptima encontrada hasta lograr el número de localizaciones deseadas. Sin embargo, esta variante de solución es computacionalmente costosa y resulta muy difícil conocer a priori la cantidad real de iteraciones necesarias.

Otra variante puede ser el uso de algoritmos para solucionar problemas de optimización multi-modal [57], como son algoritmos genéticos, algoritmos evolutivos, *Particle Swarm Optimization* (PSO) [58], entre otros. Además, para lograr mayor eficiencia en los algoritmos de búsqueda, otras técnicas como las arquitecturas en cascadas pueden ser utilizadas [2].

Una vez seleccionada la técnica para la búsqueda de los objetos, en nuestro caso los rostros; dos cuestiones importantes a resolver en los métodos basados en la apariencia durante el proceso de detección son: qué características o rasgos extraer, y qué algoritmo de aprendizaje o clasificador aplicar. Por su importancia en este proceso, a continuación se revisan las principales propuestas existentes para enfrentar estas dos tareas.

Características o rasgos utilizados para representar un rostro

Uno de los tantos desafíos existentes y del cual depende la eficacia de cualquier sistema de detección es la extracción de características de la imagen o cuadro de la secuencia de video que se quiera procesar. Para la detección de rostros han sido propuestos muchos tipos de características sobre la base de varias propiedades físicas de los rostros, como la intensidad del color, la textura, los vértices, las formas y los rasgos faciales. Las características ideales deben tener poder discriminativo para diferenciar entre el rostro de las personas y los demás objetos presentes en las escenas, siendo robustas ante variaciones en la iluminación, cambios de expresión o ligeras variaciones en la pose. Además, debido a la necesidad de sistemas que operen en tiempo real, estas características deben ser rápidas de obtener o calcular. Por lo expuesto anteriormente, escoger las características o rasgos a utilizar para representar los rostros en las imágenes es un problema.

Utilizando los métodos basados en la apariencia se puede representar el rostro manera global o de manera local. La representación global trata la región del rostro como un todo, mientras que en la local el rostro es divido en partes y cada una de ellas se trata individualmente. Esta última facilita el enfrentamiento a problemas como la oclusión parcial y las variaciones de iluminación. Una de las primeras técnicas propuestas en la literatura para representar el rostro de una persona fue mediante la utilización de los valores de los píxeles como características [59], [60]. Sin embargo, históricamente trabajar solo con las intensidades de los píxeles de la imagen, introduce el problema de la alta dimensionalidad de los rasgos extraídos y hace que la tarea de detección se vuelva computacionalmente costosa. Además, existen otras razones que han motivado el uso de otras características o rasgos que describen vecindades de los píxeles en la imagen en lugar de los valores de los píxeles directamente [50], [3]. La razón más común es que este tipo de rasgos permiten codificar el conocimiento específico del dominio de interés, como por ejemplo las estructuras de bordes y líneas, lo cual es difícil de aprender utilizando las intensidades de los píxeles directamente en una cantidad finita de datos de entrenamiento. Otra razón fundamental es que los sistemas basados en estos rasgos funcionan mucho más rápido con respecto a los basados en píxeles [50]. Por últi-

mo, una tercera motivación es la sensibilidad de la representación basada en píxeles ante las condiciones de iluminación y los ruidos en las imágenes.

El uso del conjunto de funciones base *Haar wavelet* ha sido una de las propuestas más extendidas y eficaces para extraer la información de la textura que describe una clase de objetos [61]. Esta representación codifica en una imagen las diferencias de las intensidades promedio entre dos regiones rectangulares adyacentes, capturando las similitudes estructurales entre las instancias de la clase de objetos, en este caso el rostro. Motivados por esta idea, Viola y Jones [50] adoptaron el uso de las *Haar wavelet* y desarrollaron las llamadas características-*Haar*. Una característica-*Haar* considera regiones rectangulares adyacentes (blancas y negras) en un lugar específico dentro de una ventana de detección, suma las intensidades de los píxeles en estas regiones y calcula la diferencia entre ellas. Luego, esta diferencia se utiliza para clasificar las subregiones candidatas de una imagen. Dentro de cualquier ventana de detección el número total de características-*Haar* (conjunto exhaustivo) es muy grande, mucho mayor que el número de píxeles. Con el fin de garantizar una clasificación rápida, el proceso de aprendizaje debe excluir a una gran mayoría de estas características, y centrarse en un conjunto pequeño de características críticas o más significativas para la clasificación.

En la Figura 5 se muestra el conjunto base de características-*Haar*, formado por tres tipos de estas características: utilizando dos rectángulos, tres rectángulos y cuatro rectángulos. Para el caso de dos rectángulos, el valor representa la diferencia entre la suma de los píxeles dentro de las dos regiones rectangulares, que tienen el mismo tamaño y forma y pueden estar horizontal o verticalmente adyacentes. Las características de tres rectángulos calculan la suma dentro de los dos rectángulos exteriores y a este valor se le resta la suma en el rectángulo central. Por último, las características de cuatro rectángulos calculan la diferencia entre pares diagonales de rectángulos.

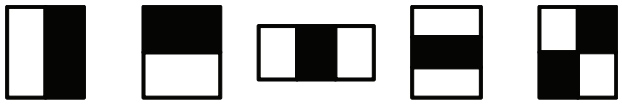


Fig. 5. Conjunto base de características- *Haar*. (A) y (B) muestran características de dos rectángulos en sentido horizontal y vertical respectivamente, (C) y (D) características de tres rectángulos en sentido horizontal y vertical respectivamente, y (E) características de cuatro rectángulos.

Para calcular las características *Haar* rápidamente a varias escalas se introduce la representación de la imagen conocida como imagen integral. La imagen integral puede calcularse a partir de una imagen con unas pocas operaciones por píxel. La imagen integral en el punto (x,y) se define como la suma de los valores de los píxeles que se encuentran por encima y a la izquierda de dicho punto en la imagen original [50], es decir:

$$ii(x,y) = \sum_{a \leq x, b \leq y} i(a,b), \tag{2}$$

donde $ii(x,y)$ representa la imagen integral y $i(a,b)$ el valor de la intensidad del píxel (a,b) en la imagen original (ver Figura 6, en la cual el recuadro completo es la imagen y la parte gris representa el valor de la imagen integral). Una vez calculada la imagen integral, esta puede ser evaluada a cualquier escala o ubicación en tiempo constante. Dado que el conjunto exhaustivo de características-*Haar* en cada ventana de búsqueda en la imagen es muy amplio, en el método propuesto por Viola y Jones se utiliza el algoritmo *AdaBoost* [62] para seleccionar un pequeño número de características distintivas de todas estas posibles

combinaciones. A pesar de los resultados obtenidos en este trabajo, el conjunto de características utilizado no es robusto ante cambios en la iluminación y debido a su simplicidad tienen un pobre nivel descriptivo. Por esta razón, se han desarrollado una serie de trabajos basados en la creación de nuevos conjuntos de características robustas para optimizar el rendimiento de la tarea de detección de rostros.

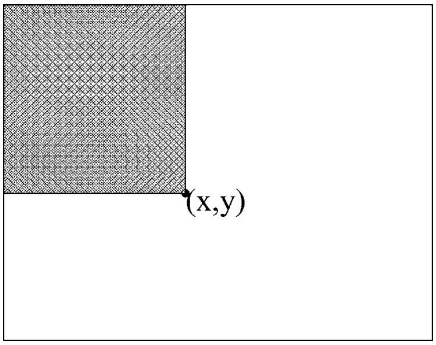


Fig. 6. Imagen integral en el punto (x,y).

La creación de nuevos conjuntos de características para representar los rostros en la tarea de detección ha estado encaminada en dos direcciones: la creación de conjuntos que extiendan el conjunto base de características-*Haar* [50], [61] y la creación de otros tipos de conjuntos [6], [7], [9].

Una de las primeras extensiones del conjunto de características-*Haar* consiste en agregar características rotadas 45° que también pueden calcularse de manera rápida y eficiente [63]. En la Figura 7 se muestran las características adicionadas, agrupadas según sus prototipos en características de borde, características de línea, características alrededor del centro y características de la diagonal. Utilizando las características adicionales se mejora significativamente el poder expresivo del sistema de aprendizaje y en consecuencia, se mejora el rendimiento del sistema de detección de rostros. En la Figura 8 se muestra el resultado de la comparación de la detección usando ambos conjuntos. Como puede observarse, el conjunto ampliado mejora la tasa de éxito con respecto al conjunto base de características-*Haar*. Sin embargo, el conjunto ampliado de características usualmente complica el aprendizaje.

Otra extensión del conjunto base de características-*Haar*, es el conocido como características *Joint-Haar*, basado en la co-ocurrencia de múltiples características-*Haar* [3]. La co-ocurrencia de características se puede definir como la búsqueda de la probabilidad conjunta de múltiples características ocurridas al mismo tiempo. Esto incrementa las diferencias entre las densidades probabilísticas de las clases y por tanto mejora la clasificación. Las combinaciones de características se obtienen mediante el método *sequential forward selection* (SFS) [64]. La cantidad de características a combinar se determina mediante heurísticas encaminadas a disminuir los errores de clasificación y dicha cantidad es acotada. Los resultados obtenidos en este trabajo muestran que este método obtiene un mayor rendimiento que el de Viola y Jones [50], incluso cuando ambos métodos utilizan el mismo número de características. Dos desventajas de este enfoque son que no se garantiza encontrar la mejor combinación de características usando SFS y que la combinación de un número grande de características podría provocar sobre entrenamiento [65]. Por lo tanto, estos dos aspectos son merecedores de estudio en aras de mejorar este enfoque.

En otra de las extensiones que se basa en la co-ocurrencia de las características-*Haar* se mantiene solamente las relaciones ordinales entre los rectángulos para obtener un conjunto de características, denominado características-*Haar* binarias, las cuales solo mantienen la información del signo descartando el valor de la diferencia absoluta entre las intensidades acumuladas [4]. Luego, múltiples características-*Haar* binarias son ensambladas juntas y su co-ocurrencia es usada como un nuevo tipo de característica.

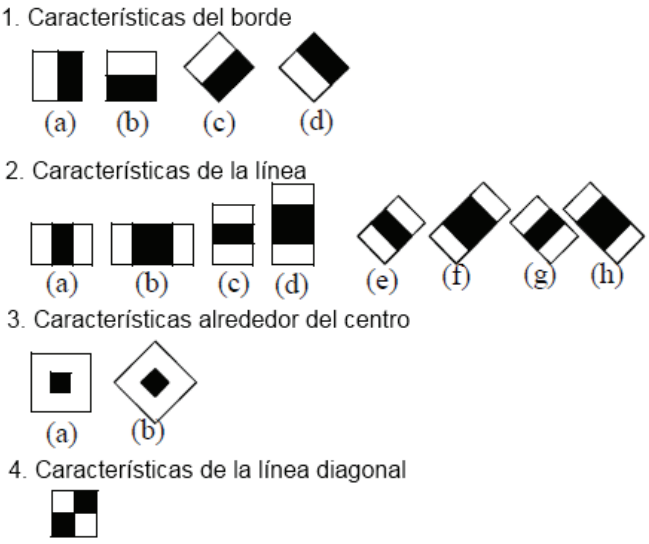


Fig. 7. Ejemplos del conjunto ampliado de características-*Haar*.

Dado que el número de características-*Haar* binarias ensambladas es enorme, se propone un conjunto reducido, llamado *Locally Assembled Binary* (LAB). Entre todo el conjunto de las características-*Haar* binarias ensambladas, las LAB son aquellas que solo combinan 8 características-*Haar* binarias de dos rectángulos localmente adyacentes con el mismo tamaño (ver Figura 9) y tienen un rectángulo centro común. En la Figura 10 se muestra un ejemplo de una característica LAB. Una desventaja de esta variante es que, con el objetivo de hacerla computacionalmente factible, restringe la cantidad y la forma en que las características-*Haar* pueden ser combinadas. El nuevo conjunto propuesto captura la estructura de la intensidad local de la imagen, siendo no sólo más robusto a las variaciones de iluminación, sino también muy discriminante en la clasificación de rostros/no rostros con un menor número de características para el aprendizaje. En los experimentos realizados por los autores [4], el conjunto propuesto resultó ser más eficiente que las características-*Haar* tanto en poder discriminante como en costo computacional.

Estudios recientes realizados en el área de minería de datos mostraron un eficiente camino para generar combinaciones válidas de características usando la técnica de minería de conjuntos frecuentes (FIM, del inglés *frequent item-set mining*) [65]. Motivados por esta idea, se propuso un nuevo conjunto de características llamado *informative Frequent Assembled Haar Feature* (iFAHF) para la detección de rostros [1]. Este conjunto, encaminado a superar las limitaciones de las características *Joint-Haar* [3] y las LAB [4], está formado por combinaciones de características-*Haar* generadas automáticamente mediante el FIM sin restricciones. Las combinaciones son generadas basadas en un análisis estadístico de las muestras de entrenamiento y por lo tanto estas generalizan bien sobre los datos de prueba. Estas nuevas características permiten capturar no solo la información de textura local sino también sus configuraciones espaciales. En la Figura 11 se muestran los resultados obtenidos por los autores de este trabajo en las comparaciones hechas sobre un mismo conjunto de datos y usando el mismo clasificador. Como se puede apreciar en esta figura, usando el nuevo conjunto de características se obtiene una mayor precisión y una menor tasa de falsos positivos con respecto a las características-*Haar* [2], las *Joint-Haar* [3] y las LAB [4].

Recientemente, el uso de una nueva extensión del conjunto de características-*Haar* junto al clasificador máquinas de vectores soporte (SVM, por sus siglas en inglés) fue explorado en la detección de rostros frontales en tiempo real [5]. El conjunto propuesto está basado en el conjunto base de características-*Haar*

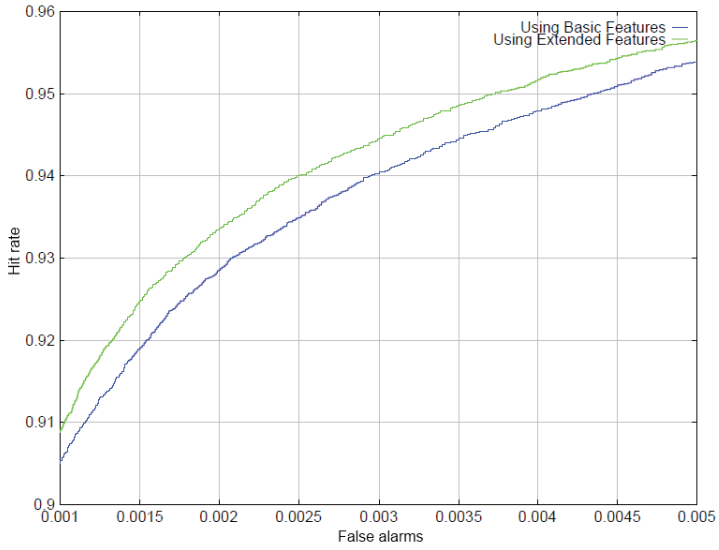


Fig. 8. Comparación entre el conjunto base de características-*Haar* (línea azul) y el conjunto ampliado (línea verde).

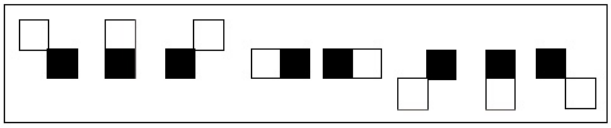


Fig. 9. Características-*Haar* binarias de dos rectángulos que se ensamblan en una característica LAB.

de Viola y Jones [50]. Para la obtención de estas características, en lugar de utilizarse los valores de los píxeles se utilizan los valores de la varianza de cada región. Estas características basadas en los valores de varianza también pueden calcularse muy rápido mediante la imagen integral. La Figura 12 muestra los resultados obtenidos en experimentos realizados en este trabajo en dos bases de datos diferentes. El sistema de detección de rostros resultante del empleo del conjunto de características propuesto con el clasificador SVM [5] mejora ligeramente al sistema basado en el conjunto base de características-*Haar* con el algoritmo *AdaBoost* [50]. Sin embargo, resulta difícil atribuir estas mejoras al uso del nuevo conjunto de características debido a que en la comparación se emplearon clasificadores diferentes.

Además de las diferentes extensiones del conjunto base de características-*Haar* otros trabajos se han encaminado en la utilización de otros tipos de conjuntos de características. El uso de conjuntos basados en los patrones binarios locales (LBP, del inglés *Local Binary Pattern*) se ha convertido en una opción más popular que las características-*Haar* para la detección de rostros [66]. Las propiedades más importantes de las características LBP: la tolerancia frente a los cambios monotónicos de iluminación y la simplicidad de cálculo [67], han permitido enfrentar la vulnerabilidad de las características-*Haar* ante las variaciones en las condiciones de iluminación, siendo útiles en entornos no controlados y logrando que el proceso de entrenamiento sea más corto.

El descriptor LBP original ha demostrado ser un poderoso descriptor de textura [66]. El operador LBP básico [68] asigna una etiqueta a cada píxel de una imagen en una vecindad de 3x3, usando como umbral

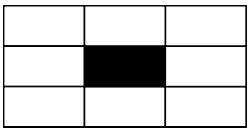
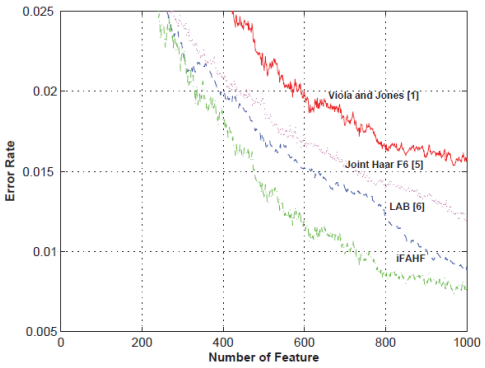
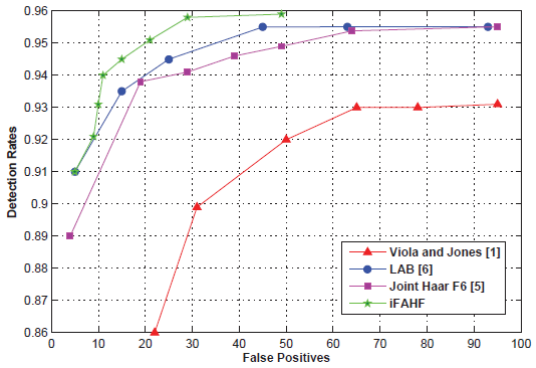


Fig. 10. Ejemplo de una característica LAB.



(a) *Curvas de los errores de prueba en la detección de rostros frontales*



(b) *Curvas ROC en la detección de rostros frontales.*

Fig. 11. Comparación presentada por los desarrolladores de las características iFAHF [1], entre estas y el uso de las características-*Haar* [2], las *Joint-Haar* [3] y las LAB [4].

el píxel del centro. Luego, el resultado obtenido se considera como un número binario. La forma decimal del código de 8 bits resultante se puede expresar como:

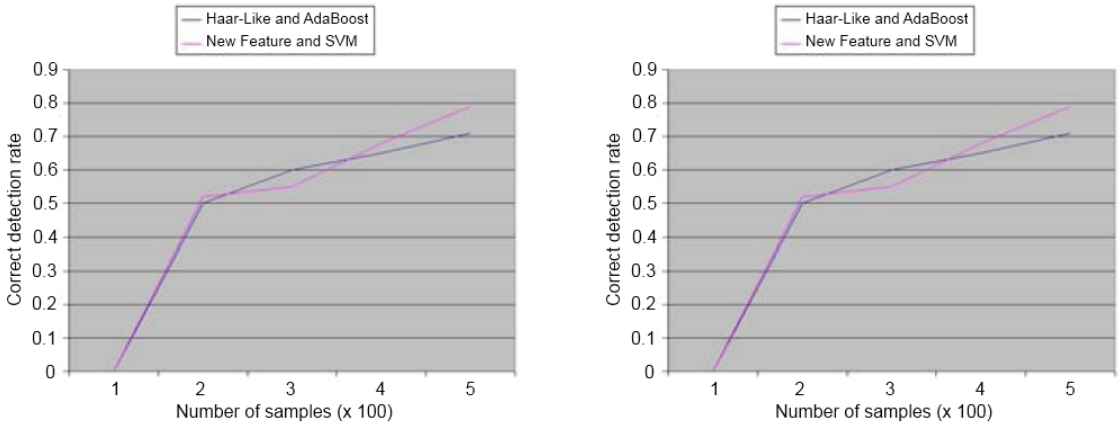
$$LBP(x,y) = \sum_{n=0}^{7} f(v_i - v_c)2^n, \tag{3}$$

donde v_c corresponde al valor de intensidad del píxel central (x,y) y v_i son los valores de intensidad de los 8 píxeles que rodean al píxel central. La función $f(v_i - v_c)$ se define como:

$$f(v_i - v_c) = \begin{cases} 1 & \text{si } v_i - v_c \geq 0 \\ 0 & \text{si } v_i - v_c < 0 \end{cases}$$

Un ejemplo del proceso de obtención de la etiqueta LBP de un píxel se muestra en la Figura 13. Una vez que se tienen todas las etiquetas LBP de todos los píxeles se construye un histograma con esos valores, que es utilizado como descriptor de la textura de la imagen.

Distintas extensiones del LBP básico han sido propuestas con el objetivo de mejorar su rendimiento en la detección de rostros. Por ejemplo, codificando regiones rectangulares mediante el operador LBP se pueden obtener características más distintivas que las tradicionales características-*Haar* [6]. El conjunto conocido como multibloque de patrones binarios locales (MB-LBP por sus siglas en inglés), está inspirado en las características-*Haar* [6]. El conjunto MB-LBP en lugar de comparar los valores de píxeles codifica las regiones rectangulares mediante la comparación de los valores medios del rectángulo central con los de sus rectángulos vecinos, lo que puede hacerse en tiempo constante usando la imagen integral. En la Figura 14 se muestra una descripción detallada del proceso. Los resultados obtenidos en este trabajo muestran que las características MB-LBP capturan mayor información acerca de la estructura de la imagen y la



(a) *base de datos de rostros YaleB* (b) *base de datos de rostros CalTech.*

Fig. 12. Comparación presentada en [5] entre el método propuesto y el sistema de Viola y Jones [2].

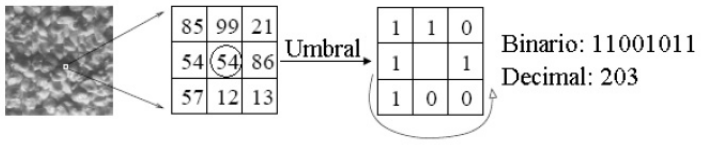


Fig. 13. Asignación de etiqueta a un píxel mediante el operador LBP básico.

clasificación basada en este conjunto logra ser más distintiva con respecto al uso de las características-*Haar* y los LBP. Además, el número de características MB-LBP (rectángulos a varias escalas, ubicaciones y radios) del conjunto exhaustivo es mucho menor que el de las características-*Haar*, lo que hace que el proceso de entrenamiento sea menos costoso. La Figura 15, muestra las comparaciones realizadas en sus experimentos entre las características MB-LBP, las características-*Haar* [69] y los LBP [6]. Como se puede apreciar en ambas gráficas el uso de las características MB-LBP obtuvo tanto la menor tasa de error 15(a) como el mayor rendimiento en la clasificación 15(b). Además, el detector en cascada obtenido (9 capas y 470 características MB-LBP) es más eficiente que el de Viola y Jones [69] (32 capas y 4297 características-*Haar*), con menor niveles de cascada y cantidad de características.

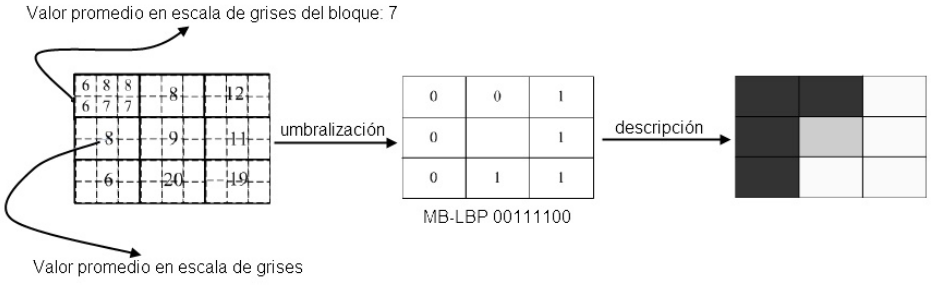
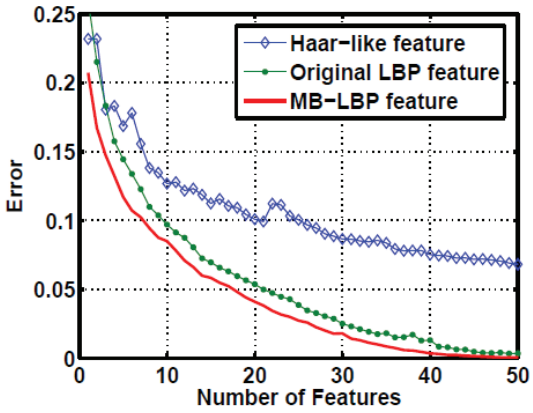
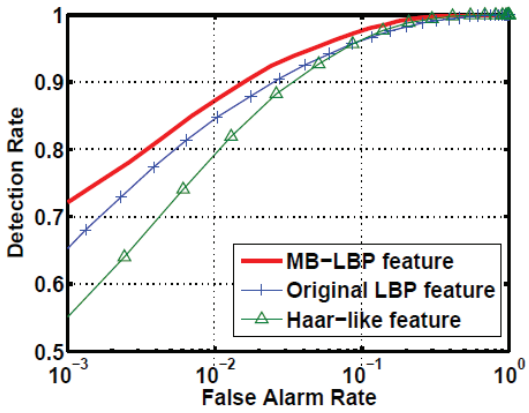


Fig. 14. Ejemplo de región codificada utilizando MB-LBP.



(a) *Curvas de la tasa de error en función de la cantidad de características seleccionadas en el proceso de entrenamiento.*



(b) *Curvas ROC del rendimiento de clasificación de los tres clasificadores en el conjunto de la prueba.*

Fig. 15. Comparación de los resultados obtenidos con las características MB-LBP, Haar y LBP original [6].

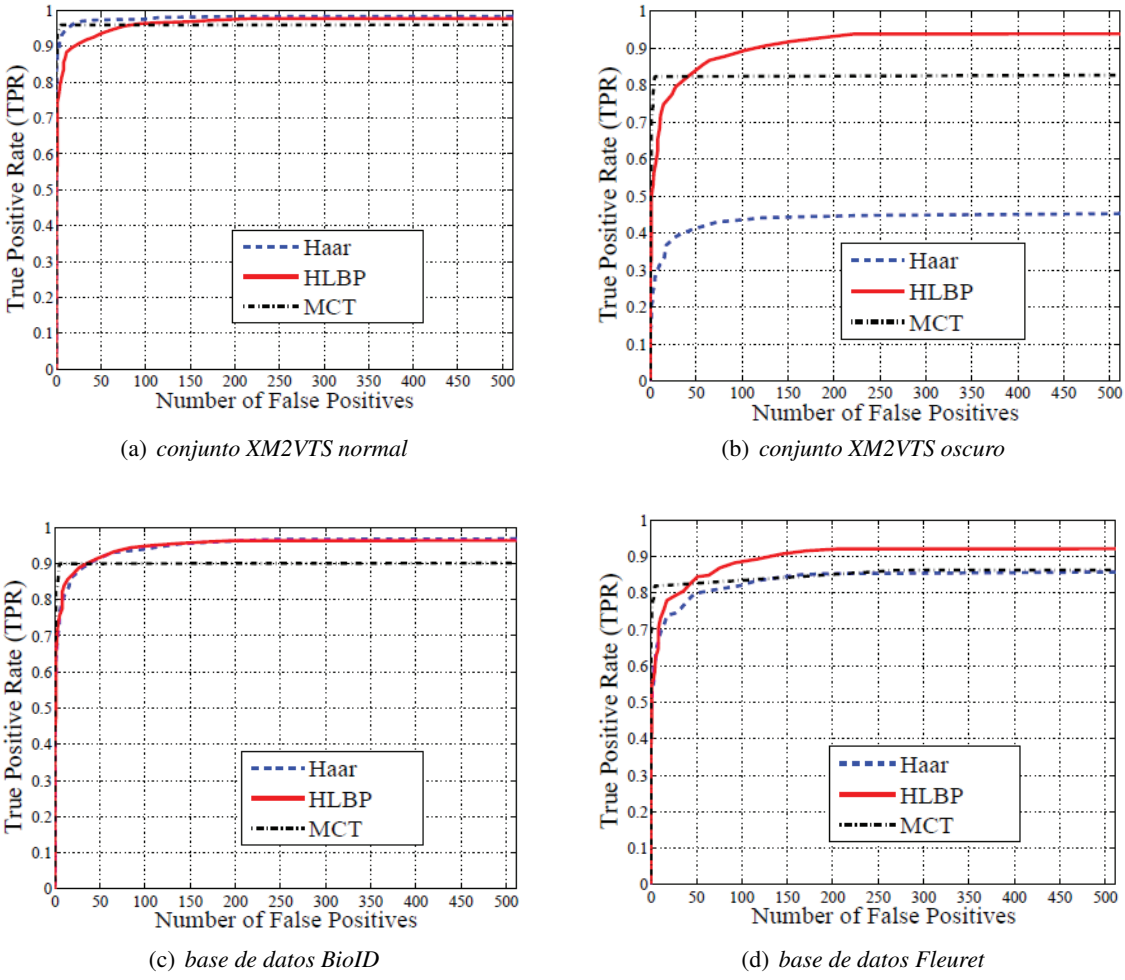
Otras variantes del LBP como el *Integral* LBP (INTLBP) y el *Dominant* LBP (DLBP) fueron también propuestas recientemente para la tarea de detección de rostros [70] y [71]. El conjunto INTLBP, combina dos descriptores de imagen existentes, los histogramas integrales [52], [72] y los LBP para obtener una nueva representación con propiedades importantes como procesamiento rápido en tiempo constante, rotación e invariancia a la escala. El enfoque propuesto se basa en el cálculo de histogramas integrales de imágenes LBP. Para una imagen de $n \times m$, el histograma integral se representa mediante $(n+1) \times (m+1)$ arreglos de longitud L, donde L representa el número de posibles valores del histograma. De esta manera, el histograma integral $H_{x,y}[p]$ en la posición (x,y) se define según el histograma de la imagen arriba y a la izquierda de (x,y) como:

$$H_{x,y}[p] = \sum_{a \leq x, b \leq y} \delta(a,b), \tag{4}$$

donde $\delta(a,b) = 1$, o $\delta(a,b) = 0$, si el valor de intensidad del píxel (x,y) pertenece o no al p-ésimo *bin* del histograma. De esta manera, en cada desplazamiento de la ventana de búsqueda en la imagen, primeramente se crea una imagen LBP mediante la sustitución los valores de los píxeles por el correspondientes códigos LBP y luego, de esta se calcula el histograma integral. Por otra parte, el conjunto DLBP se define como los patrones ocurridos más frecuentemente en una imagen de rostro. Estas características son discriminatorias y robustas en un rango de resoluciones de la imagen facial, lo que es muy útil para aplicaciones reales.

La combinación de las características-*Haar* y los LBP ha sido utilizada para capturar la variación de los patrones locales de textura en regiones específicas [7]. El conjunto denominado características-*HaarLBP*, se basa en la comparación de cantidades de etiquetas LBP en dos subregiones adyacentes de la imagen. Estas subregiones están representadas por un conjunto de rectángulos o máscaras similares a las *Haar* [50]. En otras palabras, estas características indican si el número de veces que una determinada etiqueta LBP se produce en una región es mayor o menor que el número de veces que se produce en otra región, compensado por un determinado umbral. El cálculo de estas características mediante una variación del método histograma integral [73] permite que el algoritmo pueda emplearse en tiempo real. De esta

manera, el nuevo conjunto de características obtenido es relativamente robusto a grandes variaciones de iluminación, de pose y de fondo, y también a pequeñas variaciones en la postura. En los experimentos realizados por los autores [7] las características-*HaarLBP* fueron comparadas con una de las variantes del LBP, el *Modified Census Transform* (MCT) [74], conocido también como LBP modificado (mLBP), el cual compara cada uno de los píxeles de un bloque de 3x3 con la media de los valores de intensidad dentro de esa cuadrícula, en lugar del píxel central como en el LBP. Otras características que se usaron en la comparación fueron las características-*Haar* [50]. La Figura 16 muestra los resultados obtenidos en la detección de rostros en cuatro conjuntos de datos diferentes. Cada gráfico representa las curvas de porciento de detección correcto contra la cantidad de falsos positivos detectados. Como se observa en esta figura, usando el conjunto de características-*HaarLBP* se obtienen los mejores resultados en los cuatro casos, siendo a su vez menos costoso computacionalmente. En el conjunto oscuro de la base de datos XM2VTS, en el cual las imágenes presentan variaciones en las condiciones de iluminación, se puede notar la marcada superioridad de estas características sobre las otras, demostrando mayor robustez ante este tipo de variaciones.



(a) *conjunto XM2VTS normal*

(b) *conjunto XM2VTS oscuro*

(c) *base de datos BioID*

(d) *base de datos Fleuret*

Fig. 16. Comparación presentada en [7] de los resultados de detección de rostros en diferentes conjuntos de datos utilizando las características Haar, HLBP y MCT.

Con el objetivo de complementar el sistema original de codificación binaria del LBP fueron propuestos otros dos esquemas: el *Transition Local Binary Patterns* (tLBP) y el *Direction coded Local Binary Pattern* (dLBP) [8]. El esquema tLBP se basa en la comparación de píxeles vecinos en la dirección de las agujas del reloj, para todos los píxeles excepto el central (ver Figura 17(a)). Mientras que el dLBP proporciona una mejor información del patrón local, comparando los píxeles en cuatro direcciones bases que cruzan el píxel central de una región (ver Figura 17 (b)). La variación de la intensidad a lo largo de estas direcciones se codifica en dos *bits*, logrando la misma longitud que en el LBP original.

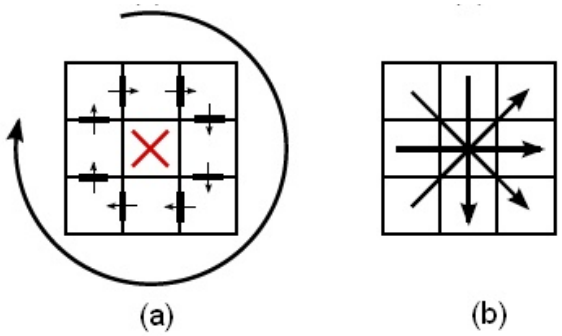
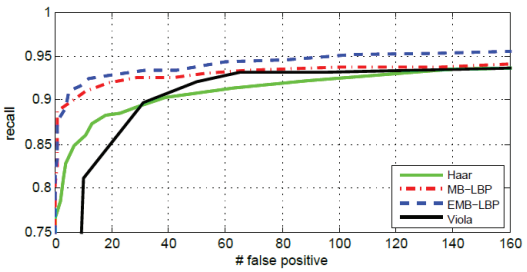
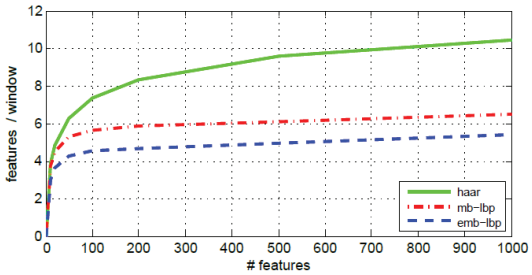


Fig. 17. Extensiones del sistema original de codificación binaria del LBP: (a) *Transition Local Binary Patterns* (tLBP) y (b) *Direction coded Local Binary Pattern* (dLBP).

Estas reglas aumentan el poder descriptivo del operador, preservando una propiedad importante del LBP: la invariancia a las transformaciones monótonas de la intensidad y por otra parte, no aumentan la complejidad de la evaluación del modelo. Su utilidad ha sido probada en la detección de distintos tipos de objetos [8]. En este caso solo analizaremos la detección de rostros, ya que es nuestra tarea de interés. Para este caso, los autores utilizaron el algoritmo de aprendizaje *WaldBoost* [75] y realizaron las pruebas en la base de datos MIT-CMU. En los experimentos realizados por los autores los esquemas de codificación propuestos y el LBP modificado (mLBP) [74] fueron evaluados utilizando el conjunto MB-LBP, ya que este conjunto ha superado al LBP estándar. Por lo que el conjunto extendido (EMB-LBP) incluye los conjuntos MB-LBP, mMB-LBP, tMB-LBP y dMB-LBP. En la Figura 18 se muestran las gráficas que reflejan los resultados obtenidos, donde el conjunto extendido (EMB-LBP) obtuvo un rendimiento superior en la clasificación (Figura 18(a)), utilizando una menor número promedio de características en cada posición escaneada (Figura 18(b)) con respecto a las características-*Haar* [50] y el MB-LBP estándar [6].



(a) *Curvas ROC en la base de datos MIT-CMU*

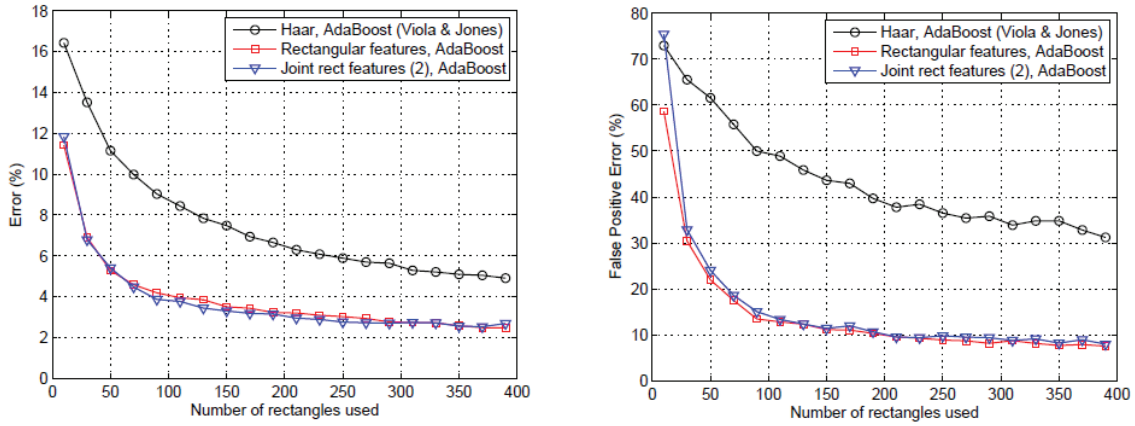


(b) *Número promedio de características usadas por posición escaneada.*

Fig. 18. Resultados de los experimentos en de la detección de rostros frontales [8].

Otro conjunto de características locales diferente, conocido como características rectangulares, combina la capacidad discriminativa de los histogramas de gradientes orientados (HoG, por sus siglas en inglés) [76] y la fuerza del operador LBP, para la detección de rostros [9]. El HoG cuenta las ocurrencias de orientación del gradiente en partes localizadas de una imagen. Para esto se utiliza una rejilla densa de los histogramas de gradientes orientados calculados en bloques de diferentes tamaños. Cada bloque está formado por un número de celdas uniformemente espaciadas. En este trabajo se considera el cambio de intensidades de los píxeles en una zona pequeña de la imagen para proporcionar una medición de los gradientes locales dentro de cada región rectangular. En el caso del HoG, el número de celdas en cada bloque se establece como uno y cada bloque puede tener diversos tamaños rectangulares. Con el objetivo de extraer de manera rápida estas características, el ángulo de gradiente se cuantifica en dos orientaciones (horizontal y vertical) y se construye un histograma que contiene ambos gradientes con y sin signo. En el caso del LBP, se obtiene un histograma de patrones binarios por cada región rectangular. De esta manera, para cada bloque rectangular, se normaliza el HoG y el histograma LBP por separado y luego se concatenan para obtener el descriptor de bloque final. Este nuevo descriptor obtenido es fácil de calcular y tiene propiedades como la tolerancia a los cambios de iluminación y la robustez al ruido de la imagen. Basados en la idea de que una sola característica no es lo suficientemente discriminativa para la clasificación y que el uso de la co-ocurrencia de características ha demostrado un mayor rendimiento [3], [4], [77], en este mismo trabajo se extiende el conjunto de características rectangulares a un conjunto más discriminativo basado en la co-ocurrencia de estas características, denominado características rectangulares conjuntas.

Los resultados obtenidos por los autores muestran que estas características no sólo superan a las características-*Haar*, sino que también obtienen mejor rendimiento cuando el entrenamiento se realiza sobre datos ruidosos [9]. Sin embargo, en cuanto al tiempo de evaluación requieren un mayor tiempo que las características-*Haar* debido a la sobrecarga en el cálculo de la imagen integral (8 imágenes integrales en memoria). En la Figura 19(a) y 19(b) se puede apreciar que las características rectangulares obtienen una menor tasa de error y de falsos positivos que las características-*Haar*.



(a) *Curvas de tasas de error en función de la cantidad de rectángulos usados.*

(b) *Curvas de falsas alarmas en la cantidad de rectángulos usadas.*

Fig. 19. Comparación entre las características-*Haar* y las características rectangulares combinadas (*Joint*) y sin combinar presentadas en [9].

Las curvas de la Figura 20 muestran que las características rectangulares superan a las características-*Haar* en todos los índices de falsos positivos. En las gráficas también se muestran la curvas correspon-

dientes a las características rectangulares conjuntas (*Joint*) que obtienen resultados similares y en algunas ocasiones mejores que las rectangulares y por ende mejores que las características-*Haar*. La cantidad de características combinadas por los autores en las características rectangulares conjuntas solo fue de dos, ya que una cantidad mayor no mejora el rendimiento más allá del alcanzado.

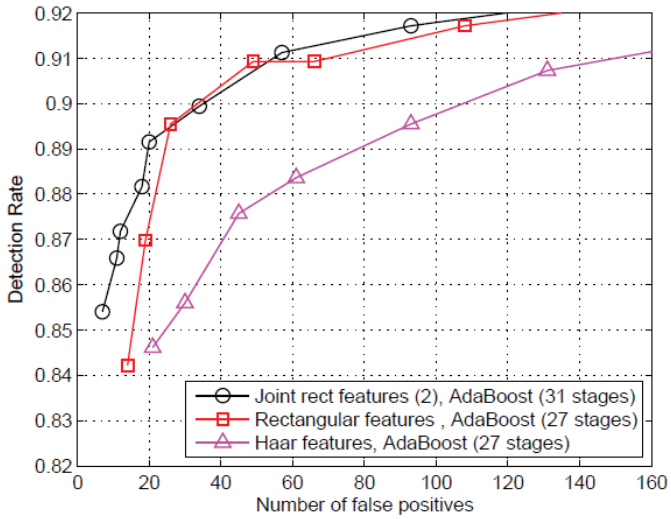


Fig. 20. Comparación de rendimiento entre las características rectangulares propuestas, las características rectangulares Joint y características-*Haar* en cascadas de fuertes clasificadores.

Una de las extensiones más recientes de los LBP se basa en la co-ocurrencia de múltiples características LBP rotacionales [77]. El conjunto conocido como Co-ocurrencia de características LBP (CoLBP) utiliza el LBP rotacional [78] con todas las posibles resoluciones en las ventanas de escaneo examinadas, con el objetivo de obtener la máxima textura posible de la ventana. El operador LBP rotacional se denota como $LBP_{P,R}$ donde P y R corresponden con el número de puntos y radio respectivamente. Los puntos se consideran como el número de puntos equidistantes que construyen el operador LBP y el radio es la distancia de los puntos al píxel central. El LBP rotacional puede ser extraído con diferentes radios y puntos en una vecindad circular. En este caso las características usadas son los valores LBP de los píxeles, a diferencia de la mayoría de las extensiones del LBP donde se usan como características los *bins* de los histogramas de los valores LBP. En este caso, la co-ocurrencia o probabilidad conjunta de los múltiples LBP rotacionales ocurridos al mismo tiempo, se seleccionan utilizando una modificación del algoritmo original SFS [64]. La extracción de las características CoLBP es computacionalmente eficiente y produce una alta tasa de rendimiento. Estas características se utilizaron para implementar un detector de rostros frontales aplicados en una secuencia de vigilancia de baja resolución [77]. Los resultados alcanzados en este trabajo prueban la capacidad de la co-ocurrencia de características para aumentar el poder discriminativo de las características LBP. Además, las características CoLBP pueden tolerar una amplio rango de iluminación y cambios borrosos. Los experimentos realizados por los autores [77] muestran que las características CoLBP superan las características-*Haar* y varias extensiones del LBP. A pesar de los resultados obtenidos en este trabajo un detalle a tener en cuenta, mencionado anteriormente, es que al usar el algoritmo SFS no se puede garantizar encontrar la mejor combinación de características; por lo que el uso de otros algoritmos de búsqueda podría mejorar los resultados.

Recientemente, medidas ordinales han sido usadas como una variante para representar estructuras locales de una imagen, al igual que los LBP. Una característica ordinal codifica una relación ordinal entre dos conceptos. Estas características son invariantes a transformaciones lineales en imágenes y flexibles para representar diferentes estructuras lineales de distinta complejidad. En la detección de rostros se ha mostrado que varias medidas ordinales en las imágenes faciales, como las que existen entre los ojos y la frente y entre la boca y la mejilla, son invariantes con diferentes personas y condiciones de imagen [79] y Schneiderman utiliza una representación ordinal para la detección de rostros [80]. Sin embargo, la característica ordinal original sólo puede mostrar la información de contraste entre dos regiones de una imagen; mientras que usando la combinación de varias medidas ordinales pueden ser representadas una mayor cantidad de estructuras locales de la imagen [81]. Un nuevo conjunto, llamado características ordinales estructuradas (SOF, del inglés *Structured Ordinal Features*), ofrece una manera más eficiente y flexible para la representación de objetos basada en apariencia [81]. De manera similar a las características MB-LBP, las SOF codifican a través de una cadena binaria la combinación de 8 bloques ordinales contenidos en círculo de manera simétrica. Los valores escalares de los promedios de los bloques también se pueden calcular de manera muy rápida mediante la imagen integral. Adicionalmente, para construir una representación de la imagen más completa se extiende el conjunto SOF a un conjunto multiescala SOF (MSOF, por sus siglas en inglés) que codifica tanto la microestructura como la macroestructura de los patrones de la imagen. En sus experimentos muestran que esta nueva representación es más eficiente que los LBP.

Otro conjunto prometedor que podría ser empleado en la detección de rostros se basa en una modificación hecha recientemente del LBP con las características-*Haar* denominado *Structured Local Binary Haar Pattern* (SLBHP, por sus siglas en inglés) [82]. Las SLBHP, propuestas originalmente para la recuperación de gráficos, adoptan cuatro tipos de características-*Haar*, que capturan los cambios de valores de gris a lo largo de la dirección horizontal, la dirección vertical y las diagonales. En las SLBHP solo se involucra la polaridad de las características-*Haar*, descaratándose la magnitud. A diferencia de las características-*Haar* tradicionales, en este enfoque los rectángulos se superponen con un píxel. Inspirados en el LBP y el hecho de que una sola característica-*Haar* binaria no tiene el suficiente poder discriminativo, los autores combinan esta característica-*Haar* binaria al igual que en el LBP. En la Figura 21 se muestra un ejemplo del proceso de obtención de una característica SLBHP. Resultados experimentales en la recuperación de gráficos mostraron que el poder discriminativo de las SLBHP es mejor que las característica-*Haar* y los LBP, incluso en condiciones ruidosas.

El surgimiento de los métodos de extracción de características locales invariantes como *Scale Invariant Feature Transform* (SIFT) [83], *Speeded-Up Robust Features* (SURF) [84], Harris-Affine [85], [86], *Hessian-Affine* [85], [86] y MSER (*maximally stable extremal regions*) [87] implicó un gran avance en el reconocimiento de objetos específicos. Estos métodos son capaces de encontrar estructuras locales que estarán presentes en distintas vistas de la imagen, además de que permiten obtener una descripción de dichas estructuras en gran medida invariante a cambios en la imagen como traslación, rotación, escala, deformaciones afines, iluminación y punto de vista. Estas propiedades de invariabilidad junto a los buenos resultados obtenidos en el reconocimiento de objetos específicos, han motivado, durante la última década, el uso de estos métodos en la categorization y detección de objetos [88], [89]; incluyendo la detección de rostros y personas. Los métodos propuestos en esta área han tratado el problema de la detección de objetos de manera general, por lo que no toman en cuenta las características específicas de cada objeto. Las potencialidades de los métodos de extracción de características locales invariantes no han sido totalmente explotadas en una categoría específica de objetos (e.g los rostros); por tanto su extensión y combinación con otras características más discriminativas podría dar muy buenos resultados en la detección de rostros, así como en otras areas de interés nuestro como el seguimiento.

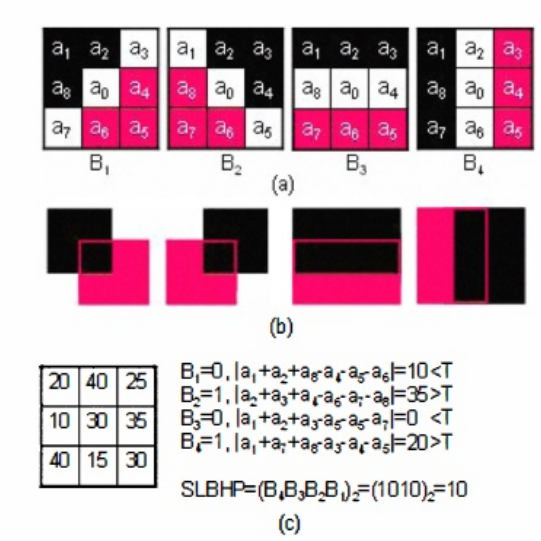


Fig. 21. Un ejemplo del proceso de obtención de una característica SLBHP. (a) Cuatro tipos características-*Haar* usadas, (b) superposición de las características-*Haar* representadas en (a), (c) ejemplo para calcular los valores de SLBHP utilizando estas características con el umbral T = 15.

Muchas extensiones o variaciones se han propuesto tanto de las características-*Haar* como de LBP, tanto para la tarea de la detección de rostros como para otras tareas. El uso de la representaciones del rostro de manera local mediante los LBP representa el estado del arte en la detección de rostros. Esto se debe principalmente a su robustez ante los cambios en las condiciones de iluminación, factor común en entornos no controlados. Conjuntos de características como SOF-MSOF, SLBHP y características locales invariantes (SIFT, SURF, Harris-Affine, Hessian-Affine, MSER, etc.) han mostrado resultados superiores a los LBP en la tarea de la representación. A pesar de que estos resultados no han sido obtenidos en el proceso de detección de rostros, una línea de investigación tentadora sería la prueba de estos conjuntos de características para representar los rostros en este proceso.

La Tabla 5 se muestra, a partir del análisis hecho, un resumen de los principales conjuntos de características propuestos en la literatura para la detección de objetos, en específico la detección de rostros, los cuales se basan principalmente en las características Haar y LBP.

Clasificadores

Como se mencionó, además del tipo de características a utilizar, el otro aspecto importante a tener en cuenta para la detección de rostros es el clasificador o algoritmo de aprendizaje a emplear. Dado que la detección de rostros puede ser vista como un problema de clasificación binaria (rostro/no rostro), a continuación se analizan los clasificadores de dos clases que han sido usados en esta tarea. Motivado por lo complejo que resulta modelar y tener un conjunto representativo de la clase ”no rostro”(dada su variedad), se analizan también los clasificadores de una clase.

Clasificadores de una clase

Los clasificadores de una clase (OCC, del inglés *one-class classifier*) como descriptores de clase son capaces de aprender dominios restringidos en un espacio de patrones multi-dimensional. Estos clasificadores tratan de distinguir una clase de objetos (clase objetivo, en nuestro caso rostros) de todos los demás objetos posibles (valores atípicos), mediante el entrenamiento de un conjunto que sólo contiene muestras de

Tabla 5. Conjuntos de características para la detección de rostros/objetos.

Tipo de Características	**Trabajos representativos**
Basadas en características-*Haar*	Características-*Haar* [50] Características-*Haar* rotadas [63] Características-*Joint Haar* [3] Características-*Variance Haar* [5] Características-LAB [4] Características-iFAHF [1]
Basadas en los patrones binarios locales (LBP)	Características MB-LBP [6] Características INTLBP [70] Características DLBP [71] Características CoLBP [77]
Combinadas	Características-*HaarLBP* [7] Características rectangulares [9]

la clase objetivo. La descripción del modelo de esta clase debe ser lo suficientemente flexible para aceptar la mayor cantidad de nuevos objetivos, pero a la vez lo suficientemente discriminante para rechazar la mayoría de los valores atípicos. Por esta razón, este tipo de clasificadores se debe utilizar cuando se ha logrado modelar la clase objetivo distinguiéndola de todos los posibles ejemplos. Usualmente, los clasificadores de una clase asumen que durante el entrenamiento solo se tienen muestras de la clase objetivo, aunque una característica importante es su robustez frente a la existencia de unos pocos valores atípicos en los datos de entrenamiento. En los OCC también aparecen los problemas que se encuentran en los problemas de clasificación convencionales, tales como la estimación del error de clasificación, la medición de la complejidad de una solución, la maldición de la dimensionalidad, la generalización del método de clasificación.

Varios métodos han sido propuestos para resolver el problema de la clasificación basada en una clase. Estos métodos pueden ser agrupados en tres enfoques principales: métodos de densidad, métodos de frontera y métodos de reconstrucción [90]. El enfoque más sencillo para obtener un clasificador de una sola clase es estimar la densidad probabilística de las muestras de entrenamiento y establecer un umbral en dicha densidad [91]. Para esto, varias distribuciones pueden ser asumidas, como la distribución Gaussiana o la distribución de Poisson. La principal desventaja de estos métodos es la difícil estimación de densidades, especialmente cuando se cuenta con una cantidad limitada de datos. Por otro lado, cuando se tiene un número suficiente de objetos se puede lograr un buen desempeño. En los métodos de frontera sólo es optimizado el contorno cerrado alrededor del conjunto de la clase objetivo. Estos métodos aunque requieren un menor número de muestras que los métodos de densidad, dependen en gran medida de las distancias entre las muestras, lo que los hace sensibles a la escala de las características. Por tal razón estos métodos le otorga una gran importancia a la definición de estas distancias. Por último, los métodos de reconstrucción no han sido desarrollados principalmente para la clasificación de una clase, sino más bien para la modelación de los datos. Estos métodos escogen un modelo y lo ajustan a los datos mediante el uso de un conocimiento previo sobre los mismos y suposiciones acerca de las características de la agrupación de los datos o de su distribución en subespacios. Una descripción más profunda de estos enfoques puede encontrase en la tesis de doctorado de David Tax del 2001 [90].

La necesidad de utilizar la clasificación basada en una clase viene de muchas aplicaciones prácticas como la detección de fallos, la detección de objetos de interés en imágenes, la detección de anomalías, la detección de enfermedades, la identificación de personas, entre otras [90]. Los clasificadores de una clase pueden ayudar a disminuir las diferencias actuales que existen entre los supuestos teóricos bajo los cuales

son diseñados los clasificadores estadísticos (usualmente suponen datos con distribuciones estacionarias, bien definidas, buen muestreo) y las características de los datos que se obtienen en condiciones reales (los cuales contienen ruido, datos faltantes, valores atípicos y son generados bajo distribuciones atípicas y no estacionarias).

Entre los principales algoritmos OCC existentes, uno de los más utilizados en la detección de rostros son los basados en las máquinas de vectores soporte de una clase (OSVMs). Los OSVMs son adaptaciones de las máquinas de vectores soporte estándar originalmente introducidas para problemas de clasificación de dos clases [92]. Las máquinas de vectores soporte (SVM) son algoritmos de aprendizaje basados en funciones núcleos que, dado un conjunto de entrenamiento, realizan la proyección de los datos a un espacio de características de mayor dimensión en el cual determinan el hiperplano óptimo de separación de las clases. Este hiperplano es aquel que maximiza la distancia (margen) entre dos hiperplanos definidos por los vectores soporte [92]. Un elemento clave en la formulación del SVM son las funciones núcleos, ya que permiten la construcción del hiperplano de separación óptimo en el espacio de mayor dimensión sin realizar explícitamente los cálculos en dicho espacio. Además, el problema de optimización a resolver es un problema cuadrático convexo [93] por lo que la solución que se obtiene es única y óptima (no óptimos locales).

Uno de los clasificadores de una clase basado en las OSVM más utilizado en la detección de rostros es el conocido *como support vector data description* (SVDD) [90]. La idea básica de esta extensión es encontrar en el espacio de características una hiper-esfera (descrita por un centro a y un radio R) con un radio mínimo, que contenga todas (o la mayoría de) las muestras de la clase objetivo. Al igual que en las SVM, las funciones núcleos (*kernels*) son utilizadas para lograr descripciones más flexibles a través de modelaciones no lineales.

Basado en este clasificador, fue propuesto un nuevo método de detección de rostros llamado máquinas de vectores soporte de una clase (OCSVM, del inglés one class support vector machines) [94]. En el proceso de aprendizaje fue utilizado el método de análisis de componentes principales (PCA) [95] con el objetivo de transformar el espacio original de características a un subespacio de baja dimensión. En el proceso de detección de rostros la imagen es dividida en 9 bloques iguales. Luego se usan 18 reglas para predecir por mayoría de votos las sub-ventanas con rostros candidatos. Una vez obtenida las posibles regiones de rostros el clasificador OCSVM es aplicado. Los resultados obtenidos en este trabajo mostraron que es posible resolver el problema de la detección de rostros solo con patrones de una clase. Sin embargo, factores como la selección de la función núcleo y sus parámetros, así como la estrategia para reducir la dimensionalidad del espacio de características influyen en el rendimiento del método propuesto. La optimización de estos factores ayudaría a minimizar el error generalizado del detector propuesto.

Otro trabajo que trata el problema de la detección de rostros frontales con el clasificador SVDD se basa en las representaciones de las muestras de entrenamiento utilizando PCA (*eigenfaces*), para obtener una frontera de decisión en torno a los datos sin necesidad de utilizar la información de los ejemplos negativos [96]. El rendimiento del SVDD fue comparado con los límites elípticos obtenidos con el clasificador de distancia en el espacio de características (DIFS). DIFS define, en un subespacio F, elipses concéntricas de puntos que son equidistantes de la media de la muestra (utilizando la distancia de Mahalanobis). Aplicando un umbral en esta distancia, se obtiene un contorno elíptico de la clase rostro. También fue analizada la influencia de la dimensionalidad del espacio de características en el desempeño del SVDD. En altas dimensiones, la mayoría de los puntos se convierten en vectores de soporte y esto sugiere que se requieran más datos de entrenamiento para encontrar una estimación confiable de la frontera. Los resultados experimentales [96] muestran que cuando los datos de entrenamiento es una muestra representativa de la clase objetivo el clasificador SVDD encuentra una descripción de límites más flexible y supera al DIFS. Una desventaja del método propuesto en este trabajo es el conjunto de características usado (*eigenfaces*).

Aunque reducen la dimensionalidad del espacio de características original, existen varios factores que afectan la robustez y eficacia de estas características. El principal problema es que para obtener un buen resultado necesitan que las imágenes estén perfectamente alineadas. Además, solo es capaz de aprender las variaciones intra-clase; por lo que se ve afectado ante cualquier pequeño cambio de apariencia ya sea por la iluminación, la pose, la oclusión e incluso las expresiones faciales. Por ultimo existe un considerable esfuerzo computacional involucrado en la generación de valores y vectores propios de las matrices de covarianza. Por lo tanto, una mejora a este trabajo podría ser el uso de conjuntos de características más robustos, preferiblemente de características locales como algunos de los analizados anteriormente.

Clasificadores de dos clases

El problema de la detección de rostros utilizando los clasificadores de dos clases puede ser tratado como un problema de clasificación binaria, donde una clase representa los ejemplos positivos (en este caso el rostro) y la otra los ejemplos negativos (todo lo que no es rostro). De esta forma, el conjunto de entrenamiento está formado por imágenes tanto de rostros como de otros objetos. Por esta razón, el conjunto de datos a colectar es mucho mayor y el universo que ocupa la clase no rostro, al ser muy amplio, se hace más difícil de abarcar, por lo que el problema se vuelve más complejo. No obstante, el empleo de este tipo de clasificadores en la tarea de detección de rostros ha obtenido un gran auge. Algunos de los clasificadores de este tipo son: las redes neuronales [97], [98], *eigenfaces* [99], *SNoW* [100], clasificador Bayesiano [101] y SVM [102], [103]. El uso de estos clasificadores en la tarea de detección de rostros proporciona resultados precisos con pocas falsas alarmas. Sin embargo, requieren de gran cantidad de tiempo para procesar una imagen, lo que representa una limitación para las aplicaciones en tiempo real [49].

En el 2001, Viola y Jones proponen el primer detector de rostros que alcanza una velocidad de detección en tiempo real y una alta precisión comparable con métodos previos del estado del arte [50]. Este trabajo es uno de los primeros en utilizar los esquemas *boosting* para la detección de rostros. El detector propuesto hace tres contribuciones fundamentales. La primera contribución de este trabajo es la introducción de la representación de la imagen llamada imagen integral que permite calcular muy rápido las características utilizadas. La segunda contribución es un método para construir un clasificador mediante la selección de un pequeño número de características importantes de un gran conjunto de características potenciales usando el algoritmo *AdaBoost* [62]. La tercera contribución es un método para la combinación de clasificadores en una estructura en cascada que aumenta drásticamente la velocidad del detector.

Los resultados obtenidos por este trabajo motivó el desarrollo de diferentes extensiones o modificaciones del sistema propuesto encaminadas principalmente en 2 direcciones: el desarrollo de nuevos conjuntos de características (analizados anteriormente) y nuevos métodos basados en *boosting*. A continuación se describen con más detalles algunas de las variantes pertenecientes a la última linea de mejoras, partiendo de sus orígenes.

Métodos basados en boosting

El algoritmo *boosting* es uno de los más importantes desarrollos en la metodología de clasificación, originado por Schapire en 1989 [104] y desde entonces aceptado como uno de los principales métodos en el campo de las máquinas de aprendizaje. *Boosting* es un método general que permite obtener valores para las probabilidades de detección y las falsas alarmas muy competitivos, mediante la combinación efectiva de un conjunto de reglas sencillas (o clasificadores débiles), que por si solas no consiguen un porcentaje de detección aceptable, pero que combinadas eficientemente producen un sistema de decisión altamente potente. Para esto, mediante un algoritmo denominado como *WeakLearn* se genera todo el conjunto de reglas sencillas posibles para la clase de objeto de interés. Luego, el *Boosting* prueba cada una de las reglas generadas sobre cada una de las muestras del conjunto de entrenamiento. Cada regla asigna una etiqueta a cada muestra según el resultado de la clasificación. Con esto, solo se seleccionan aquellas reglas que pre-

sentan un menor error, es decir, las que mejor representan el objeto. Así, dando mayor importancia a una reglas sobre las otras según el error obtenido, se construye el detector de objetos. Hasta este punto existen fundamentalmente dos cuestiones que se deben determinar para conseguir aplicar este algoritmo con éxito. En primer lugar, hallar un método que nos permita encontrar reglas simples en base a las muestras, y en segundo lugar, lograr combinar eficientemente las mejores reglas en una sola más potente.

En base a esto Freund y Schapire en 1995 [62] abordan una paso previo y muy importante en este proceso que es cómo determinar qué ejemplos son los más indicados para que en base a ellos, se puedan encontrar las mejores reglas débiles. Esto sirve para hacer una clasificación de todas las muestras del conjunto en más fáciles o más difíciles de clasificar o etiquetar; es decir, asignar un peso indicando su dificultad. Para esto, introducen el algoritmo *AdaBoost* con el objetivo de construir un clasificador fuerte mediante la combinación lineal ponderada de múltiples clasificadores débiles.

Supongamos que tenemos un conjunto de entrenamiento de n muestras pertenecientes al espacio de características X $Z_n = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, donde $y_i \in \{-1; +1\}$ representa la etiqueta de la clase asociada a la muestra $x_i \in \mathbb{R}^n$. Tomando como entrada el conjunto Z_n cada clasificador débil es entrenado por fases con el objetivo de minimizar el error empírico en una determinada distribución re-ponderada según los errores de clasificación del clasificador previamente entrenado. La distribución inicial de los pesos sobre las muestras del conjunto se asigna de manera uniforme. Luego, en cada iteración estos pesos son actualizados de manera tal que muestras mal clasificadas en una iteración previa tendrán asociado un mayor valor, con el objetivo de que el algoritmo se centre en los ejemplos difíciles. En este proceso cada clasificador débil es añadido de manera secuencial al *ensemble AdaBoost* y son combinados a través de una regla de combinación para obtener la clasificación final.

En el sistema de Viola y Jones el algoritmo *AdaBoost* fue utilizado tanto para seleccionar las características como para entrenar el clasificador [50]. En este trabajo el algoritmo *AdaBoost* es adaptado para la solución de tres problemas fundamentales en un procedimiento *boosting*: 1) el aprendizaje de forma incremental de las características fundamentales de un gran conjunto de características, 2) la construcción de clasificadores débiles basados en una sola características seleccionada, y 3) la construcción de un clasificador fuerte ($H(x)$) a partir de la combinación lineal de clasificadores débiles ponderados durante el proceso de entrenamiento. Viola y Jones dieron la regla de combinación como:

$$H(x) = \text{sign}\left[\sum_{t=1}^{T} \alpha_t h_t(x)\right] = \begin{cases} +1 & \text{si } \sum_{t=1}^{T} \alpha_t h_t(x) > 0 \\ \\ -1 & \text{en caso contrario,} \end{cases} \tag{5}$$

donde $H(x)$ es el clasificador fuerte resultante, $h_t(x)$ las reglas o clasificadores débiles combinados y α el peso correspondiente a cada una de estas reglas. De ahí que la entrada x es clasificada como rostro si $H(x) = +1$ y como no-rostro si $H(x) = -1$. De ahí que los clasificadores débiles pueden ser vistos como un voto ponderado y el clasificador fuerte como el voto ponderado de la mayoría. De manera general este algoritmo es fácil de implementar y tiene una generalización bastante buena, aunque es sensible al ruido en los datos y los valores atípicos.

Numerosas variantes del algoritmo *AdaBoost* original (conocido también como *Discrete AdaBoost*) han sido propuestas y aplicadas en la tarea de la detección de rostros. Dos variantes muy conocidas son el *Real AdaBoost* [105] y el *Gentle AdaBoost* [106]. Estas variantes son idénticas en cuanto a complejidad computacional desde la perspectiva de la clasificación, pero se diferencian fundamentalmente en la forma de combinar las reglas débiles en una sola.

El *Real AdaBoost* es una versión generalizada del *Discrete AdaBoost*, el cual utiliza los niveles de confianza de cada clasificador débil en lugar de las salidas binarias solamente. A diferencia del *Discrete AdaBoost*, el espacio de salida de los clasificadores débiles en este caso no está restringido a las etiquetas

$-1,+1$, sino que puede tomar valores reales en el espacio $\mathbb{R}$. Para el diseño de los clasificadores débiles los autores proponen realizar las predicciones sobre la base de una partición del dominio X. Más específicamente, cada uno de los clasificadores débiles, se asocia con una partición de X en bloques disjuntos $X_1, X_2, ..., X_N$, que cubren todo del espacio X. Para la división del espacio se hace uso del método de intervalos de igual ancho, mediante el cual se agrupan el número de particiones seleccionadas de manera arbitraria [105]; por lo que con este método es difícil predecir el número adecuado de particiones de antemano.

Por su parte el *Gentle AdaBoost* es una versión modificada del *Real AdaBoost* que pone menos énfasis en los valores extremos [106]. La principal diferencia entre estos dos algoritmos es la forma en que actualizan los pesos de los clasificadores débiles. En el caso del *Real AdaBoost* se utiliza el logaritmo de la razón de las probabilidades de pertenencia a las clases, mientras que en el *Gentle AdaBoost* solo se emplean las restas de estas probabilidades. Esta idea se debe a que la variante empleada en el *AdaBoost* real puede ser numéricamente inestable, dando lugar a valores de actualización muy grandes, mientras que la actualización de GAB se encuentra en el rango [-1,1].

Un análisis empírico de estos tres algoritmos demostró que el *Gentle AdaBoost* supera tanto *Discrete AdaBoost* y *Real AdaBoost* en la tarea de detección de rostros [107]. Sin embargo, el *Real AdaBoost* genera clasificadores que requieren menos cálculos de características y por tanto, es más eficiente.

Con el objetivo de resolver la limitación del algoritmo *Real AdaBoost* a la hora de estimar el número adecuado de particiones del espacio de características, fue propuesto un nuevo esquema *boosting* que utiliza un método de discretización basado en la entropía [108] para dividir el espacio de entrada en subespacios [10]. El esquema propuesto conocido como *Ent-Boost* permite estimar el número óptimo de particiones de manera automática, ya que a través de la medida de entropía, toma en cuenta la información de clase y la distribución de los datos de entrada en el proceso de división. Para seleccionar el mejor clasificador débil del conjunto de entrada, se elige aquel que maximice la divergencia *Kullback-Leibler* (KL) simétrica [109] entre dos distribuciones de muestras positivas y negativas. Así, cada clasificador débil es construido y entrenado en un conjunto de muestras de entrenamiento ponderadas de manera similar al *Real AdaBoost*. En los experimentos realizados, los autores compararon el rendimiento de clasificadores fuertes entrenados mediante diferentes esquemas boosting, como son *DiscreteAdaBoost*, *Real AdaBoost* y *Ent-Boost*. En la Figura 22 se muestran los resultados obtenidos de la comparación.

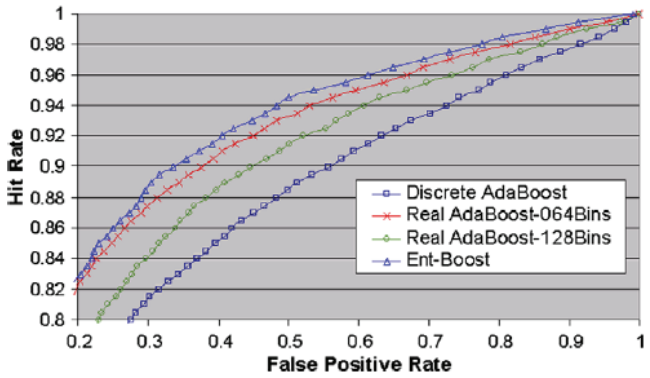


Fig. 22. Comparación de varios esquemas *boosting* [10].

En el caso del *Real AdaBoost* el número de particiones seleccionadas de forma arbitraria es 64 y 128. Las curvas mostradas en la gráfica indican que tanto el rendimiento del *Real AdaBoost* como del *Ent-Boost*

es superior al del *Diacrete AdaBoost*. En cuanto al espacio de almacenamiento, el clasificador basado *Ent-Boost* sólo utiliza 6,79 particiones como promedio, que es mucho menor que el número utilizado por los clasificadores basados en el *Real AdaBoost*. De manera general, el clasificador fuerte basado en el *Ent-Boost* logra un mejor desempeño y un almacenamiento más compacto.

Debido a que en el *AdaBoost* los clasificadores débiles se agregan de manera secuencial, no se garantiza que durante el entrenamiento se seleccione la combinación óptima de los clasificadores débiles. Otra variante del algoritmo *AdaBoost*, conocida como *FloatBoost*, trata de resolver esta limitación incorporando la idea de la búsqueda flotante [110] en el algoritmo *AdaBoost* [111]. El uso de la búsqueda flotante de manera iterativa permite no sólo agregar características durante el entrenamiento, sino que también examina las ya seleccionadas con el objetivo de eliminar las menos significativas. Esto logra que clasificadores débiles que no sean efectivos en términos de tasa de error puedan ser eliminados. Sin embargo, *Float-Boost* a pesar de mostrar mejores resultados que el *AdaBoost* requiere un mayor tiempo de entrenamiento (5 veces más), es inestable y computacionalmente costoso ante problemas de aprendizaje complicados. En la Figura 23 y 24 se muestran los resultados obtenidos en la comparación del *FloatBoost* y *AdaBoost*. Como puede observarse en ambas figuras el algoritmo *FloatBoost* obtiene las mejores tasas de detección y produce un clasificador más fuerte con menos clasificadores débiles logrando bajar los índices de error de las falsas alarmas.

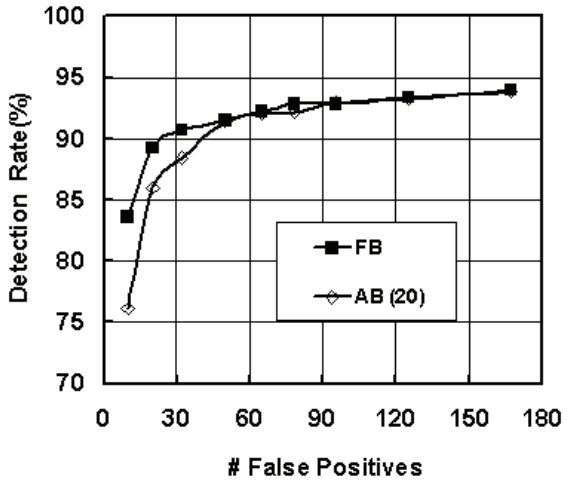


Fig. 23. Comparación de las tasas de detección de los métodos FloatBoost y AdaBoost en el conjunto de prueba MIT + CMU.

Sobre la base del algoritmo *Adaboost* fue desarrollada una extensión del sequential probability ratio test (SPRT) [112] para datos no independientes e idénticamente distribuidos [75]. El algoritmo propuesto, denominado *WaldBoost*, estima la razón de verosimilitudes de cada una de las clases a través de estimaciones de las densidades condicionales. Estas densidades se estiman usando el método de *Parzen* [113] y un conjunto de validación externo (no el conjunto de entrenamiento). Este enfoque fue aplicado y evaluado en el problema de detección de rostros en el base de datos MIT+CMU. Los resultados alcanzados en este trabajo resultaron ser superiores a los del estado del arte en cuanto tiempo de evaluación promedio y comparable en tasas de detección. El único método que superó al algoritmo *WaldBoost* en la calidad de la detección fue el propuesto en el 2004 por Bo Wu [114]. Los autores plantean que esto pudo haber sido causado por el uso de características diferentes, subconjuntos de datos de la base de datos MIT+CMU o

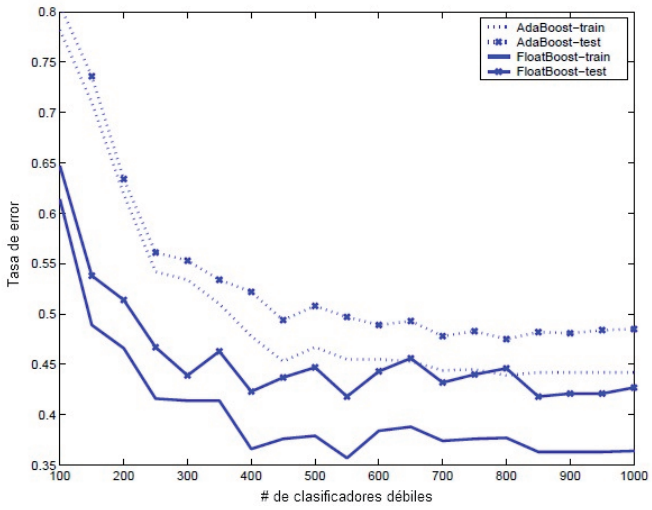


Fig. 24. Tasa de error de falsas alarmas de los algoritmos *FloatBoost* y *AdaBoost* en conjuntos de entrenamiento y prueba de rostros frontales en función del número de clasificadores débiles.

algún otro detalle de la implementación. En la gráfica de la Figura 25 se muestran los resultados obtenidos en sus experimentos, donde el algoritmo propuesto obtiene mayor rendimiento en comparación con el *AdaBoost* y el *FloatBoost*.

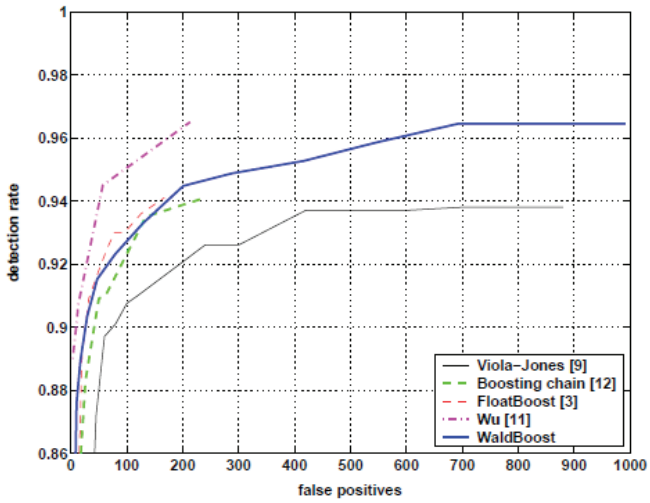


Fig. 25. Comparación de las curvas ROC del algoritmo *WaldBoost* con los métodos del estado del arte.

Los métodos basados en *boosting*, de manera general, muestran muy buenos resultados en la detección de rostros tanto en términos de precisión como de velocidad y son muy adecuados para aplicaciones en tiempo real. Sin embargo, estos métodos para la clasificación de rostros requieren muchas muestras negativas que traten de representar el resto del universo (no-rostro); por lo que consumen un gran cantidad de tiempo en la etapa de entrenamiento y pueden tardar días en la etapa de aprendizaje. Además, son

sensibles al problema de la oclusión parcial y los cambios de iluminación en las imágenes de rostros y el diseño óptimo de arquitecturas en cascadas es muy complejo. Una descripción más detallada de estos trabajos, así como otras propuestas se encuentra en un reporte técnico publicado recientemente [41].

3.1.5 Métodos basados en la información contextual

Experimentos psicofísico realizados han mostrado que el contexto es una señal crucial en el proceso de detección de objetos [115]. Por esta razón, enfoques más recientes aluden el uso de la información contextual como una dirección prometedora tanto para la localización como para la detección de rostros [41]. El conocimiento del contexto se puede definir como cualquier información que no es producida directamente por la apariencia de un objeto y puede ser obtenido a partir de los datos de una imagen cercana en una secuencia, de etiquetas de imágenes o de las anotaciones; así como por la presencia y ubicación de otros objetos.

Las características del contexto se pueden agrupar en tres categorías: contexto semántico (probabilidad), contexto espacial (posición) y contexto escala (tamaño) [116]. El contexto semántico corresponde a la probabilidad de que un objeto se encuentra en algunas escenas pero en otras no. El contexto espacial puede ser definido a través de la probabilidad de encontrar un objeto en una posición y no en otras, con respecto a otros objetos en la escena. Por último, el contexto escala brinda la relación contextual basada en la escalas de un objeto con respecto a los demás; estableciendo que existe un conjunto limitado de relaciones de tamaño entre los objetos en la escena. En las tareas de clasificación, el contexto espacial y escala son los más usados, ya que involucran el uso de todas las formas de información contextual en la escena.

En las imágenes, el uso de la información del contexto puede ser considerado a nivel global o local. En el contexto global se consideran las estadísticas de la imagen como un conjunto, donde se tienen interacciones contextuales entre objetos y escenas. Por el contrario, en el contexto local se considera la información del contexto de las áreas vecinas del objeto y la interacción contextual se puede agrupar en tres tipos diferentes: interacciones de píxeles, de regiones y de objetos. El problema de la integración de la información contextual en un marco de clasificación de objetos es una tarea difícil, ya que debe combinar la información de la apariencia de los objetos con las restricciones contextuales impuestas a los objetos de la escena dada.

Con el fin de utilizar la información contextual en la detección de rostros, fue propuesto un detector que utiliza activamente el contexto local como una señal de predicción para ganar robustez con respecto a los clasificadores tradicionales que se centran en el propio objeto [117]. Para esto, el detector empleado utiliza una cascada de clasificadores *boosting* [50] junto al conjunto de características-*Haar* ampliado [63]; donde las imágenes del conjunto de entrenamiento contienen además del rostro, el cuello y la parte superior del cuerpo. Durante la detección, la ubicación real del rostro se infiere asumiendo una posición fija dentro de la ventana de detección. Luego, el tamaño y la ubicación del mismo son directamente calculados a partir del ancho y la altura del contexto local detectado. Los resultados obtenidos en este trabajo mostraron que el uso del contexto local incrementa significativamente el número de de detecciones correctas. Los experimentos realizados por los autores evidencian que el enfoque propuesto es robusto ante variaciones en la pose y puede operar a bajas resoluciones; lo que brinda una mayor rapidez en el proceso de búsqueda.

Otro trabajo que evidencia la importancia del contexto en la localización de objetos propone un método que integra la información del contexto global y local en clasificadores basados en funciones núcleo (*kernels*) [118]. Como parte del proceso de entrenamiento se aprende de manera automática y eficiente la importancia relativa de las contribuciones de los diferentes contextos. Para esto combinan diferentes modelos del contexto en un solo clasificador discriminativo a través de *kernels* de contexto local y global. Luego el resultado obtenido se integra con la reciente propuesta de búsqueda de sub-ventana eficiente

(ESS) para la localización de objetos [54],[55], que permite una evaluación extremadamente eficiente. Entonces, un *kernel* de contexto local se define como un *kernel* para imagen en la región alrededor del objeto de interés; donde la extensión espacial de esta región define la cantidad de contexto local para su uso. Mientras que un *kernel* de contexto global es aquel que incorpora la información de toda la imagen, pero que no depende de las coordenadas del cuadro delimitador. Al especificar el contexto directamente con un enfoque de aprendizaje de funciones núcleo, se consigue una alta precisión de localización con una representación sencilla y eficaz. Dada la flexibilidad de los *kernels* locales y globales los autores recomiendan varias líneas futuras de trabajo. Se conoce que el incremento del número de datos de entrenamiento y tipos de características a combinar, mejora el rendimiento de los sistemas de localización de objetos. Por lo que, los autores suponen que el uso de más de dos kernels de contexto puede mejorar el rendimiento de la localización. Adicionalmente, plantean que la extensión espacial de las regiones que contienen la información de contexto es, con una alta probabilidad, dependiente de la clase, lo cual sugiere que pueda ser incluido en el procedimiento de aprendizaje.

Una nueva técnica de post-procesamiento basada en el contexto fue propuesta para la detección de objetos [119]. En este trabajo el contexto es definido como la distribución de la detección alrededor de una ventana candidata, mediante la variación de su escala y posición, y comprobando con un clasificador. El contexto se describe usando múltiples características como su densidad, la distribución geométrica de los ejes de escala y posición, y algunas estadísticas al respecto. De esta manera el modelo creado selecciona automáticamente las mejores características de la información contextual y optimiza sus parámetros internos. El primer paso del enfoque propuesto es escanear la imagen de entrada a través de la técnica de desplazamiento de ventana y usando un clasificador de rostro. Luego, las detecciones obtenidas se comparan con el modelo basado en el contexto para eliminar las falsas alarmas y una nueva colección de sub-ventanas es generada; así de manera repetida hasta converger al refinamiento deseado. Por último, las detecciones que convergen cerca son fusionadas utilizando el algoritmo *Mean Shift* Adaptado (AMS, del inglés *Adaptive Mean Shift*) [120]-[121], con el objetivo de eliminar múltiples detecciones sobre una misma instancia. Con este trabajo, los autores han mostrado que mediante la utilización de la información contextual es posible construir un modelo no paramétrico simple que aprende a distinguir entre falsas alarmas y detecciones verdaderas con gran precisión, generando una detección más precisa.

A modo de resumen, dentro de los métodos que se basan en la apariencia se analizaron tres cuestiones importantes para el proceso de la detección de rostros como son la técnica de búsqueda de todos los posibles rostros en la imagen, las características o rasgos a extraer y el clasificador o algoritmo de aprendizaje a aplicar.

En cuanto a las técnicas de búsqueda revisadas, el desplazamiento de ventanas es una de las más utilizadas en la detección de rostros. Sin embargo, esta técnica es computacionalmente muy costosa y genera un gran numero de detecciones falsas que generalmente necesitan ser procesadas. En la detección de objetos en general se han propuesto otras técnicas más eficientes que podrían ser extendidas al caso específico de los rostros, teniendo en cuenta los requerimientos de cada método para el caso específico.

Respecto a los conjuntos de características existentes, se pudo observar que las características-*Haar* y los LBP, así como sus extensiones, representan el estado del arte para una amplia gama de problemas de detección de rostros; teniendo en cuenta que la combinación de características es una opción más discriminativa que el uso de una sola característica en el aprendizaje. Además, se analizó el posible uso de otras características como las SOF-MSOF, SLBHP y características locales invariantes, que aunque no se han empleado directamente en la tarea de detección han mostrado buenos resultados e incluso superiores a los LBP y las características-Haar.

Por otra parte, los clasificadores analizados fueron divididos en dos grupos: los basados en una clase y los basados en dos clases. Entre estos dos grupos los clasificadores de dos clases basados en esquemas

boosting han dominado los avances más recientes en términos de precisión como de velocidad y además, son muy adecuados para aplicaciones en tiempo real. Sin embargo, estos métodos para la clasificación de rostros requieren muchas muestras negativas que traten de representar el resto del universo (no-rostro); por lo que consumen un gran cantidad de tiempo en la etapa de entrenamiento y pueden tardar días en la etapa de aprendizaje. Por lo que el uso de los clasificadores de una clase podría ser explotado y quizás obtener resultados óptimos con menor costo computacional.

De manera general los métodos basados en la apariencia junto al uso de la información del contexto han permitido un gran avance en la detección de rostros. Sin embargo, por supuesto que no es el final de la investigación en la detección de rostros sobre todo en términos de precision de la clasificación, velocidad y costo computacional, especialmente en situaciones complicadas donde influyen factores críticos como la pose y la resolución. Incluso la optimización de la búsqueda de los rostros en la imagen aun exige más investigación.

4 Seguimiento de rostros

El seguimiento visual de objetos es un área que permite explotar la correspondencia temporal existente entre los cuadros de un video. El objetivo de esta tarea es seguir la localización de uno o varios objeto en cada cuadro de una secuencia de video. El seguimiento visual de objetos de interés, específicamente de rostros, ha recibido gran atención dentro del área de la visión por computadoras y el reconocimiento de patrones, debido a su importancia en aplicaciones prácticas como la video-vigilancia, las video-conferencias, entre otras. Después de la localización de los rostros en los cuadros de un video, el seguimiento de estos es una tarea difícil pero que permite extraer un conjunto de características relevantes que posteriormente pueden servir como entrada al clasificador de rostro. Por lo tanto, se puede decir que la eficacia del seguimiento depende de la exactitud de la detección, así como la precisión del seguimiento influye directamente en la capacidad de reconocer sujetos en video.

Un algoritmo de seguimiento de rostros ideal debe cumplir un conjunto de requisitos. Debe ser robusto a la rotación del rostro y los cambios de ambiente como la iluminación. Debe ser lo suficientemente rápido como para ser empleado en aplicaciones en tiempo real. Por último, debe proporcionar información detallada no solo sobre el movimiento del rostro sino también del cambio de la forma y la textura [122]. Existen diferentes factores que perjudican la eficacia de los algoritmos de seguimiento. Algunos de estos factores son: la pérdida de información causada por la proyección del mundo 3D en una imagen 2D, las variaciones de la apariencia, la oclusión, los cambios de iluminación de la escena, los posibles movimientos y formas complejas de los objetos, su naturaleza no rígida o articulada, la entrada y salida de estos en las escenas, así como los requisitos del procesamiento en tiempo real [123].

A continuación se analizan los principales enfoques para el seguimiento de rostros. Posteriormente se hace una revisión de los trabajos más importantes reportados recientemente en la literatura para el seguimiento de rostros en video.

4.1 Principales enfoques para el seguimiento de rostros

El objetivo del seguimiento de rostros es obtener la trayectoria de uno o varios rostros sobre el tiempo mediante la localización de la posición de estos en cada cuadro de la secuencia de video. Un factor adicional a enfrentar en el seguimiento de múltiples rostros es que estos rostros pueden ocluir uno al otro, dependiendo así de varios factores, tales como cuán similares son los rostros, cuánto tiempo dura la oclusión, y en qué por ciento un rostro está ocluido por otro.

Para llevar a cabo el seguimiento de rostros en una secuencia de video existen dos enfoques fundamentales. El primer enfoque se basa en la detección de los posibles rostros en cada cuadro de la secuencia usando un método de detección de rostros (ver sección 3) y luego, mediante un algoritmo de seguimiento se realizan las correspondencias de los rostros detectados a través de los cuadros. Este enfoque no tiene en cuenta el comportamiento de los objetos durante su trayectoria en la secuencia. Por el contrario, en el segundo enfoque se realiza la detección de rostros en el primer cuadro y seguidamente estos son seguidos a lo largo de toda la secuencia de video. En ambos enfoques, los rostros pueden ser representados mediante la forma y/o la textura.

Varios métodos han sido propuestos para el seguimiento de rostros basados en las diferentes representaciones del rostro [122], [124]. Señales o características como el color de la piel, el movimiento, los rasgos faciales, la intensidad de los píxeles, los bordes, la forma, la apariencia, entre otras, así como la combinación de ellas, han sido empleadas por estos métodos. En la Figura 26 se muestra una taxonomía en la que se dividen los métodos de seguimiento según las representaciones más significativas que han sido empleadas para este objetivo.

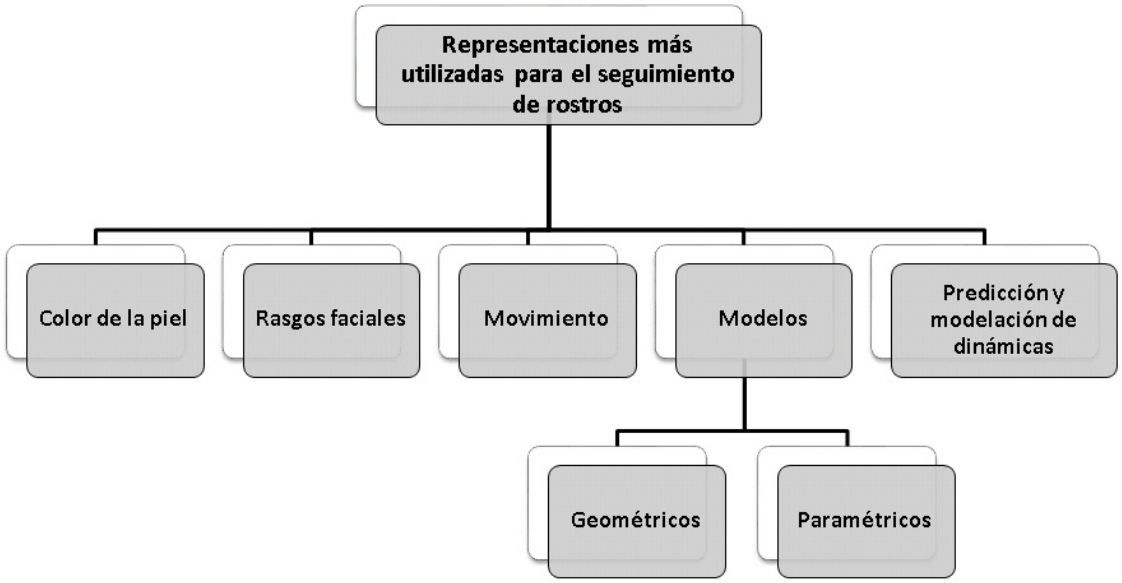


Fig. 26. Taxonomía propuesta para los métodos de seguimiento de rostros en videos.

Entre todas las señales utilizadas para llevar a cabo el seguimiento de rostros, el color es una de las más usadas [123]. Los métodos que se basan en el color de la piel son fáciles de implementar, rápidos y de bajo costo computacional. Sin embargo, estos métodos fallan cuando las condiciones de iluminación cambian drásticamente o cuando el fondo de la imagen tienen un color similar al de la piel del rostro. Incluso, si en la imagen existen regiones con colores parecidos al de la piel, como las manos u otras partes del cuerpo, entonces estas regiones inevitablemente son identificadas erróneamente confundiéndolas con rostros.

El seguimiento de los rasgos faciales es otro de los enfoques que permite seguimiento de rostros. Para el seguimiento de las características existen dos pasos importantes. El primero es decidir qué características seguir y el segundo es el seguimiento en sí mismo. Una de las primeras investigaciones del seguimiento de la cabeza con rasgos faciales se basó en el seguimiento de las esquinas de los ojos y la boca. Luego, otras características faciales tales como el iris, la frente, las mejillas y características transitorias han sido utilizadas también para el seguimiento de rostros [124]. El algoritmo Lucas-Kanade (LK) [125] y las variantes surgidas a partir de este como Kanade-Lucas-Tomasi (KLT) [126] y sus extensiones [127], [128], [129], [130] son unos de los métodos más conocidos para el seguimiento de las características faciales. El seguidor Lucas-Kanade iterativamente trata de minimizar la diferencia entre la imagen y una plantilla deformable. Esta técnica puede ser usada además, para la alineación de la imagen, el análisis de flujo óptico y la estimación de movimiento. De manera general, el seguimiento de las características faciales a pesar de ser preciso y fiable en el plano de movimiento, es muy sensibles a la oclusión, la escala y los cambios de resolución. Además estos métodos requieren la detección previa de los rasgos faciales.

Otro enfoque que también ha sido usado para el seguimiento de rostros es el basado en el movimiento. Los métodos pertenecientes a este enfoque dependen de un método robusto para agrupar movimientos visuales de manera consistente sobre el tiempo. Aunque estos métodos son rápidos no garantizan que las regiones seguidas tengan algún significado semántico [124].

Enfoques basados en modelos tanto geométricos como paramétricos se han usados también para el seguimiento de rostros. Los modelos geométricos son utilizados generalmente para el seguimiento de la cabeza, o sea, cuando se tiene interés en los movimientos rígidos (rotación y traslación) del rostro. Formas como el cilindro, el elipsoide o formas 3D se usan para recuperar el movimiento global de la cabeza. Estos métodos no usan parámetros de la forma, pues asumen que la forma del modelo de la cabeza no cambia durante el seguimiento. Por otra parte, modelos paramétricos como los modelos de la forma activa (ASM, del inglés *Active Shape Model*) y los modelos de apariencia activa (AAM, del inglés *Active Appearance Model*) han sido ampliamente usados en el seguimiento de rostros, así como en el reconocimiento [131], [132].

Los ASMs son métodos estadísticos que incluyen dos tipos de submodelos: un modelo de la forma y un modelo del perfil [133]. Estos modelos son deformados de manera iterativa para adaptarse a un ejemplo de un objeto en una nueva imagen, el cual es representado por un conjunto de puntos. Para esto solo se basan en restricciones de forma, lo que no permite aprovechar toda la información disponible en la imagen como es la textura del objeto destino. Además, el uso de estos métodos requiere una buena inicialización del modelo y su precisión puede disminuir en gran medida en condiciones de movimientos rápidos, grandes rotaciones y oclusión temporal.

Por su parte, el AAM es una técnica rápida para optimizar los parámetros de un modelo estadístico de la apariencia, que muestran tanto las variaciones de la forma como de la apariencia [134]. A partir del entrenamiento del AAM un modelo de rostro puede ser construido. Una vez construido este modelo, el seguimiento del rostro se logra mediante el ajuste del modelo aprendido a una secuencia de entrada. El seguimiento de rostros basado en los AAMs tiene una exacta alineación, alta eficiencia, y eficacia para el manejo de deformaciones del rostro. Sin embargo, el AAM es sensible a la forma inicial y su estabilidad puede ser reducida ante imágenes con fondos desordenados.

Otras técnicas usadas para el seguimiento de rostros son técnicas de predicción y modelación de las dinámicas de objetos como el filtro de Kalman [135], y sus extensiones, el algoritmo de condensación [136], *mean-shift* [137], *cam-shift* [138], entre otras.

En diferentes trabajos se describen y analizan de manera detallada los distintos enfoques y métodos tradicionales utilizados para el seguimiento de objetos y en específico de rostros [16], [123], [124]. En varios de estos estudios se ha visto que el uso de una sola característica o señal no es suficiente para lograr un

seguimiento robusto y preciso [124]. Por lo tanto, la combinación de varias características con diferentes técnicas de predicción y modelación ha sido propuesta con el objetivo de obtener mejores resultados. A continuación se describen algunos de los principales trabajos desarrollados recientemente que evidencian esta propuesta.

4.2 Combinación de características para el seguimiento

La combinación de la información del color y la forma junto a técnicas como el *mean-shift* y filtros de partículas ha sido muy empleada para el seguimiento de rostros. Por ejemplo, al utilizar la información del color y la forma para definir con precisión los contornos del rostro con un algoritmo de seguimiento basado en el *mean-shift*, es posible lograr que este se adapte a las variaciones en la escala del objeto y los cambios de iluminación y fondo [139]. Otro ejemplo es la combinación de una extensión del algoritmo *mean-shift*, el cual integra el filtro de Kalman en el seguimiento del rostro para predecir el centro del rostro en la ventana de búsqueda, con el ASM para enfrentar problemas como las vistas no frontales o casos de movimientos de la cabeza a pequeña escala, así como movimientos rápidos u oclusión temporal [131]. Otro trabajo que evidencia este enfoque fue presentado en el 2007 por Liyue Zhao y Jianhua Tao, quienes desarrollaron un método de seguimiento de rostros basado en un filtro de partículas de múltiples señales que incluye tanto la información del color como la de los bordes [140]. Además, en este trabajo se propone un modelo de distribución de puntos y un algoritmo eficiente de actualización. El modelo de distribución de puntos propuesto permite restringir los resultados del seguimiento y evitar el fallo ante la oclusión. Por otra parte, el método propuesto para la actualización evita el seguimiento de errores acumulados. Los experimentos realizados comparan el método propuesto con el seguidor KLT ampliado y un filtro de partículas basado solo en el color. Los resultados obtenidos indican que el método combinado es más robusto a la oclusión temporal y largos movimientos faciales.

Otro conjunto de métodos para combinar características como la forma y la apariencia han sido usando el AAM en el seguimiento de rostros. Una de las extensiones más recientes de este algoritmo le agrega dos nuevas restricciones para mejorar el rendimiento del seguimiento basado AAM, con respecto a la robustez en el ajuste y la estabilidad en fondos desordenados [141]. La primera restricción es la incorporación de la correspondencia temporal en el ajuste del AAM que impone una restricción de apariencia local entre los cuadros de la secuencia. La segunda restricción es el empleo de una segmentación del rostro basada en el color como una restricción suave. A pesar de los buenos resultados obtenidos, el sistema propuesto no es robusto ante la presencia de vistas de rostros de perfil con grandes ángulos y el manejo de grandes oclusiones.

Adicionalmente, la velocidad y la precisión de la búsqueda tanto en el ASM como en el AAM son dos preocupaciones que han atraído a los investigadores para su mejora. Por ejemplo, el uso de un algoritmo de seguimiento basado en un modelo 3D de apariencia construido mediante nueve puntos fiduciales semánticos ha sido una de ellas. Además, se introduce el uso del esquema de optimización *stochastic meta-descent* (SMD) para acelerar el proceso de búsqueda del modelo de apariencia, logrando mejorar la eficiencia y precisión del seguimiento del rostro. El método propuesto fue comparado contra los AAM convencionales y el *cam-shift*. Los experimentos realizados mostraron que el algoritmo propuesto los supera en términos de eficiencia y precisión en el seguimiento de movimientos rápidos, rotados y escalados de los rostros en una secuencia de video. Sin embargo, este enfoque falla al ser empleado en el seguimiento de múltiples rostros y especialmente en presencia de oclusión.

Otro requerimiento para el uso del AAM para el seguimiento es que se necesita contar con un modelo de la forma a priori o modelos temporales para las dinámicas. Tales restricciones no son requeridas en el uso de predictores lineales para el seguimiento de características faciales utilizando solamente información

de la intensidad [142]. Cada predictor lineal proporciona un mapeo de la información a nivel de pixel para el vector desplazamiento de una característica facial seguida. Luego, múltiples predictores lineales pueden ser agrupados en una unidad rígida para el seguimiento de un punto único de la característica con una mayor robustez y precisión. Los resultados experimentales muestran que el método propuesto es más robusto y preciso que los AAMs, sin usar ningún tipo de información de la forma a priori y con menor cantidad de ejemplos de entrenamiento.

Otra vía utilizada para el seguimiento de las características faciales ha sido el seguimiento basado en la combinación del flujo óptico y plantillas [143]. Este esquema fue desarrollado para el seguimiento de movimientos de la cabeza y la captura de los movimientos de la boca y las cejas de varias personas en tiempo real. Esta combinación, junto con un esquema de distribución, hace que el algoritmo de seguimiento propuesto sea muy robusto contra movimientos rápidos y el desenfoque del movimiento. Para reducir la influencia de la oclusión parcial de la cabeza, en este trabajo se identifican y se excluyen los puntos ocluidos, tanto en el flujo óptico como en el gestor de plantillas. Una vez que se conoce la posición orientación de la cabeza, se utiliza esta información junto con un modelo 3D para encontrar los movimientos de la boca y las cejas. El seguimiento de la boca es realizado mediante la reconstrucción de la plantilla inicial de la boca, en lugar de seguir las partes de la boca; de esta forma, se eliminan la mayoría de las restricciones de la forma de la boca. Sin embargo, este seguidor no es robusto ante la presencia de oclusiones y grandes rotaciones de la cabeza. Para el seguimiento de las cejas se realiza el seguimiento de tres partes de estas: las dos esquinas y el centro. Para hacer esto, se minimiza la función de error que depende de la similitud de la plantilla y de la restricción de la forma de la ceja.

Un sistema para el seguimiento de múltiples rostros fue desarrollado basado en las máquinas de vectores de relevancia (RVM, del inglés, Relevance Vector Machine) y aprendizaje *boosting*. En este sistema, un detector de rostros basado en el algoritmo de aprendizaje *boosting* se utiliza para detectar los rostros en el primer cuadro, y un modelo del movimiento del rostro y un modelo de color son creados. El modelo de movimiento del rostro se compone de un conjunto de RVMs que aprenden la relación entre el movimiento del rostro y su apariencia, mientras que el modelo de color del rostro es el histograma 2D de la región del rostro en el espacio de color CrCb. En el proceso del seguimiento diferentes métodos (seguimiento RVM, búsqueda local, give up seguimiento) son utilizados de acuerdo a los diferentes estados de los rostros, los cuales cambian de acuerdo a los resultados de seguimiento. En la búsqueda de la imagen los autores introducen la matriz de similitud para ayudar a la correspondencia eficiente de los rostros. Los resultados obtenidos en este trabajo demostraron propiedades importantes del sistema propuesto como: encontrar de forma automática la aparición de nuevos rostros, la robustez ante la oclusión y eficiencia computacional, corriendo a unos 20 cuadros/segundos.

Un sistema novedoso que rompe el esquema descrito anteriormente es el propuesto por Krystian Mikolajczyk y colaboradores en el 2010 [130]. El sistema desarrollado para el seguimiento de rostros humanos a largo plazo en videos en entornos no controlados, se basa en una extensión del método *Tracking-Learning-Detection* (TLD) [144] conocida como Face-TLD. Este sistema combina información a priori sobre la clase objetivo a partir de la información del video. Para esto, cada cuadro de la secuencia es procesado por un *tracker* y un detector, cuyas salidas pasan a ser integradas para estimar la ubicación del objeto. En este caso el detector consta de dos partes: un detector de caras genérico entrenado *off-line* para localizar rostros frontales [145] y un algoritmo de validación entrenado *on-line* que decide que rostros corresponden con los sujetos seguidos, mediante el almacenamiento de todos los patrones positivos y negativos que han sido coleccionados durante el seguimiento. Para la selección de estos patrones los autores proponen varias estrategias de aprendizaje basadas en la similitud entre patrones. Por otra parte el *tracker* usado es una adaptación del propuesto en [144], con el objetivo de lograr resistencia ante oclusiones parciales. Como resultado el sistema propuesto es resistente a las oclusiones parciales, los cambios de apariencia y muy

conveniente para aplicaciones en tiempo real. Un aspecto a tener en cuenta es que los autores no consideran la actualización del *tracker*, lo cual podría aumentar la robustez y eficacia del sistema Face-TLD.
En resumen en esta sección se han descrito brevemente los principales enfoques y métodos existentes para realizar el seguimiento de rostros, así como sus principales ventajas y desventajas, teniendo en cuenta que el seguimiento de múltiples rostros es diferente al seguimiento de una solo rostro ya que estos rostros pueden ocluir uno al otro. Como se puede observar, en los trabajos analizados se han usado señales como el color de la piel, la apariencia, la forma, la intensidad, entre otras, para lograr el seguimiento. Además, se vio que una forma de aumentar la eficacia de estos métodos es el empleo de la combinación de estas señales junto con la aplicación de técnicas de modelación y predicción de las dinámicas de los objetos. Un factor que afecta considerablemente el rendimiento de la mayoría de estos métodos es la oclusión entre objetos.

5 Conclusiones

En este reporte se ha hecho un análisis del estado actual de los métodos de detección y seguimiento de rostros existentes. Estos tienen una gran importancia en el proceso de reconocimiento de rostros en video, ya que la calidad de sus resultados influye directamente en la respuesta de la clasificación. De los métodos analizados para la detección de rostros, los métodos basados en la apariencia junto al uso de la información contextual han mostrado los resultados más prometedores. No obstante, tres cuestiones importantes que influyen en las tasas de detección, los por cientos de falsas alarmas aceptadas, así como el costo computacional de los métodos existentes y que todavía requieren de más investigación son: la técnica de búsqueda en la imagen, las características a extraer y el clasificador a emplear.

De las técnicas de búsqueda propuestas en la literatura, el desplazamiento de ventanas es una de las más utilizadas para la detección de rostros. Estos métodos se basan en una búsqueda exhaustiva sobre todas las posibles regiones de la imagen, por lo que generalmente obtienen múltiples detecciones sobre una misma instancia y presentan un alto costo computacional. Además se deben seleccionar parámetros como el tamaño óptimo de la ventana y el paso del desplazamiento.

La búsqueda de subventanas eficiente usando el esquema de optimización *branch and bound*, es otra alternativa de métodos de búsqueda, que ha sido utilizada para la localización de objetos en general, que podría ser aplicada al caso específico de rostros. Este método resulta muy rápido y al mismo tiempo, requiere muchas menos evaluaciones del clasificador; permitiendo el uso de clasificadores más complejos como las SVM. Sin embargo, este método solo obtiene la mejor ubicación de un objeto en una imagen. Por lo que su aplicación en la tarea de la detección de rostros requeriría encontrar:

- una estrategia para la obtención de las ubicaciones de todos los rostros presentes en la imagen. Para esto se ha propuesto la repetición del algoritmo n veces eliminando en cada iteración la región encontrada en la iteración anterior. Esta vía no es eficaz pues es difícil saber a priori el número de iteraciones necesarias y es, además, computacionalmente costosa.
- la formulación del esquema *branch and bound* cuando se emplean otros conjuntos de características o clasificadores más apropiados para rostros. O sea, el uso de características más específicas y discriminativas, así como la posibilidad de emplear otros clasificadores, requiere de la búsqueda de una buena función acotadora para el proceso de optimización.

Otras propuestas han sido el uso de arquitecturas de cascada para el aumento de eficiencia de los algoritmos de búsqueda y el empleo de métodos de optimización multimodales para podar el espacio de búsqueda. Por otra parte, se han empleado mapas de probabilidades basados en el color de la piel para

seleccionar las regiones candidatas; donde la elección del espacio de color, la sensibilidad ante los cambios en las condiciones de iluminación, así como la semejanza del color del tono de la piel con otros objetos, representan retos a enfrentar.

Respecto a los conjuntos de características analizados para la representación de los rostros, las características-*Haar* y los LBP, así como sus extensiones son las más usadas en la literatura, debido a su poder descriptivo, robustez y facilidad de uso. No obstante, existe un grupo de características con las que se han obtenido resultados satisfactorios en la representación de imágenes, como son las SOF-MSOF, SLBHP y características locales invariantes, que no han sido utilizadas directamente en la tarea de la detección de rostros.

Además, los rostros podrían ser representados en otros espacios como el de disimilitud. Este enfoque ha mostrado ventajas con datos de alta dimensionalidad como es el caso de las imágenes [146], [147]. Con el uso de una medida de disimilitud adecuada se puede incluir la información necesaria para obtener una mejor descripción de los objetos y por tanto, lograr una mayor discriminación entre las clases.

Por otra parte, entre los clasificadores analizados los métodos basados en *boosting* han sido uno de los más utilizados debido a su precisión y velocidad, lo que los hace muy adecuados para aplicaciones en tiempo real. Debido a que estos métodos modelan el problema de la clasificación como un problema de dos clases, requieren de un gran conjunto de entrenamiento para representar las clases involucradas (rostro/no-rostro). Por tal motivo estos algoritmos consumen una gran cantidad de tiempo en la etapa de entrenamiento, lo que los hace computacionalmente muy costosos. Una posible vía de solución a este problema podría ser el uso de clasificadores de una clase donde solo se requiere un conjunto de muestras de la clase objetivo, sin necesidad de emplear ejemplos negativos. Esto junto al uso de la información del contexto podría mejorar significativamente los por cientos de falsos positivos obtenidos con los métodos actuales de detección.

En cuanto al seguimiento de rostros, se revisaron los principales enfoques, entre los que dominan la combinación de características junto con la aplicación de técnicas de modelación y predicción de las dinámicas de los objetos. Sin embargo, la robustez ante la oclusión y el seguimiento de múltiples rostros son tareas que todavía exigen de más investigación. El movimiento es una señal importante para el seguimiento de objetos, que podría vincularse con la información del contexto para obtener un significado semántico.

A pesar de que no se analizó profundamente los métodos para la clasificación (identificación/verificación) de los rostros en video, se pudo apreciar las numerables ventajas de este escenario sobre el reconocimiento en imágenes fijas. Además se mostró la importancia del uso de técnicas de súper-resolución como un paso de preprocesamiento en la clasificación de las imágenes de rostros. De manera general, los métodos revisados para la clasificación de los rostros fueron agrupados en tres categorías: los basados en técnicas para imágenes fijas, los basados en la información temporal y los basados en señales híbridas. De estos métodos, dado que solo estamos contando con el rostro como rasgo biométrico, los que utilizan la información temporal son los más apropiados para el reconocimiento de rostros en video. Sin embargo, estos métodos aún presentan dificultades como:

- en ocasiones suponen coherencia temporal entre imágenes consecutivas, cuando pudiera ocurrir que las imágenes coleccionadas se hayan obtenido desde varias vistas y sobre largos períodos de tiempo,
- se les dan iguales pesos a todas características espacio-temporales cuando en realidad algunas de ellas contribuyen más que otras en el reconocimiento,
- no explotan al máximo la información local de los rostros, cuando se ha demostrado su importancia para el reconocimiento

Referencias bibliográficas

1. Zhang, B., Ye, G., Wang, Y., Wang, W., Xu, J., Herman, G., Yang, J.: Informative frequent assembled feature for face detection. In: Proceedings of the 16th IEEE international conference on Image processing. ICIP'09, Piscataway, NJ, USA, IEEE Press (2009) 1201–1204
2. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision **57** (2004) 137–154
3. Mita, T., Kaneko, T., Hori, O.: Joint haar-like features for face detection. In: Proceedings of the Tenth IEEE International Conference on Computer Vision. ICCV '05, Washington, DC, USA, IEEE Computer Society (2005) 1619–1626
4. Yan, S., Shan, S., Chen, X., Gao., W.: Locally assembled binary (lab) feature with feature-centric cascade for fast and accurate face detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.(CVPR 2008). (2008)
5. Khac, C.N., Park, J.H., Jung, H.Y.: Face detection using variance based haar-like feature and svm. Engineering and Technology **60** (2009)
6. Zhang, L., Chu, R., Xiang, S., Liao, S., Li, S.: Face detection based on multi-block lbp representation. In: Proceedings of IAPR/IEEE International Conference on Biometrics (ICB2007). (2007)
7. Roy, A., Marcel, S.: Haar local binary pattern feature for fast illumination invariant face detection. In: Proceedings of the British Machine Vision Conf., 2009. (2009)
8. Trefný, J., Matas, J.: Extended set of local binary patterns for rapid object detection. In: Proceedings of the Computer Vision Winter Workshop 2010. (2010)
9. Paisitkriangkrai, S., Shen, C., Zhang, J.: Face detection with effective feature extraction. In: Proceedings of the 10th Asian conference on Computer vision - Volume Part III. ACCV'10, Berlin, Heidelberg, Springer-Verlag (2011) 460–470
10. Le, D.D., Satoh, S.: Ent-boost: Boosting using entropy measure for robust object detection. In: Proceedings of the 18th International Conference on Pattern Recognition - Volume 02. ICPR '06, Washington, DC, USA, IEEE Computer Society (2006) 602–605
11. Group, I.B.: Biometrics market and industry report 2009-2014 (Enero 2011)
12. Jain, A.K., Flynn, P., Ross, A.A.: Biometrics: Personal Identification in Networked Society. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1998)
13. Bolle, R., Pankanti, S.: Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society. Kluwer Academic Publishers, Norwell, MA, USA (1998)
14. Li, S.Z., Jain, A.: Handbook of Face Recognition. Springer-Verlag (2005)
15. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surveys **35**(4) (2003) 399–458
16. Wang, H., Wang, Y., Cao, Y.: Video-based face recognition: A survey. World Academy of Science, Engineering and Technology **60** (2009) 293–302
17. Li, B., Chellappa, R.: A generic approach to simultaneous tracking and verification in video. IEEE Transactions on Image Processing **11** (May 2002) 530–544
18. Küblbeck, C., Ernst, A.: Face detection and tracking in video sequences using the modified census transformation. Image Vision Comput. **24** (June 2006) 564–572
19. Zheng, W., Bhandarkar, S.M.: Face detection and tracking using a boosted adaptive particle filter. Journal of Visual Communication and Image Representation **20**(1) (2009) 9 – 27
20. Cristóbal, G., Gil, E., Sroubek, F., Flusser, J., Miravet, C., Rodríguez, F.: Superresolution imaging: a survey of current techniques. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Volume 7074 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. (2008)
21. Babu, R.S., Murthy, K.S.: A survey on the methods of super-resolution image reconstruction. International Journal of Computer Applications **15** (2011) 1–6
22. Arachchige, S.: Face recognition in low resolution video sequences using super resolution. Master's thesis, Kate Gleason College of Engineering (2008)
23. Arandjelovic, O., Cipolla, R.: An illumination invariant face recognition system for access control using video. In: In Proc. British Machine Vision Conference. (2004) 537–546
24. Roy-Chowdhury, A., Xu, Y.: Pose and illumination invariant face recognition using video sequences. In Hammoud, R., Abidi, B., Abidi, M., eds.: Face Biometrics for Personal Identification. Signals and Communication Technology. Springer Berlin Heidelberg (2007) 9–25
25. Matta, F., Dugelay, J.L.: Person recognition using facial video information: A state of the art. Journal of Visual Languages & Computing **20**(3) (2009) 180 – 187 ADVANCES IN MULTIMODAL BIOMETRIC SYSTEMS - Multimodal Biometrics.
26. Schonfeld, D., Shan, C., Tao, D., Wang, L.: Video Search and Mining. 1st edn. Springer Publishing Company, Incorporated (2010)
27. Gorodnichy, D.O.: On importance of nose for face tracking. In: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition. FGR '02, Washington, DC, USA, IEEE Computer Society (2002)

28. Shakhnarovich, G., Fisher, J., Darrell, T.: Face recognition from long-term observations. In Heyden, A., Sparr, G., Nielsen, M., Johansen, P., eds.: Computer Vision - ECCV 2002. Volume 2352 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2002) 851–865
29. Liu, X., Chen, T.: Video-based face recognition using adaptive hidden markov models. In: Proceedings of the 2003 IEEE computer society conference on Computer vision and pattern recognition. CVPR'03, Washington, DC, USA, IEEE Computer Society (2003) 340–345
30. Hadid, A., Pietikäinen, M.: Combining appearance and motion for face and gender recognition from videos. Pattern Recogn. **42** (November 2009) 2818–2827
31. Zhou, X., Bhanu, B.: Integrating face and gait for human recognition. In: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop. CVPRW '06, Washington, DC, USA, IEEE Computer Society (2006) 55–
32. Micheloni, C., Canazza, S., Foresti, G.L.: Audio-video biometric recognition for non-collaborative access granting. J. Vis. Lang. Comput. **20** (December 2009) 353–367
33. Patil, A., Kolhe, S., Patil, P.: 2d face recognition techniques: A survey. International Journal of Machine Intelligence **2** (2010) 74–83
34. Hadid, A., Pietikäinen, M.: From still image to video-based face recognition: an experimental analysis. In: Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition. FGR'04, Washington, DC, USA, IEEE Computer Society (2004) 813–818
35. Arandjelovic, O., Cipolla, R.: Face recognition from face motion manifolds using robust kernel resistor-average distance. In: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 5 - Volume 05. CVPRW '04, Washington, DC, USA, IEEE Computer Society (2004) 88–
36. Arandjelovic, O., Shakhnarovich, G., Fisher, J., Cipolla, R., Darrell, T.: Face recognition with image sets using manifold density divergence. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01. CVPR '05, Washington, DC, USA, IEEE Computer Society (2005) 581–588
37. O'Toole, A.J., Roark, D.A., Abdi, H.: Recognizing moving faces: a psychological and neural synthesis. Trends in Cognitive Sciences **6**(6) (2002) 261 – 266
38. Chen, S., Mau, S., Harandi, M.T., Sanderson, C., Bigdeli, A., Lovell, B.C.: Face recognition from still images to video sequences: a local-feature-based framework. J. Image Video Process. **2011** (January 2011) 11:1–11:14
39. Hjelmås, E., Low, B.K.: Face detection: A survey. Computer Vision and Image Understanding **83**(3) (2001) 236–274
40. Yang, M., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE **24**(1) (2002) 399–458
41. Zhang, C., Zhang, Z.: A survey of recent advances in face detection. Technical report, Microsoft Research (2010)
42. Yang, G., Huang, T.S.: Human face detection in a complex background. Pattern Recognition **27**(1) (1994) 53 – 63
43. Peer, P., Solina, F.: An automatic human face detection method. In: Proceedings of the 4th Computer Vision Winter Workshop (CVWW). (1999) 122–130
44. Yow, K.C., Cipolla, R.: Feature-based human face detection. Image and Vision Computing **15**(9) (1997) 713–735
45. Sobottka, K.and Pitas, I.: Face localization and facial feature extraction based on shape and color information. In: Proc. of International Conference on Image Processing. (1996)
46. Craw, I., Tock, D., Bennett, A.: Finding face features. In Sandini, G., ed.: Computer Vision - ECCV92. Volume 588 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (1992) 92–96
47. Yuille, A.L., Hallinan, P.W., Cohen, D.S.: Feature extraction from faces using deformable templates. International Journal of Computer Vision **8** (1992) 99–111 10.1007/BF00127169.
48. Tuytelaars, T., Mikolajczyk, K.: A survey on local invariant features. (2008)
49. Rodriguez, Y.: Face detection and verification using local binary patterns. PhD thesis, Ecole Polytechnique Federale de Lausanne (2006)
50. Viola, P., Jones, M.: Robust real-time object detection. In: International Journal of Computer Vision. (2001)
51. Derpanis, K.G.: Integral image-based representations. Technical report, Department of Computer Science and Engineering, York University, (2007)
52. Porikli, F.: Integral histogram: A fast way to extract histograms in cartesian spaces. (2005)
53. Chum, O., Zisserman, A.: An exemplar model for learning object classes. Volume 0., Los Alamitos, CA, USA, IEEE Computer Society (2007) 1–8
54. Lampert, C.H., Blaschko, M.B., Hofmann, T.: Beyond sliding windows: Object localization by efficient subwindow search. In: In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR. (2008) 1–8
55. Lampert, C.H., Blaschko, M.B., Hofmann, T.: Efficient subwindow search: A branch and bound framework for object localization. IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (2009) 2129–2142
56. Clausen, J.: Branch and bound algorithms - principles and examples. Computer (1999) 1–30

57. Goldberg, D.E., Richardson, J.: Genetic algorithms with sharing for multimodal function optimization. In: Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application, Hillsdale, NJ, USA, L. Erlbaum Associates Inc. (1987) 41–49
58. Marami, E., Tefas, A.: Using particle swarm optimization for scaling and rotation invariant face detection. In: Proceedings IEEE Congress on Evolutionary Computation (CEC), year = 2010, pages = 1–7,
59. Rowley, H.A., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Trans. Pattern Anal. Mach. Intell. **20** (January 1998) 23–38
60. Sung, K., Poggio, T.: Example-based learning for view-based human face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998) 39–51
61. Papageorgiou, C.P., Oren, M., Poggio, T.: A general framework for object detection. In: Proceedings of the Sixth International Conference on Computer Vision. ICCV '98, Washington, DC, USA, IEEE Computer Society (1998) 555–
62. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: European Conference on Computational Learning Theory. (1995) 23–37
63. Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: IEEE ICIP. (2002) 900–903
64. Whitney, A.W.: A direct method of nonparametric measurement selection. IEEE Trans. Comput. **20** (September 1971) 1100–1103
65. Cheng, H., Yan, X., Han, J., wei Hsu, C.: Discriminative frequent pattern analysis for effective classification. In: In ICDE. (2007) 716–725
66. Huang, D., Shan, C., Ardebilian, M., Chen, L.: Face recognition by computers and humans. IEEE Transactions on Image Processing (2011)
67. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In Pajdla, T., Matas, J., eds.: Computer Vision - ECCV 2004. Volume 3021 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2004) 469–481
68. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition **29**(1) (1996) 51 – 59
69. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **1** (2001) 511
70. Pereira, E., Gomes, H., de Carvalho, J.: Integral local binary patterns: A novel approach suitable for texture-based object detection tasks. In: Proceedings of 23rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI 2010). (2010)
71. Bamini.A.M, A., Kavitha.T: Dominant local binary pattern based face feature selection and detection. International Journal of Engineering and Technology (IJET) **2** (2010) 77–80
72. Haijing, Li, P., Zhang, T.: Proposal of novel histogram features for face detection. In Singh, S., Singh, M., Apte, C., Perner, P., eds.: Pattern Recognition and Image Analysis. Volume 3687 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2005) 334–343
73. Wang, H., Li, P., Zhang, T.: Histogram feature-based fisher linear discriminant for face detection. Neural Comput. Appl. **17** (November 2007) 49–58
74. Fröba, B., Ernst, A.: Face detection with the modified census transform. In: Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition. FGR' 04, Washington, DC, USA, IEEE Computer Society (2004) 91–96
75. Sochman, J., Matas, J.: Waldboost "learning for time constrained sequential detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02. CVPR '05, Washington, DC, USA, IEEE Computer Society (2005) 150–156
76. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05. (2005) 886–893
77. Louis, W., Plataniotis, K.N.: Co-occurrence of local binary patterns features for frontal face detection in surveillance applications
78. Ojala, T., Pietikaiinen, M., M, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 971–987
79. Sinha, P.: Toward qualitative representations for recognition. In: In Proceedings of the Second International Workshop on Biologically Motivated Computer Vision. (2002) 249–262
80. Schneiderman, H.: Toward feature-centric evaluation for efficient cascaded object detection. In: In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2004) 1007–1013
81. Liao, S., Lei, Z., Li, S.Z., Yuan, X., He, R.: Structured ordinal features for appearance-based object representation. In: Proceedings of the 3rd international conference on Analysis and modeling of faces and gestures. AMFG'07, Berlin, Heidelberg, Springer-Verlag (2007) 183–192
82. Su, S.Z., ZiLi, S., Chen, S.Y., Li, S.A., Duh, D.J.: Toward feature-centric evaluation for efficient cascaded object detection. In: In Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS). (2010) 670 – 674

83. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110 10.1023/B:VISI.0000029664.99615.94.
84. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). Computer Vision and Image Understanding **110**(3) (2008) 346 – 359 Similarity Matching in Computer Vision and Multimedia.
85. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In Heyden, A., Sparr, G., Nielsen, M., Johansen, P., eds.: Computer Vision - ECCV 2002. Volume 2350 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2002) 128–142
86. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. International Journal of Computer Vision **60** (2004) 63–86 10.1023/B:VISI.0000027790.02288.f2.
87. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing **22**(10) (2004) 761 – 767 British Machine Vision Computing 2002.
88. Mikolajczyk, K., Leibe, B., & Schiele, B.: Local features for object class recognition. In: Proceedings of the IEEE Tenth IEEE International Conference on Computer Vision (ICCV '05), year = 2005, pages = 792–1799,
89. Chang, L., Duarte, M., Sucar, L., Morales, E.: Object class recognition using sift and bayesian networks. In Sidorov, G., Hernández Aguirre, A., Reyes García, C., eds.: Advances in Soft Computing. Volume 6438 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2010) 56–66
90. Tax, D.M.J.: One-class classification. PhD thesis, Delft University of Technology (2001)
91. Tarassenko, L., Hayton, P., Brady, M.: Novelty detection for the identification of masses in mammograms. In: Proc. of the Fourth International IEE Conference on Artificial Neural Networks. Volume 409. (1995) 442–447
92. Vapnik, V.N.: Statistical Learning Theory. Wiley-Interscience (September 1998)
93. Scholkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA (2001)
94. Jin, H., Liu, Q., Lu, H.: Face detection using one-class-based support vectors. In: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR04). (2004)
95. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometrics and Intelligent Laboratory Systems **2**(1-3) (1987) 37–52
96. Vilaplana, V., Marqués, F.: Support vector data description based on pca features for face detection. (2008)
97. Mostafa, L., Abdelazeem, S.: Face detection based on skin color using neural networks. In: Proceedings of GVIP Conference. (2005) 19–21
98. Féraud, R., Bernier, O.J., Viallet, J.E., Collobert, M.: A fast and accurate face detector based on neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (2001) 42–53
99. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cognitive Neuroscience **3** (January 1991) 71–86
100. M.H. Yang, D.R., Ahuja, N.: A snow-based face detector. In Advances in Neural Information Processing Systems (2000) 855–861
101. Cootes, T., Walker, K., Taylor, C.: View-based active appearance models. In: In Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition. (2000) 227–232
102. Osuna, E., Freund, R., Girosi, F.: Training support vector machines: an application to face detection. In: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97). CVPR '97, Washington, DC, USA, IEEE Computer Society (1997) 130–
103. Ratsch, M., Romdhani, S., Vetter, T.: Efficient face detection by a cascaded support vector machine using haar-like features. In Rasmussen, C., Bulthoff, H., Scholkopf, B., Giese, M., eds.: Pattern Recognition. Volume 3175 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2004) 62–70
104. Schapire, R.E.: The strength of weak learnability. Machine Learning **5** (1990) 197–227 10.1007/BF00116037.
105. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Machine Learning **37** (1999) 297–336 10.1023/A:1007614523901.
106. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: A statistical view of boosting. The Annals of Statistics **28** (1995)
107. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: Proceedings in DAGM 25th Pattern Recognition Symposium. (2003)
108. Fayyad, U., Irani, K.: Multi-interval discretization of continuous valued attributes for classification learning. In: In Proc. Internat. Joint Conference on Artificial Intelligence (IJCAI). (1993) 1022–1027
109. Liu, C., Shum, H.: Kullback-leibler boosting. In: In Proc. Internat. Conf. on Computer Vision and Pattern Recognition (CVPR). Volume 1. (2003) 587–594
110. Pudil, P., Novovicová, J., Kittler, J.: Floating search methods in feature selection. Pattern Recognition Letters **15**(11) (1994) 1119 – 1125
111. Li, S.Z., Zhang, Z.: Floatboost learning and statistical face detection. IEEE Trans. Pattern Anal. Mach. Intell. **26** (September 2004) 1112–1123

112. Wald, A.: Sequential tests of statistical hypotheses. Annals of Mathematical Statistics **16** (1945) 117–186
113. Scott, D.W.: Multivariate Density Estimation : Theory, Practice, and Visualization. Wiley Series in Probability and Mathematical Statistics (1992)
114. Wu, B., Ai, H., Huang, C., Lao, S.: Fast rotation invariant multi-view face detection based on real adaboost. In: Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition. FGR' 04, Washington, DC, USA, IEEE Computer Society (2004) 79–84
115. tephen, P.: The effects of contextual scenes on the identification of objects. Memoryamp Cognition **3** (1975) 519–526
116. Galleguillos, C., Belongie, S.: Context based object categorization: A critical survey. Computer Vision and Image Understanding **114**(6) (2010) 712 – 722 Special Issue on Multi-Camera and Multi-Modal Sensor Fusion.
117. Kruppa, H., Santana, M.C., Schiele, B.: Fast and robust face finding via local context. (2003)
118. Blaschko, M., Lampert, C.: Object localization with global and local context kernels. In: In: BMVC. (2009)
119. Atanasoaei, C., McCool, C., Marcel, S.: On Improving Face Detection Performance by Modelling Contextual Information. Technical report (2010)
120. Comaniciu, D., Meer, P.: Mean shift analysis and applications. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. (1999) 1197–1204
121. Comaniciu, D.: An algorithm for data-driven bandwidth selection. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 281–288
122. Daijin, K.: Automated Face Analysis: Emerging Technologies and Research. Information Science Reference - Imprint of: IGI Publishing, Hershey, PA (2009)
123. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. **38** (December 2006)
124. Wang, J.J., Singh, S.: Video analysis of human dynamics–a survey. Real-Time Imaging **9**(5) (2003) 321 – 346
125. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. (1981) 674–679
126. Tomasi, C., Kanade, T.: Detection and tracking of point features. Technical report, International Journal of Computer Vision (1991)
127. Bourel, F., Chibelushi, C.C., Low, A.A., Dg, S.S.: Robust facial feature tracking. In: Proc. 11th British Machine Vision Conference. (2000) 232–241
128. Shi, J., Tomasi, C.: Good features to track. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994 (CVPR '94). (1994) 593–600
129. Chen, J., Tiddeman, B.: A robust facial feature tracking system. In: Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance, 2005. (AVSS 2005). (2005)
130. Kalal, Z., Mikolajczyk, K., Matas, J.: Face-tld: Tracking-learning-detection applied to faces. In: Proceedings of the 17th IEEE International Conference on Image Processing (ICIP), 2010. (2010)
131. Pu, B., Liang, S., Xie, Y., Yi, Z., Heng, P.A.: Video facial feature tracking with enhanced asm and predicted meanshift. In: Proceedings of the 2010 Second International Conference on Computer Modeling and Simulation - Volume 02. ICCMS '10, Washington, DC, USA, IEEE Computer Society (2010) 151–155
132. Li, Z., Chen, J., Chong, A., Yu, Z., Schraudolph, N.N.: Using Stochastic Gradient-Descent Scheme in Appearance Model Based Face Tracking. In: Proc. Intl. Workshop Multimedia Signal Processing (MMSP), Cairns, Australia, IEEE (2008)
133. Cootes, T.F., Taylor, C.J.: Active shape models - their training and application. Journal of Computer Vision and Image Understanding **61**(1) (1995) 38–59
134. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Proceedings of the 5th European Conference on Computer Vision-Volume II. ECCV '98, London, UK, Springer-Verlag (1998) 484–498
135. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans. Pattern Anal. Mach. Intell. **27** (May 2005) 699–714
136. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. International Journal of Computer Vision **29** (1998) 5–28
137. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000. (2000) 142–149
138. R., B.G.: Computer vision face tracking for use in a perceptual user interface. Intel Technical Journal (1998)
139. Vilaplana, V., Varas, D.: Face tracking using a region-based mean-shift algorithm with adaptive object and background models. Image Analysis for Multimedia Interactive Services, International Workshop on **0** (2009) 9–12
140. Zhao, L., Tao, J.: Fast facial feature tracking with multi-cue particle filter. In: Proceedings of Image and Vision Computing New Zealand 2007. (2007)
141. Zhou, M., Liang, L., Sun, J., Wang, Y.: Aam based face tracking with temporal matching and face segmentation. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on **0** (2010) 701–708
142. Ong, E.J., Lan, Y., Theobald, B.: Robust facial feature tracking using selected multi-resolution linear predictors. In: Proceedings of IEEE 12th International Conference on Computer Vision, 2009. (2009)
143. Gast, E.R.: A framework for real-time face and facial feature tracking using optical flow pre-estimation and template tracking. Master's thesis, LIACS, Leiden University (2010)

144. Kalal, Z., Matas, J., Mikolajczyk, K.: Online learning of robust object detectors during unstable tracking. In: Proceedings of the IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), 2009. (2009) 1417 – 1424
145. Kalal, Z., Matas, J., Mikolajczyk, K.: Weighted sampling for large-scale boosting. In Everingham, M., Needham, C.J., Fraile, R., eds.: BMVC, British Machine Vision Association (2008)
146. Pekalska E, D.R.: The Dissimilarity Representation For Pattern Recognition. Foundations and Applications. World Scientiic (2005)
147. Porro-Muñoz, D., Duin, R.P.W., Talavera, I., Orozco-Alzate, M.: Classification of three-way data by the dissimilarity representation. Signal Process. **91** (November 2011) 2520–2529

RT_046, enero 2012
Aprobado por el Consejo Científico CENATAV

Editor: Lic. Lucía González Bayona
Diseño de Portada: Di. Alejandro Pérez Abraham
RNPS No. 2142
ISSN 2072-6287
Indicaciones para los Autores:
Seguir la plantilla que aparece en www.cenatav.co.cu
C E N A T A V
7ma. No. 21812 e/218 y 222, Rpto. Siboney, Playa;
La Habana. Cuba. C.P. 12200
Impreso en Cuba