Incidencia en la seguridad de los aspectos organizativos y de operación ligados al sistema ferroviario(*)

María Asunción VICENTE RIPOLL[a]
César FERNÁNDEZ PERIS[a]
Ángela María COVES SOLER[a]

RESUMEN: En el estudio realizado se pretende comprobar la viabilidad de un sistema de vigilancia semiautónomo basado en reconocimiento de caras para estaciones de ferrocarril. El funcionamiento básico de este sistema de vigilancia es el siguiente: en determinados puntos de la estación de ferrocarril se sitúan cámaras de video que registran contínuamente los accesos a la misma. Un ordenador procesa las imágenes capturadas por las cámaras para detectar las caras de los sujetos que aparecen en ellas y compara estas caras con las de los sospechosos almacenadas en una base de datos.

Cada vez que se produce una coincidencia, el sistema envía un aviso a un operario humano (vigilante de seguridad) que debe decidir si se trata una falsa alarma o no.

Los dos aspectos evaluados han sido la detección de caras en las imágenes globales, y el reconocimiento de los sujetos frente a la base de datos.

En cuanto a detección, ha quedado demostrado que la tecnología actual es capaz de ofrecer resultados apropiados para un sistema de vigilancia. En cuanto al reconocimiento, si bien la tecnología actual ha mejorado sustancialmente respecto a la de hace pocos años, los resultados sólo permitirían una aplicación práctica en sistemas de vigilancia si la colocación de las cámaras cumpliese unos requisitos muy estrictos, de modo que la calidad de las imágenes fuese muy elevada. En otro caso, los porcentajes de falsas alarmas serían excesivos para ser manejados por un solo operario.

1. INTRODUCCIÓN

El objetivo del estudio es analizar una posible acción para la mejora de la seguridad en estaciones de ferrocarril. La idea básica es dotar a las estaciones de un sistema de detección y reconocimiento de caras basado en visión artificial, a instalar en los accesos a los trenes y que pueda reconocer a sospechosos o personas buscadas (delincuentes, personas secuestradas, etc.). La alternativa que proponemos en este proyecto consiste en instalar cámaras de video y comparar las imágenes de las personas que acceden al tren o a los andenes con las personas de interés (delincuentes, personas secuestradas, etc.) almacenadas en la base de datos del sistema de reconocimiento de caras.

En función de la fiabilidad del reconocimiento, el sistema se podrá utilizar

[a] Profesores Titulares del Área de Ingeniería de Sistemas y Automática y del Área de Teoría de la Señal.
Universidad Miguel Hernández. Departamento de Ingeniería de Sistemas y Automática.

(*) El artículo corresponde al resumen del proyecto elaborado en el marco de las subvenciones para la realización de estudios y acciones de difusión relacionados con transporte, sus infraestructuras y demás competencias del Ministerio de Fomento (Orden FOM/2219/2008, de 22 de julio).

de modo automático, generando una alarma cada vez que se detecta a un sujeto buscado o de interés; o bien de modo supervisado. En el modo supervisado, cada vez que se detecta un sujeto buscado se envían las imágenes a un operario de seguridad que podría comprobar si se trata de una falsa alarma o no. La figura 1 muestra el diagrama de flujo de los dos métodos de funcionamiento.

Para un correcto funcionamiento del sistema es necesario realizar dos operaciones: detección de la cara en la imagen global de la escena, esto es, determinar el número de caras y su posición; y reconocimiento del sujeto al que corresponde esa cara.

Los sistemas de reconocimiento de caras [1] han evolucionado de forma importante en los últimos años, alcanzando niveles de fiabilidad y velocidad de respuesta elevados. El objetivo del estudio es determinar, mediante pruebas exhaustivas, si estos niveles son suficientes para garantizar la seguridad en los accesos a los trenes sin ocasionar molestias y demoras a los viajeros.

Existen estudios y comparativas previas que intentan establecer las prestaciones de los algoritmos de reconocimiento actuales. Entre ellos destacamos las comparativas FRVT (*Face Recognition Vendor Test*) realizadas en los años 2000, 2002 y 2006 [2-4] por el NIST *(National Institute of*

Figura 1. Comparación de los métodos automático y supervisado

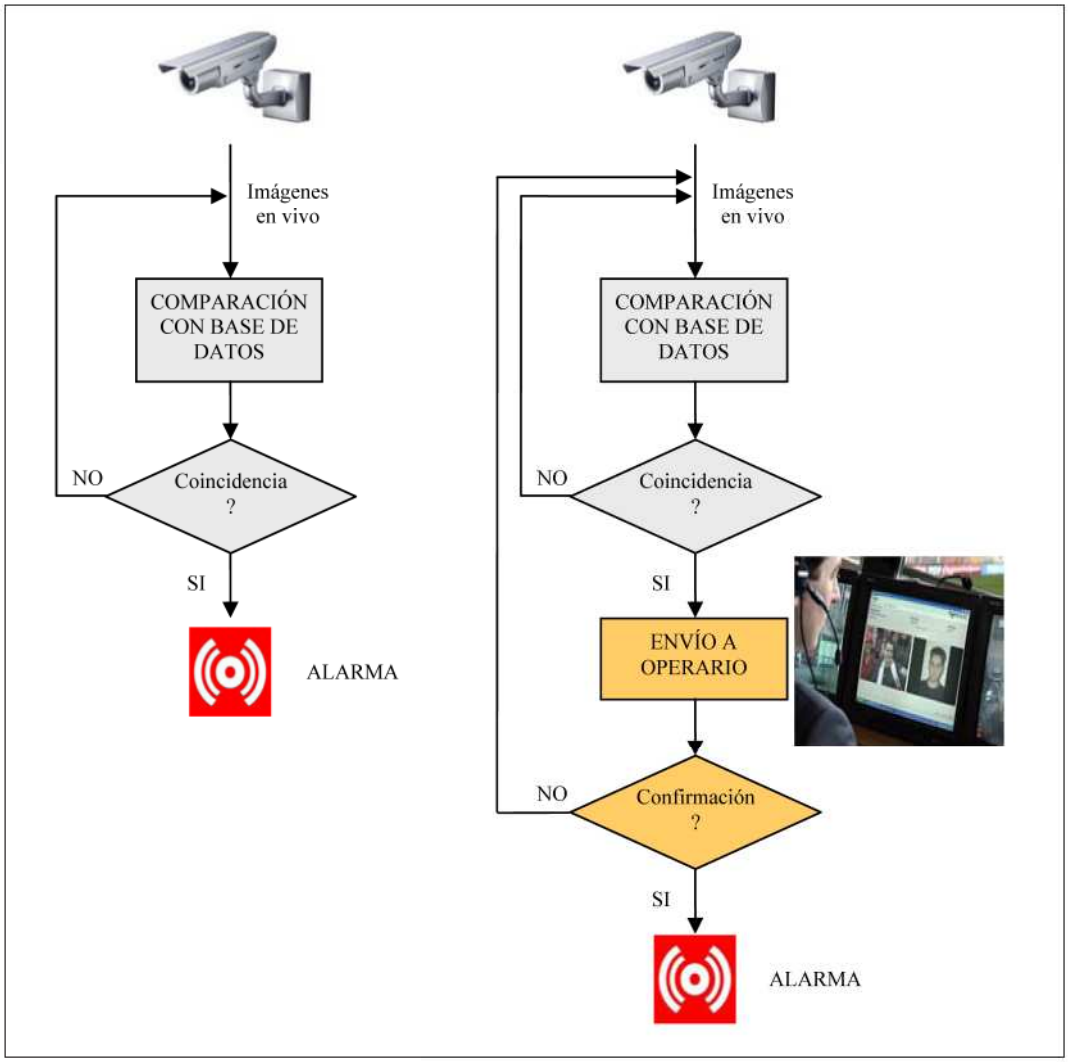
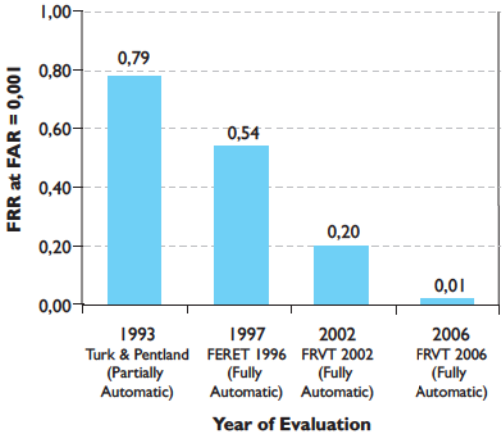


Figura 2. Resultados del estudio FRVT2006 (publicados en [2])



Standards and Technology). Estos estudios han evaluado las prestaciones de las tecnologías existentes en el campo del reconocimiento de caras mediante la comparativa de las tasas de éxito de los diferentes algoritmos, procedentes de universidades, centros de investigación y sistemas comerciales. La figura 2 (extraída del informe FRVT2006 [2]) muestra el progreso que se ha conseguido en estos algoritmos en los últimos años. Los valores que aparecen en el gráfico muestran las tasas de error en el reconocimiento como falsos rechazos (número de veces que una persona buscada no es detectada por el sistema) cuando el nivel de falsos positivos (número de veces que una persona no buscada es erróneamente identificada) se fija en 0,001 (0,1%). En el gráfico puede verse como en los últimos años este indicador ha mejorado notoriamente.

Sin embargo, una evaluación de tecnologías como la realizada en FRVT2006 se realiza en un entorno controlado y no tiene en cuenta aspectos específicos de la aplicación para la que se desea utilizar el reconocimiento de caras. Es posible que los resultados mostrados en el gráfico no puedan lograrse en un entorno como el de una estación de ferrocarril, puesto que las condiciones de iluminación y calidad de las imágenes no son óptimas. El objetivo de nuestro estudio es evaluar estas tecnologías en condiciones similares a las de la aplicación real. Para ello proponemos realizar pruebas sobre múltiples bases de datos de caras obtenidas en entornos no controlados.

2. SOFTWARE UTILIZADO

2.1. Software de detección

Los diferentes programas software de detección de caras que hemos tenido en cuenta han sido los siguientes:

- **Fraunhofer Institute Real Time face Detection [5]:** se trata de un software gratuito de detección de caras. A pesar de que anuncia buenos resultados, realizamos numerosas pruebas y no resultaron satisfactorias.
- **Viola-Jones Face Detection Algorithm [6]:** también se trata de un software no comercial. Es el más utilizado en las aplicaciones e investigaciones actuales. Las pruebas preliminares realizadas con este software dieron muy buenos resultados.
- **FaceSDX de Luxand [7]:** se trata de un software comercial, que comprende tanto la detección como el reconocimiento. Hemos probado la versión de demostración y los resultados en detección son inferiores a los ofrecidos por el software no comercial de Viola-Jones.
- **Pittsburg Pattern Recognition [8]:** software comercial que realiza conjuntamente las tareas de detección y reconocimiento. Las pruebas sobre la versión de evaluación resultaron comparables en prestaciones al detector de Viola-Jones. Como inconveniente, el resultado de la detección incorpora información sobre la orientación de la cara, que no es compatible con algunos programas de reconocimiento. Dado este inconveniente, y dado que no mejora los resultados del software de Viola-Jones, se descartó su utilización.

En conclusión, optamos por elegir el algoritmo de **Viola-Jones**.

Sus principales ventajas son su velocidad de procesado (puede operar en tiempo real) y su fiabilidad (es capaz de detectar caras en múltiples condiciones de iluminación, enfoque e incluso con oclusiones parciales).

Su principal limitación es que no detecta caras que no miren frontalmente a la cámara (aproximadamente); pero en una aplicación como la prevista esto no es un inconveniente, dado que sólo las imágenes frontales son adecuadas para el reconocimiento posterior.

2.2. Software de reconocimiento

Los programas comerciales que hemos tenido en cuenta han sido los siguientes:

- **FaceSDX de Luxand:** como hemos comentado anteriormente, se trata de un software comercial, que comprende tanto la detección como el reconocimiento. Dado que en las pruebas de la fase de detección los resultados no fueron satisfactorios, y que es imposible independizar ambas fases, se descartó este software.
- **Pittsburg Pattern Recognition:** software comercial que realiza conjuntamente las tareas de detección y reconocimiento. Las pruebas sobre la versión de evaluación resultaron comparables en prestaciones al software libre seleccionado. Como inconveniente, está basado en que el resultado de la detección indique la orientación frontal de la cara, con lo cual no acepta los datos devueltos por el detector de Viola-Jones seleccionado. Por esta razón, no se consideró este software.
- **FaceVACS SDK de Cognitec [9]:** se trata de un software comercial que permite trabajar sobre los resultados de detección de Viola-Jones, o bien utilizar un detector propio. Esta característica, junto con los resultados obtenidos en el test FRAV2D (donde fue uno de los algoritmos mejor situados en todas las pruebas realizadas), hace que lo consideremos el software comercial más apropiado.

En conclusión, hemos elegido el software de Cognitec (FaceVACS).

Adicionalmente, también hemos contemplado para nuestras pruebas algunos algoritmos no comerciales. Entre ellos, cabe destacar los siguientes:

- **PCA o análisis de componentes principales [10]:** técnica de compresión de imágenes que busca extraer la máxima información de las mismas (busca las direcciones de máxima varianza). Se utiliza en multitud de aplicaciones e investigaciones para el reconocimiento de caras.
- **LDA o análisis discriminante lineal [11]:** técnica de compresión similar a PCA pero con la ventaja añadida de buscar las direcciones que teóricamente permitirían distinguir mejor entre caras de diferentes personas. En la práctica, este aumento de complejidad no está justificado, porque sus resultados no son superiores a los ofrecidos por PCA.
- **ICA o análisis de componentes independientes [12]**: procedimiento de compresión parecido a PCA pero con la intención de encontrar direcciones que maximizan la independencia de los datos. También añade complejidad a PCA y no aporta mejoras.
- **Métodos basados en descriptores SIFT [13]:** se trata de métodos que buscan caracterizar las caras por detalles locales de las mismas (puntos de interés). Se trata de una estrategia completamente diferente a la de PCA, con lo que es razonable incluirlo en nuestras pruebas.
- **AAM o modelos activos de apariencia [14]:** se trata de un método novedoso para el reconocimiento de caras. Dado que por ahora no existen aplicaciones fiables disponibles, no se considera este software.

De entre todos ellos, hemos elegido dos algoritmos que cubren las dos técnicas básicas de reconocimiento de caras: reconocimiento basado en la apariencia global de la imagen y reconocimiento basado en la apariencia local (descripción de puntos de interés). Los algoritmos elegidos han sido los siguientes:

- Apariencia global: PCA (principal component analysis).
- Apariencia local: SIFT (scale invariant feature transform).

PCA se ha utilizado durante mucho tiempo para el reconocimiento de caras, como técnica de compresión de la información presente en la imagen, de modo global. Han surgido otros métodos como LDA (linear discriminant analysis) o ICA (independent component analysis), más sofisticados, pero que no han demostrado en la práctica ofrecer mejores resultados. En concreto, en uno de nuestros trabajos anteriores [15] demostramos teóricamente y experimentalmente la equivalencia de los métodos PCA e ICA para esta tarea.

SIFT es una transformada que permite describir inequívocamente puntos de interés dentro de las imágenes. Para nuestras pruebas, hemos utilizado esta transformada junto con un desarrollo propio [16] que la adapta al reconocimiento de caras.

3. BASES DE DATOS EMPLEADAS

A continuación se indican las bases de datos consultadas, junto con una breve descripción de las mismas (en [17] se puede encontrar un listado exhaustivo):

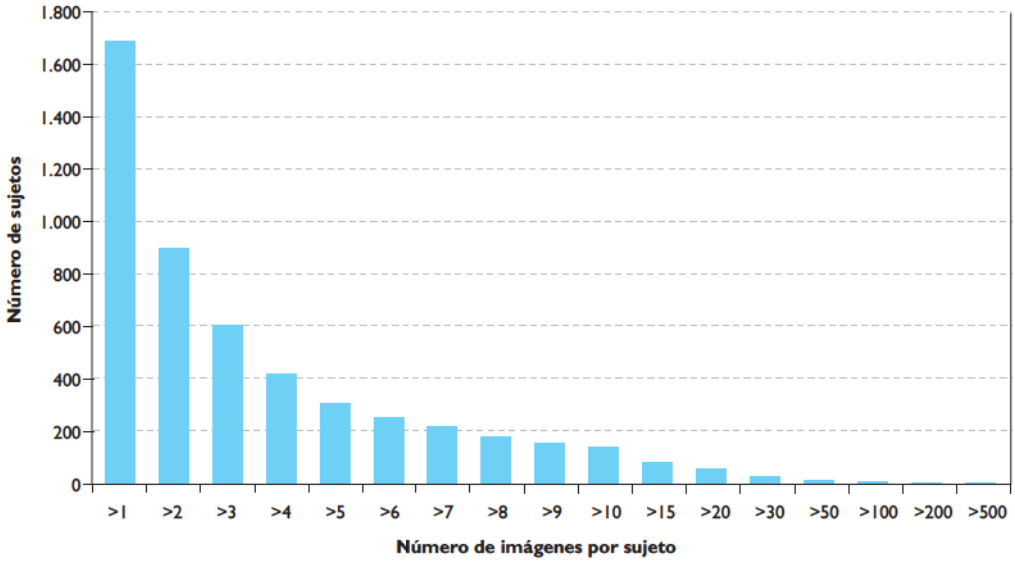
- **AR face database [18]**: contiene 126 sujetos, con un número variable de imágenes por sujeto (4000 imágenes en total). Las imágenes fueron capturadas en condiciones controladas de laboratorio.
- **FERET database [19]:** contiene un total de 14126 imágenes para 1199 sujetos (el número de imágenes por sujeto es variable). Las imágenes fueron capturadas en condiciones controladas de laboratorio.
- **AT&T database [20]:** contiene 400 imágenes de 40 sujetos (10 imágenes por sujeto), capturadas en condiciones controladas de laboratorio.
- **LFW (labeled faces in the wild) database [21]:** contiene más de 13000 imágenes de 5749 sujetos, con un número muy variable de imágenes por sujeto. La ventaja respecto de las bases de datos anteriores es que todas las imágenes se obtuvieron de Internet, con lo que las condiciones son muy variables entre ellas. Es la base de datos más similar a un entorno real, por la falta de homogeneidad de las imágenes.
- **Otras bases de datos:** o bien son similares a las anteriores, o contienen información 3D, voz y otras fuentes de información que no son de aplicación al presente proyecto.

La mayor parte de las bases de datos se han obtenido en condiciones controladas (los sujetos siempre mantienen la misma distancia a la cámara, los fondos son uniformes, la iluminación está controlada, etc.). Estas bases de datos no son apropiadas, porque las condiciones están muy alejadas de las que se pueden encontrar en una estación de ferrocarril real. No obstante, utilizaremos la base de datos de AT&T como referencia para comparar los resultados que se obtengan. Esta base de datos contiene 40 sujetos y 10 imágenes por sujeto.

Entre las bases de datos capturadas en condiciones no controladas, LFW (Labeled Faces in the Wild) es la más apropiada para nuestro estudio. En esta base de datos, todas las imágenes se han obtenido de Internet, bajo todo tipo de condiciones diferentes de iluminación, fondos, etc. Hemos utilizado esta base de datos para nuestras pruebas, aunque hemos encontrado algunas limitaciones:

- La base de datos aconseja un método de utilización concreto para obtener estadísticas de funcionamiento de los algoritmos de reconocimiento. Este método se basa en comparar imágenes dos a dos y decidir si corresponden o no a la misma persona, bajo la condición de que ninguna otra imagen de esas dos personas se haya utilizado durante el entrenamiento del sistema. Esta estrategia de evaluación no es apropiada para nuestro estudio, porque la situación en un entorno de vigilancia es diferente: se dispone de imágenes de entrenamiento para los sospechosos que se quiere detectar. Por lo tanto, hemos utilizado la base de datos LFW pero hemos modificado la estrategia de obtención de resultados.
- La base de datos incluye 5749 sujetos, pero el número de imágenes por sujeto

Figura 3. Distribución de la base de datos LFW



no es homogéneo. Este número varía desde 530 imágenes para el sujeto más popular hasta una sola imagen para los sujetos menos populares. En la figura 3 se muestra la distribución de la base de datos (se han eliminado los 4069 sujetos para los que existe una sola imagen para poder mostrar mejor el resto de datos). Para nuestro proyecto, hemos utilizado los 158 sujetos para los que existen al menos 10 imágenes, y en todos los casos hemos utilizado únicamente las 10 primeras imágenes de cada sujeto (incluso para los sujetos que tienen un número de imágenes mucho mayor). La idea es no introducir un sesgo en los clasificadores (si la mayor parte de las imágenes perteneciesen al mismo sujeto, los clasificadores mostrarían un sesgo hacia ese sujeto).

- La base de datos LFW sólo sirve para probar los algoritmos de reconocimiento de caras, pero no los de detección de caras. La razón es que todas las imágenes han sido preprocesadas con el algoritmo de detección de Viola-Jones. Dado que ninguna de las otras bases de datos disponibles resultan adecuadas (por ser excesivamente homogéneas) para realizar experimentos de detección de caras, hemos utilizado un conjunto de imágenes específico.

4. ADAPTACIÓN DEL SOFTWARE Y DE LAS BASES DE DATOS

4.1. Entorno de trabajo utilizado

El objetivo es conseguir que todos los algoritmos de detección probados y todas las bases de datos utilizadas puedan ser manejados desde un entorno común, y facilitar de este modo la elaboración de estadísticas.

Se ha decidido utilizar el entorno Matlab, por las facilidades que ofrece para la realización de experimentos de manera sencilla y fiable, y por disponer de este entorno en la Universidad Miguel Hernández. A continuación detallamos los trabajos que se han realizado sobre las bases de datos y el software:

4.2. Base de datos AT&T

La base de datos AT&T contiene 40 sujetos, para cada uno de los cuales se dispone de 10 imágenes. Las imágenes se encuentran en formato pgm en niveles de gris, con lo que hemos previsto las conversiones de formato necesarias para que puedan ser utilizadas por los diferentes algoritmos de reconocimiento.

Dado que cada imagen muestra un único sujeto, situado siempre frente a la cámara y

a una distancia uniforme de la misma, no es necesario ningún preprocesamiento adicional.

4.3. Base de datos LFW

La base de datos LFW contiene un total de 5749 sujetos, pero el número de imágenes por sujeto no es homogéneo, como se ha comentado anteriormente. El primer trabajo de adaptación de esta base de datos ha consistido en seleccionar aquellos sujetos para los que existen al menos 10 imágenes, y descartar las restantes imágenes por encima de la décima, en el caso de que existiesen para alguno de los sujetos. El resultado es un subconjunto de la base de datos con 158 sujetos y 10 imágenes para cada sujeto, exactamente la misma estructura que la base de datos AT&T. Esto permite hacer experimentos comparativos fácilmente.

En segundo lugar, a pesar de que la base de datos está generada utilizando el detector de Viola-Jones (también se ha comentado anteriormente), las imágenes no se limitan únicamente al rostro del personaje de interés, sino que cubren un área muy superior; en algunos casos incluso llegan a aparecer varias personas en la misma imagen. Para permitir el correcto funcionamiento de los algoritmos de reconocimiento de caras, hemos preprocesado las imágenes volviendo a pasar sobre ellas el detector de Viola-Jones y manteniendo únicamente el área donde aparece el rostro del personaje de interés. Este preprocesamiento no altera el resultado de los experimentos, dado que en el entorno real (estación de ferrocarril) también planteamos utilizar como primer paso el detector de Viola-Jones.

Adicionalmente, las imágenes se encuentran en formato jpeg en color, por lo que se convierten a niveles de gris para homogeneizarlas con las de la otra base de datos utilizada.

4.4. Software de detección basado en el método de Viola-Jones

Hemos utilizado la versión de este software disponible en la librería de visión artificial OpenCV. Dado que el software no estaba preparado para ser llamado desde un entorno externo como Matlab, ha sido necesario recompilarlo con ligeras modificaciones. Adicionalmente, hemos desarrollado las funciones de Matlab adecuadas para poder llamar al software de detección desde cualquier experimento.

4.5. Software de reconocimiento basado en PCA

Dado que este software de reconocimiento está desarrollado sobre el entorno Matlab, las modificaciones requeridas han sido menores. Aún así, ha sido necesario desarrollar las funciones Matlab que permiten llamar a las rutinas de reconocimiento desde cualquier experimento.

4.6. Software de reconocimiento basado en SIFT

El software de cálculo de la transformada SIFT no es accesible directamente desde Matlab, ni es modificable, dado que se trata de un programa cerrado. El método que se ha utilizado para poder acceder a este software desde Matlab se ha basado en el acceso a los resultados del programa mediante transferencia de ficheros (un fichero de entrada con la imagen a analizar y un fichero de salida con los resultados). Se trata de un proceso que ralentiza ligeramente las pruebas pero que funciona adecuadamente.

El resto del algoritmo de reconocimiento de imágenes basado en SIFT (la parte desarrollada por nosotros mismos) está realizada en el entorno Matlab, con lo cual no ha habido problemas de adaptación. Es posible llamar a este método desde cualquier experimento.

4.7. Software de reconocimiento comercial FaceVACS

El software FaceVACS dispone de un entorno de desarrollo que permite su utilización de múltiples maneras: por una parte, dispone de librerías que permiten crear aplicaciones a medida; y por otra parte dispone de ejecutables que se pueden utilizar desde la línea de comandos. Para

homogeneizar las pruebas que se realizan con los demás algoritmos, hemos decidido crear rutinas de Matlab que se comunican con los ejecutables de FaceVACS mediante ficheros. De este modo, la realización de las pruebas con el software comercial es muy similar a la realización de las pruebas con los algoritmos basados en PCA y SIFT.

5. PRUEBAS EXPERIMENTALES

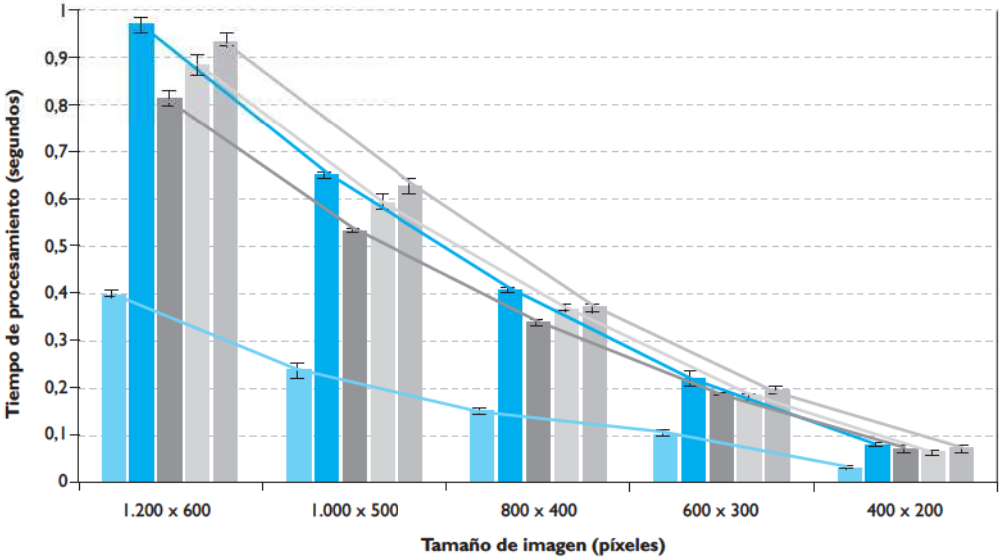
5.1. Detector de caras de Viola-Jones

En primer lugar, se ha evaluado el tiempo de procesamiento; para ello hemos realizado experimentos con imágenes de diferentes tamaños y que contienen diferente número de caras. El objetivo es encontrar relaciones entre estos factores y el tiempo de procesamiento. Cada una de las cinco imágenes utilizadas (figura 4) se ha submuestreado a cinco resoluciones diferentes para las pruebas. Los experimentos se realizaron en un PC estándar (procesador Intel Quad core 2.0 GHz con 2GB de RAM) bajo el sistema operativo Windows, y los resultados se han promediado para 3 repeticiones del experimento. El gráfico de la figura 4 muestra tales resultados (tiempo medio junto con los tiempo máximo y mínimo de cada experimento).

Por una parte, queda claro que, para un tamaño fijo de imagen, el tiempo de procesamiento depende de la imagen en concreto, pero no del número de caras presentes en la imagen. Otros factores como el fondo de la imagen son más relevantes (esto se puede comprobar al comparar las dos primeras imágenes de la figura 5, ambas con una sola cara pero con fondos de complejidad muy diferente). Por otra parte, tal y como esperábamos, el tiempo de procesamiento se ve muy influenciado por el tamaño de la imagen. Hay una relación aproximadamente cuadrática que hace recomendable la utilización de imágenes de tamaño reducido.

Figura 4. Imágenes utilizadas para la medida del tiempo de detección (método Viola-Jones)

Figura 5. Tiempo de procesamiento requerido por el algoritmo de Viola-Jones



Sin embargo, el tamaño no se puede reducir por encima de un cierto límite, dado que los resultados del algoritmo de detección se deterioran rápidamente. Para comprobar este efecto, se realizó un nuevo experimento con el objetivo de buscar el tamaño mínimo admisible. Utilizando las mismas imágenes del experimento previo, medimos los errores obtenidos a diferentes resoluciones. En la figura 6 se muestran los resultados; y en la figura 7 se muestra un ejemplo: la imagen superior (1200 x 600) contiene caras con una dimensión media de 49 píxeles; en la imagen intermedia (1000 x 500) la dimensión media de las caras es de 41 píxeles; y en la imagen inferior esta dimensión es de 33 píxeles (nos

Figura 6. Errores de detección en función del tamaño de las caras (en píxeles)

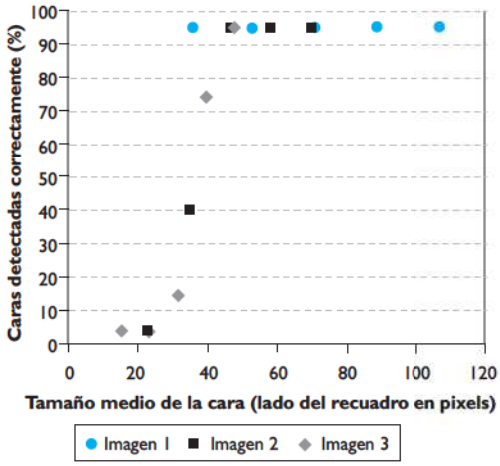
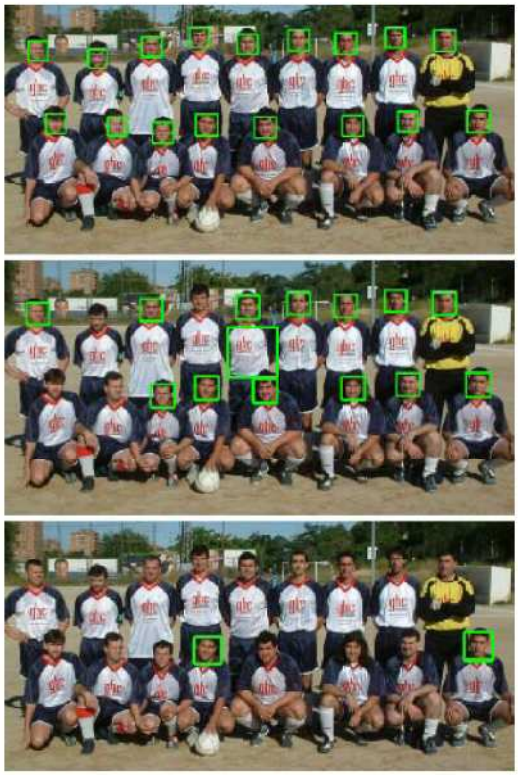


Figura 7. Ejemplos de errores de detección para pequeños tamaños de cara



referimos al lado del cuadrado que engloba a la cara en la imagen, tal y como lo devuelve el detector de Viola-Jones). Como conclusión, sólo es posible detectar caras fiablemente cuando están descritas por al menos 40 x 40 píxeles. En cualquier caso, el hecho de que algunas de las caras presentes en la imagen no se detecten correctamente no es un problema grave en la aplicación que proponemos, dado que la misma persona aparecerá en más de una imagen, aunque las imágenes se capturen a una velocidad relativamente baja (para la mayor parte de los cálculos de este trabajo se supone una frecuencia de procesamiento de imágenes de 1 Hz).

Considerando que se utilizan tamaños de cara de aproximadamente 40 x 40 píxeles, y suponiendo que la cámara debe cubrir un área mucho más amplia que únicamente la persona de interés, es razonable suponer que la imagen debe tener al menos 5 veces la dimensión de la cara verticalmente y 10 veces horizontalmente. Esto representa imágenes de al menos 400 x 200 píxeles. El tiempo medio requerido para procesar una imagen de este tamaño, de acuerdo con la figura 5, es de 63 ms.

El otro condicionante importante del detector es la orientación de la cara respecto de la frontal. De acuerdo con la pruebas realizadas, no puede establecerse un umbral exacto, pero como criterio general podemos establecer un rango de +30º a -30º respecto de la frontal. Dentro de ese rango (y si las condiciones en cuanto a número de píxeles son adecuadas), las caras se detectan con fiabilidad. Fuera de ese rango no se puede garantizar la detección. En cualquier caso, esta condición en cuanto a la desviación respecto de la frontal no representa un problema, sino que sirve para facilitar el trabajo de los sistemas de reconocimiento, que tendrían dificultades si las orientaciones variasen en un rango superior.

5.2. Sistema de reconocimiento basado en PCA

Hemos realizado pruebas tanto de tiempo de cómputo como de fiabilidad del reconocimiento.

La velocidad de procesamiento está muy relacionada con el tamaño (en número de sujetos) de la base de datos. Suponiendo un funcionamiento real de la aplicación, cada imagen que capture el sistema en tiempo real debe ser analizada para buscar sospechosos, por lo tanto debe ser comparada con todos los sospechosos almacenados en la base de datos. Hemos realizado experimentos con las bases de datos AT&T (40 sujetos) y LFW (158 sujetos), y los resultados se pueden ver en la tabla 1.

Tabla 1

TIEMPO DE PROCESAMIENTO PARA EL RECONOCIMIENTO DE CARAS UTILIZANDO PCA

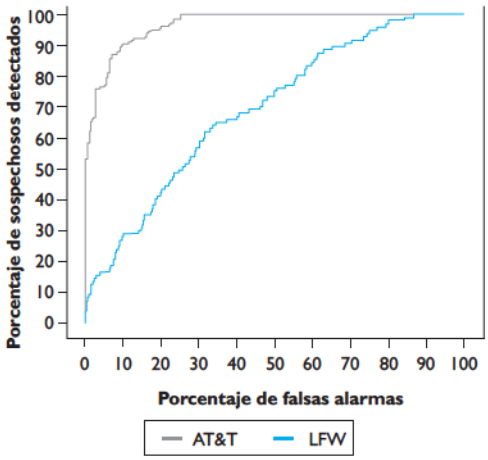
Base de datos	LFW	AT&T
Número de sujetos	158	40
Número de imágenes por sujeto	4	4
Número total de comparaciones	632	160
Tiempo de procesamiento medio PCA (segundos)	0,73	0,19

En todos los casos, suponemos que de las 10 imágenes de que disponemos para cada sujeto, sólo cuatro de ellas se consideran imágenes de entrenamiento (lo que en la aplicación real constituirían las imágenes almacenadas en la base de datos para cada sospechoso o persona fichada). Las otras 6 imágenes de cada sujeto se utilizan como imágenes de test, lo que en la aplicación real serían las imágenes capturadas en vivo en la estación de tren. Este número de imágenes de entrenamiento lo consideramos razonable para un correcto funcionamiento de la aplicación: si cada sospechoso estuviese descrito por una única imagen, probablemente no fuera reconocido adecuadamente en las imágenes en vivo; y si el número de imágenes fuese mucho más elevado, se plantearían problemas de tiempo de cómputo.

Tal y como esperábamos, el tiempo de procesamiento aumenta linealmente con el tamaño de la base de datos. Aún así, el método PCA podría ser empleado sin problemas manteniendo una frecuencia de procesamiento de 1Hz, incluso sobre una base de datos del tamaño de LFW. Adicionalmente, sería posible reducir el tiempo de cómputo optimizando el software.

Para evaluar la fiabilidad del reconocimiento, se han utilizado las curvas

Figura 8. Resultados de reconocimiento utilizando el método PCA



ROC (receiver operating characteristic), manteniendo la estructura de 4 imágenes de entrenamiento y 6 imágenes de test. Para cada imagen de test, se mide la distancia a todas las imágenes de entrenamiento (160 imágenes en el caso de la base de datos AT&T y 632 en el caso de la base de datos LFW). La imagen cuya distancia es mínima es la que indica cuál es la clasificación más probable para la imagen de test. La estrategia mencionada constituye una aplicación del método del vecino más cercano. Se probó también una estrategia diferente, basada en promediar la distancia a todas las imágenes de entrenamiento de cada sujeto, y elegir el sujeto cuya distancia media fuese menor para clasificar la imagen de test. Dado que las tasas de reconocimiento obtenidas fueron inferiores para ambas bases de datos (AT&T y LFW), se descartó esta segunda estrategia.

La figura 8 muestra los resultados obtenidos. La primera conclusión que cabe extraer es que la base de datos LFW es mucho más exigente para un algoritmo de reconocimiento de caras.

5.3. Sistema de reconocimiento basado en SIFT

Las pruebas relativas al tiempo de procesamiento han sido similares a las realizadas con el método PCA. Los resultados obtenidos en este caso se muestran en la tabla 2.

Tabla 2

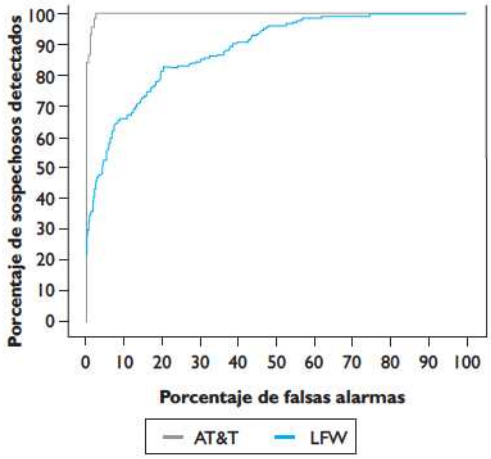
TIEMPO DE PROCESAMIENTO PARA EL RECONOCIMIENTO DE CARAS UTILIZANDO SIFT

Base de datos	LFW	AT&T
Número de sujetos	158	40
Número de imágenes por sujeto	4	4
Número total de comparaciones	632	160
Tiempo de procesamiento medio SIFT (segundos)	3,20	0,84

Estos tiempos de procesamiento serían excesivos, en el caso de la base de datos LFW, para la aplicación del sistema en tiempo real a una frecuencia de 1Hz. No obstante, la paralelización del proceso (utilización de varios procesadores) podría hacer factible la aplicación de este método.

Para evaluar la fiabilidad del reconocimiento, se han utilizado igualmente las curvas ROC (receiver operating curve), manteniendo la estructura de 4 imágenes de entrenamiento y 6 imágenes de test y utilizando la estrategia del vecino más cercano. Los resultados se muestran en la figura 9, donde se aprecia la superioridad del método SIFT respecto del método PCA.

Figura 9. Resultados de reconocimiento utilizando el método SIFT



5.4. Sistema de reconocimiento FaceVACS

Las pruebas relativas al tiempo de procesamiento han sido similares a las realizadas con los métodos PCA y SIFT. Los resultados obtenidos en este caso se muestran en la tabla 3.

Tabla 3

TIEMPO DE PROCESAMIENTO PARA EL RECONOCIMIENTO DE CARAS UTILIZANDO EL SOFTWARE COMERCIAL FACEVACS

Base de datos	LFW	AT&T
Número de sujetos	158	40
Número de imágenes por sujeto	4	4
Número total de comparaciones	632	160
Tiempo de procesamiento medio FaceVACS (segundos)	0,201	0,177

Se trata de tiempos de procesamiento suficientemente rápidos como para aplicar el método en tiempo real a una frecuencia de procesamiento de imágenes de 1Hz. La fiabilidad del reconocimiento se ha evaluado también en términos de las curvas ROC. La figura 10 muestra los resultados obtenidos sobre las dos bases de datos.

Globalmente, las figuras 11 y 12 (detalle) muestran los resultados de los tres métodos sobre la base de datos AT&T. La figura 13 muestra los resultados sobre la base de datos LFW.

Figura 10. Resultados de reconocimiento utilizando el método FaceVACS

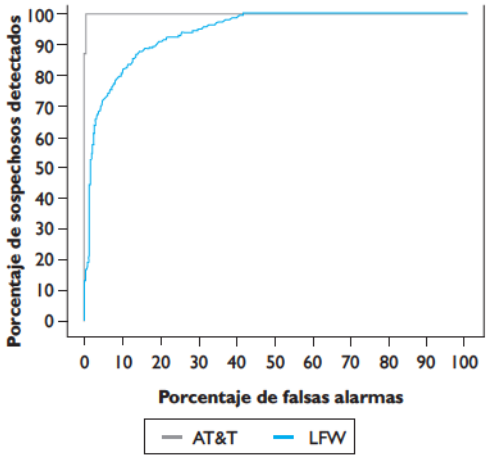


Figura 11. Resultados comparados sobre la base de datos AT&T

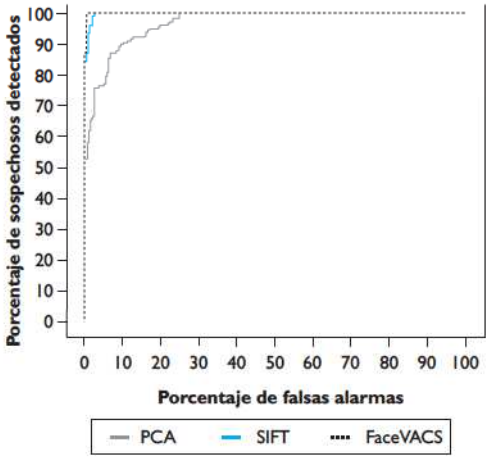
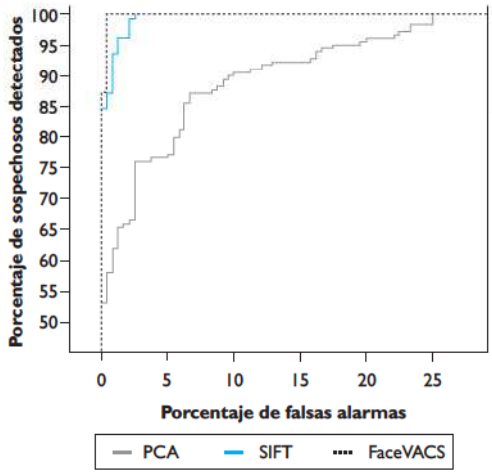


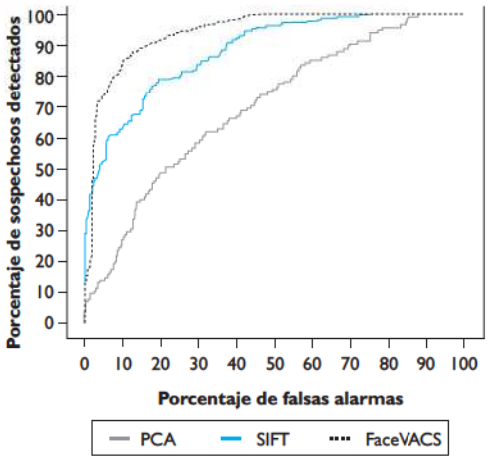
Figura 12. Detalle de los resultados comparados sobre la base de datos AT&T



A la vista de los gráficos mencionados anteriormente, queda patente la superioridad del método FaceVACS sobre el método SIFT, particularmente cuando la calidad de las imágenes es baja (base de datos LFW). También queda patente la superioridad del método SIFT sobre el método PCA, en ambas bases de datos. Como conclusión, se elige el método FaceVACS para el resto de análisis del presente trabajo.

Una vez seleccionado el método FaceVACS, es importante evaluar la influencia de la calidad de las imágenes en los resultados obtenidos. Por una parte,

Figura 13. Resultados comparados sobre la base de datos LFW



queda clara la diferencia entre los resultados obtenidos para ambas bases de datos:

- Para la base de datos AT&T los resultados son prácticamente óptimos. La curva ROC alcanza el 100% de sospechosos detectados con un 0,4% de falsas alarmas, y si se lleva el nivel de falsas alarmas a cero, es esperable un porcentaje de sospechosos detectados del 87%.
- Sin embargo, los resultados para la base de datos LFW son considerablemente peores. El 100% de sospechosos detectados sólo se alcanza si el nivel de falsas alarmas sube hasta un 43%; y si se mantiene el nivel de falsas alarmas cercano a cero, sólo es esperable reconocer a un 10% de los sospechosos.

No es sencillo que la calidad de las imágenes obtenidas en un entorno real de una estación de ferrocarril alcance los niveles de la base de datos AT&T, pero sí es posible que las imágenes sean de mejor calidad que las de la base de datos LFW. Para la base de datos LFW, se han ordenado las imágenes de acuerdo a la calidad de las mismas para un algoritmo de reconocimiento (se ha utilizado un criterio subjetivo) desde la de mejor calidad (imagen 1) hasta la de peor calidad (imagen 10). Este ordenamiento se ha tenido en cuenta a la hora de realizar las pruebas de los algoritmos de reconocimiento, de modo que las imágenes de entrenamiento han sido siempre las 4 mejores imágenes de las 10 disponibles; y las 6 imágenes restantes se han utilizado como imágenes de test. El objetivo es intentar reproducir las condiciones que se encontrarían en un funcionamiento real del sistema de reconocimiento:

- Las imágenes de entrenamiento serían las imágenes almacenadas para cada sospechoso que se desea localizar (por ejemplo, las imágenes de su ficha policial). Es razonable que estas imágenes tengan una buena calidad, porque normalmente se han tomado en condiciones controladas, e incluso se han podido seleccionar las mejores imágenes si hay muchas imágenes disponibles.
- Las imágenes de test serían las que el sistema de supervisión captaría en la propia estación de tren, en situaciones no controladas, con lo cual es de suponer que su calidad sería inferior.

En la base de datos AT&T no se ha considerado necesario realizar este reordenamiento, dado que todas las imágenes tienen una calidad similar para un sistema de reconocimiento.

Concentrándonos en la base de datos LFW, las que hemos considerado peores imágenes de cada sujeto presentan generalmente uno o varios de los siguientes problemas que pueden suponer dificultades para el algoritmo de reconocimiento de caras:

- Punto de vista excesivamente alejado de la vista frontal (en algunos casos, la imagen está tomada completamente de perfil).
- Orientación no vertical de la cara: en algunos casos este efecto es muy acusado.
- Oclusiones que impiden ver gran parte del rostro, por manos u otros elementos.
- Gorras.
- Gafas de sol.
- Varias personas en la misma imagen.
- Gestos muy forzados.

Para evaluar de modo preciso el efecto de la eliminación de las peores imágenes, se realizan nuevas pruebas de reconocimiento. En estas pruebas, se calcula la curva ROC obtenida por el método FaceVACS sobre la

base de datos LFW pero eliminando las peores imágenes de cada sujeto, según el siguiente esquema:

- Experimento 1: no se elimina ninguna imagen de test; se mantienen 5+6+7+8+9+10.
- Experimento 2: se elimina la peor imagen; se mantienen 5+6+7+9.
- Experimento 3: se eliminan las 2 peores imágenes; se mantienen 5+6+7+8.
- Experimento 4: se eliminan las 3 peores imágenes; se mantienen 5+6+7.
- Experimento 5: se eliminan las 4 peores imágenes; se mantienen 5+6.
- Experimento 6: se eliminan las 5 peores imágenes; se mantiene sólo la imagen 5.

Figura 14. Resultados método FaceVACS eliminando las peores imágenes

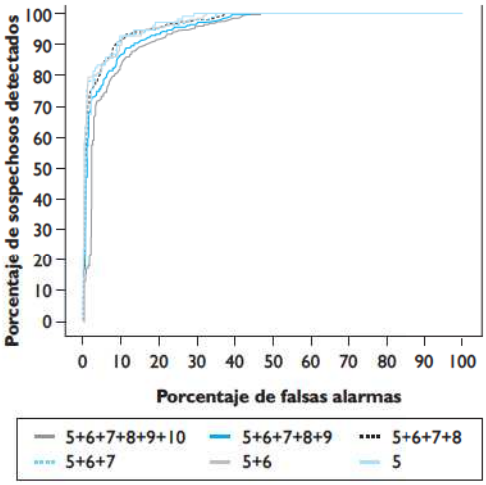
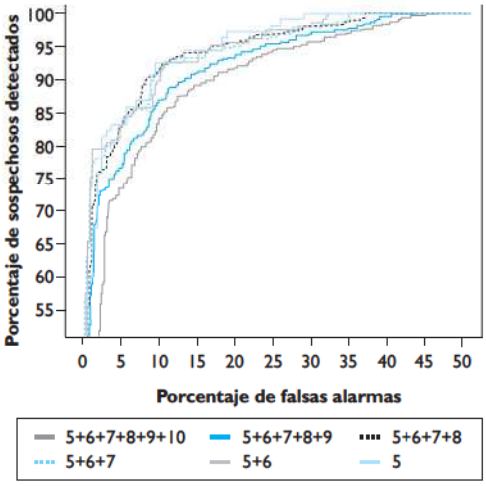


Figura 15. Resultados método FaceVACS eliminando las peores imágenes (detalle)



Los resultados se muestran en las figuras 14 y 15 (detalle).

Los resultados muestran como, efectivamente, si se elimina la que subjetivamente se considera la peor de las imágenes (imagen 10) para todos los sujetos, los resultados mejoran notablemente. Si adicionalmente se elimina la segunda peor imagen (imagen 9), los resultados vuelven a mejorar apreciablemente. Si se continúan eliminando imágenes, los resultados apenas varían.

Esta información puede ser de gran utilidad para determinar cuál debe ser la calidad de las imágenes necesaria para un adecuado funcionamiento del sistema.

6. VIABILIDAD DEL SISTEMA PROPUESTO

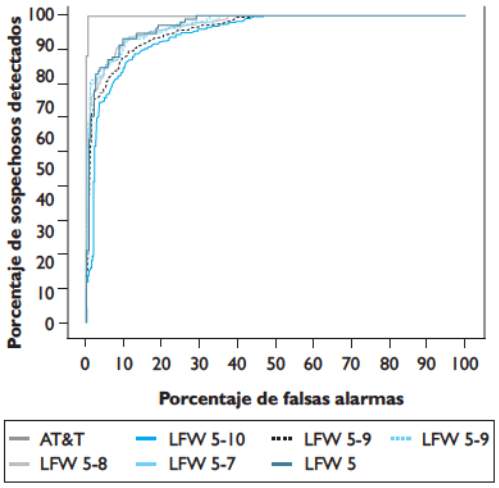
6.1. Criterios definidos

En primer lugar, hemos fijado una tasa de procesamiento de 1Hz. Esto quiere decir que se procesa una imagen cada segundo. De este modo, suponiendo un ritmo normal en el tránsito de las personas frente a la cámara, cada persona aparecerá en al menos 3 imágenes. Este dato ha de ser tenido en cuenta a la hora de calcular la tasa de sospechosos que pueden pasar inadvertidos: la probabilidad de que un sospechoso pase inadvertido en las tres imágenes que se capturan de él será el producto de las probabilidades individuales.

Por ejemplo, si el sistema se ajusta para una tasa de reconocimiento de sospechosos del 70%, la probabilidad de que un sospechoso pase inadvertido en una imagen sería del 30%; y la probabilidad global de que pase inadvertido sería $0{,}3^3 = 0{,}027 = 2{,}7\%$ (el 97,3% de los sospechosos serían detectados).

La figura 16 muestra conjuntamente los resultados obtenidos por el mejor método probado (método FaceVACS) sobre la base de datos AT&T y sobre la base de datos LFW, en este último caso contemplando la posibilidad de eliminar las peores imágenes. En los apartados siguientes, se determinan los posibles escenarios a contemplar en función

Figura 16. Resultados de reconocimiento utilizando el método FaceVACS



de los requerimientos que se deseen imponer y se estudian los resultados que se podrían obtener sobre detalles de la figura anterior.

6.2. Escenario 1: mínimas falsas alarmas

En este escenario de funcionamiento, el requisito fundamental es que el número de falsas alarmas sea cero o el mínimo posible. La idea fundamental es no cargar de trabajo al vigilante, que en condiciones normales tiene que verificar cada una de las falsas alarmas.

Sobre el gráfico de la figura 16, buscamos el punto correspondiente al 0% de falsas alarmas o, en el caso de no existir, el punto mínimo. Pueden apreciarse los siguientes resultados, en función de la base de datos contemplada (se resumen en la tabla 4):

- Sobre la base de datos AT&T, es posible conseguir un 87% de reconocimiento de sospechosos manteniendo el nivel de falsas alarmas en cero.
- Sobre la base de datos LFW completa (incluyendo las peores imágenes), no es posible conseguir un nivel cero para las falsas alarmas. El valor mínimo sería un 0,1% de falsas alarmas, correspondiente a una falsa alarma cada 1000 segundos (16 minutos). En cualquier caso, para este valor sólo sería esperable reconocer a un 10% de los sospechosos.
- Si se eliminan las peores imágenes de la base de datos LFW, la situación mejora considerablemente:

  — Eliminando la peor imagen, es posible mantener el nivel de falsas alarmas a cero y conseguir un 15% de reconocimientos.
  — Eliminando las dos peores imágenes, este porcentaje es del 16%.
  — Eliminando las tres peores imágenes, el porcentaje sube al 19%.
  — Eliminando las cuatro peores imágenes, se obtiene un 20%.
  — Por último, eliminando las 5 peores imágenes se obtiene un 23%.

A la vista de los resultados de la tabla, queda claro que podría ponerse a funcionar el sistema con una tasa de falsas alarmas nula, pero sólo en el caso de que las imágenes tuviesen una calidad similar a las de la base de datos AT&T (suponiendo que cada persona aparece en tres imágenes diferentes mientras pasa delante de la cámara, el porcentaje de sospechosos no detectados sería únicamente del 0,2%). Dado que imágenes con tan alta calidad no son esperables en un entorno real, nos

Tabla 4

RESULTADOS PRIMER ESCENARIO (FALSAS ALARMAS NULAS O MÍNIMAS)

Base de datos	Falsas alarmas	Sospechosos reconocidos	No detectados (1 imagen)	No detectados (3 imagen)
AT&T	0%	87%	13%	0,2%
LFW (5)	0%	23%	77%	45,6%
LFW (5-6)	0%	20%	80%	51,2%
LFW (5-7)	0%	19%	81%	53,1%
LFW (5-8)	0%	16%	84%	59,3%
LFW (5-9)	0%	15%	85%	61,4%
LFW (5-10)	0,1%	10%	90%	72,9%

planteamos proponer escenarios menos restrictivos en cuanto al porcentaje de falsas alarmas.

6.3. Escenario 2: falsas alarmas a nivel 0,25%

En primer lugar, veamos qué frecuencia media de falsas alarmas sería esperable en este escenario. Tal y como se definió en el apartado 7.5, la tasa de procesamiento se establece en 1Hz (una imagen procesada por segundo). Por tanto, un 0,25% de falsas alarmas quiere decir que en término medio se generará una falsa alarma cada 400 imágenes procesadas, esto es, cada 400 segundos o cada 6 minutos y 40 segundos.

Consideramos que esta tasa de falsas alarmas es completamente admisible para el personal de vigilancia. Debe ser tenido en cuenta que una falsa alarma aparecería en la pantalla de ordenador del personal de vigilancia de un modo que sería fácil de verificar: la imagen capturada en tiempo real se mostraría frente a las imágenes almacenadas para el sospechoso que se ha creído detectar. En cuestión de segundos, el operario podría comprobar si se trata o no de una falsa alarma. Los resultados esperables en cuanto a porcentaje de sospechosos reconocidos se resumen en la tabla 5.

A la vista de los datos de la tabla, la situación en este escenario, en cuanto a tasa de reconocimiento es similar a la del escenario anterior. Sólo podría aplicarse el sistema si las imágenes tuviesen una calidad similar a la de la base de datos AT&T, lo cual es poco realista. Como dato a tener en cuenta, en una de las versiones simplificadas de la base de datos LFW (en la que se eliminan las imágenes 8, 9 y 10, las 3 con una calidad aparentemente más baja) se consiguen resultados que harían al sistema aceptable (sólo un 7,9% de sospechosos no detectados). Sin embargo, este dato no puede considerarse fiable, dado que cuando se eliminan incluso más imágenes (cuando se considera únicamente la imagen 5 o bien las imágenes 5 y 6) estos buenos resultados no se mantienen.

En conclusión, este escenario sólo podría emplearse con imágenes de muy alta calidad, similares a las de la base de datos AT&T.

6.4. Escenario 3: falsas alarmas a nivel 0,50%

Como en el caso anterior, lo primero es traducir el porcentaje de falsas alarmas a términos de frecuencia. Un nivel 0,50% equivale a decir que se produciría una falsa alarma cada 200 segundos o, lo que es lo mismo, cada 3 minutos 20 segundos (en promedio). También consideramos que se trata de una tasa razonable para que un vigilante de seguridad la pueda asumir. Los resultados resumidos los mostramos en la tabla 6.

Para este nivel de falsas alarmas, los resultados sobre la base de datos AT&T harían esperable el correcto reconocimiento de todos los sospechosos. En cuanto a la base de datos LFW, descartando el caso del utilizar sólo la imagen 5 de esta base de datos (resultado poco fiable, que no sigue el patrón de los demás), podrían considerarse resultados adecuados los obtenidos siempre y cuando no se utilicen las dos peores imágenes (imagen 9 e imagen 10). Esto quiere decir que se podría aplicar el sistema de vigilancia a condición de que la calidad de las imágenes sea al menos la misma que la presente en las imágenes 5 a 8 de la base de

Tabla 5

RESULTADOS SEGUNDO ESCENARIO (FALSAS ALARMAS A NIVEL 0,25%)

Base de datos	Falsas alarmas	Sospechosos reconocidos	No detectados (1 imagen)	No detectados (3 imagen)
AT&T	0,25%	87%	13%	0,2%
LFW (5)	0,25%	23%	77%	45,6%
LFW (5-6)	0,25%	21%	79%	49,3%
LFW (5-7)	0,25%	57%	43%	7,9%
LFW (5-8)	0,25%	18%	82%	55,1%
LFW (5-9)	0,25%	17%	83%	57,2%
LFW (5-10)	0,25%	13%	87%	65,8%

Tabla 6

RESULTADOS TERCER ESCENARIO (FALSAS ALARMAS A NIVEL 0,50%)

Base de datos	Falsas alarmas	Sospechosos reconocidos	No detectados (1 imagen)	No detectados (3 imagen)
AT&T	0,50%	100%	0%	0%
LFW (5)	0,50%	23%	77%	45,6%
LFW (5-6)	0,50%	57%	43%	7,9%
LFW (5-7)	0,50%	63%	37%	5,1%
LFW (5-8)	0,50%	50%	50%	12,5%
LFW (5-9)	0,50%	18%	82%	55,1%
LFW (5-10)	0,50%	14%	86%	63,6%

Tabla 7

RESULTADOS CUARTO ESCENARIO (FALSAS ALARMAS A NIVEL 0,75%)

Base de datos	Falsas alarmas	Sospechosos reconocidos	No detectados (1 imagen)	No detectados (3 imagen)
AT&T	0,75%	100%	0%	0%
LFW (5)	0,75%	60%	40%	6,4%
LFW (5-6)	0,75%	65%	35%	4,3%
LFW (5-7)	0,75%	64%	36%	4,7%
LFW (5-8)	0,75%	56%	44%	8,5%
LFW (5-9)	0,75%	41%	59%	20,5%
LFW (5-10)	0,75%	17%	83%	57,2%

datos LFW, y los resultados esperables serían un reconocimiento del 50% de sospechosos (considerando una sola imagen) o del 88% (considerando las 3 imágenes). En estas condiciones consideramos que el sistema sería aplicable y resultaría beneficioso para la seguridad, pero en cualquier caso, en los apartados siguientes se analizan otros escenarios.

6.5. Escenario 4: falsas alarmas a nivel 0,75%

Continuando con la suposición de que la frecuencia de procesamiento de imágenes es de 1Hz, un 0,75% se corresponde con un promedio de una falsa alarma cada 133 segundos o, lo que es lo mismo, algo menos de una alarma cada 2 minutos. Sigue siendo una frecuencia razonable para ser atendida por el vigilante, aunque su nivel de atención debe ser mayor (debe dedicarse casi con exclusividad a vigilar las falsas alarmas).Los resultados de reconocimiento, que se resumen de nuevo en la tabla 7.

A la vista de los resultados de la tabla 7, un sistema de vigilancia de este tipo sería viable siempre que la calidad de las imágenes fuese al menos equivalente a la de las imágenes 5-8 de la librería LFW (en el peor de los casos, dejaría de detectarse al 8,5% de los sospechosos). Esta empieza a ser una situación bastante factible con una adecuada colocación de las cámaras de video en la estación de ferrocarril.

En cualquier caso, en los siguientes apartados se estudian situaciones con mayor porcentaje de detección de sospechosos, aún a cambio de aumentar las tasas de falsas alarmas.

6.6. Escenario 5: falsas alarmas a nivel 1,00%

La equivalencia en frecuencia promedio de falsas alarmas es, en este caso, de una falsa alarma cada 100 segundos (aproximadamente, cada minuto y medio). Consideramos que sigue siendo asumible por un vigilante de seguridad aunque, evidentemente, su dedicación debe ser exclusiva a esta tarea. Los resultados resumidos se muestran en la tabla 8.

A la vista de los resultados de la tabla, el sistema se muestra válido para todas las imágenes salvo para la décima imagen de la

Tabla 8

RESULTADOS QUINTO ESCENARIO (FALSAS ALARMAS A NIVEL 1,00%)

Base de datos	Falsas alarmas	Sospechosos reconocidos	No detectados (1 imagen)	No detectados (3 imagen)
AT&T	1,00%	100%	0%	0%
LFW (5)	1,00%	60%	40%	6,4%
LFW (5-6)	1,00%	75%	25%	1,6%
LFW (5-7)	1,00%	73%	27%	2,0%
LFW (5-8)	1,00%	62%	38%	5,5%
LFW (5-9)	1,00%	53%	47%	10,4%
LFW (5-10)	1,00%	17%	83%	57,2%

Tabla 9

RESULTADOS SEXTO ESCENARIO (FALSAS ALARMAS A NIVEL 1,50%)

Base de datos	Falsas alarmas	Sospechosos reconocidos	No detectados (1 imagen)	No detectados (3 imagen)
AT&T	1,50%	100%	0%	0%
LFW (5)	1,50%	67%	33%	3,6%
LFW (5-6)	1,50%	80%	20%	0,8%
LFW (5-7)	1,50%	78%	22%	1,1%
LFW (5-8)	1,50%	71%	29%	2,4%
LFW (5-9)	1,50%	60%	40%	6,4%
LFW (5-10)	1,50%	20%	80%	51,2%

Tabla 10

RESULTADOS SÉPTIMO ESCENARIO (FALSAS ALARMAS A NIVEL 2,00%)

Base de datos	Falsas alarmas	Sospechosos reconocidos	No detectados (1 imagen)	No detectados (3 imagen)
AT&T	2,00%	100%	0%	0%
LFW (5)	2,00%	76%	24%	1,4%
LFW (5-6)	2,00%	80%	20%	0,8%
LFW (5-7)	2,00%	78%	22%	1,1%
LFW (5-8)	2,00%	76%	24%	1,4%
LFW (5-9)	2,00%	70%	30%	2,7%
LFW (5-10)	2,00%	40%	60%	21,6%

librería LFW (la peor de todas). Si no existiesen imágenes de tan mala calidad, cabría esperar que sólo el 10,4% de los sospechosos pasarían sin ser detectados, en el peor de los casos.

6.7. Escenario 6: falsas alarmas a nivel 1,50%

La frecuencia media de falsas alarmas es, en este caso, de una alarma cada 67 segundos (aproximadamente, una alarma por minuto). Todavía puede tratarse de un escenario admisible, aunque cerca del límite. Los resultados de reconocimiento se muestran en la tabla 9.

En este escenario, las conclusiones son similares a las del escenario anterior: es posible implementar el sistema con éxito siempre que las imágenes sean de calidad superior a la peor imagen de la librería LFW. En estas condiciones, cabe esperar que la tasa de sospechosos sin detectar no supere el 6,4%.

6.8. Escenario 7: falsas alarmas a nivel 2,00%

La frecuencia de falsas alarmas se eleva en este caso a una falsa alarma cada 50 segundos (más de una falsa alarma por minuto). Aunque se trata de una tasa muy

Tabla 11

RESULTADOS OCTAVO ESCENARIO (FALSAS ALARMAS A NIVEL 2,50%)

Base de datos	Falsas alarmas	Sospechosos reconocidos	No detectados (1 imagen)	No detectados (3 imagen)
AT&T	2,50%	100%	0%	0%
LFW (5)	2,50%	76%	24%	1,4%
LFW (5-6)	2,50%	80%	20%	0,8%
LFW (5-7)	2,50%	78%	22%	1,1%
LFW (5-8)	2,50%	76%	24%	1,4%
LFW (5-9)	2,50%	73%	27%	2,0%
LFW (5-10)	2,50%	59%	41%	6,9%

Tabla 12

SOSPECHOSOS NO DETECTADOS EN CADA ESCENARIO Y CADA BASE DE DATOS

Escenario	1	2	3	4	5	6	7	8
Porcentaje de falsas alarmas	0%	0,25%	0,50%	0,75%	1,00%	1,50%	2,00%	2,50%
Segundos entre falsas alarmas	Infinito	400 s	200 s	133 s	100 s	67 s	50 s	40 s
AT&T	0,2%	0,2%	0%	0%	0%	0%	0%	0%
LFW (5)	45,6%	45,6%	45,6%	6,4%	6,4%	3,6%	1,4%	1,4%
LFW (5-6)	51,2%	49,3%	7,9%	4,3%	1,6%	0,8%	0,8%	0,8%
LFW (5-7)	53,1%	7,9%	5,1%	4,7%	2,0%	1,1%	1,1%	1,1%
LFW (5-8)	59,3%	55,1%	12,5%	8,5%	5,5%	2,4%	1,4%	1,4%
LFW (5-9)	61,4%	57,2%	55,1%	20,5%	10,4%	6,4%	2,7%	2,0%
LFW (5-10)	72,9%	65,8%	63,6%	57,2%	51,2%	51,2%	21,6%	6,9%

elevada, se muestran los datos de reconocimiento obtenibles en la tabla 10.

Los valores de la tabla 10 muestran que, llegado este escenario, sería incluso factible un sistema de supervisión en el que las imágenes pudieran tener una calidad tan escasa como las peores imágenes de la librería LFW. Los peores resultados esperables serían una tasa de sospechosos no reconocidos del 21,6% (el sistema detectaría al 78,4% de los sospechosos).

6.9. Escenario 8: falsas alarmas a nivel 2,50%

Éste es el último escenario que consideramos, con una frecuencia promedio de aparición de falsas alarmas equivalente a una alarma cada 40 segundos. No se consideran escenarios con mayor tasa de falsas alarmas por considerarlos completamente inviables. Los resultados de reconocimiento se muestran en la tabla 11.

Queda patente que en este escenario las tasas de reconocimiento serían suficientemente fiables para trabajar con cualquier tipo de imágenes, incluso de muy baja calidad como las peores de la librería LFW (la peor tasa de sospechosos no detectados sería de un 6,9%, lo cual quiere decir que se detectarían el 93,1% de los sospechosos).

6.10. Resumen de todos los escenarios

Como resumen de todos los escenarios, se muestra la tabla 12, donde únicamente se indican los porcentajes esperables de sospechosos no detectados en función de la base de datos contemplada y de la tasa de falsas alarmas. En esta tabla se han marcado en negrita todos aquellas situaciones que se consideran viables para un sistema de vigilancia. Hemos establecido un límite del 10% de sospechosos sin detectar. Consideramos que un sistema de ayuda a la vigilancia que detecte al menos al 90% de los sospechosos ya resulta una herramienta de gran utilidad, por supuesto, como

complemento de otras medidas de seguridad.

De acuerdo con los datos de la tabla, podríamos concluir lo siguiente:

- Si se consigue que las imágenes captadas por las cámaras de seguridad en la estación de ferrocarril sean tan homogéneas y de calidad tan elevada como las de la base de datos AT&T, entonces el sistema de vigilancia sería viable incluso con una tasa de falsas alarmas igual o muy próxima a cero.
- Si al menos se consigue que las imágenes tengan una calidad similar a la de las imágenes 5 a 8 de la base de datos LFW (se excluyen las dos peores imágenes) entonces el sistema sería viable desde una tasa de falsas alarmas de un 0,75% o, lo que es lo mismo, con un periodo esperable entre falsas alarmas algo superior a los dos minutos (133 segundos). Consideramos que se trata de un valor razonable, y que un operario de vigilancia podría manejar tales falsas alarmas sin dificultad.
- Si se consigue que la calidad de las imágenes sea similar a la de las imágenes 5 a 9 de la base de datos LFW (sólo se excluye la peor imagen), entonces el sistema sólo sería viable a partir de una tasa de falsas alarmas del 1,50%, lo que equivale a 67 segundos entre dos falsas alarmas. Es posible que un operario pueda manejar estas falsas alarmas, pero es posible que en algunos casos se presentasen dificultades (dado que el periodo entre falsas alarmas es un valor medio, sería probable que en algunas ocasiones las falsas alarmas se encontrasen más próximas entre sí, dificultando la labor del operario).
- Por último, si la calidad de las imágenes llega a ser tan baja como la de las peores imágenes de la librería LFW (imagen número 10), entonces el sistema sólo sería viable a partir de una tasa de falsas alarmas del 2,5%, que equivale a un periodo promedio de sólo 40 segundos entre dos falsas alarmas. Consideramos que esta tasa de falsas alarmas no sería asumible por un operario.

6.11. Viabilidad en cuanto a tiempo de cómputo

Otro de los factores a tener en cuenta para establecer la viabilidad del sistema de vigilancia propuesto es el tiempo de cómputo. El tiempo de procesamiento total es la suma del tiempo empleado en la detección de la cara en la imagen más el tiempo necesario para el reconocimiento de la misma. En todos los casos nos ceñiremos a las prestaciones obtenidas sobre un ordenador estándar (en concreto, un PC estándar con procesador Intel Quad core 2.0 GHz, 2GB de RAM y sistema operativo Windows XP).

En cuanto al tiempo de detección, los resultados indicados en el apartado 5.5 del presente informe nos indican que sobre un tamaño de imagen adecuado (400x200 píxeles) el tiempo de procesamiento medio es de 63 ms.

En cuanto al tiempo de detección, y de acuerdo con los valores indicados en la tabla 3, el algoritmo FaceVACS necesita en término medio 201 ms para comparar una imagen frente a la base de datos LFW y 177 ms para comparar una imagen frente a la base de datos AT&T.

Por lo tanto, en el peor de los casos evaluados en este informe, el tiempo total de procesamiento sería de 63+201 = 264 ms; suficientemente rápido como para procesar las imágenes no sólo a 1Hz como estaba previsto sino a frecuencias cercanas a los 4Hz (4 imágenes procesadas por segundo). Si mantenemos la frecuencia de 1Hz como velocidad adecuada de funcionamiento, estos resultados nos indican que podríamos trabajar sobre una base de datos de sospechosos muy superior al tamaño de la base de datos LFW utilizada (158 sujetos). En principio, y suponiendo un aumento lineal del tiempo de cómputo en función del número de sujetos (suposición razonable), el tiempo total de procesamiento alcanzaría 1000 ms para un número de sujetos superior a 3600.

7. CONCLUSIONES

De acuerdo con los resultados mostrados en los apartados anteriores, la viabilidad del sistema depende altamente de la

calidad de las imágenes que se puedan captar en la estación de tren, con lo cual este estudio debe ser contrastado con experimentos in situ donde se capten imágenes en el entorno real de la estación de tren. No obstante, y a partir de las bases de datos utilizadas, nuestra conclusión es que el sistema sería viable con una colocación adecuada de las cámaras de video en puntos estratégicos de la estación (en las puertas de entrada, en el acceso a andenes, en las taquillas, o incluso sobre los arcos de detección de metales).

Consideramos que debería realizarse un estudio posterior en el que se analizasen todas las ubicaciones posibles para las cámaras de video en las estaciones de tren y se registrasen imágenes durante el funcionamiento habitual de la estación, para poder comparar la calidad de estas imágenes frente a las utilizadas en el presente estudio.

Por otra parte, si estas imágenes de prueba se capturasen de forma metódica, con una serie de sujetos identificados que harían las veces de sospechosos, se podría utilizar el trabajo realizado en el presente estudio (en cuanto a la preparación de los algoritmos de detección y reconocimiento de imágenes) para obtener nuevas estadísticas de funcionamiento, esta vez sobre imágenes reales. Los resultados obtenidos en este estudio posterior servirían para comprobar sobre el entorno real la viabilidad de la solución propuesta.

8. REFERENCIAS

[1] ZHAO, W.; CHELLAPPA, R.; ROSENFELD, A. y PHILLIPS, P.J. (2003). Face Recognition: A Literature Survey, ACM Computing Surveys, pp. 399-458.

[2] FRVT 2006 and ICE 2006 Large-Scale Results. PHILLIPS, P.J.; TODD SCRUGGS, W.; O'TOOLE, Alice J.; FLYNN, Patrick J.; BOWYER, Kevin W.; SCHOTT, Cathy L. y SHARPE, Matthew. www.frvt.org.

[3] The NIST HumanID Evaluation Framework, by MICHEALS, R.J.; GROTHER, P.J. y PHILLIPS, P.J. Proceedings 4th International Conference on Audio Visual Based Person Authentication (2003).

[4] Face Recognition Vendor Test 2002 Performance Metrics, by GROTHER, P.J.; MICHEALS, R.J. y PHILLIPS, P.J. Proceedings 4th International Conference on Audio Visual Based Person Authentication (2003).

[5] Fraunhofer Institute, www.iis.fraunhofer.de.

[6] VIOLA, P. y JONES, M.J. (2004). Robust Real-Tine Face Detection. International Journal of Computer Vision, 57, 2, 137-154, ISSN 0920-5691.

[7] Luxand development, www.luxand.com.

[8] Pittsburg Pattern Recognition, www.pittpatt.com.

[9] Cognitec, www.cognitec.com.

[10] TURK, M. y PENTLAND, A. Eigenfaces for recognition. J. Cognitive Neuroscience, 3, 1, 71-86, ISSN 0898-929X (1991).

[11] BELHUMEUR, P.N.; HESPANHA, J.P. y KRIEGMAN, D.J. (2002). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Analysis and Machine Intelligence, 19, 7, 711-720, ISSN 0162-8828.

[12] LIU, C. Face Enhanced independent component analysis and its application to content based face image retrieval. IEEE. Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, 34, 2, 1117-1127, ISSN 1083-4419 (2004).

[13] BICEGO, M.; LAGORIO, A.; GROSSO, E. y TISTARELLI, M. (2006). On the use of SIFT features for face authentication, Proceedings of Conf. on Computer Vision and Pattern Recognition Workshop, pp. 35, ISBN 0-7695-2646-2, New York, NY, USA, IEEE.

[14] IVAN, Paul (2007). Active Appearance Models for Face Recognition. PhD Thesis,Vrije Universiteit, Amsterdam.

[15] VICENTE, M.A.; HOYER, P.O. y HYVARINEN, A. (2007). Equivalence of Some Common Linear Feature Extraction Techniques for Appearance-Based Object Recognition Tasks. IEEE Trans. Pattern Analysis and Machine Intelligence, 29, 5, 896-900, ISSN 0162-8828.

[16] FERNANDEZ, C.; VICENTE, M.A. (2008). «Face recognition using multiple interest point detectors and SIFT descriptors», 8th IEEE International Conference on Automatic Face and Gesture Recognition.

[17] Face Recognition Homepage, www.face-rec.org.

[18] The AR Face Database, http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html.

[19] FERET Face Database, http://face.nist.gov/colorferet.

[20] AT&T, The Database of Faces (formerly «The ORL Database of Faces»), www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[21] University of Massachusetts, Labeled Faces in the Wild, vis-www.cs.umass.edu/lfw.