Ingeniería Investigación y Tecnología, volumen XVII (número 2), abril-junio 2016: 257-271
ISSN 1405-7743 FI-UNAM
(artículo arbitrado)
doi:10.1016/j.riit.2016.06.010

Re-identificación de personas a través de sus características soft-biométricas en un entorno multi-cámara de video-vigilancia

People Re-Identification Based on Soft-Biometrics Features in Multi-Camera Surveillance Scenario

Moctezuma-Ochoa Daniela Alejandra
Catedrática CONACYT
Laboratorio Nacional de Geo-Inteligencia Territorial
Centro GEO, México
Correo: dmoctezuma@centrogeo.edu.mx

Información del artículo: recibido: mayo de 2015, reevaluado: agosto de 2015, aceptado: septiembre de 2015

Resumen

Los sistemas de video vigilancia inteligente se convirtieron en un importante tema de investigación en los últimos años debido a su importancia para el sector de la seguridad. En este artículo se presenta un sistema completo de video vigilancia inteligente. Este sistema comprende dos importantes módulos: detección de personas y re-identificación de personas mediante características soft-biométricas. Estas dos fases en conjunto proporcionan un sistema completo para la re-identificación de personas en escenarios de video vigilancia, específicamente en entornos multi-cámara. Los resultados de cada etapa del sistema, así como los resultados finales, muestran la importancia de las características soft-biométricas, así como un gran resultado en tasa de reconocimiento en diversas bases de datos públicas y estándares en la comunidad científica de video vigilancia inteligente. Además se presenta una base de datos propia adquirida en el aeropuerto de Madrid, Barajas.

Descriptores:

- video vigilancia inteligente
- re-identificación de personas
- características soft-biométricas
- detección de personas
- aprendizaje incremental

Abstract

In last years, the intelligent video surveillance systems have become an important research topic due to relevance for security sector. In this paper, a full intelligent video surveillance system is presented. This system is conforming by two important modules: people detection and people re-identification based on soft-biometrics features. These two modules provide a whole system for people re-identification in multi-camera surveillance environment. The results show the relevance and potential of soft-biometrics features as well as a promising identification rate result in several standard and public databases and moreover with an own database acquired at International Barajas airport in Madrid.

Keywords:

- *intelligent video surveillance systems*
- *people re-identification*
- *soft-biometrics features*
- *people detection*
- *incremental learning*

Introducción

La necesidad de incrementar la seguridad se deja sentir en todo el mundo, no solo por compañías privadas sino también por los gobiernos y las instituciones públicas. Debido a esto, últimamente los sistemas de video vigilancia inteligente se han convertido en una importante área de investigación gracias a su aplicación en el sector de la seguridad. La video vigilancia es, de hecho, una tecnología clave para la lucha contra el terrorismo y el crimen, además es de gran ayuda en la seguridad pública. Tratando de responder a estas necesidades de seguridad, la comunidad científica se centra en detectar, seguir e identificar a las personas, así como identificar su comportamiento (Alahi *et al*., 2010). La video vigilancia tradicional consiste en monitorear el comportamiento, actividades y otros cambios en el entorno, por medio de operadores visuales del sistema que intentan proporcionar seguridad a la zona vigilada. La video vigilancia es muy útil para el gobierno, el cumplimiento de las leyes y para mantener el control social, reconocer y monitorear amenazas y prevenir o investigar actividades criminales (Union, 2009). En términos generales, la video vigilancia de una zona amplia y crítica requiere de un sistema de múltiples cámaras que sirvan para monitorear a las personas de forma constante (Huang *et al*., 2008).

Por otro lado, la video vigilancia inteligente surgió en los últimos años debido al aumento en el número de cámaras instaladas en diferentes edificios, así como a la necesidad de mayor seguridad y a la aceptación, cada vez mayor, de la sociedad hacia este tipo de sistemas. A diferencia de la video vigilancia tradicional, la video vigilancia inteligente es más que un conjunto de monitores conectado a varias cámaras, de hecho, actualmente esta se puede considerar como una poderosa tecnología para el control de la seguridad. La principal diferencia entre el sistema tradicional y el sistema de video vigilancia inteligente radica en el análisis automático de la escena, dicho análisis automático puede comprender diferentes tareas importantes para la seguridad en general, algunas de estas tareas son: detección, seguimiento e identificación de personas, detección de objetos abandonados, análisis del comportamiento de las personas, análisis de las trayectorias, entre otras (García y Martínez, 2010). La identificación de personas es una de las principales tareas de un sistema de video vigilancia inteligente, para ello, normalmente se utilizan características biométricas. Las características biométricas son aquellos rasgos fisiológicos que hacen únicos a los seres humanos como la cara, la huella dactilar, la voz, la retina, etcétera. Sin embargo, recientemente se definieron otro tipo de características, las cuales se consideran adecuadas para las condiciones críticas a las que se enfrentan los sistemas de video vigilancia inteligente. Estas características son las soft-biométricas, que a pesar de no tener el alto poder discriminatorio de las biométricas, son de gran ayuda para definir la identidad de las personas. Algunas de estas características son por ejemplo: el color de piel, de cabello, la ropa, la altura, la forma de andar, cicatrices, tatuajes, etcétera. Por lo tanto, las características soft-biométricas son aquellas que proveen cierta información sobre las personas, pero que carecen de suficiente permanencia y de un alto nivel distintivo para diferenciar a un individuo de otro, especialmente cuando se utiliza cada una de forma separada (Jain *et al*., 2004). Es decir, a mayor cantidad de características soft-biométricas utilizadas, mejor será el rendimiento del sistema. Por ejemplo, el color de cabello por sí solo no ayuda a definir la identidad de una persona, pero si se combina el color de cabello con la altura, aumentará el poder distintivo para diferenciar a una persona de otra. Estas características son idóneas para las situaciones complejas que se encuentran en las zonas de video vigilancia, como es la adquisición a distancia de las características, el hecho de que la persona no está consciente de que es observada (adquisición no intrusiva), la baja resolución y la mala calidad de las imágenes, los cambios de iluminación, el gran cambio de apariencia entre las diversas cámaras, el zoom, la variación en la perspectiva, entre otros. En la figura 1 se muestran algunas imágenes de un sistema de cámaras de video vigilancia. En esta figura, se puede observar la mala calidad en las imágenes que conlleva un sistema multi-cámara de este tipo.

En este artículo se presenta un sistema completo de video vigilancia inteligente. Este sistema realiza la re-identificación de personas en un entorno de múltiples

Figura 1. Imágenes de ejemplo donde se puede observar la baja calidad, mala iluminación, zoom, etcétera

cámaras; es decir, una vez que se ha identificado en un inicio (o en un punto determinado) a una persona, se busca volver a identificarla a lo largo de su estancia o recorrido en la zona vigilada. Para llevar a cabo esta labor, el desarrollo y funcionamiento de este sistema comprende dos principales etapas de desarrollo. Primero, se propone un método de detección de personas, ya que en primer lugar se debe determinar qué de la escena es una persona y qué no. En segundo lugar, se propone un modelo de apariencia para la identificación de las personas, basado en sus características soft-biométricas. Para los experimentos se utilizaron varias bases de datos públicas y estándar en la comunidad científica de video vigilancia inteligente: PETS 2006, PETS 2007, PETS 2009 y CAVIAR. Además, se adquirió una base de datos propia, MUBA (*Multi-Camera Barajas Airport*), en el aeropuerto de Madrid, Barajas, con la colaboración del personal de AENA y de la guardia civil española.

Antes de pasar a la descripción de los métodos propuestos y comparados, en la tabla 1 se muestran las principales características de las bases de datos utilizadas en este trabajo. Aquí se detallan varias características importantes como: la resolución, escenario y número de imágenes utilizadas de cada base de datos. El número de imágenes utilizadas varía dependiendo el tipo de entorno, mono o multi-cámara, en el caso de la detección de personas solo se utilizaron las imágenes mono-cámara. Para la etapa de identificación con soft-biométricos se utilizaron las imágenes del entorno multi-cámara.

Sistema propuesto

Como se mencionó previamente, para el desarrollo de este sistema se llevaron a cabo dos principales etapas: detección de personas y la generación de un modelo de apariencia basado en soft-biométricos. Todos los métodos propuestos en este sistema se probaron con diversas bases de datos públicas y conocidas a nivel internacional por la comunidad científica de video vigilancia inteligente. Además, se adquirió una base de datos en el aeropuerto Internacional de Madrid, MUBA, con un total de ocho cámaras colocadas a lo largo de la terminal 4.

Detección de personas

El método que se propone para la detección de personas se basa en la combinación del método HOG (*histograma de orientaciones del gradiente*) y en los filtros de Gabor. Se decidió utilizar el método HOG porque es uno de los métodos de detección de objetos más utilizados, es una herramienta base para proponer nuevas técnicas en la literatura. Por otro lado, la utilización de los filtros de Gabor se propone para realzar las características principales que representan a la figura humana.

En esta sección se describen los métodos alternativos que se seleccionaron como base para una comparación del desempeño del método propuesto. Estos métodos alternativos se seleccionaron porque que reportan buenos resultados y se utilizan en la literatura. Para las pruebas de todos los métodos se emplearon cuatro bases de datos públicas: PETS 2006, PETS 2007, PETS 2009 y CAVIAR. A continuación se describe cada aspecto del método de detección de personas propuesto, HoGG (*histograma de orientaciones del gradiente con Gabor*).

Método HoGG

El método HoGG, se basa en la combinación del método *Histogram of Oriented Gradients* (Dalal y Triggs, 2005) y los filtros de Gabor (Lades *et al*., 1993). Los efectos del pre-procesamiento, utilizando los filtros de Gabor, se analizan a detalle, así como la mejora experimentada con el realce de la información en las imágenes y la influencia que se ejerce sobre las características extraídas. En la figura 2, se presenta el esquema general del funcionamiento del método HoGG. Se pueden ver los diferentes pasos para detectar a una persona.

Primero se detectan los objetos en movimiento en la imagen, posteriormente se hace la substracción del fondo de la imagen, una vez que se tienen los objetos en movimiento se utilizan los filtros de Gabor para preprocesarlos, después se extraen las características con el algoritmo HOG, y finalmente con un clasificador simple se determina si el objeto en movimiento es una persona o no.

Para entender mejor el funcionamiento del método propuesto HoGG es necesario conocer tanto los filtros de Gabor como el método base HOG. Más adelante se describirán con más detalle los dos componentes del método propuesto.

Tabla 1. Descripción de las bases de datos utilizadas

Base de datos	Resolución	Escenario	Núm. de imágenes utilizadas	
			Mono-cámara	Multi-cámara
PETS 2006	720 x 576 píxeles	Estación de tren	11,204	4,703
PETS 2007	720 x 576 píxeles	Aeropuerto	13,207	-
PETS 2009	768 x 576 píxeles	Estacionamiento (exterior)	2,625	1,837
CAVIAR	348 x 288 píxeles	Centro Comercial	7,786	-
MUBA	640 x 480 píxeles	Aeropuerto	-	5,336
Totales			34,822	11,876

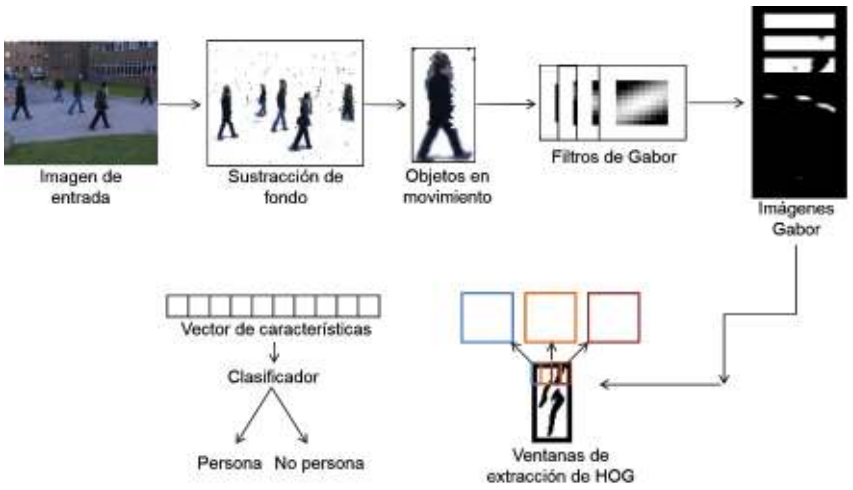

Figura 2. Esquema general del método de detección de personas: HoGG

Filtros de Gabor

Tomando en cuenta la complejidad de la detección de personas en escenarios reales, los filtros de Gabor se seleccionaron para transformar la imagen de entrada. Los filtros de Gabor son exitosamente empleados en muchas aplicaciones de análisis de imágenes como análisis de textura, verificación facial, reconocimiento de caracteres y recuperación de imágenes por contenido. Una de las propiedades más importantes de los filtros de Gabor es que realizan la localización óptima tanto en el dominio espacial como en el de la frecuencia (Yuan *et al*., 2003). Los filtros de Gabor trabajan como una especie de detector de bordes en una base no-ortogonal, por lo tanto, cada característica extraída por un filtro se correlaciona con otra característica generada por otro. Es decir, la información se resalta en algunas orientaciones, pero tanto la relación de vecindad y de orientación se mantienen en las características extraídas. Los filtros de Gabor enfatizan las características de la imagen sobre los componentes de la misma frecuencia, mientras que rechaza otros componentes discriminando subpartes de los objetos y extrayendo características locales invariantes a cambios de escala, rotación, traslación e iluminación (Kyrki *et al*., 2004). Todo lo anterior es especialmente adecuado en el entorno de video vigilancia, donde hay grandes cambios en la apariencia de los objetos. Además, los filtros de Gabor capturan información en una área local y combinan la respuesta de varios filtros a diferentes orientaciones, frecuencias y escalas, por lo tanto, son adecuados para representar objetos complejos, ya que la información del objeto completo se mantiene y enriquece debido a la combinación de la respuesta de todos los filtros (Krüger, 2002). Lo planteado por Lades *et al*. (1993), se consideró un banco de filtros de Gabor de 5 frecuencias y 8 orientaciones. La expresión analítica de los filtros de Gabor se puede escribir de acuerdo con la ecuación 1, donde

$$\psi(x, y; x_0, y_0, f_0, \sigma_x, \sigma_y, \theta, \phi) =$$

$$\frac{f_0^2}{\pi\sigma_x\sigma_y} exp\left[-f_0^2\left(\frac{x_r^2}{\sigma_x^2}+\frac{y_r^2}{\sigma_y^2}\right)\right]$$

$$\exp\left(2\pi i f_0 x_r + i\phi\right) \quad (1)$$

donde

$x_r = (x - x_0) = \cos\theta + (y - y_0)\sin\theta$ y y_r
$= -(x - x_0)\sin\theta + (y - y_0)\cos\theta$;
x_0 y y_0 = la posición en el espacio de la wavelet
f_0 = frecuencia central de la onda plana
σ_x = determina el ancho del eje mayor de la envolvente Gaussiana
σ_y = determina el ancho del eje menor de la envolvente Gaussiana
θ = ángulo (contrario de las manecillas del reloj) entre la dirección de propagación de la onda *y* el eje *x*
ϕ = desplazamiento de la fase de la onda (Gabor, 1946).

Con el conjunto de filtros de Gabor generados para el pre-procesamiento de las imágenes de entrada, el rendimiento del método HOG se mejora. Como se comentó previamente, se generaron 40 filtros de Gabor (5 frecuencias y 8 orientaciones); sin embargo, el uso de los 40 filtros requiere de un alto tiempo de procesamiento, lo que hace ineficiente su utilización para la detección de personas en tiempo real y en entornos de video vigilancia. Por ello se realizó una selección de los filtros que obtuvieran una mejor respuesta empleando un subconjunto de las bases de datos utilizadas, esto se explica a detalle en la sección de experimentos.

Histograma de las orientaciones del gradiente: HOG

Como se mencionó, debido a que el método del *histograma de las orientaciones del gradiente* (HOG) se ha empleado satisfactoriamente en muchos trabajos de la literatura, se empleó como parte del método de detección de personas

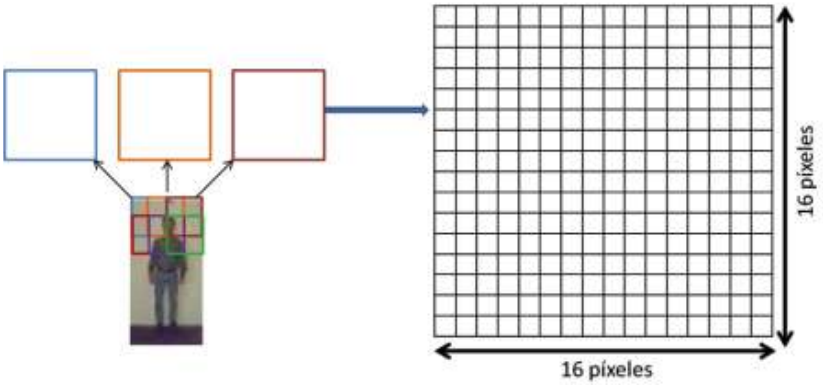

Figura 3. Establecimiento de ventanas de extracción sobre la imagen con el método HOG

propuesto en este trabajo. El método HOG se basa en un histograma que acumula las orientaciones del gradiente en cada una de las ventanas de extracción que se definen en la imagen. Este histograma acumulará el número de orientaciones, de 0 a 180 grados, contabilizadas en cada ventana. Con el algoritmo HOG la imagen se divide en celdas de dimensión 16×16 con un solapamiento de 8 píxeles entre ellas. Dado que las dimensiones de las imágenes con las que se trabajó son 32 × 64 píxeles, se establecieron 21 ventanas de extracción por cada imagen. Un ejemplo del establecimiento de estas ventanas de extracción se muestra en la figura 3.

Una vez que se establecen estas ventanas de extracción se procede a normalizarlas, tal que su media sea 0 y su varianza 1 para que todos los valores integren un rango similar. Para lograr esta normalización se procede a generar la imagen integral y la imagen integral al cuadrado. Sea *img(x, y)* el valor correspondiente a la ventana de extracción de la imagen *img* en la columna *x* y renglón *y*, la imagen integral se calcula como indica la ecuación 2

$$ii(x,y) = \sum_{x' \le x, y' \le y} img(x', y') \qquad (2)$$

La imagen integral al cuadrado se calcula elevando al cuadrado cada elemento de la imagen integral. Una vez calculadas la imagen integral y la imagen integral al cuadrado, la normalización de cada ventana de extracción se realiza como se indica la ecuación 3

$$img(x,y) = \frac{img(x,y) - \mu}{\sigma} \qquad (3)$$

donde μ es la media y σ la desviación estándar. Posteriormente se realiza el cálculo de la orientación y dirección del gradiente de cada píxel en la imagen. El cálculo del gradiente se consigue filtrando la imagen mediante dos máscaras unidimensionales: la horizontal [−1, 0, 1] y la vertical $[-1, 0, 1]^T$. Para el cálculo del histograma de orientaciones se emplea una matriz de 2×2 celdas y 9 orientaciones, por lo que de cada ventana de extracción se obtendrá un vector de 2 × 2 × 9 = 36 características. Esta matriz de 2 × 2 se utiliza para calcular la posición que ocupa cada una de las características extraídas en el vector. El rango de orientaciones va desde 0 hasta 180 grados. De esta forma, el valor acumulado en cada posición del vector de características es el valor de la magnitud del gradiente multiplicado por un peso situado en una matriz de pesos. Para el cálculo de esta matriz de pesos se utiliza una función Gaussiana con sigma igual a 8, que es la mitad del tamaño de la ventana de extracción. La función Gaussiana se calcula a partir de las dimensiones de la ventana de extracción y el resultado es una matriz de pesos de dimensión de 16×16 celdas. Al finalizar todo el procedimiento se obtiene un histograma con 9 orientaciones por cada una de las 4 celdas en las que se divide cada ventana de extracción. El vector final de características es el resultado de la concatenación del histograma de orientaciones de cada ventana de extracción; es decir, un vector de 756 características dado que se emplean 21 ventanas de extracción en toda la imagen. Para mayor información técnica de este algoritmo ver la publicación original en Dalal y Triggs (2005).

Experimentos y análisis de los resultados

Para la comparación del método propuesto, HoGG, se seleccionaron varios algoritmos alternativos para la detección de personas, estos son: *Rapid Object Detection using a Boosted Cascade of Simple Features* (Viola y Jones, 2001), *Improvements of Object Detection using Boosted Histograms* (Laptev, 2006), e *Histograms of Oriented Gradients for Human Detection* (Dalal y Triggs, 2005). En el caso del algoritmo de Viola y Jones, se utilizaron dos variantes, en la primera el algoritmo se entrenó con un conjunto estándar de imágenes (que vienen disponibles

en la implementación de OpenCV (2013)), mientras que en la segunda variación se utilizó un conjunto específico de imágenes para el entrenamiento a partir de las bases de datos utilizadas en este trabajo.

El método de Viola y Jones (2001), hace uso de las características tipo *haar* y del aprendizaje con *AdaBoost*, el método de Laptev (2006) utiliza los histogramas de orientación del gradiente en regiones rectangulares definidas en la imagen en combinación con el aprendizaje tipo Boosting. El método HOG (Dalal y Triggs, 2005) contabiliza las orientaciones del gradiente en cada una de las ventanas de extracción definidas en toda la imagen.

El objetivo del método Viola y Jones (2001) es detectar objetos de forma rápida y precisa. Para ello, introducen una nueva representación de la imagen llamada imagen integral, que se utiliza también en el método HOG. Este método utiliza un algoritmo de aprendizaje basado en AdaBoost, el cual selecciona un pequeño número de características visuales a partir de un conjunto mayor. Además combina, de manera incremental, clasificadores complejos en forma de cascada, lo que permite a las regiones del fondo de la imagen descartarse rápidamente para que de esta manera se enfoque el cálculo computacional, solo en las regiones más prometedoras. El procedimiento de detección utiliza tres características simples: el doble rectángulo, el triple rectángulo y el cuádruple rectángulo. El doble rectángulo es la diferencia entre la suma de los píxeles contenidos en regiones de dos rectángulos; las regiones tienen el mismo tamaño y forma y están horizontal o verticalmente adyacentes. El triple rectángulo sirve para calcular la suma dentro de los dos rectángulos de los extremos menos la suma del rectángulo central. Las características que se extrajeron del cuádruple rectángulo, se calculan con la diferencia de las partes diagonales del rectángulo. La clasificación por medio de cascadas requiere poco tiempo de detección, es decir, funciona rápido. La clasificación por medio de AdaBoost es un sistema que construye una regla de clasificación final usando varios clasificadores menores, denominados "débiles" por su sencillez y escasa precisión. Por sí solos, estos clasificadores débiles, no constituyen un sistema de clasificación eficaz debido a su alta inexactitud, pero al usarlos en conjunto es posible construir un clasificador mucho más preciso. De este método se utilizó la implementación disponible en OpenCV (2013). OpenCV es una librería de visión artificial libre y originalmente desarrollada por Intel. En la implementación de OpenCV de este método, vienen disponibles diversos archivos de entrenamiento, de caras, de personas, de la parte superior e inferior del cuerpo humano, entre otras. Por lo tanto, del método Viola y Jones (2001) se utilizaron dos variaciones, en la primera se empleó el archivo de entrenamiento para la detección de personas, disponible en la implementación de OpenCV, a esta variación se le llamará V&J-1 para futuras referencias. Para la segunda variación, se generó un nuevo archivo de entrenamiento con las imágenes contenidas en las bases de datos INRIA (INRIA, 2005) y MIT (MIT, 2000), a esta segunda variación se le llamará V&J-2. Cabe señalar que los archivos de entrenamiento generados con nuevas bases de datos surgieron con herramientas de OpenCV creadas para ese propósito.

Otro método utilizado para esta comparación fue el propuesto por Laptev (2006) donde se toma en cuenta un conjunto completo de regiones rectangulares dentro de la imagen para posteriormente aplicar el algoritmo HOG. Este método tiene como objetivo la detección de objetos, centrándose en el problema específico de la detección de personas mediante la combinación de un histograma de características locales con un clasificador AdaBoost. Aquí, se escoge la posición y la forma del histograma de características para minimizar el error en la etapa de entrenamiento. Además, se considera un conjunto completo de regiones rectangulares en una ventana normalizada del objeto y se calcula el histograma de las orientaciones del gradiente (HOG) para varias direcciones de cada región. Posteriormente se aplica el procedimiento del clasificador AdaBoost para seleccionar las características del histograma previamente calculado y aprender el objeto que se desea clasificar.

Para la detección se utiliza una técnica de ventana de escaneo que se aplica en toda la imagen, esta ventana de escaneo calcula una medida para cada parte de la imagen y el clasificador se aplica en las ventanas que mayores medidas obtuvieron. Una importante contribución de este método es que se adapta a un marco de trabajo tipo *boosting*, el cual intenta responder a la pregunta de si un conjunto de clasificadores débiles pueden crear, en conjunto, a un clasificador fuerte.

El tercer método comparado es el método HOG, el cual se describió previamente (ver sección *Histograma de las orientaciones del gradiente: HOG*). Por lo tanto, la comparación del método propuesto HoGG se realiza contra cuatro métodos diferentes: V&J-1, V&J-2, Laptev y HOG.

En los experimentos se lleva a cabo una evaluación exhaustiva orientada a contar personas. Como medida de evaluación del rendimiento del sistema se utilizó la medida F (*F-measure*). La medida F utiliza el valor de *Precision* y *Recall* y es muy útil para evaluar el nivel de balance de los sistemas; por ejemplo, un sistema de alto nivel de seguridad que detecta pocas personas con un bajo nivel de falsas detecciones tendrá una *pre-*

cision muy alta pero un *recall* muy bajo. Por el contrario, un sistema que detecta muchas personas correctamente, pero también comete muchas falsas detecciones, tendrá un *precision* baja y un *recall* alto. El valor óptimo de la medida F es 1, lo cual se da cuando *Precision* = *Recall* = 1 y su peor resultado es 0 (*Precision* = *Recall* = 0). Para el cálculo de estas medidas se utilizan las ecuaciónes 4, 5 y 6

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (4)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (5)$$

$$Medida\ F = 2 * \frac{Precision * Recall}{Preciision + Recall} \quad (6)$$

Selección de los mejores filtros de Gabor

Debido a que la utilización de los 40 filtros de Gabor es ineficiente para detectar personas en entornos de video vigilancia, donde se requiere una respuesta rápida, se realizó una prueba para determinar cuáles eran los filtros que mejor resultado obtenían. Esta prueba consistió en lograr un resultado (en *precision* y *recall*) para cada filtro de manera individual, es decir, se pre-procesó la imagen con cada filtro, posteriormente se extrajeron las características con el método HOG y se entrenó al clasificador con ellas. De esta manera se obtuvo un resultado por cada uno de los 40 filtros de Gabor utilizados. En la tabla 3 se muestran los resultados, tanto en *precision* y *recall*, con cada uno de los 40 filtros, en la primera columna están las diversas orientaciones consideradas, y por cada orientación se consideraron 5 escalas diferentes. En la tabla 3 se puede observar que los filtros que obtuvieron mejores resultados son: filtro con 157.5 grados y escala de 5.6; filtro de 67.5 grados y escala de 8, filtro con 90 grados y escala de 11.31 y filtro con 112.5 grados y escala de 16. El quinto mejor filtro tiene un rendimiento menor, por eso solo se consideraron los primeros cuatro mejores filtros. Sin embargo, como un experimento adicional se probaron las combinaciones de los primeros seis mejores filtros, en la tabla 2 se muestran los resultados de esta combinación, aquí se puede ver que al combinar, por ejemplo, el mejor filtro con el segundo mejor el resultado disminuye; sin embargo, el resultado más alto se alcanzó con la combinación de los primeros 4 mejores filtros (0.749 de medida F).

Evaluación y comparativa con otros métodos

En esta sección se detallan los resultados obtenidos con el método propuesto, es decir, con los métodos alternativos descritos anteriormente. En la tabla 4 se observan los resultados, en valores de *Precision*, *recall* y medida F, por cada método comparado y con las cuatro bases de datos públicas utilizadas para la evaluación.

Como se observa en la tabla 4 , los mejores resultados en medida F se obtuvieron por el método propuesto HoGG, el segundo mejor método fue el método Laptev y los peores resultados se obtuvieron con el algoritmo V&J-1 con una medida F promedio de 0.27, mientras que el mejor método obtuvo una medida F con valor de 0.70.

Para ver un ejemplo de los resultados, en la figura 4 se muestran las diferentes imágenes de resultado de cada uno de los métodos comparados, así como del mé-

Tabla 2. Rendimiento de la combinación de los primeros 6 mejores filtros de Gabor

Número de filtros combinados	1	2	3	4	5	6
Parámetros del filtro: escala/orientación	5.6/157.5	11.31/90	16/112.5	8/67.5	11.31/135	16/22.5
Medida F	0.742	0.732	0.717	0.749	0.511	0.491

Tabla 3. Resultados para la selección de los mejores filtros de Gabor

O	S														
	4			5.6			8			11.3			16		
	Pre	Rec	F	Pre	Rec	F	Pre	Rec	F	Pre	Rec	F	Pre	Rec	F
0	0.81	0.57	0.67	0.8	0.62	0.7	0.76	0.64	0.7	0.751	0.662	0.704	0.741	0.269	0.394
22.5	0.77	0.57	0.66	0.79	0.6	0.68	0.8	0.57	0.67	0.800	0.005	0.010	0.722	0.740	0.731
45	0.77	0.5	0.61	0.78	0.54	0.64	0.83	0.56	0.67	1.000	0.013	0.025	0.728	0.644	0.683
67.5	0.83	0.57	0.68	0.77	0.35	0.480	**0.85**	**0.65**	**0.74**	0.744	0.631	0.683	0.727	0.518	0.605
90	0.84	0.44	0.57	0.78	0.42	0.55	0.82	0.62	0.71	**0.757**	**0.717**	**0.737**	0.756	0.634	0.690
113	0.85	0.32	0.46	0.74	0.34	0.47	0.82	0.59	0.68	0.799	0.573	0.668	**0.745**	**0.729**	**0.737**
135	0.83	0.23	0.35	0.79	0.48	0.6	0.85	0.62	0.72	0.747	0.715	0.731	0.763	0.608	0.677
158	0.8	0.61	0.69	**0.82**	**0.68**	**0.74**	0.81	0.15	0.25	0.739	0.674	0.705	0.751	0.512	0.609

Tabla 4. Resultados obtenidos con cuatro bases de datos públicas y con cuatro métodos de referencia

Método	PETS 2006			PETS 2007			PETS 2009			CAVIAR			F promedio
	Prec	Recall	F	Prec	Recall	F	Prec	Recall	F	Prec	Recall	F	
V&J-1	0.92	0.07	0.13	0.79	0.07	0.13	0.96	0.33	0.48	1.00	0.22	0.37	0.27
V&J-2	0.93	0.27	0.42	0.97	0.17	0.29	0.99	0.49	0.61	0.99	0.21	0.35	0.41
Laptev	0.99	0.26	0.36	0.99	0.41	0.58	1	0.46	0.59	0.99	0.40	0.56	0.52
HOG	0.65	0.41	0.50	0.81	0.42	0.56	0.89	0.35	0.50	0.87	0.26	0.39	0.48
HoGG	0.73	0.60	**0.66**	0.86	0.58	**0.70**	0.91	0.67	**0.77**	0.92	0.58	**0.69**	**0.70**

todo propuesto HoGG. Aquí se puede observar como el método propuesto es el que mayor número de detecciones correctas realiza, por el contrario el método V&J-1 es el que menor número de personas detecta en la imagen.

Para una mejor visualización del desempeño, en la figura 5 se muestran las posiciones promedio de cada método en el ranking con su correspondiente 95% de intervalo de confianza. En esta gráfica se puede ver que el método HoGG logró la mejor posición con valor de 1, 12 ± 0, 34. De hecho, el método HoGG fue el mejor (posición igual a 1) en 14 de un total de 16 secuencias. En las dos secuencias en las que no fue el mejor, el método HoGG obtuvo la segunda mejor posición en el ranking. Por lo tanto, se puede concluir que el buen funcionamiento del método HoGG es independiente de las situaciones del entorno, por lo que es robusto y estable.

Además de evaluar los métodos comparados en términos de medida F y su posición en un ranking, se analizó el tiempo que cada uno de los métodos tardaba en procesar cada imagen, es decir, el número de imágenes que cada método procesa por segundo. Por medio de este análisis se puede concluir que las diferencias en tiempo de procesamiento son mínimas, prácticamente todos los métodos procesan a la misma velocidad ≈ 25 imágenes por segundo. A pesar de que el método propuesto HoGG utiliza un pre-procesamiento con los filtros de Gabor, es igualmente capaz de procesar 25 imágenes por segundo, ya que la imagen de entrada solo se convoluciona con 4 filtros en vez de 40 como tradicionalmente se hace.

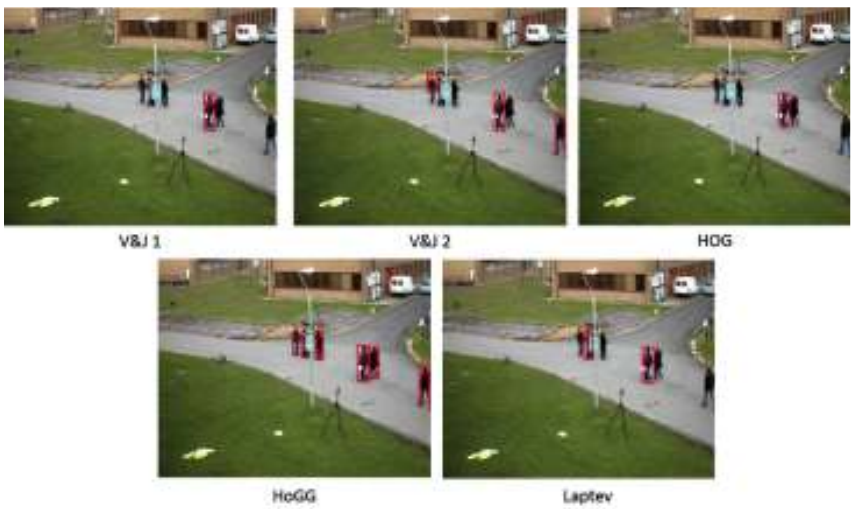

Figura 4. Imágenes de resultado de todos los métodos

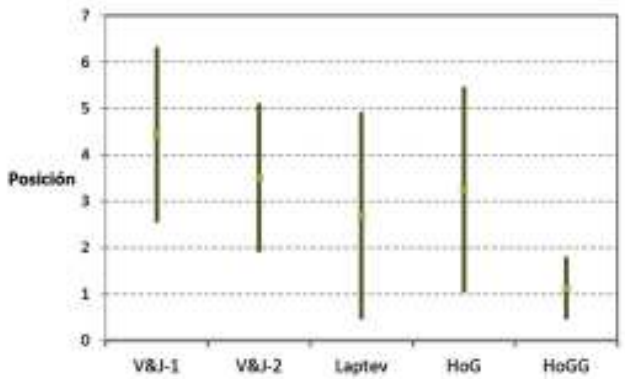

Figura 5. Posición media (cuadrado verde) e intervalo de confianza de cada uno de los métodos

Como una comparación adicional a los resultados obtenidos con otros métodos de la literatura, se buscaron trabajos más recientes, entre ellos se encuentra el trabajo presentado por Zweng (2012), donde se propone un método para la detección de personas utilizando un modelo con características relacionales en combinación con funciones de similaridad de histogramas, como la intersección de histogramas, la correlación de histogramas, etcétera. Estas características relacionales son combinadas con las características extraídas con el método HOG. Para las pruebas se utilizó la base de datos PETS 2009 y los resultados no llegan ni a 70% de reconocimiento, a diferencia del método propuesto HoGG que con la misma base de datos logra 77% de reconocimiento.

Otro método en la literatura es el propuesto por García (2015), donde se presentan dos tareas de post-procesamiento con el objetivo de mejorar los resultados en la detección de personas. Primero, se propone el uso de un filtro de segmentación de personas respecto al fondo, posteriormente se evalúa la combinación de diferentes métodos de detección de personas con el objetivo de proporcionarles mayor robustez en la detección y así mejorar los resultados. Para los experimentos se utilizaron dos bases de datos: la PDds y PETS 2009. Comparando los resultados obtenidos con la base de datos PETS 2009, se puede observar que en el trabajo propuesto por García (2015) se obtuvo un rendimiento de 85% de tasa de reconocimiento y con el método propuesto HoGG, se obtuvo un valor de medida-F de 77%; sin embargo, en García y Martínez (2015) solo se utilizó una cámara de la secuencia de imágenes de PETS 2009, mientras que en el trabajo propuesto se utilizaron cuatro cámaras diferentes de esa misma base de datos.

Otros trabajos como los propuestos en Milani *et al*. (2013) y Nghiem y Bremond (2014) presentan resultados mayores en tasas de reconocimiento, sin embargo, esos resultados se obtuvieron por medio de bases de datos propias, adquiridas bajo ciertas condiciones controladas en laboratorio, lo cual no puede comparare contra los resultados obtenidos, utilizando bases de datos públicas y estándares en la comunidad de video vigilancia inteligente. Por ejemplo, en Milani (2013) se alcanzan entre 50 y 100% de tasa de reconocimiento, dependiendo el escenario de prueba, unos más complejos que otros. Sin embargo, 100% se logró con imágenes adquiridas en condiciones controladas. En Nghiem (2014) se obtuvo un resultado de 99% de valor de Medida-F, lo cual es un excelente resultado; sin embargo, como ya se comentó, las imágenes utilizadas para probar los métodos son imágenes adquiridas en entornos de laboratorio con condiciones controladas que favorecen, en muchas ocasiones, a los métodos que proponen los autores.

Modelo de apariencia con soft-biométricos

Debido a que las imágenes adquiridas en la mayoría de los sistemas de video vigilancia son de baja calidad y se adquieren a distancia, es imposible obtener características biométricas, como por ejemplo, el rostro. Por lo anterior y con el objetivo de identificar a una persona etiquetada como sospechosa para su posterior re-identificación de manera no intrusiva, en este trabajo se propone la utilización de un conjunto de características soft-biométricas relacionadas con la apariencia de las personas. El proceso de re-identificación consiste en identificar a una persona a lo largo de diversas cámaras localizadas en una zona de video vigilancia. Las características que se extraen se analizan y ponderan de acuerdo con diversas técnicas propuestas. En términos generales, la solución propuesta consiste en, primero, extraer un conjunto de características soft-biométricas relacionadas con la apariencia de cada persona, posteriormente estas características se ordenan en un ranking de acuerdo a su importancia para la clasificación, posteriormente cada característica se pondera de acuerdo con su posición en el ranking, y finalmente el proceso de clasificación se lleva a cabo por medio de la distancia Euclídea. En la fase de entrenamiento, se obtienen el mejor ranking y el mejor tipo de ponderación, para posteriormente utilizarlas en la fase de prueba.

Características soft-biométricas

El conjunto de características extraídas se relacionan con la apariencia de las personas y constituyen lo que llamamos *bag-of-softbiometrics*. Estas características se seleccionaron tomando en cuenta las restricciones que imponen las malas condiciones de adquisición, como resolución baja, cambios de iluminación, localización de las cámaras, cambios de perspectiva, etcétera, presentes en los entornos reales de video vigilancia. En este trabajo se considera un conjunto de 23 características teniendo en cuenta una amplia variedad de características y así medir la relevancia de cada una de ellas para la descripción de cada individuo. Estas características se relacionan con diferentes categorías como: estadísticas de color en el modelo RGB (*Red, Green and Blue*), estadísticas en escala de grises, geometría, histograma en escala de grises, estadísticas de color del modelo HSV (*Hue, Saturation and Value*), estadísticos de texturas y LBP (*Local Binary Pattern*). Dado que la apariencia visual, a distancia, de una persona se representa básicamente por su ropa, se tomaron en cuenta un conjunto de características relacionadas al color y textura para describir la apariencia de cada persona. Para las características de color se utilizaron dos modelos: RGB (*Red, Green, Blue*) y HSV (*Hue, Saturation, Value*).

La información de las imágenes se representa comúnmente por un histograma, por ello se ha considerado para el conjunto de características algunos estadísticos que se obtienen a partir de este. A simple vista, un histograma puede dar una idea aproximada de la distribución del nivel de gris de una imagen. La forma del histograma proporciona diferentes características útiles como la media, la dispersión, el contraste, etcétera. Por ello, se calculan características globales de la imagen a partir del histograma en escala de grises, estas caracte-

rísticas son: media, desviación estándar, entropía, dispersión, energía y curtosis.

En la tabla 5 se pueden ver las ecuaciones para calcular cada una de las características generadas a partir del histograma en escala de grises.

Por otro lado, uno de los principales inconvenientes de las características basadas en histogramas es que no se considera la distribución espacial ni la variación local del color. Con el objetivo de evitar esto, se utiliza la matriz de co-ocurrencia para considerar la información espacial. La matriz de co-ocurrencia en escala de grises (GCM por *Gray Co-ocurrence Matrix*) es un método conocido para la extracción de textura en el dominio espacial (Haralick *et al*., 1973). La GCM incorpora información espacial en forma de posición relativa entre los niveles de intensidad dentro de la textura; de hecho, las GCM son histogramas bi-dimensionales. La matriz de co-ocurrencia almacena el número de veces que un pixel en relación con otro, dentro de su vecindad en la imagen, tiene una combinación particular (Vadivel *et al*., 2007). Las características extraídas de la GCM son: energía, probabilidad máxima, entropía, inercia y homogeneidad. En la tabla 6 se muestran las ecuaciones para calcular cada una de estas características de textura.

También se consideró otro enfoque local basado en *Local Binary Pattern* (LBP) (Moore y Bowden, 2011). El LBP etiqueta los píxeles *fp (p = 0, ..., 7)* de una imagen en una vecindad de *3 × 3*. Cada píxel se compara contra el píxel central *fc* y el resultado es un número binario de cada una de las 8 comparaciones (8 vecinos), el valor LBP se calcula como indica la ecuación 7

$$S(f_p - f_c) = \begin{cases} 1, & if\ f_p \geq f_c \\ 0, & de\ otro\ modo \end{cases} \quad (7)$$

La última característica extraída es la excentricidad, que representa la relación entre el alto y ancho de la persona. La excentricidad se obtiene por medio de los momentos de segundo orden (m_{20}, m_{02} y m_{11}), una vez que tenemos los momentos, la excentricidad se calcula por medio de las ecuaciones 8, 9 y 10. La ecuación 8 calcula el R_{min} que es la distancia mínima y la ecuación 9 calcula el R_{max} que es la distancia máxima, a partir de estas dos se calcula el valor de excentricidad.

$$R_{min} = \frac{\sqrt{m_{20} + m_{02}} - \sqrt{4m_{11}^2 + (m_{20}^2 - m_{02}^2)}}{2} \quad (8)$$

$$R_{max} = \frac{\sqrt{m_{20} + m_{02}} + \sqrt{4m_{11}^2 + (m_{20}^2 - m_{02}^2)}}{2} \quad (9)$$

$$Excentricidad = \frac{R_{min}}{R_{max}} \quad (10)$$

En resumen, se consideró un conjunto de 23 características relacionadas al color, textura, geometría y rasgos locales, principalmente. Finalmente estas características se normalizan entre 0 y 1 con fines comparativos.

Métodos para medir la relevancia de las características extraídas: posición y ponderación

En esta sección se describen los cuatro métodos propuestos para medir la relevancia de cada una de las características extraídas. Estas técnicas se proponen teniendo en cuenta el hecho de que la relevancia de cada una de las 23 características soft-biométricas ex-

Tabla 5. Ecuaciones para calcular las características extraídas del histograma en escala de grises

Media	Desviación estándar	Entropía
$\mu = \frac{\sum_{i=1}^{L} h(i)}{L}$	$\sigma = \frac{\sqrt{\sum_{i=1}^{L} (h(i) - \mu)^2}}{L}$	$\sum_{i=1}^{L} h(i) \log_2[h(i)]$
Dispersión	**Energía**	**Curtosis**
$\sum_{I=1}^{L} abs(h(i) - \mu)$	$\sum_{i=1}^{L} h(i)^2$	$\sigma^{-4} \sum_{i=1}^{L} (h(i) - \mu)^4 p(i)$

Tabla 6. Ecuaciones para calcular las características extraídas a partir de la GCM

Energía	Máxima probabilidad	Entropía
$\sum_{i=1}^{N} \sum_{j=1}^{N} GCM(i,j)^2$	$max_{i,j} GCM\ (i, j)$	$\sum_{i=1}^{N} \sum_{J=1}^{N} GCM(i,j) \log_2[GCM(i,j)$
Inercia		**Homogeneidad**
$\sum_{i=1}^{N} \sum_{j=1}^{N} (i-j)^2 GCM(i,j)$		$\sum_{i=1}^{N} \sum_{J=1}^{N} \frac{p(i,j)}{1+(i,j)^2}$

traídas es diferente. Estos métodos se basan en los enfoques de PCA (*Principal Component Analysis*) (Jolliffe, 1973), medidas de disimilaridad y alineación del kernel (*kernel alignment*) (Cristianini *et al*., 2010). Cada uno de estos enfoques se seleccionó por los buenos resultados que han obtenido en el área de reconocimiento de patrones. A partir de estos métodos se generó otro mediante la combinación de los mismos. Los dos primeros métodos propuestos se basan en el PCA. La idea principal detrás del PCA es reducir la dimensionalidad de los datos que contengan un gran número de variables interrelacionadas (Jolliffe, 1973). Primero, se propone un enfoque que considera la importancia de la presencia de cada característica en cada uno de los 23 componentes principales. A este método se le llamará *PCA-feature-presence* (PCA-FP). En segundo lugar se propone un método con el que se pueda obtener un valor cuantitativo a partir de la salida generada por el PCA. Este segundo método considera el valor del *score* de cada característica en cada autovector multiplicado por la correspondiente proporción de varianza del componente. A este método se le llamará *PCA-feature-value* (PCA-FV). En el caso del método PCA-FP, se tiene en cuenta el hecho de que la varianza de cada componente decrece progresivamente. El método PCA-FP es un método secuencial que ordena a cada característica de acuerdo con la importancia de cada componente. Como la primera componente tiene mayor varianza se le considera de mayor importancia, y así con la segunda componente que es la segunda más importante hasta la componente 23 que es la que menos importancia se le asigna. Con el método PCA-FP se obtendrá un ranking desde 1 hasta 23 posiciones, donde la primera es la más importante y la última la menos importante. Con este método sólo se obtiene una posición y no un valor numérico como salida. Buscando un valor numérico para la salida del método, el método PCA-FV considera el valor absoluto del *score* en cada autovector para cada característica y el valor de la proporción de varianza de cada componente. El *score* de cada característica se multiplica por el valor de la proporción de varianza de cada componente. Como tercer método, se consideró un enfoque diferente basado en una medida de disimilaridad (DM para futuras referencias). Aquí, cada característica se ordena mediante la comparación de sus medias y desviaciones en cada clase (cada persona). Este ranking se calcula como indica la ecuación 11.

Donde

X_{ni} = valor de la media de cada característica *i* en cada clase *n*

S(ni) = desviación estándar de cada característica *i* para la misma clase *n*

X_{mi} = valor de la media de la característica *i* para todas las demás clases *m*

S(mi) = desviación estándar de todas las clases restantes *m*

$$Rank_i = \frac{(\left|X_{n_i} - X_{m_i}\right|)}{(S(n_i) + (S(m_i))} \tag{11}$$

El cuarto método se basa en un enfoque de *kernel*, este se usa cada vez más para modelado de datos debido a su simplicidad conceptual y su rendimiento en muchas tareas de reconocimiento de patrones (Cristianini, 2010). Es por esta razón que se considera el método de *kernel alignment* (KA para futuras referencias) en este trabajo como el cuarto método para medir la relevancia de cada una de las características extraídas. Con el método del *kernel alignment* se calculó el alineamiento entre cada característica y el *kernel* ideal (Gosselin *et al*., 2011). Cuanto mayor es este alineamiento, mayor se ajustará el *kernel* a la clase representada por los datos. El alineamiento se calcula como indica la ecuación 12.

$$A_v = \frac{\sum k_v(i,j) \cdot yy^t(i,j)}{\sqrt{\sum k_v^2(i,j) \cdot \sum_{m=1}^{i} (ns_m^2)}} \tag{12}$$

Donde k_v es una matriz que en su diagonal por bloques contiene los valores de todas las muestras de todas las personas (clases) en cada variable *v*. Es decir, la matriz k_v se construye mediante la concatenación de cada uno de los valores de cada característica de todas las imágenes de cada una de las personas. Por lo tanto, la matriz k_v es una matriz cuadrada con dimensión igual a *n* muestras (para todas las clases), donde *(i, j)* representa las posiciones de renglón y columna. El proceso de construcción de esta matriz se lleva a cabo mediante la concatenación de bloques, donde cada bloque representa a todas las muestras de la característica *v* de cada persona o clase. El *kernel* ideal se representa por yy^t, que es una matriz cuadrada que contiene 1's en su diagonal por bloques y 0's en el resto de los casos. Esto es, $\sum k_v(i, j) \cdot yy^t(i, j)$ es la sumatoria de la diagonal de la matriz k_v y *ns* es el número total de los ejemplos de cada una de las clases, se debe tener en cuenta que el número de ejemplos de cada clase puede ser diferente.

Finalmente, el último método se calcula mediante la suma de los resultados de los cuatro métodos previamente analizados, es decir, este método suma las diferentes posiciones obtenidas en cada método para cada característica, por lo que cuanto más pequeño sea el

valor de esta suma, mejor será la posición en el ranking con este método de combinación (CM para futuras referencias). Cuando las características ya están ordenadas de acuerdo con su relevancia, estas se ponderan en relación con su posición en el ranking, para esto se proponen dos maneras. En la primera, se le asigna una puntuación a cada una de las posiciones del ranking, a la primera posición se le asigna un valor de 23/276 y a la última posición un valor de 1/276. El 276 es el total de la sumatoria de 23 hasta 1, es decir, el total de puntos que se repartieron en las 23 características. Esta técnica de ponderación se llama no-paramétrica. La segunda técnica de ponderación utiliza la salida numérica generada por cada método y se le llama ponderación paramétrica. En resumen, se proponen cinco métodos para medir la relevancia de cada característica y dos técnicas de ponderación para los rankings obtenidos. Por lo que, en combinación se obtendrán ocho diferentes resultados: PCA-FP con ponderación no paramétrica, PCA-FV con ambas ponderaciones, DM con ambas ponderaciones, KA con ambas ponderaciones y CM con la ponderación no paramétrica. Además, con propósitos comparativos, también se probarán las características sin ningún tipo de ordenamiento ni ponderación (SP).

Experimentos y análisis de los resultados

Los experimentos se diseñaron con el propósito de comparar a cada sospechoso con una lista de sospechosos, sin importar el tiempo u orden de aparición. Esto es equivalente a buscar a los sospechosos en una base de datos de video vigilancia completa, tanto en las imágenes pasadas como en las imágenes actuales. Para estos experimentos se utilizaron dos enfoques de evaluación: el enfoque *watch-list* y la curva CMC (*Cumulative Match Curve*). En la evaluación *watch-list*, el propósito es identificar a un limitado y pre-establecido número de personas (los sospechosos) quienes están en una lista de búsqueda (*watch-list*) (Kamgar y Lawson, 2011). Por otro lado, el enfoque CMC se seleccionó como otra forma de evaluar a los métodos (Bolle *et al*., 2005). El enfoque CMC se utiliza como medida de comparación *1:k* (en este trabajo k es igual a 3) del rendimiento del sistema, donde *k* representa el rango de los mejores resultados considerados para la evaluación del sistema, es decir, CMC juzga la capacidad de margen de aceptación o éxito de un sistema de identificación.

Para los experimentos se utilizaron tres bases de datos multi-cámara: PETS 2006, PETS 2009 y MUBA. Se utilizaron un total de 10,978 imágenes multi-cámara, de esas imágenes, 30% se utilizó para la etapa de entrenamiento para la generación de los rankings y el restante 70% se utilizó para la etapa de prueba.

En la tabla 7 se muestran los resultados obtenidos con cada método en cada una de las bases de datos multi-cámara utilizadas (PETS 2006, PETS 2009 y MUBA). En esta tabla, NP significa No Paramétrico y P es paramétrico (ambos métodos de ponderación). En estos resultados, expresados en porcentaje de TPR (*true positive rate* o tasa de verdaderos positivos) y FAR (*falce acceptance rate* o tasa de falsos positivos), se puede apreciar que el mejor método en la base de datos PETS 2006 es el DM, y algunos de los métodos basados en PCA también obtienen buenos resultados. El mejor resultado en la base de datos PETS 2009 se obtuvo con el método PCA-FP, alcanzando 94.24% de tasa de identificación y 5.44% de tasa de falsos positivos. En la base de datos MUBA, el método PCA-FV logró el mejor resultado con 93.10% de tasa de identificación y 6.90% de tasa de falsos positivos. En general, todos los resultados obtenidos son muy prometedores, ya que todos son superiores a 90% de acierto, que es muy relevante por las difíciles condiciones no controladas que se encuentran en las imágenes multi-cámara de las bases de datos utilizadas. También se puede observar que la ponderación paramétrica tiene mayor relevancia.

En cuanto a la evaluación CMC, en la figura 6 se muestra la curva CMC para el mejor método en cada base de datos utilizada. En esta curva se puede observar, que los resultados más altos se obtienen con la base de datos PETS 2006 y, por el contrario, los resultados más bajos se lograron con la base de datos MUBA debido a la complejidad de sus imágenes. A pesar de que con $k = 1$ los resultados son mejores en PETS 2009 que en MUBA, con $k = 3$ la mejora lograda en MUBA es ampliamente superior que la lograda en PETS 2009 bajo esa condición.

Tabla 7. Resultados obtenidos con los métodos de ponderación en bases de datos multi-cámara

Base de datos	PCA-FP NP		PCA-FV P		PCA-FV NP		DM P		DM NP		KA P		KA NP		CM NP		SP	
	TPR	FAR	TPR	FAR	TPR	FAR	TPR	FAR	TPR	FAR	TPR	FAR	TPR	**FAR**	**TPR**	**FAR**	**TPR**	**FAR**
PETS 2006	94.89	5.05	95.51	4.37	94.42	5.42	**96.16**	**3.78**	96.16	3.84	95.26	7.68	94.58	5.30	94.58	5.30	94.58	5.36
PETS 2009	**94.24**	**4.44**	91.20	8.40	92.11	7.65	94.09	5.52	93.53	5.91	94.01	5.44	94.01	5.76	93.85	5.76	93.93	5.76
MUBA	90.04	9.83	**93.10**	**6.90**	90.20	9.74	91.91	8.06	91.10	8.90	92.68	7.25	92.01	7.99	90.62	9.35	90.10	9.83

Además, se puede observar que casi en todos los casos el incremento en el porcentaje de acierto es superior cuando $k = 2$ que el incremento logrado con $k = 3$.

Se puede concluir que la tasa de identificación del mejor método en cada base de datos fue superior a 93% con una tasa de falsos positivos menor a 6.90%. La ponderación paramétrica, es decir, la ponderación que utiliza la salida del método, fue más relevante que la ponderación no paramétrica. En la prueba CMC el mejor resultado fue 98.34% de identificación y 1.6% de falsos positivos, considerando los tres mejores candidatos. Analizando las características se puede concluir que la apariencia relacionada con los estadísticos de textura y con el histograma en escala de grises es más relevante en un entorno de múltiples cámaras.

En la figura 7 se puede apreciar un ejemplo del funcionamiento del sistema final, donde se identifica a una misma persona en cuatro cámaras diferentes. Aquí, se puede ver cómo las cuatro cámaras tienen diferentes entornos y niveles de complejidad, pero aun así el sistema es capaz de realizar la re-identificación de la misma persona con un buen grado de certeza.

Finalmente, en relación con el tiempo de procesamiento, a pesar de que se calculan 23 características para cada persona detectada, el procesamiento se realiza a la velocidad requerida por la mayoría de los sistemas de cámaras para funcionar en tiempo real, es decir, entre 5 y 20 imágenes por segundo. Muchos de los sistemas de cámaras existentes manejan una frecuencia de 5 imágenes por segundo debido principalmente a factores de almacenamiento; sin embargo, con las mejoras de la tecnología muchos sistemas nuevos pueden trabajar a más de 20 imágenes por segundo. El modelo propuesto para la re-identificación de personas por medio de características soft-biométricas, es capaz de procesar a esta velocidad para proporcionar un funcionamiento en tiempo real.

Conclusiones

En este trabajo se presentó un sistema completo para la re-identificación de personas en espacios con múltiples cámaras. Se abordaron todas las tareas que el sistema debe realizar para lograr una correcta identificación de las personas. En primer lugar, se presentó un método para la detección de personas, HoGG, así como su comparación con otros métodos alternativos en la literatura. El método HoGG logró los mejores resultados en todas las bases de datos utilizadas para los experimentos. Analizando los resultados de la detección de perso-

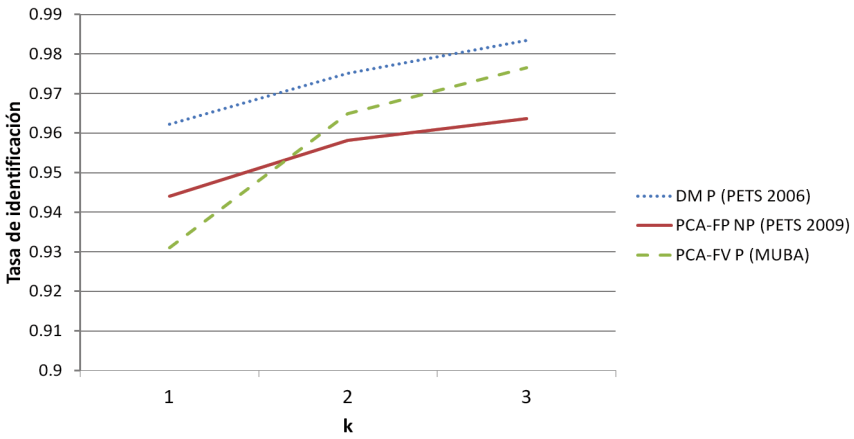

Figura 6. Curva CMC del mejor método por base de datos en el entorno multicámara. El valor de k es 1:3

Figura 7. Ejemplo de reconocimiento de un sospechoso en 4 diferentes cámaras

nas, se puede concluir que los métodos del estado del arte, han logrado resultados muy importantes y las posibles mejoras relacionadas a esta tarea cada vez se vuelven más difíciles de lograr. Sin embargo, cuando se evalúa un algoritmo de detección de personas en entornos reales, el rendimiento se ve afectado, lo que indica que el diseño experimental y la aplicación de los métodos debe ser cada vez más realista.

En segundo lugar, se propone un modelo de apariencia basado en características soft-biométricas. Las características soft-biométricas han demostrado su capacidad en el área de reconocimiento de personas, además, dichas características se pueden usar como rasgos adicionales a los biométricos clásicos. El interés en la aplicación de las características soft-biométricas no tiene el auge de las características biométricas clásicas, sin embargo, el interés sobre ellas se mantiene en constante crecimiento. Por ello, se propuso un conjunto de 23 características relacionadas al color, textura, histogramas, características locales y geométricas. También, se propusieron diversos métodos para medir la relevancia de cada una de estas características para poder ponderarlas de acuerdo con su nivel en un ranking. Con los modelos de apariencia propuestos se lograron tasas de reconocimiento superiores a 93%, con esto, el uso de las características soft-biométricas han demostrado su gran potencial.

Referencias

Alahi A., Vandergheynst P., Bierlaire M., Kunt M. Cascade of descriptors to detect and track objects across any network of cameras. *Computer Vision and Image Understanding*, volumen 114, 2010: 624-640.

Bolle R., Connell J., Pankanti S., Ratha N., Senior A. The relation between the roc curve and the cmc curve, en: Fourth IEEE Workshop on Automatic Identification Advanced Technologies, 2005, pp. 15-20.

Cristianini-Nello, Kandola-Jaz, Elisseeff-Andre, Shawe-Taylor J. On kernel-target alignment. *Advances in Neural Information Processing Systems*, volumen 194, 2002: 367-373.

Dalal N. y Triggs B. Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition,* volumen 1, 2005: 886-893.

Gabor D. Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, volumen 93, 1946: 429-441.

Garcia-Martín A. y Martínez J. Enhanced people detection combining appearance and motion information. *Electronics Letters*, volumen 49, 2013: 256-258.

García-Martín A., Martínez J.M. Post-processing approaches for improving people detection performance. *Computer Vision and Image Understanding*, volumen 133, 2015: 76-89.

Gosselin P.H., Precioso F., Philipp-Foliguet S. Incremental kernel learning for active image retrieval without global dictionaries. *Pattern Recognition*, volumen 44, 2011: 2244-2254.

Haralick R., Shanmugam K., Dinstein I. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on, SMC-3*, volumen 113, 1973: 610-621.

Huang-Yueh-MinRay X. A unified hierarchical appearance model for people re-identification using multi-view vision sensors, en: Advances in Multimedia Information Processing - PCM, volúmen 5353, pp. 553-562.

INRIA. INRIA person dataset [en línea]. Disponible en: http://pascal.inrialpes.fr/data/human/ , 2005.

Jain A.K., Dass S.C., Nandakumar K. Can soft biometric traits assist user recognition? *Proceedings in Biometric Technology for Human Identification*, 2004: 561-572.

Jolliffe I.T. Discarding variables in a principal component analysis. II: Real Data. *Journal of the Royal Statistical Society*. Series C (Applied Statistics), volumen 22, 1973: 21-31.

Kamgar-Parsi B., Lawson-W. Toward development of a face recognition system for watchlist surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volumen 33, 2011: 1925-1937.

Kyrki V., Kamarainen J.K., Kälviäinen H. Simple gabor feature space for invariant object recognition. *Pattern Recognition Letters*, volumen 25, 2004: 311-318.

Krüger V. y Sommer G. Gabor wavelet networks for efficient head pose estimation. *Image and Vision Computing*, volumen 20, 2002: 665-672.

Lades M., Vorbrüggen J.C., Buhmann J.M., Lange J., Von-Der M.C., Würtz R.P., Konen W. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, volumen 42, 1993: 300-311.

Laptev I. Improvements of object detection using boosted histograms. *Image Vision Computing*, volumen 27, 2006: 949-958.

Milani S., Bernardini R., Rinaldo R.Z. A saliency-based rate control for people detection in video, en: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), mayo de 2013, pp. 2016-2020, 26-31.

MIT. MIT pedestrian dataset [en línea]. Disponible en: http://cbcl.mit.edu/cbcl/softwaredatasets/PedestrianData.html. , 2000.

Moore S. y Bowden R. Local binary patterns for multi-view facial expression recognition. *Computer Vision and Image Understanding*, volumen 115, 2011: 541-558.

Nghiem-Tuan A.N. y Bremond F. Background subtraction in people detection framework for RGB-D cameras, en: 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), agosto de 2014, pp. 241-246.

OPENCV. Open source computer vision. 80, 2013.

UNION A.C.L. *Is the U.S. turning into a surveillance society?*, American Civil Liberties Union, 2009.

Vadivel A., Sural S., Majumdar A.K. An integrated color and intensity co-occurrence matrix. *Pattern Recognition Letters*, volumen 8, 2007: 974-983.

Viola P.A., Jones M.J. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition*, volumen 1, 2001: 511-518.

Yuan X., Sun Z., Varol Y.L., Bebis G. A distributed visual surveillance system, Conference on Advanced Video and Signal Based Surveillance, 2003, pp. 199-204.

Zweng A. y Kampel M. Improved relational feature model for people detection using histogram similarity functions, en: International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 18-21 de septimbre de 2012, pp. 422-427.

Este artículo se cita:

Citación estilo Chicago

Moctezuma-Ochoa, Daniela Alejandra. Re-identificación de personas a través de sus características soft-biométricas en un entorno multi-cámara de video vigilancia. *Ingeniería Investigación y Tecnología*, XVII, 02 (2016): 257-271.

Citación estilo ISO 690

Moctezuma-Ochoa D.A. Re-identificación de personas a través de sus características soft-biométricas en un entorno multi-cámara de video vigilancia. *Ingeniería Investigación y Tecnología*, volumen XVII (número 2), abril-junio 2016: 257-271.

Semblanza del autor

Daniela A. Moctezuma-Ochoa. Investigador cátedra CONACYT adscrita al CentroGEO. Obtuvo el doctorado en tecnologías de la información y sistemas informáticos en la Universidad Rey Juan Carlos de España con una tesis de video vigilancia inteligente, donde obtuvo mención honorífica. Tiene más de seis años de experiencia en visión artificial y reconocimiento de patrones. Sus principales líneas de investigación son: visión artificial, aprendizaje incremental, video vigilancia inteligente y reconocimiento de patrones.