

REPORTE TÉCNICO  
**Reconocimiento  
de Patrones**

**Re-identificación de personas en la  
red de cámaras de video protección**

**Edel García Reyes**

**RT\_067**

**diciembre 2014**





**CENATAV**

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

RNPS No. 2142  
ISSN 2072-6287  
Versión Digital

**SERIE AZUL**

REPORTE TÉCNICO  
**Reconocimiento  
de Patrones**

**Re-identificación de personas en la  
red de cámaras de video protección**

**Edel García Reyes**

**RT\_067**

**diciembre 2014**



## Tabla de contenido

1.	Introducción . . . . .	1
2.	Definición de la tarea de re-identificación de personas . . . . .	3
3.	Descriptores de los objetos . . . . .	4
3.1.	Rasgos globales . . . . .	5
3.1.1.	Color . . . . .	5
3.1.2.	Representación basada en contornos . . . . .	6
3.1.3.	Flujo óptico . . . . .	7
3.1.4.	Textura . . . . .	7
3.1.5.	Experiencias en el empleo de los rasgos globales . . . . .	8
3.2.	Rasgos locales . . . . .	8
3.2.1.	SIFT, SURF, Gabor . . . . .	8
3.2.2.	HOG . . . . .	9
3.2.3.	Experiencias en el empleo de los rasgos locales . . . . .	9
4.	Algoritmos del estado del arte . . . . .	11
4.1.	Symmetry-Driven Accumulation of Local Features (SDALF) . . . . .	11
4.2.	Mean Riemannian Covariance Grid (MRGC) . . . . .	11
4.3.	Uso de colores semánticos . . . . .	13
4.4.	Eficacia de los algoritmos del estado del arte . . . . .	13
5.	Conclusiones . . . . .	14
	Referencias bibliográficas . . . . .	16

## Lista de figuras

1.	Variaciones típicas de la apariencia y poses de los personas en las imágenes de baja resolución de las cámaras de vigilancia. Para la misma persona se muestran dos vistas en cada columna. Tomado de [1] . . . . .	3
2.	El diagrama muestra la estructura general de un método de re-identificación. Los peatones son detectados, seguidos y se construye la secuencia del objeto. Los colores de las dos cámaras son calibrados. Luego el vector de rasgo es construido para cada objeto. Los rasgos de cada objeto son utilizados para construir un modelo. Los rasgos son comparados usando una función de similitud, que puede tomar en cuenta la variabilidad de los objetos. (Tomado de Internet) . . . . .	4
3.	Histograma de color y moderación de la apariencia . . . . .	5
4.	Histograma de posición del color. Tomado de [2] . . . . .	6
5.	Representación del flujo óptico denso de una persona hablando[3] . . . . .	7
6.	Colección de descriptores locales . . . . .	9
7.	Descriptores locales de Gabor . . . . .	9
8.	Orientaciones locales de los gradientes [4] . . . . .	10
9.	Vista esquemática de la construcción del modelo. Tomado de [1] . . . . .	10
10.	Vista esquemática de la re-identificación. Tomado de [1] . . . . .	10
11.	Imágenes de personas y mapas de elementos sobre salientes. Tomado de [5] . . . . .	11
12.	Dos instancias de una persona consus ejes de asimetría y simetría. Tomado de [6] . . . . .	12
13.	Cálculo de la MRCG. Tomado de [7] . . . . .	12

14. (a) Imágenes originales. (b) Imágenes resultantes. Note que sólo se muestra el color con la probabilidad más alta [8]. . . . .	13
15. Evaluación del método basado en colores semántico en la base de datos VIPER [8] . . . . .	14

# Re-identificación de personas en la red de cámaras de video protección

Edel García Reyes

Equipo de Reconocimiento de Patrones, Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV),  
La Habana, Cuba  
egarcia@cenatav.co.cu

RT\_067, Serie Azul, CENATAV  
Aceptado: 20 de octubre de 2014

**Resumen.** En un ambiente de video protección es necesario detectar diferentes instancias de un individuo en las cámaras esparcidas y sin solape en los campos de observación visual (FOV - en Inglés). Este es un primer paso para adicionar al sistema mayores capacidades de recuperación de información. En este trabajo se revisan los algoritmos del estado del arte, que han sido publicados en los últimos dos o tres años, lo que viene a actualizar recientes estudios realizados sobre el tema por otros autores. Con el objetivo de ubicar al lector, que pudiera acercarse al tema mediante la lectura de este reporte, en el mismo se define la tarea de re-identificación, se abordan los descriptores mas utilizados, se muestran algunas experiencias recientes de utilización de los rasgos globales y locales. Se describen los algoritmos que actualmente reportan los mejores resultados de eficacia en las principales bases de datos de referencia. Vale la pena mencionar, que a pesar de los avances alcanzados en estos 15 años de investigaciones en re-identificación de personas, las tasas de eficacia no superan el 50 porciento en el rango 1 de las curvas CMC que se utilizan para su evaluación.

**Palabras clave:** re-identificación, video vigilancia, video protección, red de cámaras de vigilancia.

**Abstract.** In an environment of video protection is necessary to detect different individual instances in the scattered cameras without overlap in the fields of visual observation (FOV). This is a first step for adding more capacity to the system for information retrieval. In this paper we review the state of the art algorithms, which have been published in the last two or three years, updating the recent studies on the subject by other authors. In order to locate the reader, which could approach the topic by reading this report, the re-identification task is defined. The most commonly used descriptors are studied and some recent experiences of use of global and local features are showed. We describe the algorithms that currently report the most effective results in the principal benchmark databases. It is worth noting that despite the progress made in these 15 years of research on person re-identification, success rates do not exceed 50 percent in the rank-1 of CMC curves used for evaluation.

**Keywords:** re-identification, video surveillance, video protection, surveillance camera network.

## 1. Introducción

En los último dos o tres años ha existido una explosión de trabajos [9][10][11] [12] sobre la re-identificación de personas[13][14][15][16][17]. En el reporte "People re-identification in surveillance and forensics: a survey"(publicado en enero del 2013) [18] se revisan más de 100 artículos científicos recientes y se muestra una panorámica multidimensional de los enfoques de re-identificación de personas en cuanto a:

- la disposición de cámaras,
- la cantidad de muestras,

- el conjunto de rasgos utilizados,
- la forma de modelar el cuerpo humano,
- las técnicas de Aprendizaje de Máquina que se emplean, y
- el escenario de aplicación

La combinación de todas estas variables da una gran variedad de métodos que han sido desarrollados. El trabajo incluye una descripción de las bases de datos experimentales disponibles y las métricas utilizadas para evaluar la eficacia de los métodos. En el mismo se concluye que a pesar de los avances alcanzados en quince años de investigación en este tema, todavía es un problema abierto.[19]

En el presente reporte técnico nos restringimos al estudio de los métodos para la re-identificación de personas en video protección, y profundizaremos en algunos aspectos no tratados en el estudio anteriormente mencionado.

En cuanto a la disposición de las cámaras, serán considerados los métodos que se pueden utilizar cuando no es posible explotar las relaciones geométricas y temporales ya que las mismas se desconocen y no se presupone solape entre los campos de visión de las cámaras, que pudieran permitir una calibración para la reconstrucción 3D y el enfrentamiento a los problemas de oclusión. Este es el escenario más real en nuestro país, donde es prácticamente imposible mantener un registro actualizado de toda la información de las cámaras, ya que estas son movidas de lugar y sus orientaciones cambian constantemente, por necesidades reales del trabajo.

En cuanto al número de muestra que se posee de la persona, el interés se centra en los métodos que trabajan con una secuencia de fotogramas del individuo a buscar [20] [21], aunque existen métodos que se pueden usar también en el caso de contar con solamente un fotograma de la persona requerida.

Respecto a los rasgos o su combinación es interesante actualizar el conocimiento sobre todos los rasgos o sus combinaciones que están siendo utilizados para enfrentar la tarea de re-identificación de personas, ya sean rasgos de color, forma, topología, textura o biometría débil.

Es de interés estudiar los métodos que explotan el conocimiento previo de la forma genérica del cuerpo humano y su estructura en la localización los rasgos visuales extraídos, para garantizar una correcta comparación de las correspondientes partes del cuerpo y minimizar los problemas que surgen por errores en la segmentación de las siluetas y por las oclusiones.

En este reporte no se profundiza en los algoritmos de Aprendizaje de Máquina que pudieran ser utilizados en las diferentes etapas del proceso de re-identificación, ya que el interés se concentra en los métodos que no requieren muestras de aprendizaje, tiempos de entrenamiento, ni de sucesivos ajustes ante el cambio de escenarios y la aparición de nuevos objetos.

Finalmente, respecto al escenario de aplicación de la re-identificación de personas, se restringe a los métodos para la vigilancia en exteriores donde la resolución de las imágenes es insuficiente para el uso de rasgos biométricos fuertes o para utilizar modelos 3D del cuerpo muy detallados.

En cuanto al intervalo temporal se presupone que las trayectorias de corto tiempo de una cámara se pueden combinar con las de las otras cámaras, para detectar la trayectoria larga del individuo en la red de cámara.

En este reporte, con el objetivo de ubicar al lector, que pudiera acercarse al tema mediante su lectura, se define en la sección 2 la tarea de re-identificación, en la sección 3 se abordan los descriptores mas utilizados y se muestran algunas experiencias recientes de utilización de los rasgos globales y locales. En la sección 4 se describen los algoritmos que actualmente reportan los mejores resultados de eficacia en las principales bases de datos de referencia. Finalmente, se da las conclusiones del trabajo.

## 2. Definición de la tarea de re-identificación de personas

La re-identificación de personas para la búsqueda forense es la tarea de asignar un mismo identificador a todas las instancias de un individuo en un conjunto de imágenes o videos, aún después de haber transcurrido cambios en el tiempo y en el espacio. Esto significa que no se exige una continuidad temporal, que puede haber transcurrido considerable tiempo entre el momento asociado a la descripción que se tiene del individuo y las fechas de toma de las imágenes o videos donde se requiere buscarlo. Por otra parte, tampoco se restringe el lugar de la búsqueda a las imágenes y videos tomados en un mismo sitio, pues se trata de encontrar la posible aparición del individuo en otro lugar monitoreado por otra cámara. Esta tarea de extrema complejidad requiere para su solución de rasgos biométricos fuertes.

En el caso de la re-identificación de personas para la video protección inteligente se requiere restringir el lapso de tiempo entre la toma de la muestra y la captura de la base de videos donde se realiza la búsqueda [22], puesto que, por lo general, se utilizan rasgos de descripción de la apariencia, que no son persistentes por mucho tiempo, fundamentalmente, los llamados rasgos biométricos débiles.

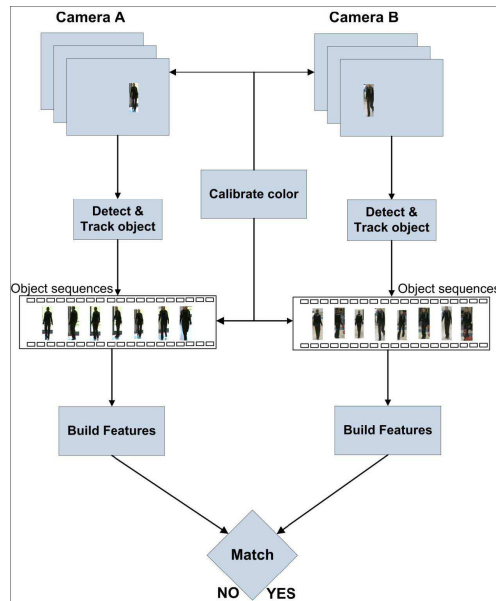
En tal tarea la apariencia de las personas que se mueven varía dependiendo de varios factores, tales como las condiciones de iluminación, el ángulo de la cámara y los cambios de pose. En la Figura 1 se muestran algunas de las variaciones de la apariencia en las imágenes de baja resolución de las cámaras de vigilancia.



**Fig. 1.** Variaciones típicas de la apariencia y poses de las personas en las imágenes de baja resolución de las cámaras de vigilancia. Para la misma persona se muestran dos vistas en cada columna. Tomado de [1]

En el contexto de esta tarea, puede ser muy útil conocer los caminos más probables seguidos por una persona en una red de cámaras. Es decir, dado un fotograma con una persona de interés, el sistema debe devolver un conjunto de posibles caminos recorridos por esta persona a través de la red de cámaras, mostrados en orden decreciente de probabilidad. Este escenario, se corresponde con la resolución de un problema fuera de línea, utilizando la información de las secuencias de videos capturadas y almacenadas por cada nodo de la red de cámaras. Es decir, el principal reto es encontrar la correspondencia entre trayectorias locales del mismo objeto extraídas de secuencias de video capturadas por diferentes cámaras en diferentes momentos. El interés es alcanzar una combinación de trayectorias locales que permita comprender las actividades realizadas por el individuo a un nivel global. En la Figura 2 se muestra la estructura general de un método de re-identificación de persona en la red amplia de cámaras de vigilancia.

Este problema de la re-identificación también está relacionado con el seguimiento de personas en la red amplia de cámaras, pues cuando una persona sale del campo visual de una cámara, los algoritmos de seguimiento, deberán determinar cuando entró en el campo visual de otro nodo para reiniciar el seguimiento local. Sin embargo, este tema está fuera del alcance de este reporte.



**Fig. 2.** El diagrama muestra la estructura general de un método de re-identificación. Los peatones son detectados, seguidos y se construye la secuencia del objeto. Los colores de las dos cámaras son calibrados. Luego el vector de rasgo es construido para cada objeto. Los rasgos de cada objeto son utilizados para construir un modelo. Los rasgos son comparados usando una función de similitud, que puede tomar en cuenta la variabilidad de los objetos. (Tomado de Internet)

### 3. Descriptores de los objetos

La eficacia de la solución a este problema de re-identificación de personas en una red amplia de cámaras depende a su vez de la robustez de la descripción de quién, de la certeza de la determinación de dónde fue detectado y cuándo fue detectado. Esta información constituye los indicios claves para realizar la tarea.

La descripción del quién se refiere a la apariencia de la persona a re-identificar, y aquí se comienzan a presentar problemas para encontrar rasgos discriminativos y robustos debido a la naturaleza compleja de la apariencia humana, puesto que es prácticamente imposible predecir como se perciben las características físicas de una persona en ambientes de observación no controlados y las actividades que realizan las personas hacen que varíen sus poses. Por otro lado, las cámaras como sensor presentan limitaciones para estimar, por sí solas determinadas características físicas como el peso corporal.

Características fisiológicas en el aspecto físico, serán visibles o no en dependencia de la distancia, la resolución, la posición de la cámara y las condiciones de iluminación. Entre estas tenemos: la altura de la persona, el tono de la piel, los órganos sexuales, lunares, pecas, color y tipo de pelo, color de los ojos, forma de la nariz, forma de los oídos, forma del cuerpo, deformaciones, mutilaciones, amputaciones, cicatrices, quemaduras y heridas, etc. En la apariencia influye mucho la ropa y el aspecto personal: vestidos, sombrero, calzado, corte de pelo, barba, bigote, maquillaje y pinturas, modificaciones corporales como piercing y tatuajes; objetos decorativos como collares, pulseras y anillos; dispositivos médicos que cambian la forma del cuerpo como dientes postizos, vendas, audífonos, collarines, lentes de contacto de colores diferentes, gafas, etc.

En el escenario que nos ocupa es más frecuente la observación a distancia de las personas, donde es imposible percibir detalles de su cara, y en dependencia de la aglomeración de personas y objetos en la



escena solo se visualizan características del cuerpo o de partes del cuerpo humano [23]. De este modo la apariencia de una persona varía mucho de una cámara a otra.

Por estas razones los algoritmos de detección de personas deben usar rasgos cuidadosamente seleccionados para obtener resultados confiables.

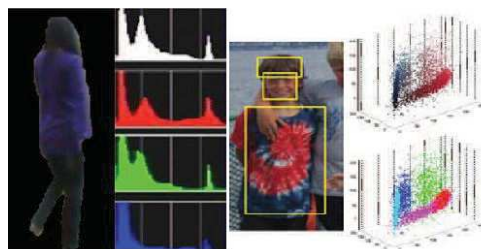
Los rasgos de bajo nivel para la descripción de la apariencia pueden ser divididos en rasgos visuales globales y rasgos visuales locales. Los globales codifican el objeto como un todo. Los rasgos locales describen el objeto como una colección de descriptores locales independientes, ejemplos, parches locales. Los rasgos globales son muy sensibles a oclusiones parciales, ruidos, cambios en los puntos de vista y cambios de iluminación. Los descriptores locales son más robustos, pero se pierde información espacial.

### 3.1. Rasgos globales

En una revisión sobre los descriptores de apariencia para la re-identificación de personas publicado en [24], (en julio de 2013), se da una visión desde dos puntos de vista de la descripción: la primera, esta relacionada con la forma en que se modela el cuerpo humano, es decir, si se considera la silueta entera, o si se detectan y describen diferentes partes del cuerpo. La segunda, se refiere propiamente a los rasgos visuales usados para describir el objeto. No obstante, se debe tener en cuenta, que cuando las personas usan ropas de colores semejantes, como uniformes o abrigos, los rasgos de apariencia deben ser acompañados con información del contexto para facilitar la re-identificación de las personas [25]. En el reporte que nos ocupa, mencionamos los rasgos comúnmente utilizados, pero el lector interesado puede leer el trabajo arriba mencionado. En esta sección, daremos una breve descripción de los rasgos y trataremos de resumir los resultados obtenidos por los investigadores que los han utilizado en la tarea de re-identificación.

#### 3.1.1. Color

El color aparente de un objeto está influenciado por la distribución de la potencia espectral del iluminante y por las propiedades de reflectancia de la superficie del objeto. Aunque, el espacio de colores RGB es ampliamente utilizado, se trata de utilizar otros espacios perceptualmente uniformes, como el  $L^* a^* b^*$ , donde la diferencia entre dos colores representados en el espacio se corresponden con la diferencia percibida por los humanos.



**Fig. 3.** Histograma de color y moderación de la apariencia

La apariencia ha sido modelada por la mezcla de modelos gaussianos, es decir; la función de densidad de probabilidad se expresa como la suma de componentes de densidad gaussianos. Cada componente paramétrica se especifica por la media y la matriz de covarianza y además se debe especificar el peso de cada componente en la combinación lineal. En la Figura 3 se muestra el histograma de color y la moderación de la apariencia del color.

A pesar de que los histogramas son robustos para objetos deformables, ellos no pueden discriminar entre apariencias que son las mismas en cuanto a la distribución del color, pero diferentes en la estructura del color, ya que ellos descartan toda la información espacial.



**Fig. 4.** Histograma de posición del color. Tomado de [2]

El histograma de posición del color es obtenido dividiendo la silueta de una persona en franjas horizontales iguales, a las cuales se le asigna el color medio de los puntos que ella contiene. En la Figura 4 se muestra el histograma de posición de color para una silueta. Se destaca que comparado con el histograma clásico contiene información espacial y usa menos memoria. Existen otras variantes que intentan preservar parte de la información espacial en el modelado del color de los objetos, entre ellos, los espaciogramas.

### 3.1.2. Representación basada en contornos

Una propiedad importante de los bordes es que son menos sensitivos a cambios de iluminación comparados con los rasgos de colores. Existe un conjunto de algoritmos que permiten obtener los bordes de los objetos que se encuentran en las imágenes, los cuales se deben representar de una forma más abstracta, para lo cual existen diferentes métodos:

- Contornos paramétricos: el perfil del contorno se representa mediante una descripción paramétrica (implicando un orden secuencial). Ejemplos: vectores  $(x(t), y(t))$ ; señales complejas  $u(t)=x(t)+jy(t)$ ; códigos de cadena de Freeman, etc.
- Contornos como conjuntos de puntos característicos: es la manera más simple de representar una forma. Esta representación, aunque pobre, permite obtener ciertos descriptores estadísticos de las formas (p.ej. centroide, eje mayor, ...).
- Aproximación de contornos mediante curvas: el contorno de un objeto se representa mediante interpolación. Resulta mejor representar un contorno por tramos que aproximen cada parte de éste mediante una primitiva geométrica (p.ej. un segmento), que describir el contorno mediante una función global. Un método sencillo de descripción son las aproximaciones poligonales. Se trata de encontrar un conjunto de puntos (vértices) sobre el contorno tales que permitan obtener una buena aproximación del contorno original. La manera clásica de tratar este problema es elegir como vértices los puntos con mayor curvatura local

### 3.1.3. Flujo óptico

El flujo óptico es el patrón del movimiento aparente de los objetos, superficies y bordes en una escena causado por el movimiento relativo entre un observador y la escena. Es un campo denso de vectores de desplazamiento que definen la traslación de cada píxel en una región de un fotograma a otro. Asumen que se mantiene constante la brillantez de píxeles correspondientes en fotogramas consecutivos.



**Fig. 5.** Representación del flujo óptico denso de una persona hablando[3]

El cálculo del flujo óptico no es un problema sencillo y computarlo para tomas realizadas en un ambiente real puede llegar a ser muy complicado. Por eso los algoritmos existentes se basan en determinadas hipótesis (intensidad constante, rigidez de los objetos, coherencia espacial, entre otras) que generalmente no se cumplen en escenarios reales. Las técnicas para calcular el flujo óptico se clasifican en:

- Diferenciales: consiste en computar la velocidad a partir de la Derivada espacio-temporal de la intensidad de la imagen o de alguna versión filtrada de la misma
- Frecuencia o energía: utilizan la Transformada de Fourier para calcular el flujo a través del dominio de la Frecuencia. Estas suelen ser adecuadas para extraer el movimiento de objetos que son difíciles de capturar para los métodos de correspondencia de bloques y diferenciales, como puede ser el caso de puntos aleatorios
- Basadas en la correlación: se basan en la búsqueda de correspondencias utilizando ventanas o patrones alrededor de cada píxel. La ventaja que tienen con respecto a los anteriores, es que se utiliza información de los vecinos que en muchos casos hacen que la búsqueda de la correspondencia sea más efectiva. Lo que se busca con estos métodos es asociar píxeles a través de regiones similares en las imágenes que se obtienen por maximización de alguna medida de similitud.

### 3.1.4. Textura

Existen varios descriptores de textura. Las texturas también son menos sensitivas a cambios de iluminación comparadas con los colores. Entre los modelos para estudiar la textura se encuentran[26]:

- Modelo estadístico: es considerado el más importante que se ha propuesto para el análisis de textura. Este modelo utiliza los rasgos estadísticos de la textura de la imagen, en la cual la textura es modelada como una función de los tonos de gris. Desde el punto de vista estadístico se observa que una textura puede definirse por un conjunto de datos estadísticos extraídos de un gran conjunto de propiedades locales de la imagen.
- Técnicas espectrales: están basadas en propiedades del espectro de Fourier y son usadas principalmente para detectar la periodicidad global presente en la imagen identificando concentración grande de la energía y picos estrechos en el espectro. En este conjunto de descriptores están considerados los descriptores derivados del filtrado digital lineal, no lineal, operadores gradientes, el filtrado digital multicanal con la transformada de Gabor.

- **Modelo estructural o geométrico:** se caracterizan porque en la definición de textura están presentes **elementos de la textura** o **primitivas**. El método de análisis normalmente depende de las propiedades geométricas de éstos elementos de la textura. Una vez que esos elementos de la textura se identifican en la imagen, hay dos enfoques principales para analizar la textura. Uno calcula las propiedades estadísticas de los elementos de la textura extraídos y utiliza éstos como descriptores de la textura. El otro enfoque, intenta extraer la regla de ubicación que describe la textura. El último puede incluir métodos sintácticos o geométricos de análisis de textura.

A pesar de que algunos autores han propuesto el uso de rasgos LBP para la modelación de la textura, por si sola no parece ofrecer buenos resultados para la tarea de la re-identificación.

### 3.1.5. *Experiencias en el empleo de los rasgos globales*

En [2] se realiza una evaluación sobre un conjunto de datos reales contenidos en una secuencia de video de 35 personas capturadas por dos cámaras instaladas abordo de un tren en diferentes posiciones. El conjunto de datos es muy difícil debido a que las cámaras están puestas en diferentes ángulos y la adquisición es influenciada por muchos factores, tales como variaciones de iluminación, reflexiones, vibraciones. Para cada pasajero en frente de una cámara, se extraen 20 cuadros regularmente espaciados en el cual la persona es completamente vista. En los experimentos se calculan tres tipos de rasgos: histogramas de color y espaciogramas con 8 bins para cada canal de color, el descriptor color/longitud de camino con 8 bins por canal de color y 8 bins por longitud de camino. Para cada silueta, es aplicado un procedimiento de invariancia a la iluminación antes de extraer los rasgos descriptores. Estos autores plantean, que existen muchos métodos publicados para hacer esto y que muchos de ellos fueron probados. De hecho, los que mejor resultados mostraron fueron:

- Normalización al Gris, la que es derivada del espacio RGB dividiendo el pixel por el promedio de la imagen en el area correspondiente a la persona que se mueve para cada canal de color.
- Normalización usando ecualización de histogramas, que es basada en el supuesto de que el orden de la respuesta del sensor es preservado a través de los cambios de iluminación de la imagen.
- Normalización Afín, donde a cada pixel se le resta el valor medio y se divide entre la desviación estándar.

Para cada pasajero en frente de una cámara, se calculó la disimilitud con respecto a las personas candidatas de la otra cámara. Los resultados de los experimentos mostraron que el rendimiento del histograma de color fue el peor, que usando espaciograma y normalización mediante ecualización de histogramas se incremento el rendimiento al 97 por ciento.

## 3.2. Rasgos locales

### 3.2.1. *SIFT, SURF, Gabor*

La imagen puede ser descrita usando una colección de descriptores locales o parches muestreados densamente o utilizando puntos de interés [27]. Ver Figura 6. Para estudiar estos métodos remitimos al lector a los trabajos [28,29,30,31,32].

Como descriptores locales son ampliamente utilizadas las descripciones de intensidad, gradiente y forma. Se destacan los rasgos SIFT [33] y los descriptores locales de Gabor, los cuales se ejemplifican en la Figura 7.



**Fig. 6.** Colección de descriptores locales



**Fig. 7.** Descriptores locales de Gabor

### 3.2.2. HOG

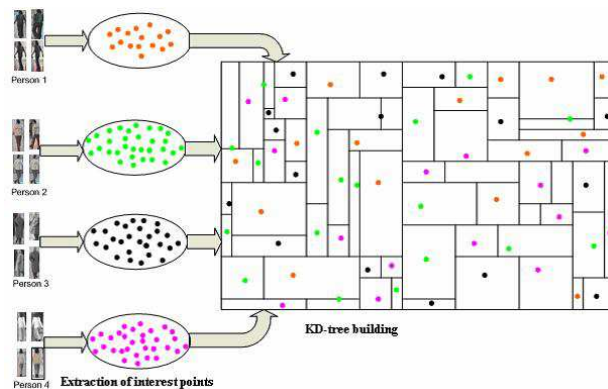
Por otra parte, diferentes variantes de los Histogramas de Orientaciones de Gradientes (HOG) constituyen los rasgos más utilizados por los algoritmos de detección de personas [4]. Ver Figura 8. La técnica cuenta las ocurrencias de los gradientes de orientación en porciones localizadas de la imagen. Este método es similar al de histogramas de orientación de bordes, al de transformación de rasgos invariantes a escala (SIFT) y a los contextos de forma, pero es diferente en que se calcula sobre una rejilla densa de celdas uniformemente espaciadas y usa solape local para la normalización del contraste y mejorar la eficacia. También se han utilizado en la detección de autos.

### 3.2.3. Experiencias en el empleo de los rasgos locales

1. Un trabajo interesante basado en puntos de interés [1], es presentado por Omar H. (2011) en su tesis de doctorado. En esta tesis, se propone y experimenta un nuevo método para la re-identificación de peatones entre cámaras disjuntas. La técnica es basada en la detección y acumulación, durante el seguimiento dentro de una cámara, de puntos de interés caracterizado por descriptores locales. Ver la vista esquemática de la construcción del modelo mostrada en la Figura 9. La detección de los puntos de interés puede ser realizada sobre las regiones de movimiento obtenidas por los métodos estándares de sustracción de fondo y detección de movimiento. La re-identificación es realizada mediante la recolección de un conjunto de puntos de interés durante una ventana de tiempo, y buscando para cada uno de ellos la correspondencia más similar entre todos los descriptores previamente almacenados en un KD-Tree. Un KD-Tree es una estructura de datos de particionado del espacio que organiza los



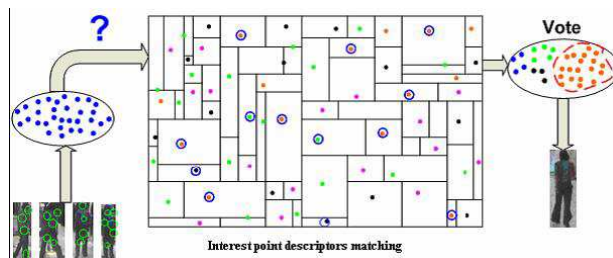
**Fig. 8.** Orientaciones locales de los gradientes [4]



**Fig. 9.** Vista esquemática de la construcción del modelo. Tomado de [1]

puntos en un Espacio euclídeo de  $k$  dimensiones. Emplea sólo planos perpendiculares a uno de los ejes del sistema de coordenadas. Todos los nodos, desde el nodo raíz hasta los nodos hoja, almacenan un punto. Para crear el KD-tree, por ejemplo, la raíz puede tener un plano alineado con el eje  $x$ , sus descendientes tendrían planos alineados con el  $y$  y los nietos del raíz alineados con el  $z$ , y así sucesivamente. En cada paso, el punto seleccionado para crear el plano de corte será la mediana de los puntos puestos en el KD-tree.

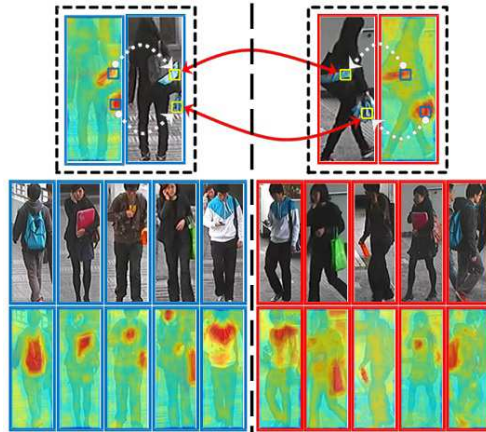
La re-identificación es realizada mediante un proceso de votación, como se muestra en la Figura 10.



**Fig. 10.** Vista esquemática de la re-identificación. Tomado de [1]

2. En [5] [34] se presenta otro trabajo interesante basado en la correspondencia de parches locales sobresalientes de las imágenes de las personas, con lo que se pretende atacar el deslineamiento de las imágenes por el problema de la pose. Intuitivamente una parte del cuerpo que se sobresale en una vista

de una cámara, también debe ser sobresaliente en la vista de otra cámara. Un conjunto de imágenes ilustrarías de este método son mostradas de la Figura 11.



**Fig. 11.** Imágenes de personas y mapas de elementos sobre salientes. Tomado de [5]

#### 4. Algoritmos del estado del arte

En esta sección mencionaremos los algoritmos con mejores resultados en las comparaciones de eficacia, que han sido reportados en los últimos 3 años. Entre ellos tenemos, el Symmetry-Driven Accumulation of Local Features (SDALF) [6], el que logra eficacia que compite con el estado del arte.

##### 4.1. Symmetry-Driven Accumulation of Local Features (SDALF)

El SDALF [6] codifica la apariencia del sujeto mediante una combinación pesada de histogramas de color (wHSV), Regiones de color de máxima estabilidad (MSCR) [35] y Parches Altamente Estructurados Recurrentes (RHRSR). Primero se encuentran dos ejes horizontales de asimetría que separan el cuerpo en partes: cabeza, torso y piernas. Sobre estas dos últimas regiones se estima un eje vertical de simetría (Ver Figura 12). Los rasgos extraídos son pesados por la distancia al eje vertical para minimizar los efectos de cambio de pose.

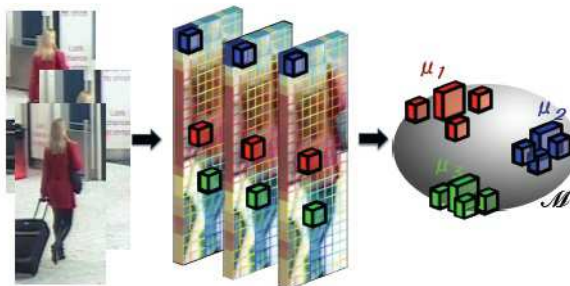
##### 4.2. Mean Riemannian Covariance Grid (MRGC)

En [7] se propone un descriptor de la covarianza, altamente discriminativo de las personas, el cual fue diseñado para tratar con las imágenes de baja resolución y en ambientes de grandes concentraciones de personas. Primeramente, para manipular la disimilitudes de colores causada por la diferencia de iluminación en las cámaras se aplica una ecualización de histogramas y la imagen es dividida en celdas con superposición o solape. Esta representación densa hace posible que la firma se comporte robusta ante oclusiones parciales y que contenga información relevante acerca de la correlación espacial entre celdas. Los autores aprovechan la ventaja del seguimiento y los resultados de la detección para combinar información de múltiples imágenes.



**Fig. 12.** Dos instancias de una persona consus ejes de asimetría y simetría. Tomado de [6]

Sea  $Cp_1, \dots, Cp_N$  un conjunto de matrices de covarianza extraídas durante el seguimiento de  $N$  fotogramas correspondientes a regiones cuadradas en la posición de la celda  $p$ . La matriz de covarianza es una matriz simétrica y definida positiva que puede ser vista como un tensor. El problema principal es que el espacio definido es una variedad (manifold) que no es un espacio vectorial con la estructura aditiva usual. Un manifold es un espacio topológico que es localmente similar a un espacio euclidiano. Esto significa que en cada punto sobre el manifold  $m$ -dimensional existe una vecindad que es homeomorfa a un subconjunto abierto de una espacio  $m$ -dimensional de dimensión menor. Realizar operaciones sobre el manifold incluye la selección de una métrica. Especificar el manifold como Riemanniano, nos permite usar la metric de Rieman, lo que automáticamente, pone a disposición del investigador una poderosa plataforma para trabajar sobre el manifold usando las herramientas de la geometría diferencial.



**Fig. 13.** Cálculo de la MRCG. Tomado de [7]

La media riemanniana de las matrices de covarianza (MRC) es definida como la covarianza media de esas matrices calculada usando el espacio de Riemann. Todas las MRC de las celdas que componen la malla completa son nombradas Malla de Covarianza Media Riemanniana (Mean Riemannian Covariance Grid - MRCG). El proceso para el calculo de la MRCG es mostrado en la Figura 13.



Posteriormente, se define una métrica para cotejar dos mallas de matrices de covarianza. En la experimentación, para cada region cuadrada se extrajo un vector de rasgos compuesto por las coordenadas  $x$ ,  $y$  de la celda, los colores en cada uno de los canales y la magnitud y la orientación del gradiente en cada canal de colores del espacio RGB.

Inicialmente, este descriptor se probó mediante la selección manual por parte del operador de la persona a buscar en la red de cámara, mostrando buenos resultados, sin embargo, en experimentos donde se realiza la detección automática, los rendimientos se vieron seriamente afectados. Posteriormente, se realizó una propuesta de LBP simplificado para realizar la detección automática de personas, pero este es un tema que requiere el desarrollo de más investigaciones.

### 4.3. Uso de colores semánticos

En [8] en lugar de usar histogramas de color simples, se propone aplicar nombres de colores semánticos para describir la imagen de una persona en el problema de la re-identificación. Los nombres de los colores básicos son: negro, azul, carmelita, gris, verde, naranja, rosado, púrpura, rojo, blanco y amarillo.

Dada una función que proyecta los valores RGB a la distribución de probabilidad sobre los nombres de los colores, se construye un histograma semántico para describir la imagen. Se debe notar que en este trabajo, se combina el histograma semántico con otros descriptores de bajo nivel utilizando el descriptor de covarianza MRCG.

El uso de colores semántico da la posibilidad de resolver el problema de la búsqueda a partir de una descripción oral o textual de la persona. Por ejemplo, una persona con camisa blanca y pantalón azul.

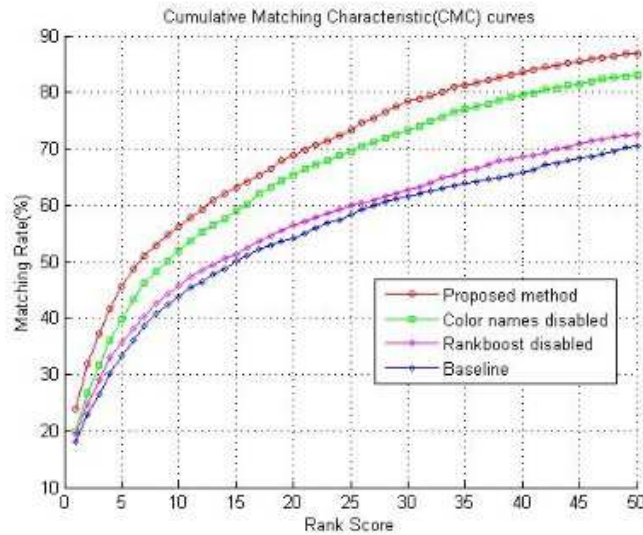


**Fig. 14.** (a) Imágenes originales. (b) Imágenes resultantes. Note que sólo se muestra el color con la probabilidad más alta [8].

### 4.4. Eficacia de los algoritmos del estado del arte

La curva de correspondencia acumulativa (cumulative match curve - CMC) es usada como una medida de la eficacia de un sistema de identificación 1: M. Esta permite juzgar la capacidad de posicionamiento del individuo buscado en la lista de candidatos que devuelve el sistema y que está ordenada por los valores de similitud entre el elemento buscado y los candidatos,

En la revisión bibliográfica realizada se observa que los algoritmos del estado del arte son evaluados usando fundamentalmente las bases de datos de referencias ETHZ, VIPER, CAVIAR, i-LIDS y en general, la eficacia medida utilizando la curva CMC está por debajo de 50 por ciento en el rango-1. En el



**Fig. 15.** Evaluación del método basado en colores semántico en la base de datos VIPER [8]

experimento al cual corresponde la Figura 15, el método propuesto alcanzó un 23.92 por ciento para  $r=1$  y el método SDALF un 19.83 por ciento para ese rango.

## 5. Conclusiones

En este reporte se ha realizado una actualización del estado de los métodos de re-identificación de personas en una red amplia de cámaras de video protección. En primer lugar se debe destacar que es un tema muy actual de investigación en los últimos años, lo que ha llevado al surgimiento cada año de nuevos algoritmos para atacar esta compleja tarea, que aún permanece sin una solución satisfactoria.

A pesar, de que no se puede hablar de que los autores están utilizando una metodología estándar para evaluar los algoritmos de forma que los resultados sean completamente comparables, los métodos basados en el descriptor de covarianza (MRCG) usado para combinar rasgos de bajo nivel como los histogramas de colores, los colores culturales, los LBP y los rasgos de Gabor, (tomados en regiones locales de la silueta del cuerpo humano) son los que están ofreciendo resultados más prometedores. No obstante, la eficacia en las curvas CMC en el rango-1 no sobrepasa el 50 por ciento en las bases de datos de competencia de cierta complejidad.

Dada la importancia que se le confiere a los colores para descripción de la apariencia en esta tarea de re-identificación de personas, es necesario continuar investigando en modelos más invariantes de los colores, en particular una opción a estudiar es el uso de los cuaterniones, para obtener histograma simples y semánticos de colores, ya que los tres canales de colores se pueden tratar como un solo objeto en el dominio de los cuaterniones, en ese mismo sentido se puede intentar obtener histogramas de orientación de gradientes de colores en este mismo espacio.

Otros problemas a enfrentar es el de la detección robusta de los objetos y la modelación de fondos, pues la calidad con que se realice estas etapas tiene implicaciones en la firma temporal que se obtiene de las personas a re-identificar.

Una línea de investigación interesante está relacionada con la conversión de una descripción textual de la apariencia en descriptores visuales para la re-identificación de personas.

## Referencias bibliográficas

1. Hamdoun, O., Moutarde, F., Stanculescu, B., Steux, B.: Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In: 2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC-08), Stanford, Palo Alto, États-Unis (2008) –
2. Cong, D.N.T., Khoudour, L., Achard, C., Meurie, C., Lezoray, O.: People re-identification by spectral classification of silhouettes. *Signal Processing* **90**(8) (2010) 2362–2374
3. David J. Fleet, Y.W.: Optical flow estimation. *Mathematical Models in Computer Vision: The Handbook* (2005) 239–258
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). (2005) 886–893
5. Zhao, R., Ouyang, W., Wang, X.: Unsupervised saliency learning for person re-identification. In: CVPR, IEEE (2013) 3586–3593
6. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, CA, USA, IEEE Computer Society (2010)
7. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Multiple-shot human re-identification by mean riemannian covariance grid. 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance **0** (2011) 179–184
8. Kuo, C.H., Khamis, S., Shet, V.: Person re-identification using semantic color names and rankboost. In: IEEE Workshop on the Applications of Computer Vision. (2013)
9. Bak, S., Brémond, F.: Person re-identification. *ERCIM News* (2013) –1–1
10. Dutra, C.R.S., Souza, T., Alves, R., Schwartz, W.R., Oliveira, L.R.: Re-identifying people based on indexing structure and manifold appearance modeling. In: SIBGRAPI - Conference on Graphics, Patterns and Images. (2013) 1–8
11. Xiong, F., Gou, M., Camps, O., Sznai, M.: Person re-identification using kernel-based metric learning methods. In: Computer Vision—ECCV 2014. Springer (2014) 1–16
12. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. (June 2014)
13. Wang, X., Zhao, R.: Person re-identification: System design and evaluation overview. In: Person Re-Identification, Springer (2014) 351–370
14. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, USA (June 2014)
15. Wu, Y., Mukunoki, M., Minoh, M.: Locality-constrained collaboratively regularized nearest points for multiple-shot person re-identification. In: Proc. of The 20th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV). (2014)
16. Chen, X., An, L., Bhanu, B.: Soft-biometrics and reference set integrated model for tracking across cameras. In Bobda, C., Velipasalar, S., eds.: Distributed Embedded Smart Cameras. Springer New York (2014) 211–230
17. Martinel, N., Micheloni, C.: Person re-identification by modelling principal component analysis coefficients of image dissimilarities. *Electronic Letters* **50**(14) (July 2014) 1000–1001
18. Vezzani, R., Baltieri, D., Cucchiara, R.: People re-identification in surveillance and forensics: a survey. *ACM Computing Surveys* (December 2013)
19. Roth, P.M., Hirzer, M., Koestinger, M., Beleznai, C., Bischof, H.: Mahalanobis distance learning for person re-identification. In Gong, S., Cristani, M., Yan, S., Loy, C.C., eds.: Person Re-Identification. Advances in Computer Vision and Pattern Recognition. Springer, London, United Kingdom (2014) 247–267 The original publication is available at [www.springer.com](http://www.springer.com).
20. LI, W., WU, Y., MUKUNOKI, M., MINOH, M.: Bi-level relative information analysis for multiple-shot person re-identification. *IEICE Transactions on Information and Systems* **E96-D**(11) (November 2013) 2450–2461
21. Munaro, M., Fossati, A., Basso, A., Menegatti, E., Gool, L.V.: One-Shot Person Re-Identification with a Consumer Depth Camera. In Gong, S., C.M.Y.S.L.C., ed.: Person Re-Identification. Springer (2014) 161–181
22. Fleuret, F., Ben Shitrit, H., Fua, P.: Re-identification for improved people tracking. In Gong, S., Cristani, M., Shuicheng, Y., Loy, C.C., eds.: Person Re-Identification. Springer (2014) 311–336
23. Mazzon, R., Tahir, S.F., Cavallaro, A.: Person re-identification in crowd. *Pattern Recognition Letters* **33**(14) (2012) 1828–1837
24. Satta, R.: Appearance descriptors for person re-identification: a comprehensive review. *CoRR* **abs/1307.5748** (2013)
25. Bialkowski, A., Lucey, P.J., Wei, X., Sridharan, S.: Person re-identification using group information. In: Digital Image Computing : Techniques and Applications (DICTA), Wrest Point Hotel, Hobart, TAS, IEEE (November 2013)
26. Rodríguez, J.L.G.: Estado actual de la representación y análisis de textura en imágenes. In: Serie Azul, Reconocimiento de Patrones, CENATAV. (2008)

27. Zhao, R., Ouyang, W., Wang, X.: Learning mid-level filters for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, USA (June 2014)
28. Tuytelaars, T., Mikolajczyk, K.: A survey on local invariant features. (2008)
29. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. *International Journal of Computer Vision* **60** (2004) 63–86 10.1023/B:VISI.0000027790.02288.f2.
30. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). *Computer Vision and Image Understanding* **110**(3) (2008) 346 – 359 *Similarity Matching in Computer Vision and Multimedia*.
31. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In Heyden, A., Sparr, G., Nielsen, M., Johansen, P., eds.: *Computer Vision - ECCV 2002*. Volume 2350 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg (2002) 128–142
32. Mikolajczyk, K., Leibe, B., & Schiele, B.: Local features for object class recognition. In: *Proceedings of the IEEE Tenth IEEE International Conference on Computer Vision (ICCV '05)*, year = 2005, pages = 792–1799,
33. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60** (2004) 91–110 10.1023/B:VISI.0000029664.99615.94.
34. Zhao, R., Ouyang, W., Wang, X.: Person re-identification by saliency matching. In: *IEEE International Conference on Computer Vision (ICCV)*, Sydney, Australia (December 2013)
35. Forssén, P.E.: Maximally stable colour regions for recognition and matching. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, IEEE Computer Society, IEEE (June 2007)

RT\_067, diciembre 2014

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2014

**Editor:** Lic. Lucía González Bayona

**Diseño de Portada:** Di. Alejandro Pérez Abraham

RNPS No. 2142

ISSN 2072-6287

**Indicaciones para los Autores:**

Seguir la plantilla que aparece en [www.cenatav.co.cu](http://www.cenatav.co.cu)

C E N A T A V

7ma. A No. 21406 e/214 y 216, Rpto. Siboney, Playa;

La Habana. Cuba. C.P. 12200

*Impreso en Cuba*

