

UNIVERSIDAD DE CONCEPCIÓN
Facultad de Ingeniería
Departamento de Ing. Civil Informática y
Ciencias de la Computación

Profesor Patrocinante:
John Atkinson Abutridy

Comisión:
Ma. Angélica Pinninghoff Junemann
Javier Vidal Valenzuela

DESARROLLO DE UN PROTOTIPO DE SISTEMA DE RE-ID DE PERSONAS PARA VIDEO-VIGILANCIA

DIEGO A. REYES MOLINA

Informe de Memoria de Título
Para optar al Título de
Ingeniero Civil Informático

Noviembre 2015

Índice general

1	Introducción	3
1.1	Objetivo general	4
1.1.1	Objetivos específicos	4
1.2	Organización de la Memoria	4
2	Métodos de Re-ID (<i>trabajo relacionado</i>)	5
2.1	Características empleadas	5
2.1.1	Color	6
2.1.2	Forma	6
2.1.3	Textura	6
2.2	Asociación de descriptores	7
2.2.1	Distancias directas	7
2.2.2	Aprendizaje supervisado	8
2.2.3	Optimización Convexa	8
3	Sistema de Re-ID de personas (<i>marco teórico</i>)	10

Índice de cuadros

2.1	Resultados obtenidos en [24]	7
2.2	Resultados obtenidos en [21]	7
2.3	Clasificación de trabajos en tipos de métodos de Re-ID [30]	9

Capítulo 1

Introducción

Por muchos años, se han utilizado cámaras de video en el área de la seguridad para vigilar zonas de interés, transmitiendo escenas en tiempo real desde distintos lugares a puestos de monitoreo centralizados, que analizan simultáneamente varios flujos de imágenes con el objetivo de detectar situaciones no deseadas. En sus inicios este análisis era llevado a cabo sólo por personas, las que debían dedicar tiempo exclusivo para monitorizar imágenes, lo que lo convertía en una labor costosa y poco eficiente.

En las últimas décadas, los avances en el área de procesamiento de imágenes y visión artificial permitieron desarrollar sistemas de vigilancia inteligente capaces de comprender una escena y detectar condiciones de riesgo sin la ayuda de humanos. En términos de seguridad, la mayoría de estos sistemas se dedican al análisis de personas, debiendo enfrentar problemas como la detección automática de personas en la escena. Luego de la detección, se requiere efectuar un seguimiento cuadro a cuadro de la ubicación de la persona a través de una o varias cámaras. Dependiendo de la ubicación de cada cámara, puede haber campos de visión (Field of View, FOV) superpuestos con otros de cámaras vecinas. En estos casos, se puede utilizar la relación existente entre la información capturada por cada cámara [19,31], pudiendo incluso inferir automáticamente la topología de la red de cámaras [16,34]. De esta forma, la existencia de áreas sin cobertura visual puede ser compensada estimando la trayectoria de una persona, basándose en su velocidad y dirección antes de desaparecer [8]. Sin embargo, en redes de cámaras con regiones ciegas demasiado extensas, es imposible predecir cuándo y dónde reaparecerá el individuo. De ahí que se requiere re-identificar personas sin utilizar características derivadas de su posición (dirección, velocidad, aceleración, etc.).

El principal interés en re-identificar una persona es lograr establecer la trayectoria recorrida y los lugares visitados dentro de toda una red de cámaras de seguridad. En video-vigilancia, la re-identificación (Re-ID) de personas consiste en reconocer si un individuo ha sido observado previamente. Formalmente, esto se puede definir como la tarea de asignar el mismo identificador (o identificadores lo suficientemente parecidos) a todas las instancias de una persona, por medio de aspectos visuales capturados desde imágenes o videos [37]. En general, Re-ID busca responder a las preguntas: ¿Dónde se ha visto a esta persona antes? y ¿A dónde fue, luego de ser vista en cierto lugar?. Respondiendo lo anterior, se puede lograr un seguimiento de personas sobre grandes extensiones de terreno resguardado por una red de cámaras [8].

Un subproblema de la Re-ID de personas es la re-identificación dentro de un solo FOV, es decir, establecer cuando un individuo reingresa al lugar monitorizado por una cámara. Una aplicación real de esto [7], es determinar la presencia reiterada de una persona en un lugar sospechoso (paradero

de autobús), con el propósito de detectar traficantes de droga.

En este trabajo, se plantea el problema de re-identificar personas dentro de un mismo FOV, con el objetivo de detectar robos de vehículos, basándose en la apariencia física global de la persona. Para esto, se empleó un modelo de sistema que re-identifique entre: personas que abandonan la escena luego de descender de un vehículo, y las que ingresan a la escena para abordar el mismo vehículo. Una alerta de posible robo se provocará cuando el sistema determine que las imágenes de estas personas corresponden a distintos individuos.

A diferencia de un sistema de Re-ID clásico que busca coincidencias para una persona entre muchos candidatos [17, 18, 20, 22, 33], el modelo propuesto compara cada persona con solo un candidato, a saber, la persona que reingresa al vehículo.

En este trabajo se desarrolló un prototipo de sistema de Re-ID, y se evaluó su efectividad en un escenario real (estacionamiento comercial).

1.1. Objetivo general

Desarrollar un prototipo de sistema Re-ID de personas para vigilancia semi-automática en estacionamientos de automóviles.

1.1.1. Objetivos específicos

- Estudiar técnicas de: detección de personas en videos, segmentación entre primer y segundo plano en videos, extracción de características relevantes en imágenes (cuadros de video), y comparación de descriptores de imágenes.
- Implementar prototipo de sistema de Re-ID de personas.
- Evaluar el prototipo con distintos tipos de características seleccionadas.

1.2. Organización de la Memoria

En el capítulo 2 se discute el trabajo previo relacionado. En el capítulo 3 se presentan los fundamentos teóricos de un clásico sistema de Re-ID de personas.

Capítulo 2

Métodos de Re-ID (*trabajo relacionado*)

En este capítulo se presenta una discusión de distintos métodos de Re-ID de personas, los que agruparemos según dos criterios: el tipo de características empleadas para construir el descriptor, y la técnica utilizada para asociar descriptores.

La efectividad de un sistema de Re-ID se evalúa mediante una curva CMC (*Cumulative Matching Characteristic*), que mide la probabilidad de que el sistema entregue correctamente el descriptor que corresponde al descriptor consultado, donde la respuesta está compuesta por k candidatos, y es considerada correcta si el descriptor que hace match con la consulta está entre dichos candidatos [14]. Luego, CMC indica la proporción de respuestas correctas obtenidas para distintos valores de k :

$$\text{Precisión} = \frac{\# \text{ respuestas correctas}}{\# \text{ consultas realizadas}} \times 100$$

2.1. Características empleadas

Color, textura y forma son las características de apariencia comúnmente usadas por los métodos del estado del arte para Re-ID de personas, siendo algunas veces combinadas para obtener un descriptor más representativo [30].



Figura 2.1: Ejemplos Data Set VIPeR: cada columna muestra una de las 632 personas.

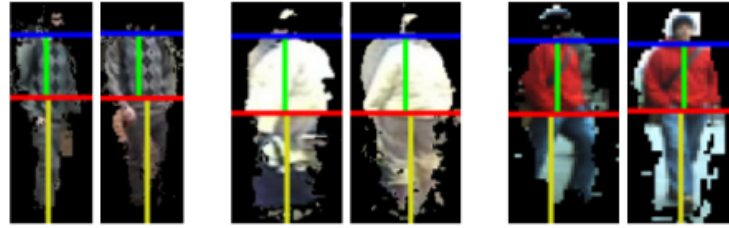


Figura 2.2: Algoritmo SDALF: Segmentación de la imagen usando ejes horizontales (partes asimétricas) y verticales (partes simétricas).

2.1.1. Color

El color como característica de apariencia se ha empleado en forma de histogramas [13, 17, 18, 20, 25, 33, 42], dada la facilidad para calcularlos. Se pueden utilizar diferentes canales de colores y sus combinaciones. Por ejemplo, del espacio de colores HSV, se ha empleado sólo el tono [13], tono y saturación [18], o los tres canales del espacio [17]. Por otro lado, histogramas del espacio de colores RGB fueron utilizados en [6, 25, 32]. Otros [20, 33, 42] han adoptado una concatenación de histogramas de los canales de los espacios RGB, YCbCr y HSV (sólo tono y saturación). En [20], presentaron un dataset (VIPeR) conformado por 632 personas, cada una de ellas fotografiadas (48x128 píxeles) en ambientes exteriores, siempre desde dos ángulos distintos y en diferentes posiciones (ver figura 2.1). Las pruebas determinan que los canales más discriminantes (en orden descendente) para Re-ID personas, son: tono, saturación, azul, rojo y verde [30]. Por último, pruebas de Re-ID con el dataset VIPeR utilizando descriptores compuestos únicamente por histogramas de colores y comparándolos con distancia euclidiana, para rangos $k = 1$ y $k = 20$ en la curva CMC, se obtiene un 6 % y 38 % de precisión de reconocimiento, respectivamente [23].

2.1.2. Forma

Se han propuesto algoritmos que hacen uso de la simetría de la figura humana [17]: *Symmetry-Driven Accumulation of Local Features*, SDALF. El método segmenta la silueta de la persona en tres partes: cabeza, torso y piernas (ver figura 2.2), comparando cada parte con su homóloga correspondiente. Luego, se buscan ejes verticales de simetría para cada una de las partes mencionadas, con el objetivo de ponderar las features (histogramas del espacio HSV de subregiones de píxeles) en relación a la distancia de éstas con el eje, destacando aquellas features que estén cercanas al eje, dado que tienen menor probabilidad de pertenecer al segundo plano de la imagen. Una evaluación de SDALF con el dataset VIPeR, obtiene para $k = 1$ y $k = 20$ en la curva CMC, un 20 % y 65 % de precisión de reconocimiento, respectivamente [17].

2.1.3. Textura

Algunos enfoques buscan Re-ID personas que abandonan y luego reingresan al FOV de una misma cámara, buscando enfrentar de forma robusta, cambios de iluminación, postura y escala [24]. Para formar el descriptor de cada persona, se emplean histogramas de colores y características de textura seleccionadas por SIFT. Las pruebas se realizaron con videos¹ de cuatro escenarios distintos, donde aparece un grupo limitado de personas que abandonan e ingresan en reiteradas

¹<http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>

Escena	personas	ocurrencias	exactitud
Salón	12	80	89.3 %
Corredor	58	38	70.4 %
Camino	10	20	95.0 %
Cancha	6	106	85.9 %

Cuadro 2.1: Resultados obtenidos en [24]

keypoint necesarios	Precisión
40	99 %
30	95 %
20	85 %
10	10 %

Cuadro 2.2: Resultados obtenidos en [21]

ocasiones, obteniendo una exactitud desde un 70.4 % hasta 95 % (ver cuadro 2.1). Un trabajo comparable a [24] es presentado en [21], donde se emplea el mismo dataset y método de extracción de características. La precisión de Re-ID obtenida en este trabajo depende de la cantidad mínima de keypoints coincidentes requeridos para establecer a sus respectivos descriptores como un match (ver cuadro 2.2)

2.2. Asociación de descriptores

La segunda fase de un sistema de Re-ID define la forma de efectuar el match entre descriptores. Éstos se pueden asociar: (a) midiendo su similitud usando métricas de distancia directa, (b) utilizando algoritmos de aprendizaje supervisado y (c) aplicando métodos de optimización [30].

2.2.1. Distancias directas

Las métricas de distancia directas estiman diferencias entre descriptores. La métrica más usada es la distancia euclidiana, utilizada sobre descriptores basados en color [1,17] y puntos de interés [18].

Sin embargo, utilizar únicamente métricas de distancia directa no siempre permite establecer una correcta similitud entre descriptores, pues cuando la mayoría de las features de un descriptor coincide con los de otro correspondiente a una persona desigual, los descriptores respectivos estarán cercanos entre sí. Del mismo modo, para una misma persona que cambia de apariencia (cambio drástico en varios features a causa de diferente iluminación, postura de la persona, punto de vista, etc.), sus respectivos descriptores se encontrarán a mayor distancia. Como consecuencia, al emplear sólo una distancia tradicional, se ignora cualquier regularidad estadística, la que podría ser estimada con algoritmos de aprendizaje supervisados [33,42]. Un sistema de Re-ID de personas que sólo emplea una métrica directa, no es robusto ante cambios de iluminación, razón por la que se requiere una calibración previa en la red de cámaras [19,25,31,34].

2.2.2. Aprendizaje supervisado

Las técnicas de aprendizaje automático (Machine Learning, ML) se han empleado en sistemas de Re-ID para distintos objetivos. Por ejemplo, un clasificador binario puede distinguir entre descriptores positivos (match) y negativos (mismatch). Un sistema de Re-ID basado en SVM, puede etiquetar un par de descriptores como match incluso si éstos presentan más diferencias que otro par compuesto por descriptores de personas distintas [33]. El método anterior obtiene, para $k = 1$ y $k = 20$ en la curva CMC, un 14 % y 68 % de precisión en reconocimiento, respectivamente.

Un algoritmo clásico (AdaBoost) [20], que selecciona las características que determinan una mayor diferencia entre pares de imágenes, con base en pares etiquetados de imágenes, obteniendo un 12 % y 60 % en la curva CMC para rangos de $k = 1$ y $k = 20$, respectivamente.

También, se ha empleado ML para aprender funciones de similitud o distancia (Distance Metric Learning o DML). El objetivo es encontrar un espacio geométrico donde descriptores de una misma persona queden a poca distancia, al mismo tiempo que descriptores de personas distintas estén a una distancia mayor [5]. Una variante de este método establece distancias de forma relativa a un tercer descriptor, por ejemplo, indicando que A está más cerca de B que de C [43]. Esto obtiene, para $k = 1$ y $k = 20$ en la curva CMC, un 15.7 % y 70.1 % de precisión de reconocimiento, respectivamente.

Otros métodos [39] utilizan el espacio geométrico obtenido con un algoritmo DML, en un clasificador kNN. Los resultados muestran un 18 % y 75 % de precisión de reconocimiento para $k = 1$ y $k = 20$ en la curva CMC, respectivamente [23]. Mejoras al enfoque anterior consideran la capacidad de rechazar del clasificador (detectar una pareja negativa) [15]. Para esto, es necesario establecer una distancia umbral, tal que, el elemento consultado es aceptado sólo si existe un vecino cuya distancia es inferior a dicho umbral. En otro caso, el elemento consultado es rechazado. Los resultados obtenidos para $k = 1$ y $k = 20$, alcanzan aproximadamente un 20 % y 80 % de precisión de reconocimiento, respectivamente ².

A diferencia de métricas directas, los métodos que utilizan DML son menos sensibles a las features seleccionadas. Sin embargo, en escenarios reales no siempre se cuenta con un conjunto de datos previamente etiquetados para entrenar el sistema.

2.2.3. Optimización Convexa

En este tipo de métodos, la búsqueda de la métrica se hace por medio de un modelo de optimización convexa, donde la función objetivo consiste en minimizar de la distancia entre pares descriptores que hagan match, con la restricción de mantener una distancia mínima entre descriptores de pares mismatch [25, 27, 31]. La principal desventaja de este enfoque es el costo computacional, debido a la cantidad de restricciones y al tamaño de cada descriptor.

²Resultados obtenidos por la mejor ejecución, a diferencia de otros trabajos donde los autores muestran un promedio de los experimentos realizados

Referencia	Características			Asociación		
	Color	Textura	Forma	Distancia	Aprendizaje	Optimización
[1, 13, 17, 18]	✓	✓		✓		
[2]		✓			✓	
[6, 38]	✓	✓	✓	✓		
[20, 33, 42]	✓	✓			✓	
[21]		✓		✓		
[22]	✓				✓	
[25, 31]	✓					✓
[27]	✓	✓	✓			✓
[32]	✓			✓		
[36]		✓			✓	

Cuadro 2.3: Clasificación de trabajos en tipos de métodos de Re-ID [30]

Capítulo 3

Sistema de Re-ID de personas (*marco teórico*)

En este capítulo, se presenta una discusión de distintos métodos empleados en un sistema de Re-ID de personas.

Re-ID se define como el proceso de establecer correspondencia entre imágenes de personas tomadas desde cámaras diferentes [4].

La identificación (descriptor) de una persona detectada es almacenada en una galería. Si el tamaño de la galería evoluciona con el tiempo, se dice que el sistema de Re-ID es de conjunto abierto, mientras que en caso contrario (galería de tamaño fijo), el sistema es de conjunto cerrado.

Cada vez que se detecta una persona en la escena, su descriptor es consultado con la galería en busca de una correspondencia. Para el caso de galería cerrada, el problema radica en asociar correctamente el descriptor consultado con su *match* correspondiente. En el caso de una galería abierta, primero se debe determinar si la persona existe en la galería (*novelty detection*), agregando el descriptor a la galería en caso de que no exista, o asociándolo con uno existente.

Un sistema típico de Re-ID tiene dos fases [4]: (1) captura de descriptores y (2) comparación de descriptores, tal como se muestra en la figura 3.1.

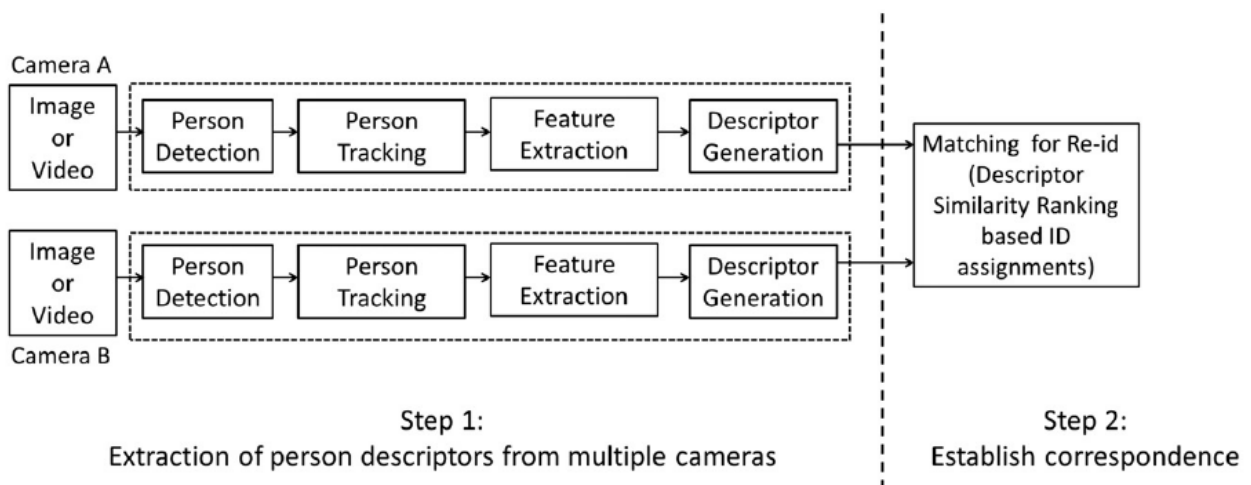


Figura 3.1: Sistema de Re-ID [4]

En una primera fase, se extrae el descriptor de una persona desde múltiples cámaras (o en

nuestro caso, distintas ocurrencias desde una misma cámara). En la segunda fase, se establece la correspondencia entre pares de descriptores, determinando si estos coinciden con la misma persona o son individuos distintos.

Las tareas típicas de la primera fase incluyen [41]:

- Detección de personas: se establece la presencia de personas en una escena. Para esto, se pueden utilizar algoritmos de aprendizaje supervisado cuando se cuenta con una base de datos con imágenes etiquetadas [12]. En el caso de un video, existen varias imágenes consecutivas (cuadros o *frames*), lo que permite detectar cuando la persona se mueve o desplaza por la escena, según los cambios producidos en cada píxel. Esto se conoce como segmentación entre primer y segundo plano (que corresponden al movimiento y regiones estáticas, respectivamente).

La detección del segundo plano permite eliminar de la imagen todo lo que no forma parte de la persona, ocultando todo el conjunto de píxeles que permanezcan sin cambios. Para esto, se utiliza un mapa de bits o matriz de ceros y unos, que se superpone a la imagen original, convirtiendo el segundo plano en un mismo valor (bit cero: color negro), dejando los píxeles del primer plano inalterados (bit uno: color original) [9]. Luego, un algoritmo de detección de bordes [11] encuentra fronteras entre regiones similares, obteniendo la silueta del objeto o persona en movimiento. En algunas ocasiones, cuando existe alto grado de contraste entre una persona y el fondo (como en una toma a la misma altura del individuo) la detección de bordes puede prescindir de la segmentación. Sin embargo, ésta permite obtener bordes con menos posibilidad de error, dado que todo el fondo tiene un mismo valor [9].

Las formas detectadas, en la práctica no siempre se producen por un movimiento real de personas. También se detectarán, por ejemplo, a causa de cambios de iluminación y/o movimiento de otros objetos. Por lo anterior, se requiere seleccionar una silueta que corresponda a una persona y no a otra cosa. Ésta silueta es reemplazada por una forma geométrica que la contenga (región de interés), la que puede tener forma de rectángulo, elipse, etc [41].

- Seguimiento de la persona: seguir la trayectoria realizada por una persona permite inferir datos como su dirección y velocidad antes de abandonar la escena. En particular, es de interés el lugar donde aparece y desaparece una persona, para determinar si es alguien que descendió o abordó a un vehículo, o es un peatón no relacionado a un automóvil.
- Extracción de características: de la región de interés, representa por un espacio de colores o una mezcla de varios espacios [20], se extraen datos (características) que formarán una firma (*signature*). Las características pueden ser de distintos tipos, entre ellos [37]:
 - Color: una imagen está compuesta por píxeles cuyos valores dependen del espacio de colores empleado. Un espacio de colores está compuesto por varios canales, cada uno de ellos representados en una matriz, cuyos valores están asociados a un píxel de la imagen. Por ejemplo, el espacio RGB está conformado por los canales rojo (R), verde (G) y azul (B). Otro espacio comúnmente usado es HSV, compuesto por los canales de tonalidad (H), saturación (S) y valor (V).
 - Forma: se refiere a datos de la silueta detectada (altura, ancho, ejes de simetría, relación entre ancho y altura, etc.).

- Posición: cuando los FOV entre cámaras están superpuestos, la posición de una persona puede ser utilizada para Re-ID, dado que el movimiento capturado en cada cámara está directamente relacionado [10, 26].
 - Textura: entrega la disposición espacial de los colores de una imagen [35]. Se ha representado por puntos seleccionados de la imagen (*keypoints*), cuyas propiedades se mantienen inviatables a cambios de rotación y tamaño [3, 29]. Generalmente, estos puntos se encuentran en la frontera entre regiones con colores muy distintos.
- Generación del descriptores: esta tarea define la forma de representar las características seleccionadas (descriptor). Generalmente se calculan estadísticas sobre las características, para generar una representación sucinta de las características (por ejemplo, histogramas). Los descriptores generados pueden ser todos de tamaño fijo o variable. Un ejemplo del primer caso son los histogramas, cuya cantidad de dimensiones es independiente del tamaño de la imagen y los colores presentes en ella, dependiendo el tamaño sólo del espacio de colores empleado. Un algoritmo típico presentado [29] (Scale-invariant feature transform, SIFT) encuentra keypoints en una imagen, generando descriptores de tamaño variable (cantidad de puntos seleccionados es variable dependiendo de la complejidad de la imagen). Sin embargo, la representación de cada punto es un vector de tamaño fijo.
 - Comparación de descriptores: aquí se establece si un par de descriptores corresponden o no a una misma persona, dependiendo de la similitud (distancia) entre descriptores [1, 12, 17, 18, 20, 28, 33, 38].

Para encontrar una coincidencia (*match*) se han propuesto varios algoritmos. Una forma es calcular la distancia euclidiana entre el descriptor consultado (*query*) con todos los posibles candidatos (vecinos) y luego elegir a los k candidatos que se encuentren a menor distancia, técnica conocida como el k -ésimo vecino más cercano (k Nearest Neighbor, k-NN). Una mejora utiliza una métrica de distancia que considera patrones estadísticos obtenidos de ejemplos en forma de tuplas $(A, B, etiqueta)$, donde A y B son descriptores, y la etiqueta indica si éstos son un match o no [15, 22, 23, 40, 42, 43].

En el caso de la comparación de dos descriptores compuestos por keypoints (ej. SIFT), se calcula la distancia de cada keypoint de un descriptor con todos los keypoints del otro descriptor, estableciendo una correspondencia positiva si existe un par de keypoints a una distancia lo suficientemente pequeña, menor a cierto umbral. Si se cuenta con varias imágenes por persona, un sistema de votación es otra alternativa para elegir el candidato que más se asemeja a la consulta. Se le asigna un voto al descriptor candidato siempre que éste posea el keypoint más similar (a menor distancia) a un keypoint del descriptor consultado, emitiendo tantos votos como keypoints tenga éste, seleccionando luego al descriptor candidato con más votos [21].

Bibliografía

- [1] BAK, S., CORVEE, E., BREMOND, F., AND THONNAT, M. Person re-identification using haar-based and dcd-based signature. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on* (Aug 2010), pp. 1–8.
- [2] BAUML, M., BERNARDIN, K., FISCHER, M., EKENEL, H. K., AND STIEFELHAGEN, R. Multi-pose face recognition for person retrieval in camera networks. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on* (2010), IEEE, pp. 441–447.
- [3] BAY, H., TUYTELAARS, T., AND VAN GOOL, L. Surf: Speeded up robust features. In *Computer vision–ECCV 2006*. Springer, 2006, pp. 404–417.
- [4] BEDAGKAR-GALA, A., AND SHAH, S. K. A survey of approaches and trends in person re-identification. *Image and Vision Computing* 32, 4 (2014), 270–286.
- [5] BELLET, A., HABRARD, A., AND SEBBAN, M. A survey on metric learning for feature vectors and structured data. *arXiv preprint arXiv:1306.6709* (2013).
- [6] BERDUGO, G., SOCEANU, O., MOSHE, Y., RUDROY, D., AND DVIR, I. Object reidentification in real world scenarios across multiple non-overlapping cameras. In *Proc. Euro. Sig. Proc. Conf* (2010), pp. 1806–1810.
- [7] BIRD, N. D., MASOUD, O., PAPANIKOLOPOULOS, N. P., AND ISAACS, A. Detection of loitering individuals in public transportation areas. *Intelligent Transportation Systems, IEEE Transactions on* 6, 2 (2005), 167–177.
- [8] BOUMA, H., BAAN, J., LANDSMEER, S., KRUSZYNSKI, C., VAN ANTWERPEN, G., AND DIJK, J. Real-time tracking and fast retrieval of persons in multiple surveillance cameras of a shopping mall. vol. 8756, pp. 87560A–13.
- [9] BOUWMANS, T., EL BAF, F., AND VACHON, B. Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science* 1, 3 (2008), 219–237.
- [10] CALDERARA, S., PRATI, A., AND CUCCHIARA, R. Hecol: Homography and epipolar-based consistent labeling for outdoor park surveillance. *Computer Vision and Image Understanding* 111, 1 (2008), 21 – 42. Special Issue on Intelligent Visual Surveillance (IEEE).
- [11] CANNY, J. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 6 (1986), 679–698.

- [12] DALAL, N., AND TRIGGS, B. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (2005), vol. 1, IEEE, pp. 886–893.
- [13] DE OLIVEIRA, I. O., AND DE SOUZA PIO, J. L. People reidentification in a camera network. In *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on* (2009), IEEE, pp. 461–466.
- [14] DECANN, B., AND ROSS, A. Can a poor verification system be a good identification system - a preliminary study. In *Information Forensics and Security (WIFS), 2012 IEEE International Workshop on* (Dec 2012), pp. 31–36.
- [15] DIKMEN, M., AKBAS, E., HUANG, T. S., AND AHUJA, N. Pedestrian recognition with a learned metric. In *Computer Vision-ACCV 2010*. Springer, 2011, pp. 501–512.
- [16] ELLIS, T., MAKRIS, D., AND BLACK, J. Learning a multi-camera topology. In *Joint IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)* (2003), pp. 165–171.
- [17] FARENZENA, M., BAZZANI, L., PERINA, A., MURINO, V., AND CRISTANI, M. Person re-identification by symmetry-driven accumulation of local features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (June 2010), pp. 2360–2367.
- [18] GHEISSARI, N., SEBASTIAN, T., AND HARTLEY, R. Person reidentification using spatiotemporal appearance. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (2006), vol. 2, pp. 1528–1535.
- [19] GILBERT, A., AND BOWDEN, R. Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. In *Computer Vision-ECCV 2006*. Springer, 2006, pp. 125–136.
- [20] GRAY, D., AND TAO, H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Computer Vision-ECCV 2008*. Springer, 2008, pp. 262–275.
- [21] HAMDOUN, O., MOUTARDE, F., STANCIULESCU, B., AND STEUX, B. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on* (2008), IEEE, pp. 1–6.
- [22] HIRZER, M., ROTH, P., AND BISCHOF, H. Person re-identification by efficient impostor-based metric learning. In *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on* (Sept 2012), pp. 203–208.
- [23] HIRZER, M., ROTH, P., KÖSTINGER, M., AND BISCHOF, H. Relaxed pairwise learned metric for person re-identification. In *Computer Vision - ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., vol. 7577 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2012, pp. 780–793.

- [24] HU, L., JIANG, S., HUANG, Q., AND GAO, W. People re-detection using adaboost with sift and color correlogram. In *Image Processing, 2008. IICIP 2008. 15th IEEE International Conference on* (Oct 2008), pp. 1348–1351.
- [25] JAVED, O., SHAFIQUE, K., RASHEED, Z., AND SHAH, M. Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views. *Computer Vision and Image Understanding* 109, 2 (2008), 146–162.
- [26] KHAN, S., AND SHAH, M. Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25, 10 (Oct 2003), 1355–1360.
- [27] KUO, C.-H., HUANG, C., AND NEVATIA, R. Inter-camera association of multi-target tracks by on-line learned appearance affinity models. In *Computer Vision–ECCV 2010*. Springer, 2010, pp. 383–396.
- [28] LIN, Z., AND DAVIS, L. Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance. In *Advances in Visual Computing*, vol. 5358 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2008, pp. 23–34.
- [29] LOWE, D. G. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* (1999), vol. 2, Ieee, pp. 1150–1157.
- [30] MAZZON, R., TAHIR, S. F., AND CAVALLARO, A. Person re-identification in crowd. *Pattern Recognition Letters* 33, 14 (2012), 1828–1837.
- [31] PORIKLI, F., AND DIVAKARAN, A. Multi-camera calibration, object tracking and query generation. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on* (2003), vol. 1, IEEE, pp. I–653.
- [32] PROSSER, B., GONG, S., AND XIANG, T. Multi-camera matching using bi-directional cumulative brightness transfer functions. In *Proceedings of the British Machine Vision Conference* (2008), BMVA Press, pp. 64.1–64.10. doi:10.5244/C.22.64.
- [33] PROSSER, B., ZHENG, W.-S., GONG, S., AND XIANG, T. Person re-identification by support vector ranking. In *Proceedings of the British Machine Vision Conference* (2010), BMVA Press, pp. 21.1–21.11. doi:10.5244/C.24.21.
- [34] STEIN, G. Tracking from multiple view points: Self-calibration of space and time. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* (1999), vol. 1, pp. 527.
- [35] STOCKMAN, G., AND SHAPIRO, L. G. *Computer Vision*, 1st ed. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2001.
- [36] TEIXEIRA, L. F., AND CORTE-REAL, L. Video object matching across multiple independent views using local descriptors and adaptive learning. *Pattern Recognition Letters* 30, 2 (2009), 157–167.

- [37] VEZZANI, R., BALTIERI, D., AND CUCCHIARA, R. People reidentification in surveillance and forensics: A survey. *ACM Comput. Surv.* 46, 2 (Dec. 2013), 29:1–29:37.
- [38] WANG, X., DORETTO, G., SEBASTIAN, T., RITTSCHER, J., AND TU, P. Shape and appearance context modeling. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* (Oct 2007), pp. 1–8.
- [39] WEINBERGER, K. Q., AND SAUL, L. K. Fast solvers and efficient implementations for distance metric learning. In *Proceedings of the 25th international conference on Machine learning* (2008), ACM, pp. 1160–1167.
- [40] WEINBERGER, K. Q., AND SAUL, L. K. Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* 10 (June 2009), 207–244.
- [41] YILMAZ, A., JAVED, O., AND SHAH, M. Object tracking: A survey. *ACM Comput. Surv.* 38, 4 (Dec. 2006).
- [42] ZHENG, W.-S., GONG, S., AND XIANG, T. Person re-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (2011), IEEE, pp. 649–656.
- [43] ZHENG, W.-S., GONG, S., AND XIANG, T. Reidentification by relative distance comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35, 3 (2013), 653–668.