Multiple View Oriented Matching Algorithm for People Reidentification

Jorge García, Alfredo Gardel, Ignacio Bravo, *Member, IEEE,* and José Luis Lázaro

***Abstract*—People reidentification is one of the most challenging tasks in computer vision, and considerable efforts have been directed toward providing solutions to this problem. The existence of extensive camera networks and surveillance systems increases the amount of people images obtained, but, on the other hand, implies the need for new algorithms to enable reidentification of people captured by the cameras. There is no one optimal model that solves the entire problem, but a set of distinctive features can be used to help in the matching process. Our proposal consists of using the orientation of each person captured in the surveillance scene to considerably improve the reidentification process. An iterative algorithm maximizes the number of successful matches and speeds up the process. A comparison with other earlier relevant studies is presented using available datasets.**

***Index Terms*—Appearance models, camera network, object recognition, people reidentification, surveillance systems.**

I. Introduction

SCENE UNDERSTANDING is commonly conceived of as a task oriented to the interpretation of a scene through video sequences [1], [2]. Different tasks must be carried out to provide a knowledge-based process, such as object recognition, motion processing, and color appearance models. Knowledge of people's identities enables a system to fully understand the scene [3], [4]. Surveillance systems are one of the most suitable applications for performing these tasks when a camera network is monitoring different, nonoverlapping areas. Typically, this type of surveillance network is composed of different cameras and involves lack of information about the space–time relationship while tracking people in the surveillance area [5]. Therefore, it is necessary to unify and share the information on detected and tracked people between different cameras in order to provide a better scene understanding.

People reidentification is the visual recognition of the same person in disjointed camera views, considering a certain set of different identities. To address this task, a feature set is extracted from each detected person on a captured image, a process known as identification. Examples of these features include color distribution, shape, texture, local attributes, etc. All the features used to model appearance are classified into two groups: global and local. The difference between them is the region of interest (ROI) from where the feature is extracted; local features focus on information at points of interest, while global features are present in a large area of a person. Typically, the set of features that defines a person is considered a signature, and the reidentification process consists of comparing different views of a signature using a similarity measure. When using nonoverlapping camera views, several problems are added to the reidentification process, since different perspectives are taken when a person is captured from disjoint views. For example, four poses are defined to model the appearance: front, back, left and right side. This aspect must be taken into account when using different views in order to create a signature, since people's appearance may be affected by lighting changes between different locations, i.e., outdoor/indoor. However, merging global and local features helps to create more robust and reliable signatures, and temporal and spatial constraints between cameras can be used to reduce potential false matches. In addition, the camera network topology must be known since each camera captures an uncontrolled environment from a certain distance, which means that recognition of biometric aspects such as face, eyes, or gait [6] does not provide sufficient reliability due to difficult segmentation, low resolution, and frame rate. Another common problem is occlusion due to movement of the people or objects within a scene. Fig. 1 shows some examples of the problems described above.

Most studies [8]–[10] have generated an appearance model-based signature from single or multiple images corresponding to the same person. Simple appearance models consist exclusively of global color features and are obtained from the general chromatic content or filter responses. However, in order to increase signature robustness, several local features can be added, such as points of interest, relevant patches, and texture segmentation. Other methods [11], [12] extract features from multiple images and use machine-learning algorithms to obtain a signature that considers perspective changes. Lastly, some studies have provided a robust distance, in an attempt to quantify and differentiate features by learning the distance-weighted function that is most likely to yield correct matches in the reidentification process [13]. As mentioned earlier, the main problem encountered in these reidentification methods is the variation in appearance that occurs when a person is captured from different perspectives. Given a pair of signatures corresponding to noticeably different viewpoints, the match between local features decreases and it is mainly the global features that provide similarity in the matching process. To mitigate this problem, here we propose a multiple view oriented model (MVOM),

Manuscript received January 19, 2014; revised April 25, 2014; accepted May 27, 2014. Date of publication June 13, 2014; date of current version August 05, 2014. This work was supported by University of Alcala through the Identificación de Personas a partir de la Reconstrucción de Imágenes Múltiples (IPRIM) project under Ref. CCG2013/EXP-064. Paper no. TII-14-0073.

The authors are with the Department of Electronics, Escuela Politécnica Superior, University of Alcala, 28871 Alcalá de Henares (Madrid), Spain (e-mail: jorge.garcia@depeca.uah.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2014.2330976

Fig. 1. Examples of VIPeR dataset [7]. Each column is one pair of images corresponding to the same person.

which represents a signature composed of different images. Each of the model's images represents a different or an updated viewpoint. Furthermore, an iterative matching process is proposed to take advantage of the various perspectives available. In general, since some trajectories are more likely than others in each camera, of the orientation of short-term tracking (STT) for each camera will have a similar orientation. The important issue here is that extension of the camera network will produce relevant information with many differently oriented trajectories, thus ensuring that the algorithm presented will yield an improvement.

This paper is organized as follows. In Section II, we review previous studies related to the subject under discussion. In Section III, we present the proposed MVOM, a method for retrieving people orientation, and the iterative matching algorithm. In Section IV, we present our experimental results obtained using three public datasets to validate our proposal. Lastly, we present our conclusion in Section V.

II. Related Work

People reidentification is currently one of the most important topics in surveillance systems, and provides a means to understand a global scene using people trajectories from images captured by a camera network. Appearance models are the methods most commonly used in order to create a signature with which to distinctively identify each person. Each signature consists of a set of features which can be classified into two groups: global and local features. Typically, global features are composed of chromatic histograms of different color spaces. In [14], the RGB color space was selected to extract a color position histogram consisting of a fixed number of horizontal bands. Thus, a person is represented by a vector with a particular color distribution. This representation is somewhat dependent on RGB values, so lighting changes between different scenes can significantly reduce the rate of reidentification. In a related study [15], a similar signature with horizontal bands was implemented but using a binary classifier based on a support vector machine (SVM). A similarity measure function based on a learning process improved both performance and accuracy in people reidentification under challenging viewing conditions [16]. Texture feature extraction is another option for characterizing the people appearance model. The major contribution of Gray and Tao [7] was the application of *Gabor* and *Schmid* filter responses. These features are more stable than color features, so the signature is more independent of viewpoint changes.

Other types of signature are complemented by local features. For example, Oliveira and Souza-Pio [17] used local histograms of HSV color space, which were determined in areas around specific points of interest. Furthermore, each ROI was used by a Speeded Up Robust Features (SURF) descriptor based on a Scale-Invariant Feature Transform (SIFT) descriptor but with improved performance. The aim of this feature was to characterize a region in a robust way that was invariant to natural viewing changes such as scale, rotation, and affine/viewpoint variance. However, when two cameras capture images from different sides of a person, the robustness of this feature is reduced. Other local features such as Haar-like features have also been extracted to define a signature, as in [18]. In addition, gradient location and orientation histogram (GLOH) features have been used, combining ideas from both SIFT and shape context [19]. The local features defined above are descriptors applied over a local interest point such as a Harris detector and Hessian-Laplace. Covariance descriptors have also been used to create a signature, as proposed in [20]. Using the dense descriptors technique, a grid structure with overlapping was applied to the image, and a cell was defined in each point of the grid where the covariance descriptor was calculated; thus, the signature was composed of a large vector of covariance descriptors. The same authors have also proposed a discrimination method to extract the relevance vector prior to implementing the grid matching process. Similar studies which have also used covariance descriptors are reported in [21] and [22]. In these cases, the covariance descriptors grid was combined with biologically inspired features and spectral clustering techniques, respectively.

All global and local features are extracted in a ROI of the image where the person appears (full body). Most of the methods assume the minimum bounding box and that some background pixels within the ROI include background information on the signature. Research such as [14] has determined the silhouette using the background/foreground update process in order to only include information about the person. Some authors have also proposed body segmentation; in [9], the body was divided into three parts: 1) legs; 2) trunk; and 3) head. Each extracted feature was weighted according to the body part and the distance between the middle of the body (vertical orientation) and the feature location. In [23] and [24], a reidentification process was proposed based on attributes, where a matching process between segmented parts was used. Another possible classification of appearance models is based on the use of a single shot [7] or multiple shots [8], [20], [25] in order to create the signature. Single shot methods only use one image where the person appears in order to extract features, whereas a set of images is required to apply multiple shot methods. The choice of a single or multiple shot method depends on the availability/use of tracking information. Multiple shot methods yield a more independent signature than single shot methods as regards the captured viewpoint or lighting changes. In contrast, fusion techniques

are necessary to combine information from multiple images. In [25], a covariance descriptor was applied to generate an over-complete descriptive statistical model. Thus, a discrimination model was applied to select the most selective features. A similar technique was proposed in [20], where the authors used a variance measure to distinguish between discriminant patterns and common patterns, assuming that common patterns belonged to the background and were not thus taken into account to construct the signature.

In [26], multiple images were used to create each individual signature. The authors introduced a novel domain for matching the reidentification of a person and panoramic appearance mapping (PAM) for feature representation. This large area enables the introduction of information captured by multiple cameras—whether overlapped or not—using the relative position of the person with respect to the camera. Thus, considering the orientation of people from the camera, the images are placed in the corresponding zone of the PAM, which in turn provides an easy and more robust mechanism for matching images for reidentification. The authors only mapped RGB points onto the PAM, without computing the features for later comparison as we do in our proposal. In the study by Baltieri *et al.* [27], a common three-dimensional (3-D) model was created for any person under analysis. The authors focused their attention on calibrated camera spaces, giving an accurate measurement of each 3-D feature, superimposed on the 3-D model of each individual. The authors overwrote the model information depending on the reliability of the current measurement, thus only one measurement was incorporated into the 3-D model at a time. Another issue was that comparison in the 3-D domain could lead to larger errors, mainly due to the image–model transformation, which is not required in our proposal. Novel depth cameras introduce more data into the reidentification problem. In [28], RGB-D information was used to set-up a 3-D descriptor for people reidentification. These authors based reidentification on a 3-D cylindrical grid that included the RGB information retrieved by the depth camera. However, they did not compute the orientation of the people with respect to the depth camera; instead, the data were incorporated into the 3-D grid and matched with other 3-D grids regardless of the orientation of the people captured. In our proposal, the orientation of people trajectories is considered an important parameter, which will be evaluated iteratively in the people reidentification process. This particular issue is analyzed in [29]. These authors studied the viewpoint invariance of person reidentification, and found that in many cases, only a color histogram could be considered a good matching feature for a view-independent description of a person's appearance. However, other features extracted from an individual were generally only view-independent to a certain degree. They introduced a set of oriented captures of each person. Our criticism is that apparently all the oriented views of a person were captured by the same camera, and only a quantized orientation range of person images was stored. In our proposal, all image information is stored if there is a reliable orientation value in the person-camera view. Another important idea from the paper cited above is that in general, symmetry exists between the left and right sides of a person, thus speeding up the reidentification process. We take this notion one step further, providing an iterative matching step that would reduce the computing time required for the process depending on the information previously obtained by the system.

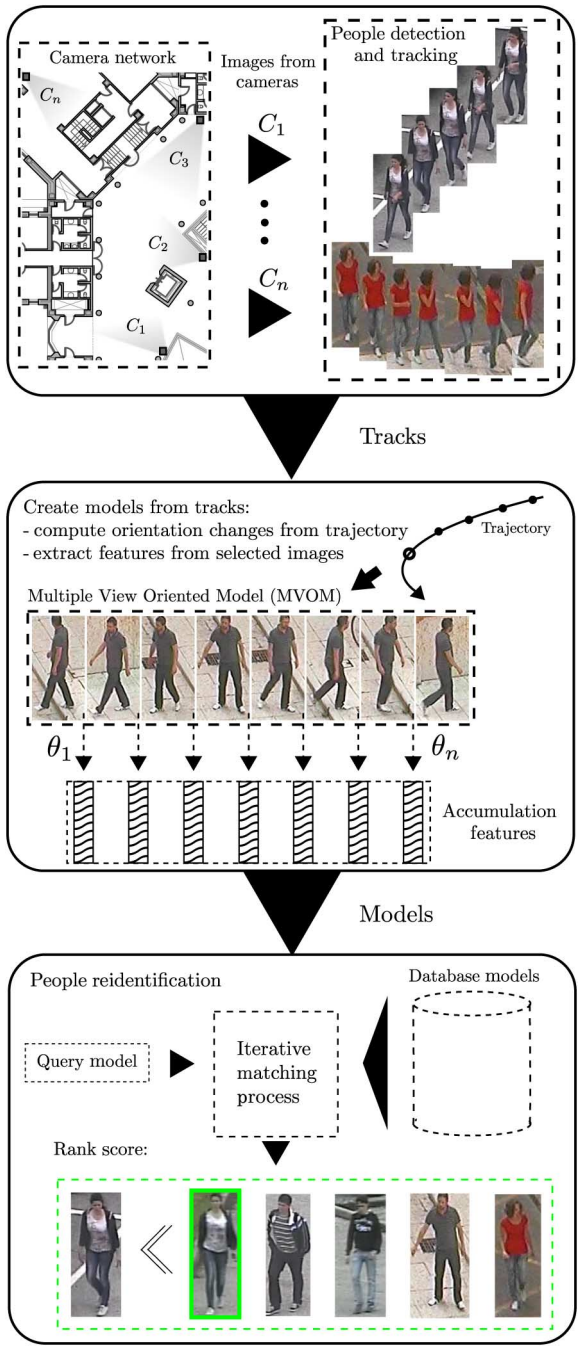

Fig. 2. System overview: steps corresponding to Multiple View Oriented Matching Algorithm for people reidentification.

III. Multiple View Oriented Matching Approach

Given a camera network with nonoverlapping fields of view located in a large surveillance area where an unknown set of people is captured, the reidentification problem can be defined as the correspondence between people across different camera images. We propose a MVOM in order to encode the appearance of the person from different perspectives with respect to the camera. An overview of our approach is shown in Fig. 2. The first

step is to compute an orientation value for each image corresponding to an STT. Then, different perspectives of the person are extracted from each STT according to his/her orientation. A set of features is used to represent each perspective. Finally, an iterative matching process is proposed in order to identify correspondences between different models taking advantage of the perspective values of the model in order to achieve a robust matching.

A. Multiple View Oriented Model

Since a camera provides an STT of the people crossing the scene [30], [31], multiple images may be used to model their appearance. These methods, which are referred to as multiple shots, merge the information extracted from all the images to create a signature, which is more suitable for perspective or people orientation changes. Often, problems can arise in such proposals when different perspectives corresponding to the same person have strong dissimilarities. A corrupt signature is computed, causing unsatisfactory correspondences in the reidentification process. We propose a MVOM that creates a signature composed of different feature vectors where each one provides an updated appearance model of the person with a defined perspective from the camera. In the same way, views with a similar perspective are captured by the MVOM so as to always have updated appearances of the people. Given an STT, two cases can arise when adding an appearance to the model, as explained below.

1) Direction Changes: The trajectory generated by a person across the scene captured by a camera may contain direction changes due to static objects which are located in the scene and cross between people or his/her own trajectory. These situations are leveraged by our MVOM to obtain different perspectives of the person and they are classified using an orientation parameter according to camera location. A new appearance sample is added to the model when a strong change is detected in a short period of time or when a weak change occurs over a long period of time.

2) Updated Perspective: Similar appearances that are complementary are incorporated into the model at every given period of time. Thus, several images with the same orientation are collected to obtain an appearance model that is less dependent on changes in perspective or camera conditions. This type of data acquisition for the appearance model leads to a larger database of possible similar images that can subsequently be refined to reduce the amount of data to be stored in the database.

Formally, the MVOM is defined as follows. Let $\mathcal{S} = \{\mathbf{s}_i\}$ be an STT captured from a camera, while $\mathbf{s}$ represents a feature vector modeling the person appearance and $i \in \mathbb{N}$ is the number of samples contained in $\mathcal{S}$. The number of samples $\mathbf{s}_i$ depends on the time that the person is within the camera field of view. In this study, we assume that the task of people tracking is already solved as proposed in [32]. Thus, the MVOM can be represented as $\mathcal{S}' \subseteq \mathcal{S}$, where $\mathcal{S}'$ is a subset of feature vectors defining different perspectives of the trajectory. The size $\mathcal{S}'$ depends on the resolution of direction changes and the update time. Given a trajectory of a person $\mathcal{T} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{v}_x, \boldsymbol{v}_y)$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are image position vectors and $\boldsymbol{v}_x$ and $\boldsymbol{v}_y$ are image velocity vectors, direction changes can be represented as the angle between the position vector and the velocity vector. However, this parameter does not provide a perspective-camera relationship in order to compare perspectives from other STTs occurring either within the same camera (intrarelationship) or with respect to other cameras (interrelationship). Assuming people walk in a forward direction, the angle between the trajectory vector and camera vector (optical axis) provides a relationship in order to compare perspectives that satisfy both restrictions. This angle θ is defined as the estimated orientation of the person with respect to the camera. An analysis of this constraint is described in the next section.

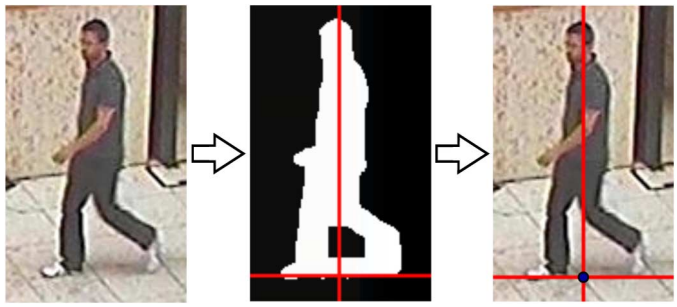

Fig. 3. Example of floor point detection through the intersection of the person vertical line with the ground plane.

From now on, we assume the vector $\boldsymbol{o}$ is known, where each element represents the estimated orientation value between two successive points of the trajectory. When the angular velocity is too large for two consecutive points, the uncertainty of θ extracted from the images is also large. The same problem can occur when the linear velocity is low for two consecutive points (stop&go situations). Similarly, projection of the scene introduces uncertainty into the estimated orientation, but it is encoded in the linear velocity. Thus, the reliability of the θ value can be defined as a function of linear and angular velocities from the person trajectory, v and ω, respectively. Given two points $(x_1 \;\; y_1 \;\; v_{x1} \;\; v_{y1})$ and $(x_2 \;\; y_2 \;\; v_{x2} \;\; v_{y2})$ from $\mathcal{T}$, v is defined as $v = \left((v_{x2} - v_{x1})^2 + (v_{y2} - v_{y1})^2\right)^{1/2}$ and $\omega = (\theta_2 - \theta_1)/T_s$, where T_s represents the camera frame rate. We propose to model the reliability ξ as a weighted function of normal distributions

$$\xi = \alpha_v \mathcal{N}_v(v_m, \sigma_v^2) + \alpha_\omega \mathcal{N}_\omega(0, \sigma_\omega^2) \tag{1}$$

where $\mathcal{N}_v$ and $\mathcal{N}_\omega$ represent normal distributions for each velocity with σ_v^2 and σ_ω^2 variances. v_m represents the average image velocity in order to obtain a low weight when the person is not walking. Lastly, α_v and α_ω are their corresponding weights.

Typically, single camera tracking provides the person position in the image as the centroid of its bounding box, head detection point, etc. We propose to use a new point defined as floor point in order to increase the precision of the orientation recognition task and the parameter values for its reliability function. This point represents the intersection of the person vertical line with the ground plane, as shown in Fig. 3. Given the foreground image m, the floor point (x, y) is computed as follows:

$$x = \min_{p \in \mathbb{N}} \left\{ \left| \sum_{i=1;j}^{p} m(i,j) - \sum_{i=p;j}^{\text{size}(m)} m(i,j) \right| \right\} \tag{2}$$

$$y = \min_{p \in \mathbb{N}} \left\{ \text{d}((x,p),(x_m,y_m))_{|_{m(x,p)=1}} \right\} \tag{3}$$

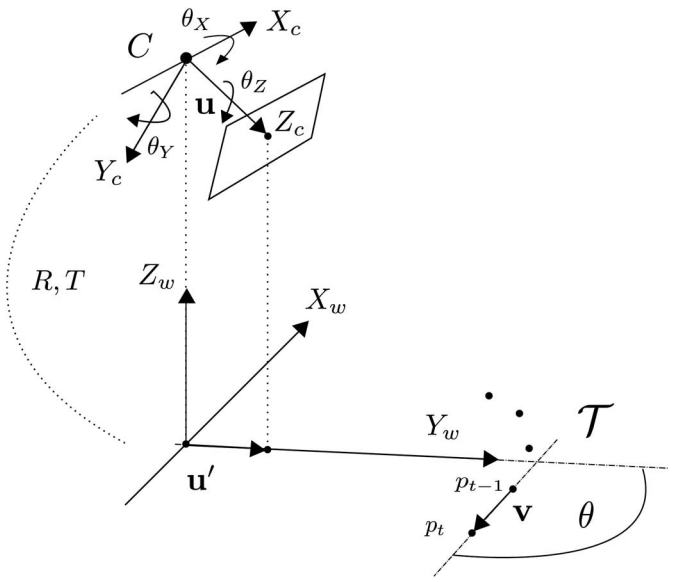

Fig. 4. System diagram related to retrieving people orientation.

where $\mathrm{d}(\cdot)$ represents the Euclidean distance and (x_m, y_m) is the middle-bottom point of the image. Lastly, the MVOM is expressed as $\mathcal{M} \equiv \mathcal{S}' = \{\mathbf{s}_i, \theta_i, \xi_i\}$. A threshold parameter ρ is defined to determine which samples are incorporated to the MVOM from $|\Delta\theta| = |\theta_b - \theta_a|$, where θ_b is the orientation value of the last sample added to the model and θ_a is the orientation value of the current sample. ρ parameter is fixed according to the real values obtained from each camera.

B. Retrieving People Orientation

Our proposal retrieves the trajectory orientation of people in the scene with respect to the camera. In this section, we show that only two calibration parameters will be necessary to retrieve the estimated orientation of people trajectories. Fig. 4 shows the global diagram used to obtain the orientation value θ for a sample of the person trajectory with respect to the camera. Vector $\boldsymbol{u}$ is the camera optical axis, $\boldsymbol{u}'$ its corresponding projection over the ground plane and $\boldsymbol{v}$ is the trajectory vector between two consecutive points. Let us make the following assumptions: first of all, the origin of the world coordinates system is located in the ground plane where the vertical line intersecting the camera coordinate system originates. Thus, the translation vector $\boldsymbol{T} = [\boldsymbol{T}_x \boldsymbol{T}_y \boldsymbol{T}_z]^\top$ will have $\boldsymbol{T}_x = 0$ and $\boldsymbol{T}_y = 0$. As noted previously, the projection of the camera optical axis $\boldsymbol{u}'$ is parallel to the Y_w axis. Thus, the X_w axis is parallel to the axis X_c which in turn ensures that in the RT transformation, $\theta_y = 0$ always. Given a point of the trajectory in world coordinates $\mathbf{P}_w = (X_w, Y_w, Z_w)$, the perspective transformation equation $\mathbf{P}_c = [RT][\mathbf{P}_w 1]^\top$ where $\mathbf{P}_c = (X_c, Y_c, Z_c)$ represents the projected point in the camera. Given that the ground point of a person trajectory has the coordinate $Z_w = 0$, the expression can be shortened

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \Delta \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \tag{4}$$

where Δ is a reduced matrix

$$\begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \cos\theta_x \sin\theta_z & \cos\theta_x \cos\theta_z & 0 \\ \sin\theta_x \sin\theta_z & \cos\theta_z \sin\theta_x & T_z \end{bmatrix}. \tag{5}$$

In our proposal, it is of interest to use the inverse transformation, from the camera points to the world coordinate points, in order to obtain the orientation θ. The reduced inverse transformation to obtain $[X_w Y_w]^\top$ is

$$\begin{bmatrix} X_w \\ Y_w \end{bmatrix} = \begin{bmatrix} \cos\theta_z & \dfrac{\sin\theta_z}{\cos\theta_x} \\ -\sin\theta_z & \dfrac{\cos\theta_z}{\cos\theta_x} \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \end{bmatrix}. \tag{6}$$

The orientation vector $\boldsymbol{o} = (o_x o_y)$ is given by the subtraction of two consecutive trajectory points p_1 and p_2, where $o_x = X_{w2} - X_{w1}$ and $o_y = Y_{w2} - Y_{w1}$. Lastly, the orientation θ for the person trajectory in this particular scenario is given by $\theta = \arctan(o_y/o_x)$. The new vector $\boldsymbol{o}$ is given by

$$o_y = \frac{s}{f_y}\left(-\sin\theta_z(u_2 - u_1) + \frac{\cos\theta_z}{\cos\theta_x}(v_2 - v_1)\right) \tag{7}$$

$$o_x = \frac{s}{f_x}\left(\cos\theta_z(u_2 - u_1) + \frac{\sin\theta_z}{\cos\theta_x}(v_2 - v_1)\right). \tag{8}$$

Using the standard assumption of zero skew and unit aspect ratio in the intrinsic camera parameters, we find that $f = f_y = f_x$ and $s = 1$. Thus, the person orientation value θ obtained from the captured camera points of the people trajectory in the floor plane is

$$\theta = \arctan\left(\frac{-\sin\theta_z(u_2 - u_1) + \dfrac{\cos\theta_z}{\cos\theta_x}(v_2 - v_1)}{\cos\theta_z(u_2 - u_1) + \dfrac{\sin\theta_z}{\cos\theta_x}(v_2 - v_1)}\right) \tag{9}$$

which shows that nonintrinsic camera parameters are necessary and it only depends on two extrinsic rotation parameters θ_x and θ_z easily retrieved from the structure of the scene captured by each camera.

C. Iterative Matching Process

Our proposal assumes that a large camera network will produce different STTs of people across each camera. As stated in Section II, although global features are independent of the person–camera orientation, these kinds of feature will produce false matches because their capacity to distinguish a large number of people is relatively reduced. That is to say, global features do provide a matching function, but this function does not provide any reliability as regards the correspondence match. Thus, when large numbers of people are introduced for reidentification, local dependent orientation features should be considered: such was our starting point for an iterative orientation aggregation algorithm considering the person–camera perspective.

Formally, let $\mathcal{C} = \{C_n\}$ be a nonoverlapping camera network where $n \in \mathbb{N}$ and assume that there are m people in the area covered by the camera network (m is not assumed to be known).

Each camera C_n provides a set of STT, where each one is captured at different time instants. Let $\mathcal{O} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{m_j}\}$ be the set of all MVOMs from all STTs captured by $\mathcal{C}$, where m_j is the total number of MVOMs. Moreover, as expressed above, each $\mathcal{M}_j$ is generated by a person from an STT, and consists of $\{\mathbf{s}_i, \theta_i, \xi_i\}$ where $i \in \mathbb{N}$, $\mathbf{s}$ is a feature vector which encodes the appearance, θ is the retrieved orientation value, and ξ is the reliability value. The subscript i depends on the number of detected direction changes and the number of updates. The people reidentification problem can be defined as follows. Let us suppose a correspondence between two MVOMs $\mathcal{M}_a$ and $\mathcal{M}_b$ written as h_a^b. There is a binary variable $\gamma_{h_a^b}$ which is true if h_a^b is a correct hypothesis, i.e., $\mathcal{M}_a$ and $\mathcal{M}_b$ correspond to the same person. We need to find a set of correspondences between MVOMs $H = \{h_a^b, \ldots\}$ such that $h_a^b \in H$ if $\gamma_{h_a^b}$ is true. The function to set the $\gamma_{h_a^b}$ variable is defined in (10), where ϕ_{match} is a decision threshold fixed according to some external restriction, balancing the number of reidentification results and the confidence of the measurement. Given two feature vectors $\boldsymbol{s}_a$ and $\boldsymbol{s}_b$, the matching distance expressed as $\hat{d}_{a,n}^{b,m}$ is a normalized function which measures the similarity between two appearances. This will be referred to again in the next section. The iterative matching process is shown in Algorithm 1

$$\gamma_{h_a^b} = \begin{cases} \text{true} & \hat{d}_{a,n}^{b,m}(\mathbf{s}_n^a, \xi_n^a, \mathbf{s}_m^b, \xi_m^b) \leq \phi_{\text{match}} \\ \text{false} & \text{otherwise.} \end{cases} \tag{10}$$

Algorithm 1: Iterative Matching Algorithm for a query multiple view oriented model (MVOM).

Data: $\mathcal{M}_a$, $\mathcal{O} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{m_j}\}$, $\phi_{\theta,\text{ini}}$, ϕ_{match}

Result: Ranking $\mathcal{R}$

Initialization: $\phi_\theta := \phi_{\theta,\text{ini}}$;

while $\phi_\theta \leq 180$ **do**

  Flag: = true;

  while: *Flag* **do**

    for $i \leftarrow 1$ **to** m_j **do**

      Find pairs $\{\mathbf{s}_n^a, \mathbf{s}_m^i\}$ where $|\theta_n^a - \theta_m^i| \leq \phi_\theta$

    end

    Compute matching distances $\mathcal{D} = \{d_{a,n}^{i,m}, \ldots\}$

    Sort matching distance results from lowest to highest distances

    for $i \leftarrow 1$ **to** size of $\mathcal{D}$ **do**

      if $\gamma_{h_a^i}$ **then**

        if $\mathcal{M}_i \notin \mathcal{R}$ **then**

          Add $\mathcal{M}_i$ to $\mathcal{R}$

          $\mathcal{M}_a := \mathcal{M}_a \cup \mathcal{M}_i$

          break;

        end

      else

        Flag: = false;

        break

      end

    end

  end

  $\phi_\theta := \phi_\theta + \phi_{\theta,\text{ini}}$;

end

Refill $\mathcal{R}$ with $\mathcal{M}_i \notin \mathcal{R}$ with the last set of matching distances

The estimated orientation makes it possible to compare and match an STT of a person obtained from one camera with another STT. The threshold parameter ρ, which is defined to select which samples are incorporated into the MVOM, provides the starting range of orientation $\phi_{\theta,\text{ini}}$ that is considered in the first iteration. Considering the range of orientation values given by the division of 180° by $\phi_{\theta,\text{ini}}$, a matching distance considering local features is executed for each range. The ranking of the matching for this MVOM is obtained, only accepting reliable values that exceed a confidence threshold ϕ_{match}. The matching distance provides an index of reliability of the aggregation performed. Once the aggregation of the current ϕ_θ value has been performed, the process continues increasing the angular value to be more flexible and consider other MVOMs with an increased distance in their orientation values. Another important aspect is the aggregation of MVOMs. In our context, the local features are stored when the distance in orientation values given for each MVOM differs in at least the uncertainty of the orientation value computation. Thus, an entire map of local features is stored in the person–camera orientation values. The algorithm ends when the orientation range ϕ_θ reaches 180 degrees, i.e., the full perspective of the person. Similar to the many-to-many approach, when the range of orientation values has finalized, the ϕ_{match} considered while computing the matching function is enlarged to consider more and more distance in the local features, which in turn will produce a less reliable matching. In this way, all MVOMs that have not been included by the iterative matching algorithm are added according to the minimum distance in order to complete the ranking.

IV. Experimental Results

In this section, we describe the experiments we conducted to evaluate the proposed MVOM. The results are presented through two common curves used in people reidentification: 1) the cumulative matching characteristic (CMC) curve; and 2) the synthetic recognition rate (SRR) curve. The CMC curve provides the recognition percentage which represents the expectation of finding the correct label within the best matched labels. Meanwhile, the SRR curve provides the probability that any label among the best matched labels is correct. Furthermore, we

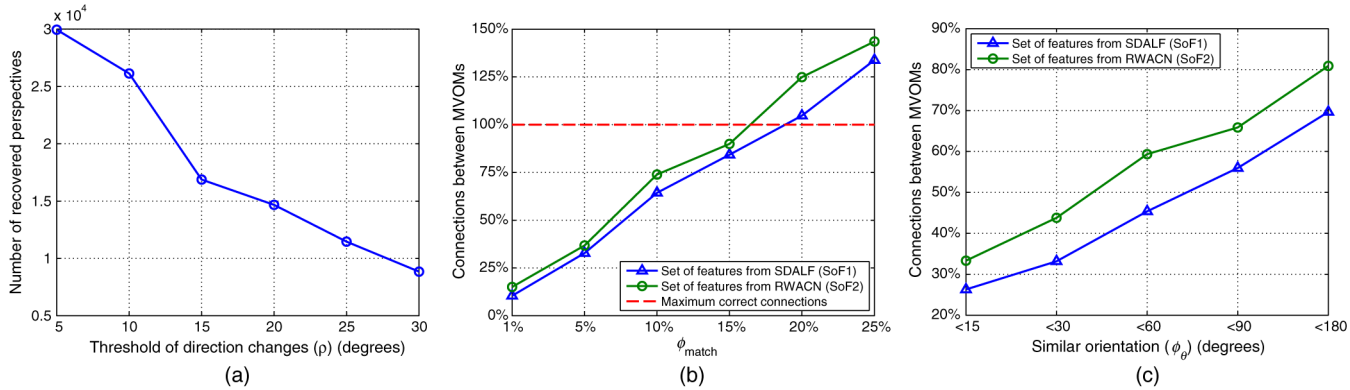

Fig. 5. Behavioral analysis of the proposed algorithm. (a) The number of retrieved perspectives varying the threshold of direction changes. (b) and (c) The percentage of connections between MVOMs for matching distance threshold and the range of considered similar distances.

conducted a behavioral analysis of the proposed algorithm. Different datasets were tested in order to achieve a reliable comparison to validate the people reidentification approach. One constraint of our proposed approach is the need for an STT associated to each person captured by a camera. Thus, datasets providing only one snapshot of the person in each camera, as VIPeR dataset proposed in [7], cannot be used to collect results for our proposal. However, the people tracking information is highly common in real situations. In this study, the 3DPeS, SAIVT, and ETHZ datasets were used to evaluate our proposed method. These datasets present several differences such as indoor/outdoor-uncontrolled environment, number of collected people, and topology of the camera network.

A. Implementation Details

Different features are accumulated in a vector in order to encode the visual appearance of a person. A feature vector $\boldsymbol{s}$ is constructed from each perspective that contains an MVOM. To compare the effectiveness of the proposed model with respect to other contributions, we implemented the two sets of features proposed by Farenzena *et al.* in [9] and Martinel and Micheloni [10]. The first feature set, hereafter referred to as SoF1, consists of weighted color histograms, maximally stable color regions (MSCR) and recurrent high-structured patches. All features were weighted with respect to the vertical axis and classified in two principal regions (legs and trunk). The area corresponding to the head was not considered to extract features because it is formed by very few pixels; therefore, there is a limited information about the person. The second feature set, hereafter referred to as SoF2, leverages three robust features: 1) SIFT features; 2) a pyramid of histograms of orientation gradients (PHOG); and 3) Haralick texture features. The same body part divisions used in [9] were implemented. SIFT features were used to extract the chromatic appearance from HSV color space at different points of interest. The PHOG feature was calculated on three levels and accumulated into a single oriented histogram. Haralick texture features were determined in the regions where most information was concentrated (leg and trunk regions). A complete explanation of all the features and their implementation, as used for all the experiments presented in this section, can be found in [9] and [10].

Fig. 6. Samples of image pairs from the 3DPeS dataset.

To set the normalized matching distance $\hat{d}^{b,m}_{a,n}$ for a perspective pair, we combined the matching distance between feature vectors with reliability values for the perspectives as $\hat{d}^{b,m}_{a,n} = \bar{\xi} d^m_n$, where $\bar{\xi}$ is the mean of the two reliability values and d^m_n is the matching distance between two feature vectors. The set of distances was weighted to obtain a reliable combination. In order to determine the reliability of a perspective, reliability weights were fixed as follows: $\alpha_v = 0.4$ and $\alpha_\omega = 0.6$. These values were estimated using 10 trajectories from each dataset and were left unchanged for all the experiments. Each d^m_n was a weighted combination of three feature distances. Feature weights were set to $\alpha_{\mathrm{WH}} = 0.4$, $\alpha_{\mathrm{MSCR}} = 0.4$, and $\alpha_{\mathrm{RHSP}} = 0.2$ for the SoF1 feature set and $\alpha_{\mathrm{WGCH}} = 0.2$, $\alpha_{\mathrm{PHOG}} = 0.4$, and $\alpha_{\mathrm{HAR}} = 0.4$ for the SoF2 feature set. These weights values were obtained from experiments proposed in [9] and [10], respectively. Fig. 5(a) shows the number of recovered perspectives when the threshold of direction changes ρ was modified. Low threshold values provided a large number of retrieved perspectives from an STT, but similar feature vectors were obtained using the normalized matching distance. The value was fixed as follows: given a set of MVOMs from different cameras, the matching distance was computed for all combinations of images. The threshold ρ was set around the mean of orientation distances from all image pairs where the matching distance was similar to ϕ_{match}. Threshold parameter ρ was set to 10° for all experiments. This value was designated as the initial value in the iterative matching algorithm, i.e., $\phi_{\theta,\mathrm{ini}}$ starts with a value equal to ρ and increased ρ with each iteration. Fig. 5(b) shows the percentage of connections between MVOMs when the decision threshold ϕ_{match} obtained different values for $\hat{d}^{b,m}_{a,n}$. It can be seen that the number of connections exceeded 100% for several ϕ_{match} values. In these cases, some connections

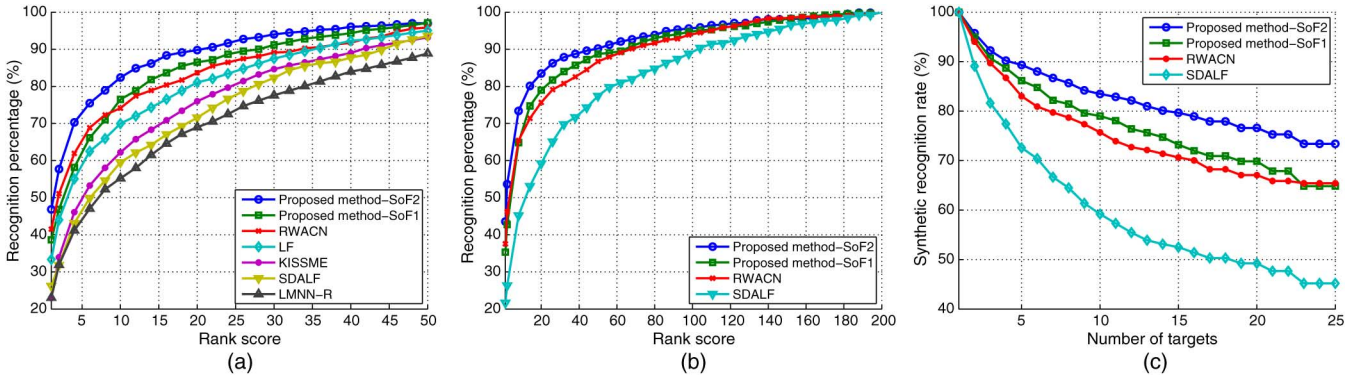

Fig. 7. Results according to 3DPeS dataset. (a) and (b) CMC curves for two different setups. SRR curves are shown in (c) corresponding to the CMC curves of (b).

were erroneous and the MVOMs corresponding to different people were connected, producing a corrupt MVOM. Since $\hat{d}_{a,n}^{b,m}$ was normalized between 0 and 1, the parameter ϕ_{match} was set to 0.12. This value provided a large number of connections and was fairly near to the number of maximum correct connections. Given that the decision threshold must be restrictive, minor variations in this parameter did not modify the overall results. Fig. 5(c) shows the percentage of connections between MVOMs when the parameter ϕ_θ increased in the iterative matching algorithm with $\phi_{\text{match}} = 0.12$. In our experiments, we ran 10 independent trials and the results given below were computed as the average of all of them.

B. 3DPeS Dataset

The 3DPeS dataset was proposed by Baltieri *et al.* in [33]. It contains different sequences of 200 people taken from a multicamera distributed surveillance system. Eight static cameras were used in an outdoor scenario, each one with different lighting conditions and calibrated parameters. People were detected multiple times with different viewpoints. The lighting conditions between cameras did not change too much, but people were captured multiple times over the course of several days, resulting in strong variations in lighting conditions in some cases. This results in a challenging dataset to evaluate people reidentification algorithms. Fig. 6 shows some examples of image pairs corresponding to the same people.

We compared the results of the proposed method with those reported of RWACN [10], LF [34], KISSME [35], SDALF [9], and LMNN-R [36]. The same setup was used, where only 95 people were randomly chosen to compute the CMC curve. Fig. 7(a) shows the performance of our method compared to the previous methods. Using SoF2, the proposed method outperformed the others, especially for low rank scores. Table I presents the top ranked matching rates, and it can be seen that our method achieved the highest recognition percentages. Fig. 7(b) and (c) shows the CMC and SRR curves for a setup where all people were used to obtain the results. Proposed method-SoF2 achieved 43% correct recognition for rank 1, whereas RWACN only obtained 37% using the same feature set. Similar results were obtained for proposed method-SoF1 and SDALF, where the correct recognition values were 35.5% and 21%, respectively. In both comparisons, a notable difference was maintained in the top positions because the MVOMs contain orientation values, which provide better matching. Higher ranking positions were more similar for both methods due to the fact that the current dataset lacked orientation information.

TABLE I
TOP RANKED MATCHING RATE (%) ON 3DPES DATASET

Methods	Rank:1	Rank:5	Rank:10	Rank:20	Rank:50
Proposed-SoF2	**46.9**	**73.3**	**82.4**	**89.8**	97.0
Proposed-SoF1	38.7	63.3	76.5	86.5	**97.2**
RWACN [10]	41.5	65.7	74.13	83.7	95.9
LF [34]	33.3	58.2	70.0	81.1	95.1
KISSME [35]	22.9	49.0	62.2	76.0	93.2
SDALF [9]	26.2	46.1	59.5	71.6	93.6
LMNN-R [36]	23.0	44.9	55.2	69.0	88.9

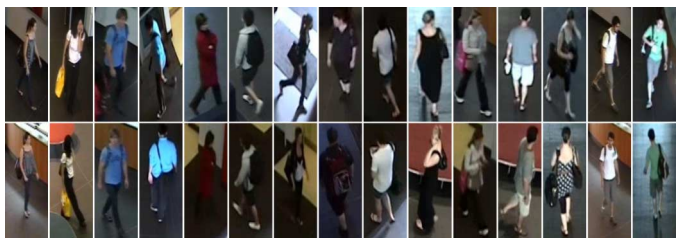

Fig. 8. Samples of image pairs from the SAIVT dataset.

C. SAIVT Dataset

This multicamera surveillance database was proposed by Bialkowski *et al.* in [37]. It was captured from an existing surveillance network to provide a real indoor scenario. The dataset consists of 150 people moving around a building environment, captured by eight different cameras with nonoverlapping fields of view. The dataset was collected in an uncontrolled manner, so it provides a highly unconstrained environment in which to test people reidentification approaches. This results in a challenging multicamera database designed for the task of people reidentification. Some image pairs corresponding to the same people are shown in Fig. 8.

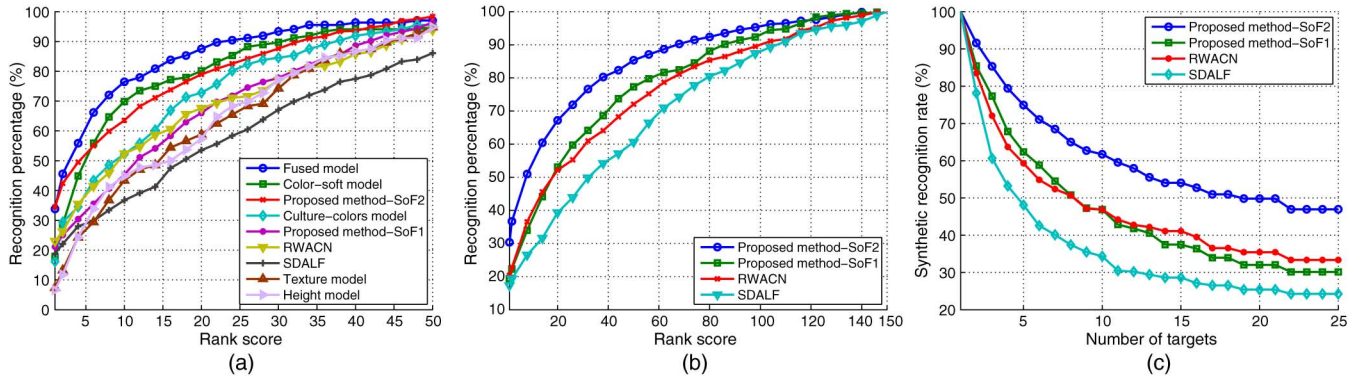

Fig. 9. Results according to SAIVT dataset. (a) and (b) CMC curves for two different setups. SRR curves are shown in (c) corresponding to the CMC curves of (b).

TABLE II
TOP RANKED MATCHING RATE (%) ON SAIVT DATASET

Methods	Rank:1	Rank:5	Rank:10	Rank:20	Rank:50
FM [37]	33.8	**61.8**	**76.5**	**87.5**	97.0
CCM [37]	17.6	51.5	69.8	80.1	94.8
Proposed-SoF2	**34.5**	52.2	63.5	78.9	**98.4**
CSM [37]	16.2	39.7	52.2	72.7	95.6
Proposed-SoF1	21.1	33.7	45.3	65.9	96.4
RWACN [10]	22.9	38.2	52.4	67.6	93.7
SDALF [9]	18.8	29.2	36.7	53.5	86.1
TM [37]	7.4	27.2	43.4	58.8	94.8
HM [37]	6.6	30.9	44.8	57.3	94.8

Fig. 10. Samples of image pairs from the ETHZ dataset.

We report the results of our method and compare them with those reported for RWACN [10], SDALF [9], Fused Model (FM), Culture-Colors Model (CCM), Color-Soft Model (CSM), Height Model (HM), and Texture Model (TM); these latter methods are presented in [37]. We adopted the same setup for a camera pair as that used in [37], which is denoted as 3-8. This camera pair contains 99 people viewed from similar perspectives. Fig. 9(a) shows the performance of our method compared to the previous methods. Using SoF2, the proposed method obtains similar results with respect to the others. Table II presents the top ranked matching rates, and it can be seen that our method achieved the highest recognition percentage for rank number 1. Fig. 9(b) and (c) shows the CMC and SRR curves for a setup where all people were used to obtain the results. Proposed method-SoF2 achieved 30.5% correct recognition for rank 1, whereas RWACN only obtained 20.4% using the same feature set. Similar results were obtained for proposed method-SoF1 and SDALF, where the correct recognition values were 19.5% and 17.6%, respectively.

D. ETHZ Dataset

The ETHZ dataset was introduced in [38] and consists of three outdoor video sequences captured from moving cameras mounted on a children's stroller. This dataset is not specifically designed for people reidentification, but some authors have used these video sequences to obtain results using a specific set of snapshots. The set of images introduces variations in appearance and lighting changes. Given the requirements of our proposal concerning calibration parameters, only sequence SEQ.#1 can be used to obtain the performance. This sequence contains 83 people and includes calibration parameters and odometry. Fig. 10 shows some examples of image pairs corresponding to the same people.

We compared the results obtained using the proposed method with those reported for eSDC_Knn [39], eSDC_ocsvm [39], PLS [40], eBiCov [41], RWACN [10], and SDALF [9]. A similar setup to that suggested in [39] was used to obtain results with our method. However, each STT was split into two parts of a similar length, enabling us to evaluate the iteration matching algorithm. Fig. 11(a) shows the performance of our method compared to previous methods. The proposed method using SoF2 outperformed the others, especially for rank scores 3, 4, and 5. Table III presents the top ranked matching rates, where it can be seen that our method achieved the highest recognition percentages, with the exception of rank score 1, where the eSDC_Knn method provided the best recognition percentage. Fig. 11(b) and (c) shows CMC and SRR curves for the full rank score. The proposed method-SoF2 achieved an 80.5% correct recognition for rank 1, whereas RWACN only obtained 57.7% using the same feature set. Similar results were obtained for the proposed method-SoF1 and SDALF, where the correct recognition values were 77.0% and 52.2%, respectively.

Our proposed method clearly reidentifies people correctly in an extensive camera network, providing better values when increasing numbers of people are considered in the SRR curve. Note that our technique performs processing independently for each individual without requiring knowledge of the full dataset. Only images with different orientations are extracted from the trajectory of the individual, and

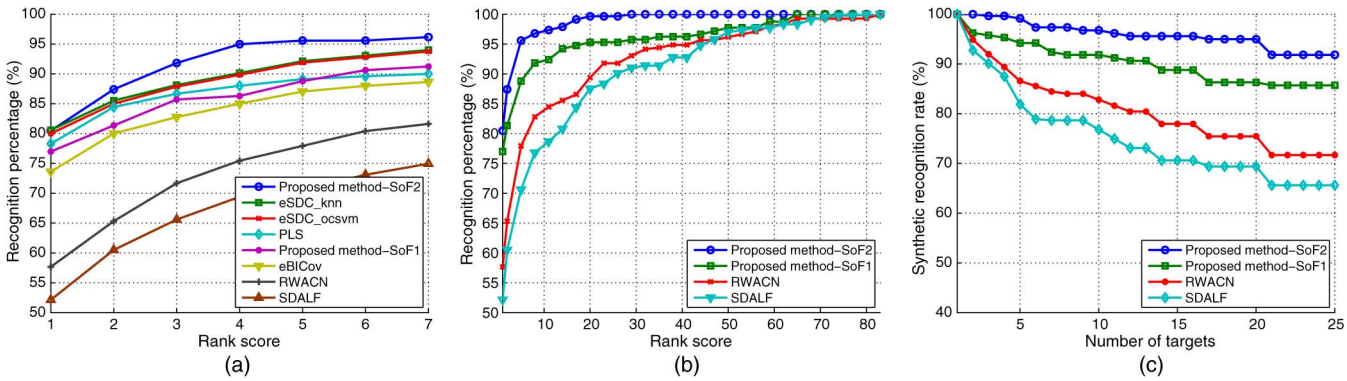

Fig. 11. Results according to ETHZ dataset. CMC curves are shown in (a) and (b). SRR curves are shown in (c) corresponding to the CMC curves of (b).

TABLE III
TOP RANKED MATCHING RATE (%) ON ETHZ DATASET

Methods	Rank:1	Rank:2	Rank:3	Rank:4	Rank:5
Proposed-SoF2	80.5	**87.4**	**91.8**	**94.9**	**95.6**
eSDC_Knn [39]	**80.6**	85.5	88.1	90.14	92.13
eSDC_ocsvm [39]	80.0	85.0	87.8	89.7	91.88
PLS [40]	78.3	84.4	86.6	88.0	89.1
Proposed-SoF1	77.0	81.3	85.7	86.3	88.8
eBiCov [41]	73.6	80.0	82.7	85.0	87.04
RWACN [10]	57.7	65.3	71.7	75.4	77.9
SDALF [9]	52.2	60.5	65.6	69.4	70.6

subsequently a set of reliable, robust, and descriptive features are extracted.

V. CONCLUSION

In this paper, we present the MVOM to carry out people reidentification process. To address variations in appearance due to the different perspectives of the person obtained from a camera network, we propose a model composed of different perspectives of the person. Each perspective is represented by a feature vector, an estimated orientation parameter and a reliability parameter, extracted from the person trajectory. An iterative algorithm maximizes the number of successful matches while speeding up the process. The proposed model does not require training stages to carry out the reidentification process, and it is not necessary to have a prior knowledge about the full dataset. To provide a reliable comparison, various experiments have been performed with three feature sets proposed by other authors. There are several problems in the reidentification process, such as lighting changes and low image resolution. However, problems related to appearance variation due to perspective changes have been reduced by increasing ranking values with respect to other proposal.

REFERENCES

[1] N. Martinel, C. Micheloni, C. Piciarelli, and G. Foresti, "Camera selection for adaptive human-computer interface," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 44, no. 5, pp. 653–664, May 2014.
[2] H. Liu, S. Chen, and N. Kubota, "Intelligent video systems and analytics: A survey," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1222–1233, Aug. 2013.
[3] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Boosted human re-identification using riemannian manifolds," *Image Vis. Comput.*, vol. 30, no. 67, pp. 443–452, 2012.
[4] T. Bai and Y. Li, "Robust visual tracking using flexible structured sparse representation," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 538–547, Feb. 2014.
[5] X. Zhou, Y. Li, B. He, and T. Bai, "GM-PHD-based multi-target visual tracking using entropy distribution and game theory," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1064–1076, May 2014.
[6] I. Bouchrika, J. N. Carter, and M. S. Nixon, "Recognizing people in non-intersecting camera views," in *Proc. 3rd Int. Conf. Crime Detect. Prev. (ICDP'09)*, Dec. 2009, pp. 1–6.
[7] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. 10th Eur. Conf. Comput. Vis. (ECCV'08)*, 2008, vol. 5302, pp. 262–275.
[8] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 898–903, 2012.
[9] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'10)*, Jun. 2010, pp. 2360–2367.
[10] N. Martinel and C. Micheloni, "Re-identify people in wide area camera network," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW'12)*, Jun. 2012, pp. 31–36.
[11] M. Hirzer, P. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV'12)*, 2012, vol. 7577, pp. 780–793.
[12] Y. Wu, M. Minoh, M. Mukunoki, W. Li, and S. Lao, "Collaborative sparse approximation for multiple-shot across-camera person re-identification," in *Proc. IEEE 9th Int. Conf. Adv. Video Signal-Based Surveill. (AVSS'12)*, 2012, pp. 209–214.
[13] W. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013 doi: 10.1109/TPAMI.2012.138.
[14] D.-N. T. Cong, C. Achard, and L. Khoudour, "People re-identification by classification of silhouettes based on sparse representation," in *Proc. 2nd Int. Conf. Image Process. Theory Tools Appl. (IPTA'10)*, Jul. 2010, pp. 60–65.
[15] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning implicit transfer for person re-identification," in *Proc. Comput. Vis. ECCV 2012 Workshops Demonstr.*, 2012, vol. 7583, pp. 381–390.
[16] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *Proc. Brit. Mach. Vis. Conf.*, 2010, pp. 21.1–21.11.
[17] I. Oliveira and J. Souza-Pio, "People reidentification in a camera network," in *Proc. 8th IEEE Int. Conf. Dependable, Auton. Secure Comput. (DASC'09)*, Dec. 2009, pp. 461–466.
[18] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Person re-identification using HAAR-based and DCD-based signature," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS'10)*, Aug. 29/Sep. 1, 2010, pp. 1–8.
[19] M. Bauml and R. Stiefelhagen, "Evaluation of local features for person re-identification in image sequences," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS'11)*, Aug. 30/Sep. 2, 2011, pp. 291–296.
[20] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Multiple-shot human re-identification by mean riemannian covariance grid," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS'11)*, Aug. 30/Sep. 2, 2011, pp. 179–184.

[21] B. Ma, Y. Su, and F. Jurie, "Bicov: A novel image representation for person re-identification and face verification," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 57.1–57.11.
[22] J. Metzler, "Appearance-based re-identification of humans in low-resolution videos using means of covariance descriptors," in *Proc. IEEE 9th Int. Conf. Adv. Video Signal-Based Surveill. (AVSS'12)*, Sep. 2012, pp. 191–196.
[23] R. Layne, T. Hospedales, and S. Gong, "Person re-identification by attributes," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 24.1–24.11.
[24] L. Ma, X. Yang, Y. Xu, and J. Zhu, "Human identification using body prior and generalized EMD," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP'11)*, Sep. 2011, pp. 1441–1444.
[25] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. 17th Scand. Conf. Image Anal.*, 2011, pp. 91–102.
[26] T. Gandhi and M. Trivedi, "Person tracking and reidentification: Introducing panoramic appearance map (PAM) for feature representation," *Mach. Vis. Appl.*, vol. 18, no. 3–4, pp. 207–220, 2007.
[27] D. Baltieri *et al.*, "Multi-view people surveillance using 3D information," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV'11)*, 2011, pp. 1817–1824.
[28] J. Oliver, A. Albiol, and A. Albiol, "3D descriptor for people re-identification," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR'12)*, Nov. 2012, pp. 1395–1398.
[29] K. Jungling and M. Arens, "View-invariant person re-identification with an implicit shape model," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS'11)*, 2011, pp. 197–202.
[30] X. Zhang, W. Hu, S. Chen, and S. Maybank, "Graph-embedding-based learning for robust object tracking," *IEEE Trans. Ind. Electron.*, vol. 61, no. 2, pp. 1072–1084, Feb. 2014.
[31] P. Vadakkepat, P. Lim, L. De Silva, L. Jing, and L. L. Ling, "Multimodal approach to human-face detection and tracking," *IEEE Trans. Ind. Electron.*, vol. 55, no. 3, pp. 1385–1393, Mar. 2008.
[32] J. Garcia *et al.*, "Directional people counter based on head tracking," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3991–4000, Sep. 2013.
[33] D. Baltieri, R. Vezzani, and R. Cucchiara, "3DPES: 3D people dataset for surveillance and forensics," in *Proc. 1st Int. ACM Workshop Multimedia Access 3D Human Objects*, Scottsdale, Arizona, USA, 2011, pp. 59–64.
[34] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2013, pp. 3318–3325.
[35] M. Kostinger, M. Hirzer, P. Wohlhart, P. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2012, pp. 2288–2295.
[36] M. Dikmen, E. Akbas, T. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Proc. 10th Asian Conf. Comput. Vis. (ACCV)*, 2011, pp. 501–512.
[37] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, and P. Lucey, "A database for person re-identification in multi-camera surveillance networks," in *Int. Conf. Digit. Image Comput. Techn. Appl. (DICTA'12)*, 2012, pp. 1–8.
[38] A. Ess, B. Leibe, K. Schindler, and L. van Gool, "A mobile vision system for robust multi-person tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'08)*, Jun. 2008, pp. 1–8.
[39] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, Jun. 2013, pp. 3586–3593.
[40] W. Schwartz and L. Davis, "Learning discriminative appearance-based models using partial least squares," in *Proc. 22nd Braz. Symp. Comput. Graphics Image Process. (SIBGRAPI'09)*, Oct. 2009, pp. 322–329.
[41] B. Ma, Y. Su, and F. Jurie, "BICOV: A novel image representation for person re-identification and face verification," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 57.1–57.11.

Jorge García received the B.S. degree in telecommunications engineering and the M.Sc. degree in electronics system engineering from the University of Alcalá, Madrid, Spain, in 2009 and 2011, respectively, and is currently pursuing the Ph.D. degree in the Electronics Department at the same university.

Since 2009, he has been with the Electronics Department, University of Alcalá,. His research interests include computer vision, video surveillance applications, scene understanding, and system based on field programmable gate arrays.

Alfredo Gardel received the M.Sc. degree in telecommunication engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1999, and the Ph.D. degree in telecommunication from the University of Alcalá, Madrid, Spain, in 2004.

Since 1997, he has been a Lecturer with the Electronics Department, University of Alcalá. His research interests include infrared and computer vision, monocular metrology, robotics sensorial systems, and design of advanced digital systems.

Ignacio Bravo (M'07) received the B.S. degree in telecommunications engineering, the M.Sc. degree in electronics engineering, and the Ph.D. degree in electronics from the University Alcala, Madrid, Spain, in 1997, 2000, and 2007, respectively.

Since 2002, he has been a Lecturer with the Electronics Department, University of Alcala. Currently, he is an Associate Professor with the University of Alcala. His research interests include reconfigurable hardware, vision architectures-based in field programmable gate arrays and electronic design.

José Luis Lázaro received the B.S. degree in electronic engineering and the M.Sc. degree in telecommunication engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1985 and 1992, respectively, and the Ph.D. degree in telecommunication from the University of Alcalá, Madrid, Spain, in 1998.

Since 1986, he has been a Lecturer with the Electronics Department, University of Alcalá, where he is currently a Professor. His research interests include robotics sensorial systems by laser, optical fibers, infrared and artificial vision, motion planning, monocular metrology, and electronics systems with advanced microprocessors.