Modeling Feature Distances By Orientation Driven Classifiers For Person Re-Identification

Jorge García[a,*], Niki Martinel[b], Alfredo Gardel[a], Ignacio Bravo[a], Gian Luca Foresti[b], Christian Micheloni[b]

[a] *Department of Electronics, University of Alcala, Alcalá de Henares (28801), Spain*
[b] *Department of Mathematics and Computer Science, University of Udine, Udine (33100), Italy*

Abstract

Most of the open challenges in person re-identification arise from the large variations of human poses as captured by different camera views. To tackle the re-identification challenges state-of-the-art methods propose to directly match image features or to learn the transformation of features that undergoes between two cameras. Other methods attack the problem by learning optimal similarity measures. However, the performance of all these methods are strongly dependent from the person pose and orientation. We focus on this aspect and introduce three main contributions ot the field: (i) propose a method to extract multiple frames of the same person with different orientation so as to capture the complete person appearance; (ii) learn the pairwise feature dissimilarities space (PFDS) formed by the subspace of pairwise feature dissimilarities computed between images of persons with similar orientation and the subspace of pairwise feature dissimilarities computed between images of persons with different orientations; (iii) within each subspace, a classifier is trained to capture the multi-modal inter-camera transformation of pairwise image dissimilarities and to discriminate between images of the same person (positive pair) and images of different persons (negative pair). To validate the proposed approach we show the superior performance of our approach to state-of-the-art methods using two publicly available benchmark datasets.

Keywords: Person Re-Identification, Pose and Orientation Recovery, Appearance Models, Surveillance Systems, Scene Understanding.

1. Introduction

The person re-identification problem, formally defined as the problem of associating a given person acquired by a camera to the same person previously acquired by any other camera in the network at any location and at any time instant, is increasingly gaining attention by the community. This challenging task is very important for surveillance applications such as inter-camera tracking, multi-camera behavior analysis,etc.. Despite the problem can be alleviated by deploying a large number of sensors such as all

[*] Corresponding author: Tel.: +0-000-000-0000; fax: +0-000-000-0000;
Email address: `author@author.com` (Christian Micheloni)

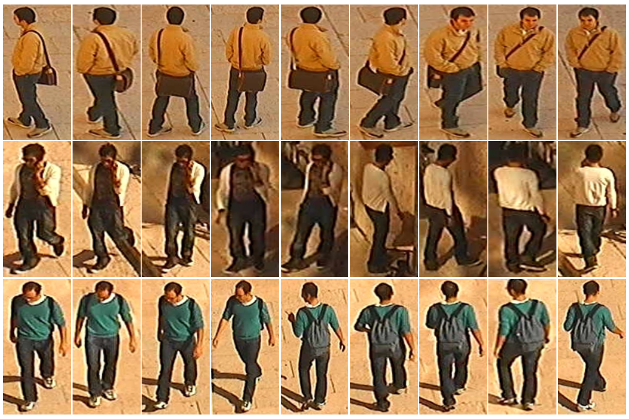


Figure 1: Images of a persons acquired by the same camera. Person appearances look very different among all the images due to the changes in pose and illumination conditions.

the areas of the monitored environment are covered by camera field-of-views (FoVs), the costs of system installation, maintenance,etc., lead to a non-feasible solution. Thus, in a real scenario, we have to deal with partial area coverage that yields to the re-identification problem.

Motivation: Despite much effort has been spent by the community to find the best signature (e.g. [1, 2, 3, 4]), to learn the feature transformation that undergoes between camera pairs (e.g. [5, 6, 7, 8]) and to find the optimal similarity measure (e.g. [9, 10, 11, 12]), re-identify a person that moves across disjoint cameras still remains an open issue. Almost all existing works assume that a uni-modal inter-camera transformation of features occurs between two camera views. However, the we believe that the deployment and the configuration of the cameras (it is a combination of view points, illumination changes, and photometric settings, etc.) together with the appearance of a person give rise to multi-modal inter-camera transformations (see Fig. 1 for an example). In particular, current methods highly suffer the strong pose and orientation changes that may occur when a person moves between cameras FoVs.

Contribution: Motivated by these, and inspired by the fact that as the transformation between appearance features is multi-modal, so is the transformation of the distances between them [13], we introduce the following contributions:

1. we build upon the idea that the transformation learned for the same person seen from different viewpoints may be less reliable than the one learned for the same person seen from the same point of view. Hence, we introduce a method to recover images of the same person with different orientations so as we can capture the multi-modal appearance of a person with higher reliability;
2. the person orientations are used to learn different models for the different transformations that exist between pairwise feature dissimilarities. We form the *pairwise feature dissimilarities space (PFDS)* and divide it into two main regions, the region for which the dissimilarities are computed between pair of images with similar orientation and the region containing all other different pairwise orientations.

3. within each region, a classifier is trained to capture all the possible multi-modal transformation of feature dissimilarities. Those are used to discriminate between pairwise images of the same person (positive pair) and pairwise images of two different persons (negative pair). This also allows to pose the re-identification as a binary classification problem.

The rest of the paper is organized as follows. A brief description of the related work is given in section 2. In section 3 a system overview and details of the modules that compose our re-identification approach are described. The superior performance of our approach over existing state-of-the-art methods is shown in section 4. Finally, conclusions are drawn in section 5.

2. Related Work

In the recent past, the community has proposed to tackle the problem of person re-identification across non-overlapping cameras using several approaches that differs from the way the person body is modeled, to which features are used (i.e. biometrics or appearance), to how matches between individuals are computed, etc. [14, 15]. While recent works in the field of person re-identification can be grouped on the basis of any of such categories, we group them as follows: i) methods that use discriminative appearance-based signatures, ii) methods that capture the transformation of features across camera pairs, and iii) methods that learn the optimal distance metric between appearance features.

Discriminative signature based methods are the most commonly explored approaches for person re-identification. In [2] particular interest has been focused on finding the best set of features that can be exploited to match persons across cameras. In [16] the objective was to model the spatial distribution of the appearance relative to each of the object part. A discriminative signature computed using the Mean Riemannian Covariance patches was used in [17]. In [1], frames were used to built a collaborative representation that best approximates the query frames. In [3], the distribution of color features projected in the log-chromaticity space was described using the shape context descriptor. In [18] an unsupervised framework was proposed to extract distinctive features, then a patch matching method was used together with adjacency constraints. In [4] a combination of biologically inspired features and covariance descriptors was proposed. In [19] local feature descriptors were encoded by fisher vectors and pooled to provide a global image representation. In [20] an articulated multiple-instance-based compositional template was proposed to model person appearance. Appearance features and similarities with a reference set of persons were used in [21]. Human saliency was also explored in [18, 22] to reject body parts that are non-discriminative for the re-identification task.

Those methods rely only on the discriminative power of appearance features to perform a pure feature matching. While no training is required and good results can be achieved when images are similar, this is still an unreliable solution as such methods generally assume that features are not transformed between cameras.

Transformation learning based methods were explored in [5] to capture the transformation across non-overlapping cameras in a tracking scenario. Similarly, the problem of capturing the non-linear transformation between features was addressed in [23]. In [24] pairwise dissimilarity profiles between categories were learned and exploited in a nearest neighbor classification framework. In [6] the implicit transformation function of features was learned by concatenating appearance feature vectors of persons viewed by different cameras. In [7] a Weighted Brightness Transfer Function that assigns unequal weights to observations based on how close they are to test observations was proposed. In [13], the error on the transformation function was modeled by a binary random forest classifier. In [8] a transformation between locally aligned feature spaces was leatned. These methods assume that there is a unique transformation between features and that it can be used to project the feature from one camera to the feature space of the other camera.

Distance learning based methods learn the best metric between appearance features of the same person across camera pairs. In [9, 10, 11] the Largest Margin Nearest Neighbour and a derivation of it were exploited. In [25] the re-identification problem was addressed in a transfer learning framework. In [26] a distance metric based on a statistical inference perspective, is learned from equivalence constraints. In [12] a novel distance learning method based on sparse pairwise constants was proposed. The approach has been extended in [27], by introducing a smooth regularizer. The problem of finding the optimal similarity measure was proposed in [28, 29]. In [30] the re-identification problem was formulated as a local distance comparison problem introducing an energy-based loss function that measures the similarity between appearance instances. Such approaches match features in the same feature space and did not consider cross-view transformations.

These three categories has been also recently exploited by re-ranking methods. In [31] a method based on nonlinear ranking with difference vectors was proposed. In [32] a multi-task support vector ranking was proposed to incorporate the training data from source domain with label information. In [33] it was shown that user intervention for re-ranking in the deployment stage improved re-identification performance.

3. Proposed Approach

An overview of our approach is shown in Fig. 2. The four underlying modules work as follows. Given a pair of cameras with non-overlapping FoVs, the first module retrieves different perspectives for each short-term tracklet corresponding to a person crossing the scene (Section 3.1). Each retrieved perspective is characterized by an orientation value and a reliability value. The next module captures the appearance of the person by extracting different global and local features based on color, texture and shape information (Section 3.2). The high-dimensional feature vectors extracted from a pair of perspectives of persons acquired by two different cameras are input to the feature dissimilarities module which computes the pairwise feature dissimilarities vector (PFD) (Section 3.3). The set of all PFD makes up the pairwise feature dissimilarities space (PFDS). The PDFS is split in two groups considering the orientation of the pairwise images. The former is composed of PFDs between images of persons with similar orientation, while the other is formed of PFDs computed for images that have different orientations. Two binary classifiers, named as similar orientation classifier (SO-Classifier) and non-similar orientation classifier (NSO-Classifier), are trained to separate between positives ans negatives PFDs in the two subspaces, respectively. Finally, the last module exploits the learned models to classify new PFDs. In particular, only PFDs associated to orientation distances with high reliability values are used to select among the two classifiers. If a PFD is extracted from images with low reliability values, the NSO-Classifier (Section 3.4) is used.

3.1. Retrieving People Perspectives

We assume that the people detection and tracking tasks in a single camera have already been achieved as in [34]. Then, each camera provides a set of short-term tracklets of persons to carry out the re-identification process. Some situations may cause direction changes in the trajectory of a person due to static objects which are located in the scene, crosses with other people or due to his/her own trajectory. We take advantage of these situations in order to obtain multiple images with different perspectives of the person. An orientation value is assigned to each perspective according to camera location. Let $\mathcal{T} = (\mathbf{x}\ \mathbf{y}\ \mathbf{v}_x\ \mathbf{v}_y)$ be a trajectory captured by a camera, where $\mathbf{x}$ and $\mathbf{y}$ are image position vectors and $\mathbf{v}_x$ and $\mathbf{v}_y$ are image velocity vectors provided by a multi-tracking task. We can define the orientation as the angle between the position vector and the velocity vector of the trajectory. However, this value does not provide a perspective-camera relationship that allows to compare perspectives from short-term tracklets computed by other cameras. Therefore, we need to find a new relationship to relate different perspectives satisfying the previous restriction. Assuming a person always walks in a forward direction, the angle between the trajectory vector and camera vector provides a common inter-camera relationship that can be used to compare against different perspectives. Notice that the camera vector (optical axis) is projected on the ground floor. We define this angle as the estimated orientation of the person with respect to the camera.

Let $\mathbf{T} = [T_x\ T_y\ T_z]$ and $\mathbf{R} = [\sigma_x\ \sigma_y\ \sigma_z]$ be the extrinsic parameters which denote the coordinate system transformation from the scene to the camera. Considering that the pan angle is null for each camera ($\sigma_y = 0$), the projection of the camera vector and axis y of the world coordinate system are parallel. Thus, we can determine the orientation vector as:

$$\theta = \arctan \frac{\mathit{\Delta}\mathbf{Y}_w}{\mathit{\Delta}\mathbf{X}_w}, \tag{1}$$

where $\mathit{\Delta}\mathbf{Y}_w$ and $\mathit{\Delta}\mathbf{X}_w$ are the coordinate difference vectors on the ground floor ($\mathbf{X}_w = 0$). We fixed the origin of world coordinates below to the camera, so the translation between coordinate systems only depends on the camera height. The translation vector is defined as $\mathbf{T} = [0\ 0\ T_z]$ where T_z is the height of the camera.

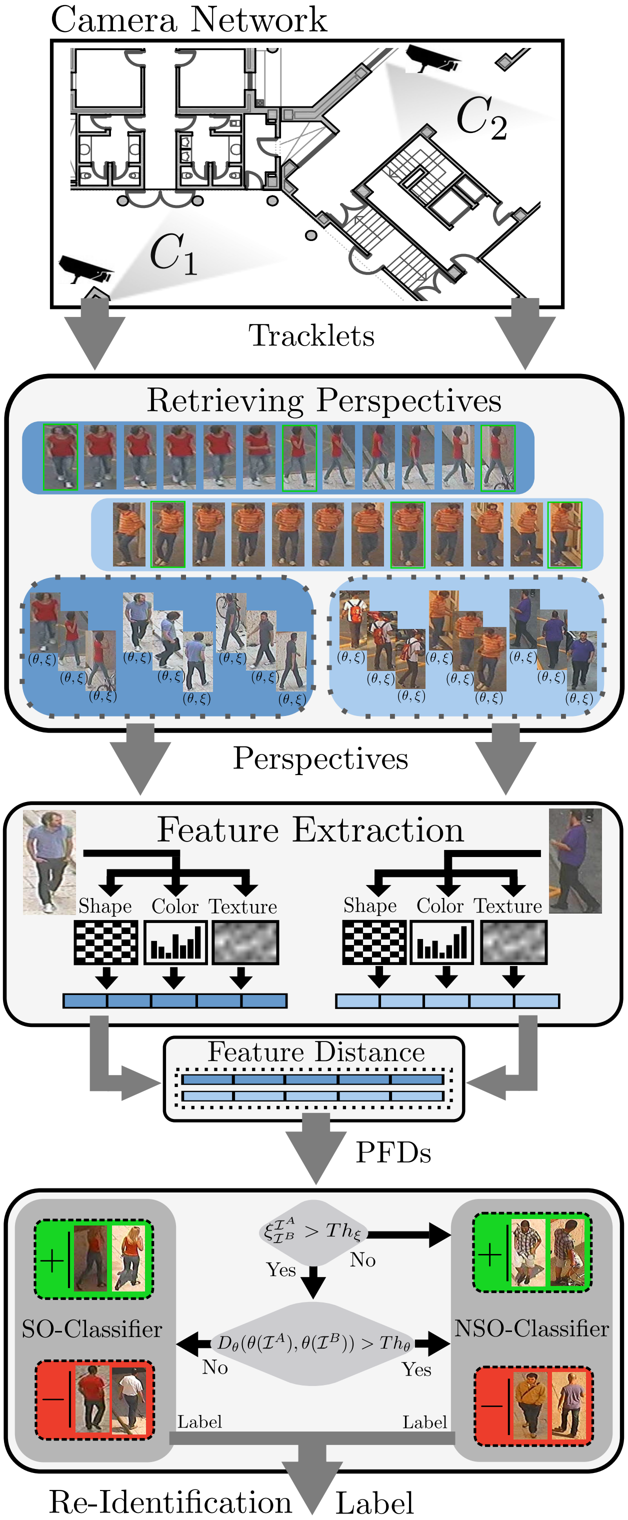


Figure 2: Overview of our approach. Given two tracklets of two persons acquired by disjoint cameras, we first recover the perspectives of the persons so as we have multiple frames of a same person viewed from different poses/viewpoints. Then, for each image shape, color and texture features are extracted and the PFD between image pairs is computed. Finally, the re-identification is performed by sending the PFD to one of the two previously learnt binary classifiers. The classifier is chosen on the basis of the orientation distance and the reliability value between the persons images used to compute the PFD.

Consequently, $\mathit{\Delta}\mathbf{Y}_w$ and $\mathit{\Delta}\mathbf{X}_w$ can be expressed as a function of coordinate difference vectors in the camera ($\mathit{\Delta}\mathbf{Y}_c$ and $\mathit{\Delta}\mathbf{X}_c$). That is:

$$\mathit{\Delta}\mathbf{Y}_w = -\sin\sigma_z \mathit{\Delta}\mathbf{X}_c + \frac{\cos\sigma_z}{\cos\sigma_x}\mathit{\Delta}\mathbf{Y}_c \tag{2}$$

$$\mathit{\Delta}\mathbf{X}_w = \cos\sigma_z \mathit{\Delta}\mathbf{X}_c + \frac{\sin\sigma_z}{\cos\sigma_x}\mathit{\Delta}\mathbf{Y}_c. \tag{3}$$

Finally, we apply the image projection model using the intrinsic parameters to obtain an expression dependent on the image position vectors ($\mathbf{x}$ and $\mathbf{y}$).

In some situations, such as the people is turning around slowly, people walking far from the camera, crowded scenarios, etc., the accuracy of the estimated orientation might be not enough. To mitigate this problem, we propose a reliability value to model the probability of the orientation error using the linear and angular velocity vectors. Previous unreliable situations correspond to trajectories where the angular velocity is higher than zero and/or the linear velocity is low. Given two consecutive steps of the trajectory $\mathcal{T}$, the linear and angular velocities are defined as $v = (\mathit{\Delta} v_x^2 + \mathit{\Delta} v_y^2)^{1/2}$ and $\omega = \mathit{\Delta}\theta/T_s$, respectively. T_s is the frame-rate of the camera and $\mathit{\Delta}$ represents the difference between two consecutive steps. The reliability is modeled as a weighted function of two normal distributions with mean (μ_v, μ_ω) and variance $(\sigma_v^2, \sigma_\omega^2)$.

$$\xi = (1-\alpha)\mathcal{N}(\mu_v, \sigma_v^2) + \alpha\mathcal{N}(\mu_\omega, \sigma_\omega^2) \tag{4}$$

We set $\mu_\omega = 0$ and μ_v as the average person velocity in order to obtain a low weight when the person is not walking. Finally, α is used to balance the influence between linear and angular velocity.

Let $\Upsilon = \{I_i | i = 1, ..., N\}$ be a short-term tracklet captured from a camera, where I represents an image of the person and N is the number of images contained in Υ. The number of retrieved perspectives depends on the resolution of direction changes expressed by $\mathit{\Delta}\theta$ between the last image added and the image under analysis. Summarizing, we provide to the next module a set of perspectives $\Omega = \{I_j, \theta_j, \xi_j | j = 1, ..., M\}$ where M is the number of retrieved perspectives.

3.2. Feature extraction

Numerous features have been used to model the person appearance and to tackle the re-identification challenges (see [2]). Following the suggestions in [2], we proposed to build the feature representation of a given image I by considering color, texture and shape features. While shape features may not be very discriminative for other methods, that is not the case for our method as the pose is much related to the orientation of a person.

Color features: We consider that most of the persons wear different colored clothes for the upper and lower body part. Accordingly, we divide the body in three salient parts: legs, torso and head, as in [35]. We discard the head region as it generally contains few and not informative pixels. RGB, HSV, YUv and Lab color spaces are used to extract the histograms $\mathcal{H}_\omega^c(I) \in \mathbb{R}^{n_c}$, for each color component $c \in \{R, G, B, H, S, V, Y, U, v, L, a, b\}$ and body part $\omega \in \{T, L\}$, where T denotes the torso and L denotes the legs.

Texture features: We consider three filter responses (Gabor, Schmid, Leung-Malik filter banks) and two texture descriptors (Local Binary Pattern [36] and Local Phase Quantization [37]).

After convolving each image with a single Gabor filter we computed the modulus of the response and we quantized it in a histogram with g bins. We denote the set of all such histograms as $\{\mathcal{G}_i(I)\}_{i=1}^{I}$, where i indicates the i^{th} Gabor filter. Similarly we get the set of histograms $\{\mathcal{S}_j(I)\}_{j=1}^{J}$, each of which has s bins, by convolving the 13 standard Schmid filters. We convolve each given image with the Leung-Malik (LM) filter bank consisting of first and second derivatives of Gaussians at 6 orientations and 3 scales, 8 Laplacian of Gaussian filters, and 4 Gaussians. The response of each filter was quantized in a histogram with m bins. $\{\mathcal{L}_k(I)\}_{k=1}^{K}$ is the set of all such histograms, where k indicates the k^{th} LM filter.

Local Binary Pattern (LBP) encode the local structure around a pixel using circular neighborhoods with radius r on grayscale image. A binary number is obtained by concatenating all binary values in a clockwise direction to form the label of each region. Then, the descriptor is composed by a histogram of all labels with p bins. We denoted the histogram as $\mathcal{B}(I)$. Finally, Local Phase Quantization (LPQ) is based on computing the short-term Fourier transform on local region of the grayscale image. The local Fourier coefficients are computed for 4 frequency points. Then, the signs of the real and imaginary part of each coefficient are quantized using a binary scalar quantizer. Results of the 8 bit binary coefficients are represented as integers using binary coding. The codes are quantized in a histogram with q bins denoted as $\mathcal{Q}(I)$.

Shape features: Pyramid Histogram of Oriented Gradients (PHOG) feature is used to capture the shape of a given person. Let $l = 0, \cdots, L$ be the level of the spatial pyramid, and 4^l the number of cells in which the image is divided at each level l. Then, the PHOG feature $\mathcal{P}(I)$ is formed by concatenating all the HOG features extracted for each cell of the pyramid. This results in a vector of size $b\sum_{l=0}^{L} 4^l$, where b is the number of bins used to compute the HOG features.

3.3. Pairwise Feature Dissimilarities

Once all the features have been extracted from a pair of images acquired by two disjoint cameras we compute the PFD as suggested in [13]. Let I^A and I^B be the two given images, then the pairwise dissimilarities are computed as:

- Color: $D_{\mathcal{H}_\omega^c}(\mathcal{H}_\omega^c(I^A), \mathcal{H}_\omega^c(I^B))$ for all ω and c.
- Gabor: $D_{\mathcal{G}}(\mathcal{G}_i(I^A), \mathcal{G}_i(I^B))$, for $i = 1, \cdots, I$.
- Schmid: $D_{\mathcal{S}}(\mathcal{S}_j(I^A), \mathcal{S}_j(I^B))$, for $j = 1, \cdots, J$.

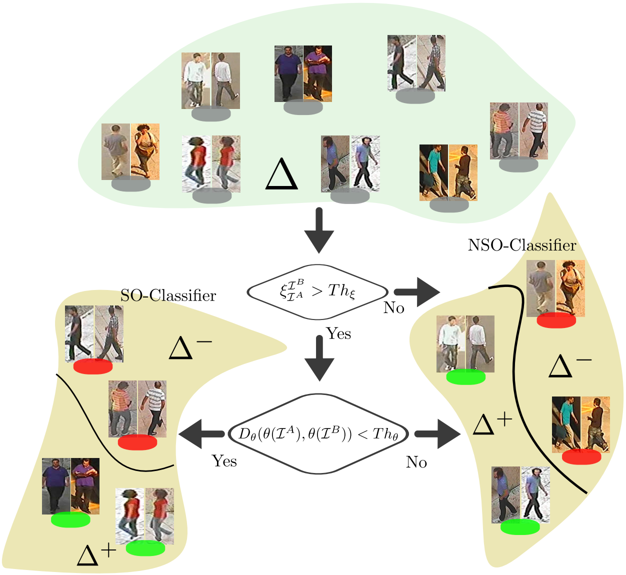

Δ
NSO-Classifier
SO-Classifier
$\xi_{\mathcal{I}^A}^{\mathcal{I}^B} > Th_\xi$
No
Yes
Δ^-
$D_\theta(\theta(\mathcal{I}^A), \theta(\mathcal{I}^B)) < Th_\theta$
Yes
No
Δ^+
Δ^-
Δ^+

Figure 3: Image representation of the proposed dual-classification. The PFDS is split into two subspaces on the basis of the orientation of pairwise images. A classifier is trained to discriminate between positive PFDs (green dots) and negative PFDs (red dots) for each subspace. The decision surface is depicted by a black line.

- LM filters: $D_{\mathcal{L}}(\mathcal{L}_k(I^A), \mathcal{L}_k(I^B))$, for $k = 1, \cdots, K$.
- LBP: $D_{\mathcal{B}}(\mathcal{B}(I^A), \mathcal{B}(I^B))$.
- LPQ: $D_{\mathcal{Q}}(\mathcal{Q}(I^A), \mathcal{Q}(I^B))$.
- PHOG: $D_{\mathcal{P}}(\mathcal{P}(I^A), \mathcal{P}(I^B)$.

Notice that here we do not specify any particular distance measure since the algorithm can be used with different metrics.

Then, we form the PFD by concatenating the all computed pairwise distances in a d- dimensional vector, denoted by $\Delta = (x_1, ..., x_d)^T \in \mathbb{R}^d$. The PFD computed for a positive pair (i.e. for pair of images of the same person) is denoted as Δ^+, while the PFD computed for a negative pair (i.e. for pair of images of different persons) is denoted as Δ^-. The set of all positive and negative PFDs forms the PFDS $\mathbf{\Delta}$.

3.4. Dual-Classification

The proposed dual-classification scheme is shown in Fig. 3. We propose train two classifiers to model the PFDS. Each classifier is selected depending on the orientation distance and the reliability values of pairwise images.

Given the PFDS, we first compute the orientation distance and the pairwise reliability for all PFDs as follows. Let $\{\theta(I^A), \xi(I^A)\}$ and $\{\theta(I^B), \xi(I^B)\}$ be the orientations of persons and the reliabilities in images

I^A and I^B, the orientation distance is defined as:

$$D_\theta\Big(\theta(I^A), \theta(I^B)\Big) = \arccos\Big(\cos(\theta(I^A))\cos(\theta(I^B)) + \sin(\theta(I^A))\sin(\theta(I^B))\Big) \tag{5}$$

and the pairwise reliability is defined as:

$$\xi_{I^B}^{I^A} = \min\{\xi(I^A), \xi(I^B)\}. \tag{6}$$

During the training phase, we select only PFDs where $\xi_{I^B}^{I^A} > Th_\xi$ from the PDFS. Th_ξ is a fixed threshold used to ensure that the orientation distances of the PFDs have enough accuracy. The resulting subset is split into two groups: PFDs with similar orientation, i.e. PFDs for which $D_\theta(\cdot, \cdot) < Th_\theta$, where Th_θ is a fixed threshold, and the rest of PFDs. Then, we train two classifiers: the similar orientation classifier (SO-Classifier) to discriminate between PFDs with similar orientation and the non-similar orientation classifier (NSO-Classifier).

In our current framework, we used two SVMs as the SO-Classifier and the NSO-Classifier. Each one learns the parameters of the decision boundary that best separates the PFDs computed for positives pairs from the ones computed for negative pairs. Given a subset of PFDs denoted as $\mathbf{x}_i$, $i = 1, ..., N$ where N is the number of training samples, the goal is to minimize the error function expressed by

$$\min_{w,b,\gamma} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \gamma_i \tag{7}$$

subject to the constraints $y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \gamma_i$ and $\gamma_i \geq 0$, where $\mathbf{w}$ is the vector of coefficients, $\phi(\mathbf{x}_i)$ is the feature map for $\mathbf{x}_i$, γ_i is the slack variable used to handle the non-separable input data and C is the regularization parameter. Once the minimization problem is solved, the decision function in the dual form is given by

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{N} y_i w_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{8}$$

where $K(\mathbf{x}_i, \mathbf{x}) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x})$ is the standard radial basis Kernel function (RBF) and $y_i \in \{1, -1\}$ is the label space.

During the classification phase, different perspectives from the same person are used to build the PFDs. They are independently processed, i.e. without tracking information, time constraints, etc. Given a test PFD, $\hat{\mathbf{x}}$, we first compute the pairwise reliability and evaluate the result with the threshold Th_ξ. If $\xi_{I^B}^{I^A} < Th_\xi$, the NSO-Classifier is used to compute the final a decision. In the other case, we compute the orientation distance as in eq.(5), then, we input the PFD to the classifier trained for the PFD group in which the test PFD rely. For both cases, the probability given from the classifier is used to tell whether the images used to compute the PFD are for the same person $f(\hat{\mathbf{x}}) \geq 1$ (positive pair) or are for different persons $f(\hat{\mathbf{x}}) = -1$ (negative pair). We average pool the probabilities computed between all the PFDs corresponding to the same person to compute the final score.

4. Experimental Results

In this section, first, we give the implementation details and report the performance of our approach on two public benchmark datasets, then we compare our results with state-of-the-art methods. All the reported results are in terms of Cumulative Matching Characteristic (CMC) curves.

4.1. Implementation Details

We have used the following settings for all the experiments presented in this section. Images from both datasets have been resized to 64 × 128 pixels. Color histograms have been computed for the RGB, HSV, YUV and Lab color spaces using $n_c = \{16\ 16\ 16,\ 36\ 25\ 20,\ 18\ 32\ 32,\ 20\ 32\ 32\}$ bins, respectively. We used Gabor filters at 8 orientations and 5 scales, the standard 13 Schmid filters and the LM filters described in section 3. For each filter response, histograms $\mathcal{G}(I)$, $\mathcal{S}(I)$ and $\mathcal{L}(I)$ have been computed with 16 bins. LBP descriptors have been computed using 8 neighbourhood points and $r = 1$. For compute LPQ descriptors, the window size has been set to $[5, 5]$. Both $\mathcal{B}(I)$ and $\mathcal{Q}(I)$ histograms have been computed with 255 bins. PHOG features have been extracted using 4 levels and 9 bins. The χ^2 distance has been used to compute all the feature distances (i.e., $D_{\mathcal{H}^c_\omega}$, $D_\mathcal{G}$,$D_\mathcal{S}$, $D_\mathcal{L}$, $D_\mathcal{B}$, $D_\mathcal{Q}$, $D_\mathcal{P}$). The SVM parameters have been separately estimated for each dataset using 4-fold cross-validation. Finally, for each experiment we select 1 positive pair and 1 negative pair for each person in the train set, while for the test we randomly selected 10 positive and 10 negative pairs per person. To make a fair comparison, for each experiment, we run 10 different trials using different person IDs and different image pairs (samples).

4.2. Datasets

Several datasets have been proposed to test re-identification algorithms such as VIPeR, CAVIAR4REID, ETHZ, WARD, etc. Typically, they provide an image or different images of the person in each camera, but none of them come with the required information we need to retrieve the people orientation from each camera. The two dataset we are using have footages acquired by a very high number of cameras and are representative of real indoor and outdoor surveillance scenarios. They also come with very strong illumination changes, occlusions, viewpoint variations, etc.

3DPeS Dataset: The 3D people surveillance dataset[1] (3DPeS) has been introduced in [38]. The dataset is composed by 8 cameras and each one presents different light conditions (clear light/shadow areas) and calibration parameters. Different sequences of 200 people have been taken from a multi-camera distributed surveillance system. Persons were detected multiple times with different viewpoints, time instants and on different days. This results in a challenging dataset with strong variation of light conditions (see Fig.4).

[1] Available at `http://www.openvisor.org/3dpes.asp`

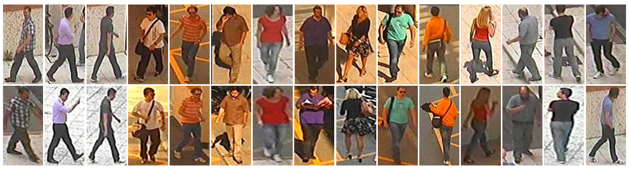


Figure 4: Image samples from 3DPeS dataset. Each column corresponds to a pair of images of the same person captured by two different cameras.

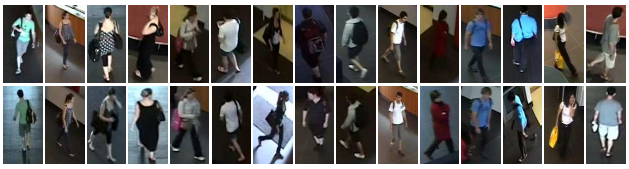


Figure 5: Image samples from the SAIVT dataset. Each column corresponds to a pair of images of the same person captured by two different cameras.

SAIVT Dataset: The SAIVT-SoftBio dataset[2] has been introduced in [39]. The dataset was captured by a real surveillance camera network in an uncontrolled fashion, so it provides a highly unconstrained environment in which to test person re-identification approaches. The 150 persons were acquired by 8 indoor cameras with non-overlapping field of views. Some example images are shown in Fig. 5.

Fig. 6 shows the orientation distances spectrum for each dataset. We select the camera pair 1-2 on 3DPeS dataset since this one has the highest number of persons crossing the captured areas. Similarly, camera pairs 3-8 and 5-8 of the SAIVT dataset have been selected. Camera views 3 and 8 have similar viewing angles, while camera views 5 and 8 have dissimilar viewing angles. The PFDS corresponding to the camera pair 1-2 on 3DPeS dataset is uniformly distributed along to the all orientation distances. However, the PFDS corresponding to camera pairs from SAIVT dataset are concentrated in two regions of the orientation distance spectrum. Notice that all camera pairs have a symmetrical shape in 90 degrees because people normally make round trajectories. Since there are enough image pairs in all orientations, the 3DPeS is used in the following to analyze the performance of our method using different algorithm parameters.

4.3. Performance of Our Approach

An extensive analysis has been carried to show the performance of our approach. We focus on the performance of the classifiers using different ranges of orientation distances with specific training settings. In addition, a set of values are considered in order to determine the proposed thresholds. Finally, we report the performance of our method using single features only, combination of features and different training/testing set sizes.

[2] Available at `https://wiki.qut.edu.au/display/saivt/SAIVT-SoftBio+Database`

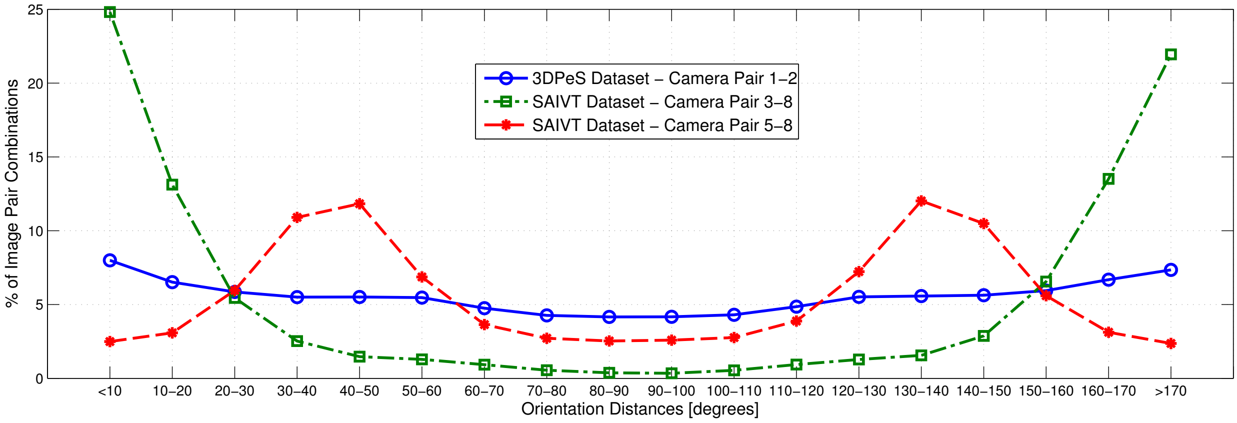


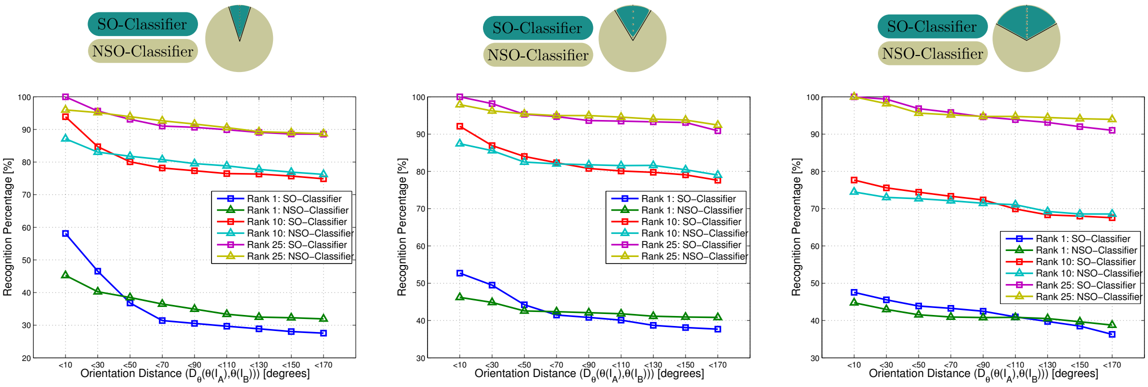
Figure 6: Orientation distances spectrum for each camera pairs used in the experiments.

Classifiers: The following tests show the performance of the proposed classifiers for different experiments varying the set used for test. In each experiment PFDs within a determined range of orientation distances are considered. The pairwise reliability $\xi_{I^B}^{I^A}$ is forced to overtake the reliability threshold Th_ξ for all the PFDs, which is set to 0.8. In this way, we guarantee that the orientation distance of the PFD corresponds to the perspectives of the images. In addition, we evaluate the classifiers separately, i.e., all the PFDs are used to re-identify in each classifier. We use three training settings corresponding to three different values for the Th_θ parameter, which are maintained throughout the experiments.

Fig. 7(a), (b) and (c) show the performance for some distinctive ranks when the Th_θ parameter is set to 15, 30 and 60 degrees, respectively. The SO-Classifier achieves a greater recognition than NSO-Classifier when the orientation distance of PFDs is below or near to the Th_θ used to train. However, the performance of the SO-Classifier remarkably decreases when the orientation distance of PFDs exceeds the Th_θ used to train, while the performance of the NSO-Classifier is more consistent. Thus, the inter-camera transformation captured by the SO-Classifier is more reliable than the one. Indeed, the SO-Classifier achieves a recognition rate of 58.5% for rank 1 with $Th_\theta = 15$ degrees, 53% with $Th_\theta = 30$ and 47.7% with $Th_\theta = 60$.

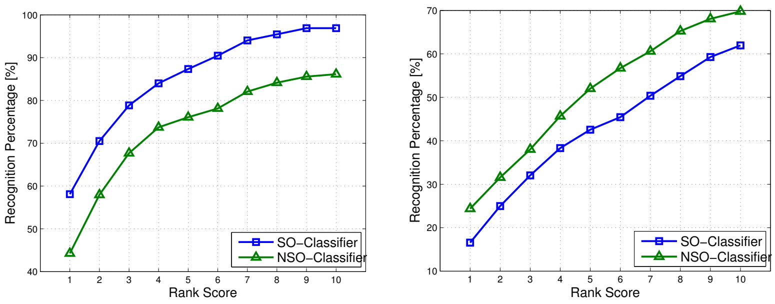
Fig. 8 shows the performance of the classifiers for two specific set of PFDs. We select the training set samples by using $Th_\theta = 30$ degrees. PFDs with similar orientation distances are used in Fig. 8(a) where the SO-Classifier achieves a greater recognition than NSO-Classifier. The opposite result is obtained in Fig. 8(b) since the set is composed of PFDs with non-similar orientation distances. The SO-Classifier achieves a recognition rate of 58% for rank 1 when PFDS with similar orientation make up the test set, while the NSO-Classifier achieves a recognition rate 24.5% with PFDs with non-similar orientation.

Parameter Th_θ: Fig. 9(a) shows the performance of the proposed method for some distinctive ranks as a function of the orientation threshold Th_θ. We provide results using three training settings corresponding to 15, 30 and 60 degrees. In this case, the parameter Th_ξ is set to 0.7 and the value of the Th_θ is used to



(a) (b) (c)

Figure 7: Performance of the classifiers varying the set of PFDs depending on orientation distances. Recognition percentage for some rank scores using three training settings corresponding to 15, 30 and 60 degrees are shown in (a), (b) and (c) respectively.



(a) (b)

Figure 8: Recognition percentage of the classifiers using two test settings: (a) Set of PFDs with similar orientation and (b) Set of PDFs with non-similar orientation.

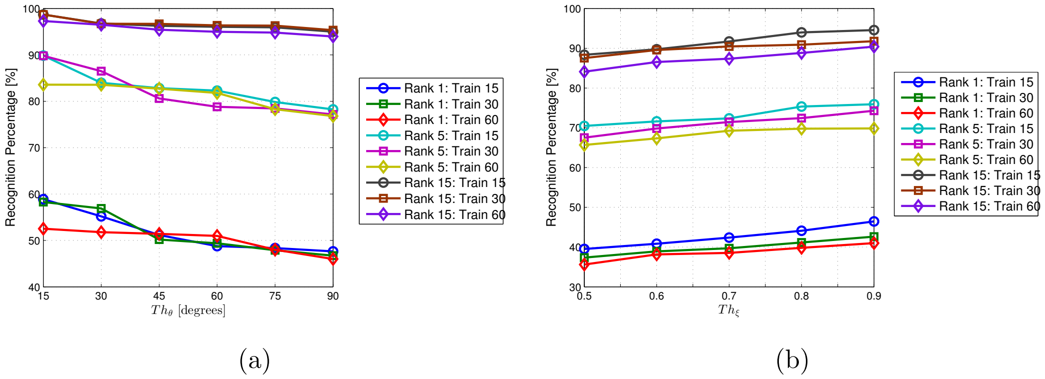


Figure 9: Performance of the proposed method: (a) varying the parameter Th_θ. (b) varying the parameter ϕ_θ.

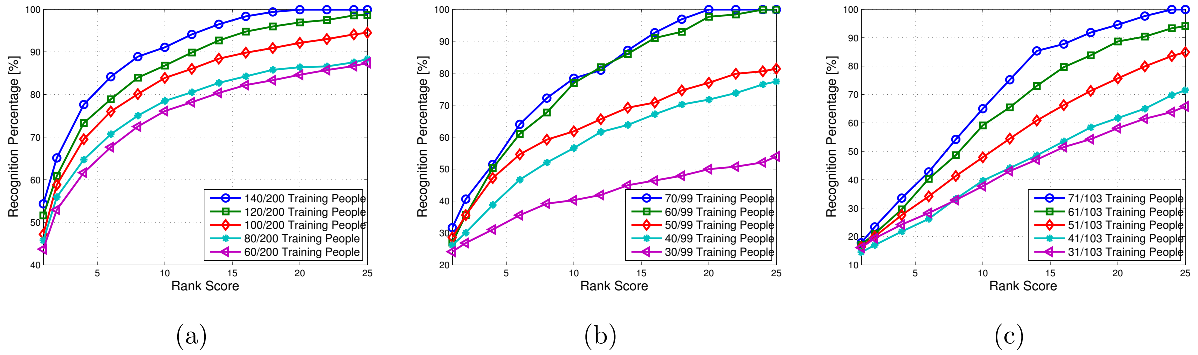


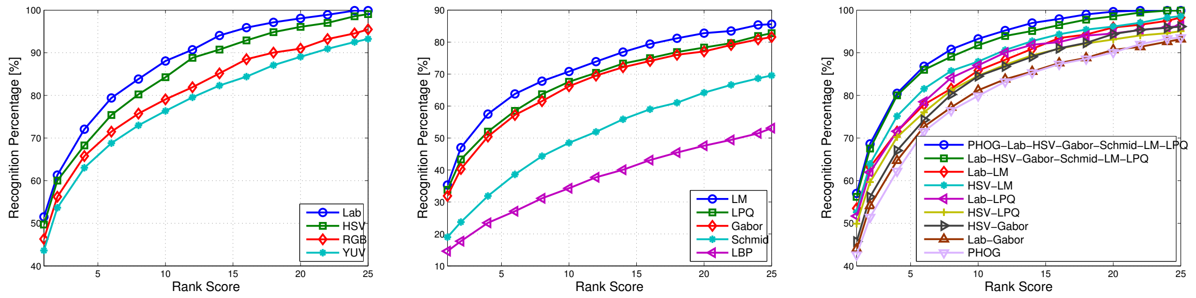
Figure 10: Performance of the proposed method varying train and test dataset sizes. Recognition performance for camera pair 1-2 on 3DPeS dataset, camera pair 3-8 and camera pair 5-8 on SAIVT dataset are shown in (a), (b) and (c), respectively.

decide which classifier is chosen. Results show that the recognition rate is constant when the value of Th_θ is below the value used for training. The recognition rate decreases when the value of Th_θ gets larger.

Parameter Th_ξ: Fig. 9(b) shows the performance of the proposed method as a function of the reliability threshold Th_ξ. We provide results using three training settings corresponding to $Th_\theta \in \{15, 30, 60\}$ degrees. Notice that the recognition rate increases when the value of Th_ξ also increases. The reason is that low values of $\xi_{I^B}^{I^A}$ provide inaccurate PFDs to the classifiers.

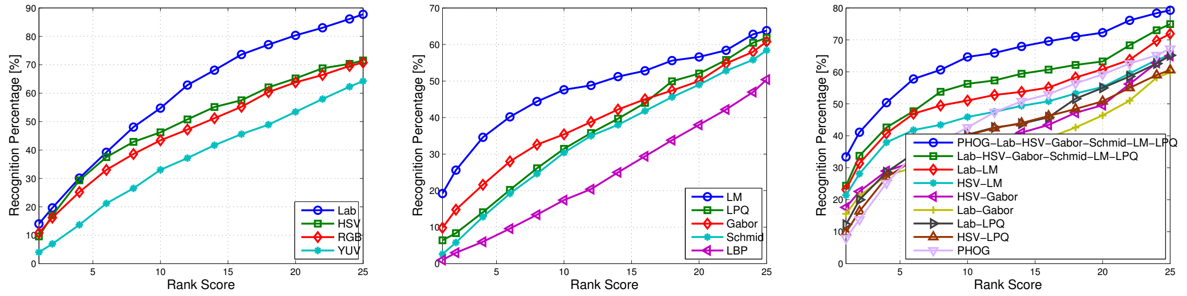
Train/Test Size: Figure 10(a), (b) and (c) show the performance of the proposed method as a function of the training/test set size for 3DPeS and SAIVT datasets. The results are obtained using $Th_\theta = 30$ degrees for the training, $Th_\xi = 0.7$ and a feature vector composed by PHOG, Lab and LM filter. The recognition percentage notably improves for all the three camera pairs when the number of persons used for training increases. We achieve a maximum recognition rate of 54.5% for rank 1 when our method is evaluated using camera pair 1-2 on 3DPeS dataset, 32% for camera pair 3-8 and 18% for camera 5-8, both on SAIVT dataset. Finally, curves on 3DPeS dataset are closer that curves on SAIVT dataset.

Features: All experiments are obtained using the training setting with $Th_\theta = 30$ degrees, $Th_\xi = 0.7$.



(a) (b) (c)

Figure 11: Performance of the proposed method on the 3DPeS dataset (camera pair 1-2) using different features. Comparison of color space features, texture features and mixed features are shown in (a), (b) and (c) respectively.



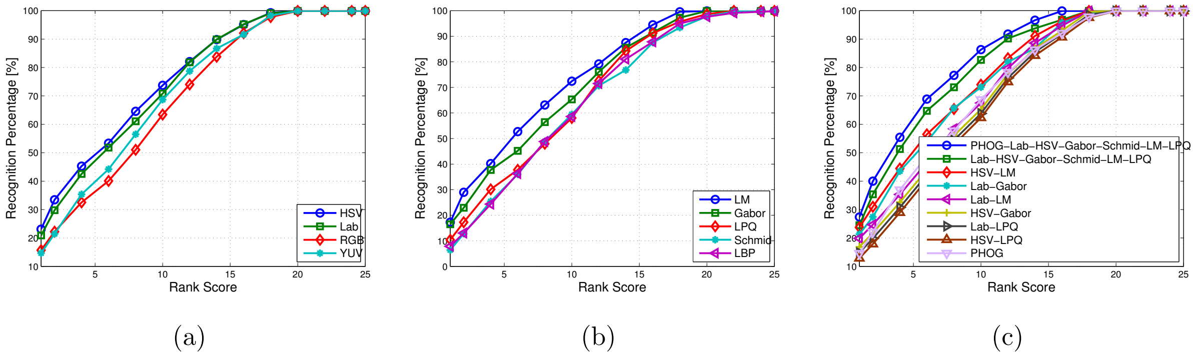
(a) (b) (c)

Figure 12: Performance of the proposed method on the SAIVT dataset (camera pair 3-8) using different features. Comparison of color space features, texture features and mixed features are shown in (a), (b) and (c) respectively.

In Fig. 11 we report the performance of our method on 3DPeS dataset for camera pair 1-2 using different features. We show the results for different color spaces in Fig. 11(a), texture features in (b) and combinations of them in (c). Lab and LM features obtain better performance that the other color and texture features by achieving a recognition rate of 51.5% and 35% for the rank 1, respectively. It is worth noticing that color features are more discriminative that texture features. However, the combination of all features is the one that achieves the best performance where the proposed method reaches a 57% for the rank 1.

Fig. 12(a), (b) and (c) show the performance of our method on SAIVT dataset for camera 3-8 using different color space features, texture features and combinations of them, respectively. Lab and LM features still obtain the best recognition percentages. A recognition rate of 14% and 19% for the rank 1 is achieved, respectively. Texture features obtain a recognition rate greater than color features, but the difference decreases respect to 3DPeS dataset. The combination of all features achieves 33.5% of recognition rate for rank 1.

Finally, Fig. 13(a), (b) and (c) show the performance of our method on SAIVT dataset for camera 5-8 using different color space features, texture features and combinations of them, respectively. In this case,



(a) (b) (c)

Figure 13: Performance of the proposed method on the SAIVT dataset (camera pair 5-8) using different features. Comparison of color space features, texture features and mixed features are shown in (a), (b) and (c) respectively.

Table 1: Top ranked matching rate (%) on 3DPeS dataset

Methods	rank:1	rank:5	rank:10	rank:20	rank:50
Proposed	**55.2**	**78.6**	**79.9**	**93.8**	**97.9**
LF	33.3	58.2	70.0	81.1	95.1
KISSME	22.9	49.0	62.2	76.0	93.2
LMNN-R	23.0	44.9	55.2	69.0	88.9

HSV and LM features achieve the best recognition percentages, which are of 21.5% and 19% for rank 1, respectively. The same features are less performing on camera pair 3-8 of the SAIVT dataset where, shape and texture features are more discriminative. For camera pair 5-8, all the features have similar performance, but the combination of all features achieves 28.5% of recognition rate for rank 1.

4.4. Comparisons with state-of-the-art methods

We compare our results with the ones of LF [29], KISSME [26] and LMNN-R [9] on the 3DPeS dataset. We follow the same experimental protocol of those and split the dataset into a training set and a test set, where each one is composed of 95 randomly selected persons.

Fig. 14 shows the performance of our method compared to LF, KISSME and LMNN-R. Our method outperforms the others especially for low ranks, which are the most representatives. We achieve 55.2% correct recognition rate at rank 1, while, for the same rank, a recognition percentage of 33.43%, 22.94% and 23.03% is achieved by LF, KISSME and LMNN-R respectively. In Table 1 recognition percentages for ranks 1, 10, 25 and 50 are reported. Our approach is the only one that achieves a recognition percentage higher than 90% for rank score 15.

We also compare our results on SAIVT dataset with those reported in [39] (i.e. Texture Model, Height Model, Culture-Colours Model, Colour-Soft Model and Fused Model) and RWACN proposed in [35]. We adopt the same protocol used in [39] and report the performance for two camera pairs, denoted as 3-8 and

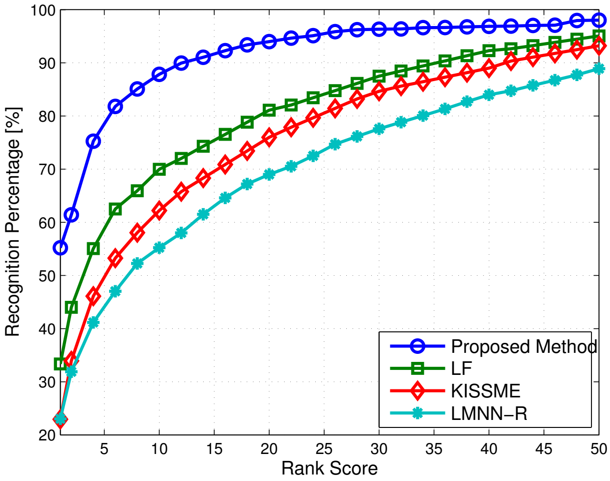


Figure 14: Comparisons of the proposed method with state-of-the-art methods on the 3DPeS dataset.

Table 2: Top ranked matching rate (%) on SAIVT dataset (Camera pair 3-8)

Methods	rank:1	rank:5	rank:10	rank:20	rank:50
Proposed	29.7	46.9	67.3	83.1	97.0
FM	**33.8**	**61.8**	**76.5**	**87.5**	**97.0**
CSM	16.2	39.7	52.2	72.7	95.6
CCM	17.6	51.5	69.8	80.1	94.8
RWACN	13.3	33.1	45.3	60.5	87.7
TM	7.4	27.2	43.4	58.8	94.8
HM	6.6	30.9	44.8	57.3	94.8

5-8. Camera pair 3-8 has 99 persons viewed by similar perspectives while camera pair 5-8 has 103 persons acquired at very different perspectives.

In Fig. 15(a) we report our results for camera pair 3-8. For such camera pair where the views are similar we achieve a performance between the Fused Model and Colour-Soft Model, outperforming all others. However, the difference with the Fused Model decreases along the rank score axis. Table 2 report recognition percentages for ranks 1, 10, 25 and 50. We achieve a recognition percentage of 29.7% for rank score 1, while recognition rate of 33.98% is achieved by the Fused Model.

In Fig. 15(b) we report our results for camera pair 5-8 where the viewpoint are dissimilars between cameras. We achieve a recognition rate of 26.5% for the rank 1, outperforming all other methods used for comparison. Our method achieves a recognition percentage higher than 90% for a rank score of 30. In Table 3 recognition percentages for ranks 1, 10, 25 and 50 are reported.

Table 3: Top ranked matching rate (%) on SAIVT dataset (Camera pair 3-8)

Methods	rank:1	rank:5	rank:10	rank:20	rank:50
Proposed	**26.5**	**45.1**	**62.0**	**79.8**	**97.8**
FM	18.5	33.6	50.0	67.8	93.5
CCM	9.29	26.4	46.4	60.7	95.7
HM	9.28	22.8	37.8	65.0	94.3
RWACN	6.24	18.8	29.8	46.7	81.1
CSM	5.0	15.7	30.7	58.5	90.0
TM	1.4	17.8	29.2	47.1	83.6

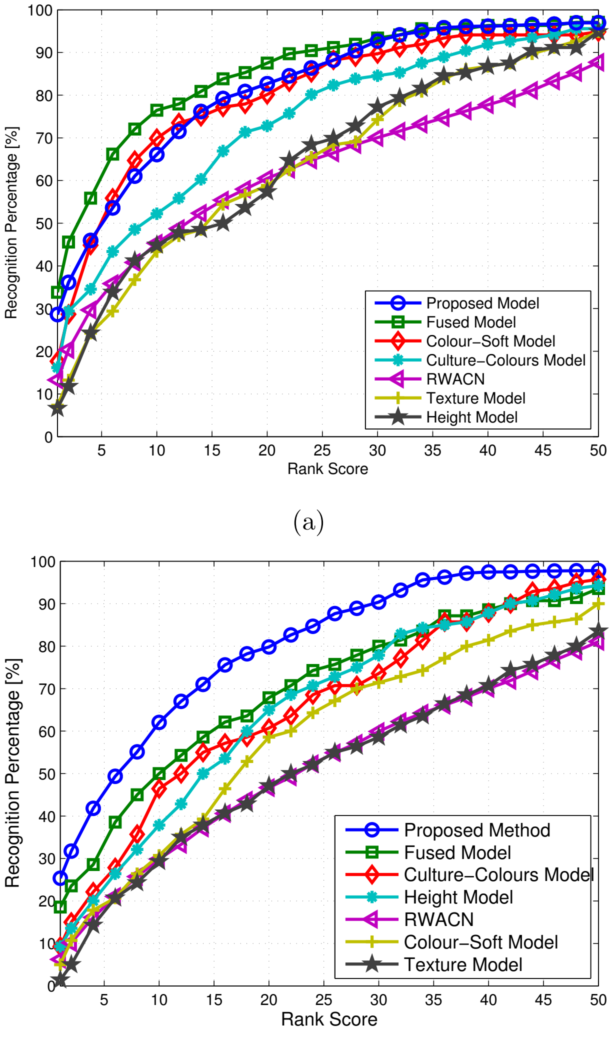


Figure 15: Comparisons of the proposed method with state-of-the-art methods on the SAIVT dataset: (a) Camera pair 3-8 and (b) Camera pair 5-8.

5. Conclusion

In this paper we have introduced a novel re-identification approach where different perspectives of the person are used to learn the transformations of pairwise feature dissimilarities across camera pairs. Towards this goal we have recovered different perspectives from short-term tracklets and proposed a person orientation value that allowed us to obtain a inter-relationship between perspectives computed for different cameras. We have also introduced a reliability value associated to each recovered perspective. This allows to model the accuracy of the orientation value. Pairwise feature dissimilarities have been used to learn the multi-modal inter-camera transformations. In particular the PFDS has been divided into two regions, one formed by PFDs computed for images with similar orientations, the other composed of PFDs computed for images with all other possible orientations. Then, for each subspace a binary classifier has been trained to discriminate between images of the same person and images from different persons. Extensive evaluations, analysis and comparisons with state-of-the-art methods have been conducted to show the superior performance of our method on two publicly available benchmark datasets.

References

[1] Y. Wu, M. Minoh, M. Mukunoki, W. Li, S. Lao, Collaborative Sparse Approximation for Multiple-Shot Across-Camera Person Re-identification, in: Advanced Video and Signal-Based Surveillance, Ieee, 2012, pp. 209–214. doi:10.1109/AVSS.2012.21.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6328018

[2] C. Liu, S. Gong, C. C. Loy, On-the-fly Feature Importance Mining for Person Re-Identification, Pattern Recognitiondoi:10.1016/j.patcog.2013.11.001.
URL http://linkinghub.elsevier.com/retrieve/pii/S0031320313004512

[3] I. Kviatkovsky, A. Adam, E. Rivlin, Color Invariants for Person Re-Identification, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (7) (2013) 1622–1634. doi:10.1109/TPAMI.2012.246.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6357194

[4] B. Ma, Y. Su, F. Jurie, Covariance Descriptor based on Bio-inspired Features for Person Re-identification and Face Verification, Image and Vision Computing 32 (2014) 379–390. doi:10.1016/j.imavis.2014.04.002.
URL http://linkinghub.elsevier.com/retrieve/pii/S0262885614000626

[5] O. Javed, K. Shafique, Z. Rasheed, M. Shah, Modeling inter-camera spacetime and appearance relationships for tracking across non-overlapping views, Computer Vision and Image Understanding 109 (2) (2008) 146–162. doi:10.1016/j.cviu.2007.01.003.
URL http://linkinghub.elsevier.com/retrieve/pii/S1077314207000100

[6] T. Avraham, I. Gurvich, M. Lindenbaum, S. Markovitch, Learning Implicit Transfer for Person Re-identification, in: European Conference on Computer Vision, Workshops and Demonstrations, Vol. 7583 of Lecture Notes in Computer Science, Florence, Italy, 2012, pp. 381–390.
URL http://www.springerlink.com/index/10.1007/978-3-642-33863-2

[7] A. Datta, L. M. Brown, R. Feris, S. Pankanti, Appearance Modeling for Person Re-Identification using Weighted Brightness Transfer Functions, in: International Conference on Pattern Recognition, no. Icpr, 2012.

[8] W. Li, X. Wang, Locally Aligned Feature Transforms across Views, in: International Conference on Computer Vision and Pattern Recognition, IEEE, 2013, pp. 3594–3601. doi:10.1109/CVPR.2013.461.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6619305

[9] M. Dikmen, E. Akbas, T. S. Huang, N. Ahuja, Pedestrian Recognition with a Learned Metric, in: Asian conference on Computer vision, 2010, pp. 501–512.

[10] M. Hirzer, P. M. Roth, K. Martin, H. Bischof, Relaxed Pairwise Learned Metric for Person Re-identification, in: European Conference Computer Vision, Vol. 7577 of Lecture Notes in Computer Science, 2012, pp. 780–793. doi:10.1007/978-3-642-33783-3.
URL http://www.springerlink.com/index/10.1007/978-3-642-33783-3

[11] M. Hirzer, C. Beleznai, M. Kostinger, P. M. Roth, H. Bischof, Dense appearance modeling and efficient learning of camera transitions for person re-identification, in: International Conference on Image Processing, IEEE, 2012, pp. 1617–1620. doi:10.1109/ICIP.2012.6467185.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6467185

[12] A. Mignon, F. Jurie, PCCA: A new approach for distance learning from sparse pairwise constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2666–2672. doi:10.1109/CVPR.2012.6247987.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6247987

[13] N. Martinel, C. Micheloni, C. Piciarelli, Learning pairwise feature dissimilarities for person re-identification, in: International Conference on Distributed Smart Cameras, IEEE, Palm Springs, CA, 2013, pp. 1–6. doi:10.1109/ICDSC.2013.6778209.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6778209

[14] A. Bedagkar-Gala, S. K. Shah, A Survey of Approaches and Trends in Person Re-identification, Image and Vision Computingdoi:10.1016/j.imavis.2014.02.001.
URL http://linkinghub.elsevier.com/retrieve/pii/S0262885614000262

[15] R. Vezzani, D. Baltieri, R. Cucchiara, People Re-identification in Surveillance and Forensics: a Survey, ACM Computing Surveys 46 (2). doi:10.1145/0000000.0000000.

[16] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, Shape and Appearance Context Modeling, in: International Conference on Computer Vision, Ieee, 2007, pp. 1–8. doi:10.1109/ICCV.2007.4409019.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4409019

[17] S. Bk, E. Corvée, F. Brémond, M. Thonnat, Boosted human re-identification using Riemannian manifolds, Image and Vision Computing 30 (6-7) (2012) 443–452. doi:10.1016/j.imavis.2011.08.008.
URL http://linkinghub.elsevier.com/retrieve/pii/S0262885611001065

[18] R. Zhao, W. Ouyang, X. Wang, Unsupervised Salience Learning for Person Re-identification, in: International Conference on Computer Vision and Pattern Recognition, 2013.

[19] B. Ma, Y. Su, F. Jurie, Local Descriptors Encoded by Fisher Vectors for Person Re-identification, in: European Conference on Computer Vision, Workshops and Demonstrations, Florence, Italy, 2012, pp. 413–422.

[20] Y. Xu, L. Lin, W.-S. Zheng, X. Liu, Human Re-identification by Matching Compositional Template with Cluster Sampling, in: International Conference on Computer Vision, no. 1, Ieee, 2013, pp. 3152–3159. doi:10.1109/ICCV.2013.391.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6751503

[21] L. An, M. Kafai, S. Yang, B. Bhanu, Reference-Based Person Re-Identification, in: Advanced Video and Signal-Based Surveillance, 2013.

[22] R. Zhao, W. Ouyang, X. Wang, Person Re-identification by Salience Matching, in: International Conference on Computer Vision, 2013.

[23] F. Porikli, M. Hill, Inter-Camera Color Calibration Using Cross-Correlation Model Function, in: IEEE International

Conference on Image Processing (ICIP), 2003, pp. 133–136.

[24] Z. Lin, L. S. Davis, Learning Pairwise Dissimilarity Profiles for Appearance Recognition in Visual Surveillance, in: International Symposium on Advances in Visual Computing, Vol. 5358 of Lecture Notes in Computer Science, Las Vegas, NV, 2008, pp. 23–34. doi:10.1007/978-3-540-89639-5.
URL http://www.springerlink.com/index/10.1007/978-3-540-89639-5

[25] W. Li, R. Zhao, X. Wang, Human Reidentification with Transferred Metric Learning, in: Asian Conference on Computer Vision, 2012, pp. 31–44.

[26] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: International Conference on Computer Vision and Pattern Recognition, no. Ldml, 2012, pp. 2288–2295. doi:10.1109/CVPR.2012.6247939.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6247939

[27] D. Tao, L. Jin, Y. Wang, Y. Yuan, X. Li, Person Re-Identification by Regularized Smoothing KISS Metric Learning, IEEE Transactions on Circuits and Systems for Video Technology 23 (10) (2013) 1675–1685. doi:10.1109/TCSVT.2013.2255413.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6490028

[28] W.-S. Zheng, S. Gong, T. Xiang, Re-identification by Relative Distance Comparison., IEEE transactions on pattern analysis and machine intelligence 35 (3) (2013) 653–668. doi:10.1109/TPAMI.2012.138.
URL http://www.ncbi.nlm.nih.gov/pubmed/22732661

[29] S. Pedagadi, J. Orwell, S. Velastin, Local Fisher Discriminant Analysis for Pedestrian Re-identification, in: International Conference on Computer Vision and Pattern Recognition, 2013, pp. 3318–3325. doi:10.1109/CVPR.2013.426.

[30] G. Zhang, Y. Wang, J. Kato, T. Marutani, M. Kenji, Local distance comparison for multiple-shot people re-identification, in: Asian conference on Computer Vision, Vol. 7726 of Lecture Notes in Computer Science, 2013, pp. 677–690. doi:10.1007/978-3-642-37431-9.
URL http://link.springer.com/10.1007/978-3-642-37431-9

[31] T. Zhou, M. Qi, J. Jiang, X. Wang, S. Hao, Y. Jin, Person Re-identification based on nonlinear ranking with difference vectors, Information Sciences (April). doi:10.1016/j.ins.2014.04.014.
URL http://linkinghub.elsevier.com/retrieve/pii/S0020025514004617

[32] A. J. Ma, P. C. Yuen, J. Li, Domain Transfer Support Vector Ranking for Person Re-identification without Target Camera Label Information, in: International Conference on Computer Vision, Ieee, 2013, pp. 3567–3574. doi:10.1109/ICCV.2013.443.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6751555

[33] C. Liu, C. C. Loy, S. Gong, G. Wang, POP: Person Re-Identification Post-Rank Optimisation, in: International Conference on Computer Vision, 2013.

[34] J. Garcia, A. Gardel, I. Bravo, J. L. Lazaro, M. Martinez, Tracking People Motion Based on Extended Condensation Algorithm, IEEE Transactions on Systems, Man, and Cybernetics: Systems 43 (3) (2013) 606–618. doi:10.1109/TSMCA.2012.2220540.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6425497

[35] N. Martinel, C. Micheloni, Re-identify people in wide area camera network, in: International Conference on Computer Vision and Pattern Recognition Workshops, IEEE, Providence, RI, 2012, pp. 31–36. doi:10.1109/CVPRW.2012.6239203.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6239203

[36] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987. doi:10.1109/TPAMI.2002.1017623.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1017623

[37] E. Rahtu, J. Heikkilä, V. Ojansivu, T. Ahonen, Local phase quantization for blur-insensitive image analysis, Image and Vision Computing 30 (8) (2012) 501–512. doi:10.1016/j.imavis.2012.04.001.
URL http://linkinghub.elsevier.com/retrieve/pii/S0262885612000510

[38] D. Baltieri, R. Vezzani, R. Cucchiara, 3DPeS: 3D People Dataset for Surveillance and Forensics, in: International ACM Workshop on Multimedia access to 3D Human Objects, 2011, pp. 59–64.

[39] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, P. Lucey, A Database for Person Re-Identification in Multi-Camera Surveillance Networks, in: International Conference on Digital Image Computing Techniques and Applications (DICTA), IEEE, 2012, pp. 1–8. doi:10.1109/DICTA.2012.6411689.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6411689