Exploiting Dissimilarity Representations for Person Re-Identification

Riccardo Satta, Giorgio Fumera, and Fabio Roli

Dept. of Electrical and Electronic Engineering, University of Cagliari
Piazza d'Armi, 09123 Cagliari, Italy
{riccardo.satta,fumera,roli}@diee.unica.it

Abstract. Person re-identification is the task of recognizing an individual that has already been observed over a network of video-surveillance cameras. Methods proposed in literature so far addressed this issue as a classical *matching* problem: a descriptor is built directly from the view of the person, and a similarity measure between descriptors is defined accordingly. In this work, we propose a general dissimilarity framework for person re-identification, aimed at transposing a generic method for person re-identification based to the commonly adopted multiple instance representation, into a dissimilarity form. Individuals are thus represented by means of dissimilarity values, in respect to common prototypes. Dissimilarity representations carry appealing advantages, in particular the compactness of the resulting descriptor, and the extremely low time required to match two descriptors. Moreover, a dissimilarity representation enables various new applications, some of which are depicted in the paper. An experimental evaluation of the proposed framework applied to an existing method is provided, which clearly shows the advantages of dissimilarity representations in the context of person re-identification.

Keywords: person re-identification, dissimilarity representation, multiple instance

1 Introduction

In video-surveillance, it is often desirable to recognize a person who has already been observed over a network of camera sensors. Such task, commonly referred to as "person re-identification", is useful for a number of practical security applications, both online (i.e. tracking a person over different, non-overlapping cameras) and offline (i.e. retrieval of all the video sequences which contain an individual of interest given as query).

Typically, the low resolution of the frames taken by the sensors of the network, and the variety of possible poses, makes face recognition techniques ineffective (see Fig. 1). A common approach is thus to look at the global appearance of the individual, building a descriptor that represents the whole body.

Person re-identification has been modeled so far as a classical *matching* problem: a descriptor is built directly from the blob containing the person, and some

Fig. 1. Example of image pairs representing the same individual taken from two different non-overlapping views, extracted from the ViPER benchmark dataset [7].

distance measure between descriptors is defined accordingly. The problem of how to build a suitable descriptor has been addressed in various ways. In fact, there is not an agreement on what features provide the best discriminant capabilities. Many of the existing methods, however, are based on the common idea of representing the human body as a *bag of instances*, defined as a set of randomly taken image patches or strips, or a set of interest points [14].

Regardless of the chosen features, often the descriptors of different people share a lot of redundant information. Their images can indeed contain similar instances, tipically associated to similar characteristics of their clothes (see Fig. 2). Our intuition is based on the above premise; instead of creating the descriptor of a person directly from its image, we propose to represent an individual by means of a vector of dissimilarity values between the bag of instances drawn from its image, and a number of pre-defined bags of instances named *visual prototypes*, each corresponding to some specific "visual" characteristics obtained from a given set of template users.

Dissimilarity-based representations for pattern recognition is a recently introduced and very promising research field [11]. In the context of person re-identification, a dissimilarity representation carries appealing advantages. In particular, in terms of the compactness of the descriptor, and of the computational requirements of the matching phase, which can be implemented as a comparison between vectors. We point out that, to the best of our knowledge, this work is the first attempt to exploit a dissimilarity representation in a *matching* task, in which only one (or a few) example per class is given, that is the case of person re-identification. The adopted representation is somewhat similar to that used in the so-called "visual words" methods, largely used in scene categorization (see for instance [17]). In visual words methods, a visual codebook is built offline, and then every sample is described in terms of the frequency (count of the occurrences) of every visual word. However, differently from visual words approaches, in the dissimilarity paradigm the *whole* sample is compared with every prototype, while in visual words approaches one looks for all the occurrences of every visual word *inside* the sample. Moreover, in a visual words method, for each visual concept the occurrences are simply counted, without considering the *degree*

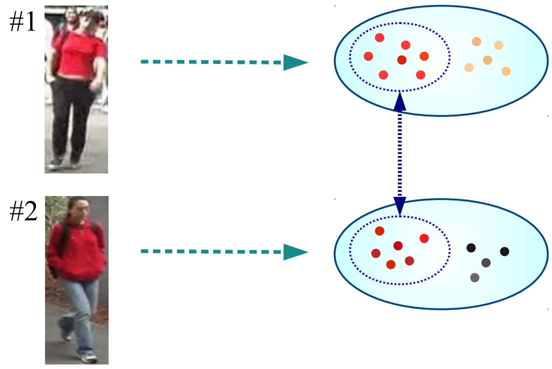

Fig. 2. An example of two pedestrians sharing clothing characteristics. Some of the instances of pedestrian #1 are similar to some of the instances extracted from pedestrian #2. Instances are represented by coloured dots. Here, only the upper body part is considered.

of presence, represented instead by a dissimilarity value. Note that a similar way to consider prototypes has been exploited in [3] for the specific task of image classification.

A dissimilarity representation also enables several new applications. An interesting one is *people grouping*, i.e. clustering individuals in the dissimilarity space so that each cluster contains only people of similar appearance, or that share the same visual characteristics. People grouping can be useful to reduce the number of candidates to be matched against a specific query, thus greatly lowering computational requirements when the number of individuals is huge, and to automatically group people in a scene, whose "role" can be inferred from their appearance (i.e. policemen, members of a sport team).

Moreover, representing individuals with vectors allows one to easily switch from a matching to a *learning* paradigm, where a classifier can be learned from a set of vectors of the same individual, for example representing different view points and poses, or of a group of individuals which share some common characteristic. A classifier is potentially able to generalize an appearance model of the individual (or group), and may represent an effective way to accumulate views taken by different frames, instead of keeping in memory all the feature vectors representing the same individual and matching every query against all of them.

The aim of this work is to provide a general dissimilarity framework for person re-identification, which we named "Multiple Component Dissimilarity" (MCD). This framework builds upon a recently proposed framework for person re-identification methods, the Multiple Component Matching (MCM) framework [14], which embeds the concept of multiple instances representation. MCM is able to frame, partially or completely, the great part of the existing methods. We will show how a generic method that can be framed in MCM can be turned into a dissimilarity-based form. We will also apply our MCD framework to an existing person re-identification method, and provide a preliminary experimental evaluation.

The paper is organized as follows. In Sect. 2 we briefly survey previous works on person re-identification, and provide details on the Multiple Component Matching framework. Then, in Sect. 3 the proposed dissimilarity framework is presented. We apply the proposed framework to an existing person re-identification method in Sect. 4 and provide an experimental evaluation. Finally, in Sect. 5 we sum up the proposed work and provide future research directions.

2 Background

In this Section, first an overview of the approaches to person re-identification available in literature is provided, then we describe the Multiple Component Matching framework for person re-identification.

2.1 Previous works on person re-identification

As mentioned in Sect. 1, person re-identification has been considered in literature as a matching problem, where the task consists in associating an individual from a probe gallery to the corresponding identity in a template gallery.

In [5], the human body is subdivided with respect to its symmetry properties: anti-symmetry separates torso and legs, while symmetry is divides left and right parts. The descriptor is made up of three local features: colour histograms, *maximally stable colour regions* (MSCR) and *recurrent high-structured patches* (RHSP), all extracted from torso and legs separately. To obtain MSCR and RHSP, several patches are sampled at random, mainly near symmetry axes; then, clustering algorithms are used to find the most significant ones. The matching distance is a combination of the distances computed on the individual features.

In [2], an human body parts detector is used to find in the body of each individual fifteen non-overlapping square cells, that have proven to be "stable regions" of the silhouette. For each cell a covariance descriptor based on colour gradients is computed. Descriptor generation and matching is performed through a pyramid matching kernel.

In [1] two methods were proposed. In the first, Haar-like features are extracted from the whole body, while in the second the body is divided into upper and lower part, each described by the MPEG7 Dominant Colour descriptor.

An approach based on harvesting SIFT-like interest points from different frames of a video sequence is described in [9]. Different frames are used also in [6], where two methods are proposed. The first one is based on interest points. The second one exploits a part subdivision of the human body based on decomposable triangulated graphs and dynamic programming to find the optimal deformation of this model for the different individuals.

In [8] the problem of defining the best descriptor for person re-identification is addressed. Different features are extracted, and their weights are computed by a boosting algorithm. Features are computed from randomly taken strips.

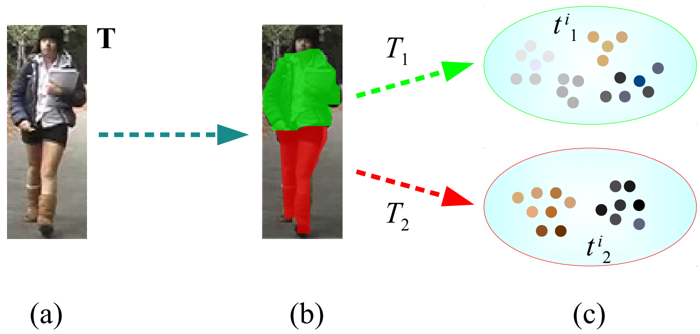

Fig. 3. An example of the MCM representation. Considering the individual in (a), a toy subdivision in two parts, upper-body (in green) and lower-body (in red), is applied (b). Every part is composed by several instances, or components (c), here represented by coloured dots.

The approach proposed in [15] is based of global color descriptors (histograms, spatiograms, color/path-length) computed from the whole blob containing the person. A graph-based method is then used to reduce the dimensionality of the descriptors.

In [13] person re-identification is considered as a relative ranking problem, exploiting a discriminative subspace built by means of an Ensemble RankSVM. Colour and texture-based features are extracted from six fixed horizontal regions.

Despite the methods summarised above exhibit many differences, it can be noted that many of them are based on a *multiple instance* representation, by taking several patches, strips, interest points. In addition, most works exploit some *part-based* model of the body, which is divided accordingly into regions/parts. These two concepts, multiple instance representation and part subdivision, provide the foundation for the Multiple Component Matching Framework [14], which is depicted in the following subsection.

In [14] the authors also proposed a direct implementation of their framework, where a two-part subdivision is adopted (torso-legs) and each part is described by a set of random and partly overlapping patches. Each patch is represented by its colour histogram.

2.2 The Multiple Component Matching framework

In this section we describe the Multiple Component Matching (MCM) framework for person re-identification. This framework has been presented in [14], and aims to provide a common foundation for existing and future methods for person re-identification. It is able to provide an unique view for the great part of the methods proposed so far in literature. Therefore, we have chosen to adopt MCM as the underlying paradigm for our proposed dissimilarity framework.

MCM is based on concepts that have found to underly most previous works, namely multiple instance representation, and part subdivision. The individual is represented by means of bags of instances, or "set of components" in MCM terminology. Such components can be any kind of local features: patches, strips, interest points. To take into account the peculiarities of the human body, MCM also embeds the concept of part subdivision. For each part, a different set of components is considered.

Formally, let $\mathcal{T} = \{\mathbf{T}_1, \ldots, \mathbf{T}_N\}$ be the *template gallery*, each corresponding to an individual. Every template $\mathbf{T}_i$ is represented as an ordered sequence of a predefined number of M sets, corresponding to the M parts into which an image is subdivided:

$$\mathbf{T}_i = \{T_{i,1}, \ldots, T_{i,M}\} \tag{1}$$

Following a multiple-instance representation, every part $T_{i,j}$ is a set of an arbitrary number $n_{i,j}$ of components (a simple example is depicted in Fig. 3), and is described by the corresponding feature vectors $\mathbf{t}^k_{i,j}$:

$$T_{i,j} = \{\mathbf{t}^1_{i,j}, \ldots, \mathbf{t}^{n_{i,j}}_{i,j}\}, \mathbf{t}^k_{i,j} \in \mathbb{X}, \tag{2}$$

where $\mathbb{X}$ denotes the feature space (assumed the same for all sets, for the sake of simplicity, and without losing generality). Given a probe $\mathbf{Q}$, which is represented as a sequence of parts as described above, the task of MCM is to find the most similar template $\mathbf{T}^* \in \mathcal{T}$, with respect to a similarity measure $D(\cdot, \cdot)$:

$$\mathbf{T}^* = \arg\min_{\mathbf{T}_i} D(\mathbf{T}_i, \mathbf{Q}). \tag{3}$$

The similarity measure D between sequences is defined as a combination of similarity measures $d(\cdot, \cdot)$ between sets:

$$D(\mathbf{T}_i, \mathbf{Q}) = f\Big(d(T_{i,1}, Q_1), \ldots, d(T_{i,M}, Q_M)\Big). \tag{4}$$

D can be any combination of the set distances, like a weighted average in which the coefficients reflect the relevance of the corresponding regions. Concerning the similarity measure d, it can be any distance measure betwen sets. A possible measure is the *k-th Hausdorff Distance* proposed by Wang and Zucker [16], which has been used in [14]. It is defined as the k-th ranked distance of the minimum distances between each element of one set and each element of the other. Comparing two sets $X = \{x_i\}$ and $Y = \{y_i\}$, we have

$$d_H(X, Y) = \max\big(h_k(X, Y), h_k(Y, X)\big) \tag{5}$$

where

$$h_k(X, Y) = \underset{x \in X}{k\text{-th}} \min_{y \in Y}(\|x - y\|) \tag{6}$$

Note that another metric has to be defined, namely the distance measure $\|x-y\|$ between the components of the sets.

Conveniently choosing the parameters of the MCM framework (part subdivision adopted, components extracted and corresponding representation, and the distance measures d and D), different specific implementations can be obtained. In particular, many of the existing methods for person re-identification can be described, fully or partially, by means of this framework.

3 The Multiple Component Dissimilarity framework for person re-identification

Here we illustrate the proposed Multiple Component Dissimilarity (MCD) framework for person re-identification. This framework builds upon the MCM framework described above, and aims at defining a dissimilarity-based version of a generic method for person re-identification which can be framed into MCM.

Consider a generic target MCM method which adopts a multiple instances representation and (possibly, but not necessary) a part subdivision, and assume that a template gallery $\mathcal{T} = \{\mathbf{T}_1, \ldots, \mathbf{T}_N\}$ is given. A probe individual $\mathbf{Q}$, which can be any element of the probe gallery, is given as well. As in MCM, the task is to find the most similar template to $\mathbf{Q}$. The proposed MCD framework requires four steps:

1. define a set of prototypes for each body part;
2. represent each element of $\mathcal{T}$ via dissimilarity vectors, one for each part;
3. represent $\mathbf{Q}$ via dissimilarity vectors, one for each part;
4. find the element of $\mathcal{T}$ which is most similar to $\mathbf{Q}$ in the dissimilarity space.

The first three steps are aimed at transposing the original problem into a dissimilarity space, while the fourth step corresponds to Eq. 3 in MCM, where this time we compare dissimilarity vectors.

Step one is to define a distinct set of *visual prototypes* for each body part. These prototypes will be used to build a dissimilarity vector for each part of each element of $\mathcal{T}$, and of $\mathbf{Q}$. The prototypes are extracted from the template gallery $\mathcal{T}$.

In MCM, each individual is represented as a set of components for each of its parts. We chose to represent each visual prototype as a set of components as well. Accordingly, the dissimilarity between a visual prototype and an individual can be computed by means of the same distance measure d between sets of components adopted by the target method (Eq. 4). This allows one to easily and directly define a dissimilarity version of any method framed in MCM, without the need of defining a new dissimilarity measure between descriptors and prototypes.

Considering the m-th body part, the procedure for defining the corresponding prototypes is the following. First, all the components belonging to the m-th part of every element of $\mathcal{T}$ are put together forming a single set of components. Then, a clustering algorithm is applied to this set; prototypes will be defined as the clusters found.

Any clustering method can be adopted, for example the well known K-Means algorithm. To reduce computational and memory requirements, it may be preferable to have prototypes made up by a reduced number of components. Thus, one can also define a two-stage clustering procedure: first, the components belonging to each individual are separately clustered; then, a second clustering is carried out on the centroids obtained at the first-stage. Note that many other algorithms to find out prototypes have been proposed in literature (see for example [12]).

This procedure ends up with a prototype gallery $\mathcal{P}$, made by M sets of prototypes, one set for each body part:

$$\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_M\} \tag{7}$$

with the m-th set of prototypes having a cardinality $N_{P,m}$

$$\mathbf{P_m} = \{P_{m,1}, \ldots, P_{m,N_{P,m}}\} \tag{8}$$

It turns out that the parameters of the clustering algorithm, which govern the number of prototypes $N_{P,m}$ for each part, are important, but not crucial: as will be shown in Sect. 4, performance does not vary drastically in respect to $N_{P,m}$.

Fig. 4 sums up the process of prototypes generation in a case where the number of parts is two.

Once prototypes have been defined, we can build a dissimilarity representation of each element of $\mathcal{T}$, and of $\mathbf{Q}$. Such dissimilarity representation is made up of a different dissimilarity vector for each part. More formally, given an individual $\mathbf{I}$ composed by m parts $I_1, \ldots, I_m$, the dissimilarity representation is the following:

$$\mathbf{I^{Dis}} = \{I_1^{Dis}, \ldots, I_m^{Dis}\} \tag{9}$$

where each I_i^{Dis} is a vector of dissimilarity measures corresponding to the i-th part:

$$I_i^{Dis} = \left[d(I_i, P_{i,1}) \ldots d(I_i, P_{i,N_{P,i}})\right] \tag{10}$$

By means of Eq. 9 and Eq. 10, all the elements $\mathbf{T}_i$ of the template gallery $\mathcal{T}$ can be described via their dissimilarity representation $\mathbf{T^{Dis}}_i$.

Once the data has been transposed into a dissimilarity space, the problem of finding the best match in the template gallery given a query $\mathbf{Q}$ can be addressed similarly to Eq. 3 of MCM:

$$\mathbf{T^{Dis}}^* = \arg\min_{\mathbf{T^{Dis}}_i} D(\mathbf{T^{Dis}}_i, \mathbf{Q}), \tag{11}$$

where the superscript **Dis** indicates a dissimilarity representation. D can be the same fusion rule of Eq. 4, this time applied to distance measures d_{Dis} between dissimilarity vectors. Considering a generic dissimilarity template $\mathbf{T^{Dis}}$ and a probe $\mathbf{Q^{Dis}}$, we have therefore:

$$D(\mathbf{T^{Dis}}, \mathbf{Q^{Dis}}) = f\Big(d_{Dis}(T_1^{Dis}, Q_1), \ldots, d_{Dis}(T_M^{Dis}, Q_M)\Big). \tag{12}$$

The distance measure d_{Dis}, can be defined as any distance measure between vectors, for example the euclidean distance.

The proposed dissimilarity representation exhibits clear advantages. First, in place of a complex descriptor, for each individual we have a set of a limited number of dissimilarity vectors, one for each part of the body, thus saving a great amount of memory for descriptors storage. Note that also the prototypes need to be stored, however the number of their elements can be conveniently reduced, for example by adopting a two-stage clustering scheme as explained previously. Furthermore, the matching becomes as simple as computing a distance between vectors, which is almost an immediate operation with modern CPUs. Such extremely fast matching can lead to several useful applications, like finding the identity of an individual among a huge number of candidates, almost in real-time.

The MCD framework we proposed can be used to define a dissimilarity version of any method which can be framed in MCM. In particular, in the following Section we apply MCD to the implementation of MCM proposed in [14].

4 Application of MCD

In this section, we provide a preliminary analysis of the application of MCD to an existing person re-identification method. We have chosen MCMimpl, a direct implementation of MCM proposed in [14] which has shown to attain state-of-the-art performance.

In MCMimpl, first the mask which separates the individual from the background is obtained by a STEL generative model [10]. The body is then divided into two parts, torso and legs, exploiting the anti-symmetry properties of the human silhouette. From each part, random and partly overlapping patches are extracted and described via a colour histogram in the HSV colour space. The distance between two sets corresponding to the same part is evaluated by the k-th Hausdorff Distance (which has been introduced in Sect. 2.2), while the final matching distance is the average of the distances of the parts.

To apply MCD, first a proper clustering algorithm to find the prototypes must be chosen. We adopted a two-stage clustering scheme, where at first patches belonging to every template are clustered with the Mean-Shift clustering algorithm [4], which does not make any assumption on the shape of the distribution nor the number of clusters. The only parameter of Mean-Shift is the bandwidth BW, which governs how spread is each cluster. The resulting centroids (actually, the real patch nearest to each centroid) are put together and clustered again, this time via the classical K-Means method. Here, the only parameter is the number of clusters K. We have chosen to adopt K-Means for the second clustering stage, since applying Mean-Shift resulted in too unbalanced clusters (many of which composed by only 1 or 2 elements). Instead, Mean-Shift has proven to be more effective in clustering the patches of a single individual.

Fig. 5 shows the result of applying this clustering algorithm to patches extracted accordingly to $MCMimpl$. A set of 10 individuals is considered, taken from the ViPER dataset [7]. Note that some prototypes look quite similar; how-

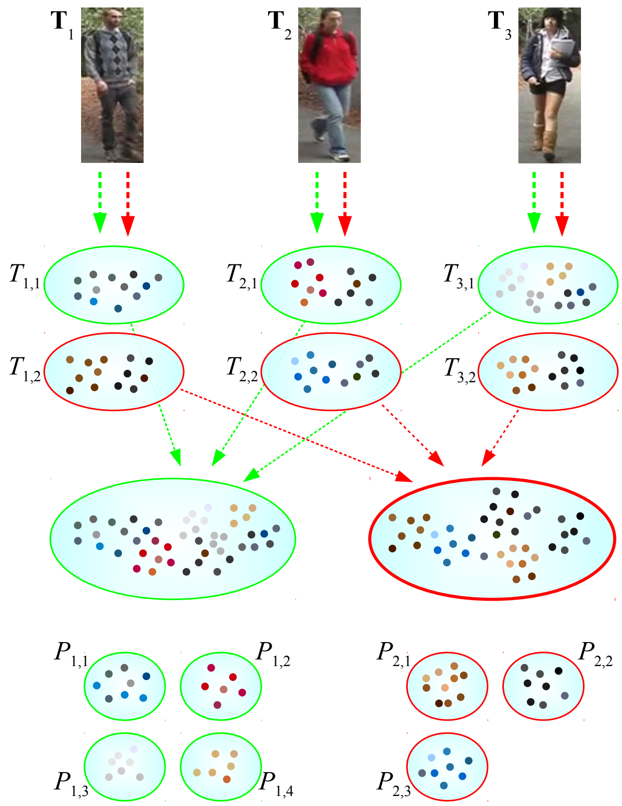

Fig. 4. Generation of the prototype gallery in MCD. Considering a template gallery of three individuals, represented as a set of components for each part according to MCM (a), all the components corresponding of each part are put together (b), then a clustering operator is applied and a number of prototypes is generated for each part (c). In this example, two parts are considered, upper (in green) and lower (in red) body.

ever, all the different visual characteristics are reasonably well captured in distinct prototypes.

Concerning the d_{Dis} distance measure (Eq. 12) between dissimilarity vectors, we adopt the euclidean distance. Finally, the overall matching distance (D in Eq. 12) is the average of the distances of the single parts.

Preliminary evaluation

A preliminary experimental evaluation of the dissimilarity version of the target method, $MCMimpl$, is provided in the following. The dissimilarity version is denoted as $MCMimpl^{Dis}$.

	$MCMimpl$	$MCMimpl^{Dis}$
Size of the descriptor	$96KB$	$640B$
Average matching time	$28.6ms$	$< 0.01ms$

Table 1. Short comparison between $MCMimpl$ and its dissimilarity version $MCMimpl^{Dis}$. The size of the descriptor is computed considering 32 bit floats values, and for $MCMimpl^{Dis}$ is referred to a number of prototypes of 80 for both torso and legs. Matching time is evaluated on a 2.4 GHz CPU, and refers for both methods to a non-optimized C++ implementation.

We changed the parameters of $MCMimpl$ originally used in [14], reducing the size of the patches and increasing their number, thus obtaining an higher granularity, that we have found to be more effective in capturing the visual characteristics. We extracted 300 random rectangular patches from each part, whose width and height are in the range $[8\%, 12\%]$ of the width and the height of the part.

The bandwidth parameter of Mean-Shift clustering was set to $BW = 0.3$ for all the experiments.

In Table 1, a short comparison between the original method and its dissimilarity version is provided. In particular, we reported the size of the descriptor and the average time required for matching.

As can be seen, the size of the descriptor for $MCMimpl^{Dis}$ is reduced by two orders of magnitude: the original descriptor, in fact, is made up of 300 different local patches for each part (torso and legs), every patch being represented by a vector of 40 features (see [14] for further details). The dissimilarity descriptor, instead, is composed by a vector of N_P elements for each part, N_P being the number of prototypes (assumed the same for all the parts).

The matching time has been evaluated as the average of 6300 single comparisons, and, as can be seen, it is also greatly reduced, being almost immediate, and leads to a matching rate of over 10^5 candidates *per second*.

We evaluated also the matching performance of $MCMimpl^{Dis}$. Given a template gallery and a probe gallery, a common way to assess the performance of a person re-identification method is the Cumulative Matching Characteristics curve, that is, the average probability of finding the correct match of the elements of the probe gallery, in the template gallery. Here, we build both the template and the probe gallery from a sub-set of the ViPER benchmark dataset [7], made up of the first 126 pedestrian. In this dataset, for every person two non-overlapping views are available. The template gallery is made up of the first view of each pedestrian, while the probe gallery is built by each second view.

In Fig. 6 we report the average CMC curve over 10 different folds of 63 pedestrians. The CMC curve of the original method $MCMimpl$ is also plotted in blue, as reference.

Performance vary in respect to the number of prototypes N_P, which in these experiments is the same for all the body parts. In Fig. 7 the performance versus N_P is evaluated by means of the area of the first 20% of the CMC curve (denoted

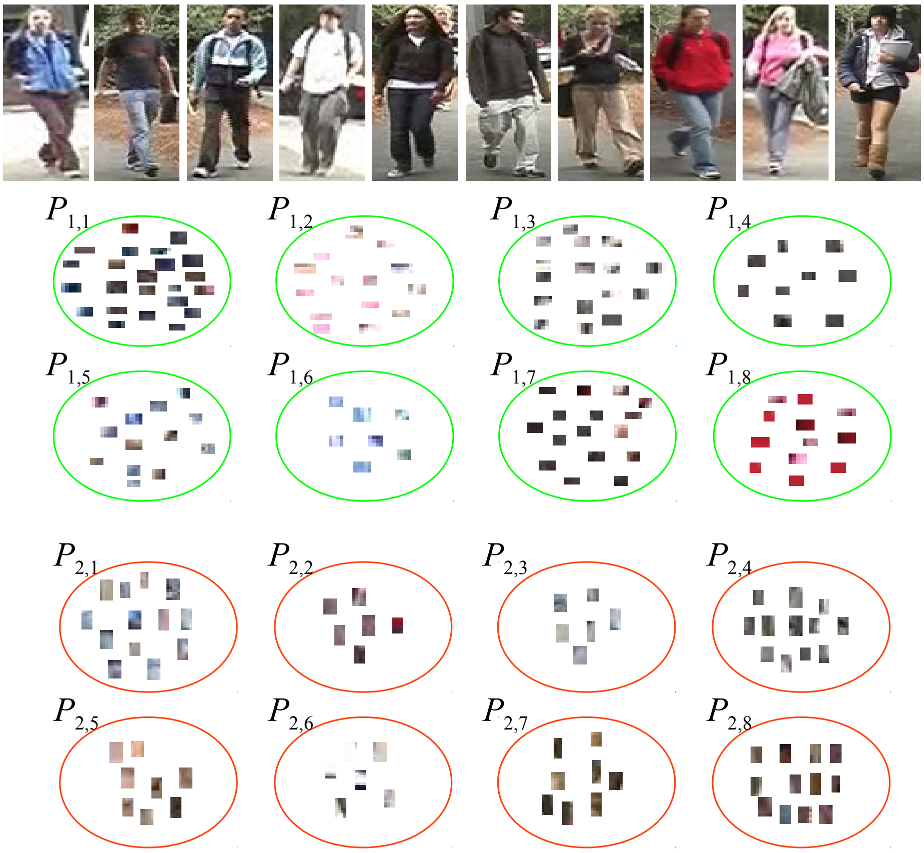

Fig. 5. Patch clustering results, for a set of 10 individuals. In green, prototypes related to the *torso* body part; in red, prototypes related to the *legs* body part. The number of prototypes is set to 8 for both the parts.

as $AUC_{20\%}$). We chose to consider the first part of the curve only, since in real application scenarios the interest is usually on the first ranks. The plot of Fig. 6 corresponds to a $N_P = 80$.

The proposed framework is aimed at taking advantages related to the compactness of the dissimilarity representation, rather than incrementing the pure matching performance. We point out that such advantages do not depend to the specific target method considered. However, note that performance attained by the dissimilarity version are comparable to that of the original method. Furthermore, the dissimilarity version slightly outperforms the original method in the first part of the curve, which as stated previously is usually the most interesting.

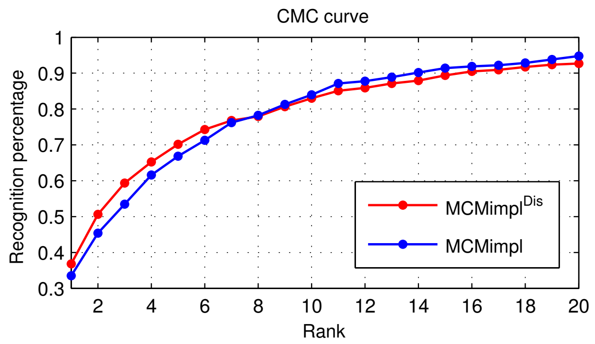

Fig. 6. Average Cumulative Matching Characteristics curve over 10 runs on a sub-set of the ViPER dataset. In blue, performance attained by the reference method $MCMimpl$; in red, performance attained by its dissimilarity version $MCMimpl^{Dis}$.

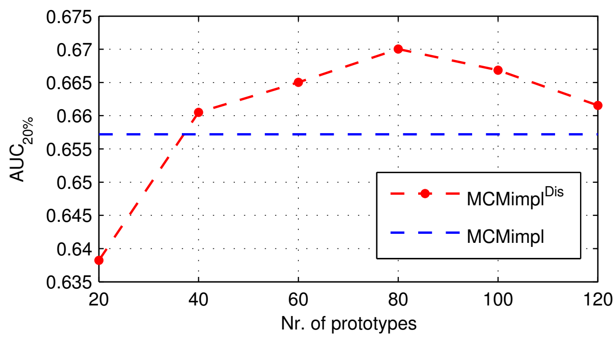

Fig. 7. $AUC_{20\%}$ attained by $MCMimpl^{Dis}$ in respect to the number of prototypes (in red). The blue line depicted as reference is the $AUC_{20\%}$ of $MCMimpl$.

5 Conclusions and future work

In this paper, we proposed a framework, named "Multiple Component Dissimilarity" (MCD), aimed at transposing a generic method for person re-identification to a dissimilarity-based form. MCD is completely general, and does not impose constraints on the specific features used by the target method considered. It only requires that the target method exploits a multiple instances representation.

Dissimilarity representations carry interesting benefits to the problem of person re-identification. The first one is the compactness of the resulting dissimilarity descriptors; regardless of the complexity of the local features adopted by the target method, the dissimilarity descriptor will be as compact as a vector of dissimilarities. The second advantage is then obvious, as once samples are described in such form, a comparison between descriptors is almost immediate, being the computation of differences between vectors extremely cheap in terms of computational requirements.

The proposed MCD framework has been applied to an existing person re-identification method, and an experimental evaluation in respect to this method has been provided. Future studies shall include a more comprehensive analysis which consider different person re-identification approaches. Methods depicted in [5, 9, 6] are good candidates to apply MCD.

A dissimilarity representation can be exploited to enable several interesting applications. Here, we briefly describe some of them.

The first possible application is *people grouping*. Once we have a set of individuals described by dissimilarity vectors, we can cluster them in the dissimilarity space, so that we obtain clusters of people sharing a similar appearance. Since it is reasonable that every individual shares different characteristics with different groups of people, a "fuzzy" of "soft" clustering should be adopted, which does not hardly assign every individual to a single cluster.

People grouping can be useful in a number of tasks. For example, it can be used as a preprocessing phase to reduce the number the candidates prior to perform matching: we can first find clusters that the query is more likely to belong to, then perform matching only against templates belonging to these clusters. This can lead to a great reduction of computational requirements when the cardinality of the template gallery is huge. Note that only the first phase (grouping) exploits dissimilarity representations, while the second phase (matching inside a single group) can be run using any person re-identification method. We can also use people grouping to perform tasks that are not strictly related to the classic person re-identification problem. For example, we can exploit it to find people that share similar appearance in a scene. Individuals whose role can be assigned in respect to their appearance (for instance, policemen, vigilantes, firemen) can be therefore grouped automatically.

Another application that a dissimilarity representation can enable, is *appearance learning*, i.e. learn the appearance of an individual from a series of dissimilarity vectors. A great practical problem in person re-identification is how to accumulate different frames of the same person in a single descriptor. Most of the techniques proposed so far deal with only one template image per person, while the few methods that consider different images adopt approaches that vary from harvesting all the information obtained from all the frames, to clustering techniques aimed at reducing the number of local features that build the final descriptor. A classifier could be a great way to build a descriptor of an individual starting from a series of frames. In fact, each frame can be described as a dissimilarity vector, and these vectors can form a training set. Then, we can train a one-class or an one-versus-all classifier to learn the appearance of the individual.

The appearance of people that show similar visual characteristics (for example policemen, firemen, sport teams) can also be learned. Furthermore, appearance learning could be applied in scenarios not related to security and surveillance, for example to recognize different traditional dressings in cultural heritage applications.

References

1. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using haar-based and dcd-based signature. In: AVSS. pp. 1–8 (2010)
2. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using spatial covariance regions of human body parts. In: AVSS (2010)
3. Carli, A., Castellani, U., Bicego, M., Murino, V.: Dissimilarity-based representation for local parts. In: Proceedings of the 2nd IEEE International Workshop on Cognitive Information Processing (CIP) (2010)
4. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24, 603–619 (2002)
5. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: CVPR (2010)
6. Gheissari, N., Sebastian, T.B., Hartley, R.: Person reidentification using spatiotemporal appearance. In: CVPR (2006)
7. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: PETS (2007)
8. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: ECCV. pp. 262–275 (2008)
9. Hamdoun, O., Moutarde, F., Stanciulescu, B., Steux, B.: Interest points harvesting in video sequences for efficient person identification. In: VS (2008)
10. Jojic, N., Perina, A., Cristani, M., Murino, V., Frey, B.: Stel component analysis: Modeling spatial correlations in image class structure. CVPR pp. 2044–2051 (2009)
11. Pekalska, E., Duin, R.P.W.: The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence). World Scientific Publishing Co., Inc., River Edge, NJ, USA (2005)
12. Pekalska, E., Duin, R.P.W., Paclik, P.: Prototype selection for dissimilarity-based classifiers. Pattern Recognition 39 (February 2006)
13. Prosser, B., Zheng, W., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: BMVA. pp. 21.1–21.11 (2010)
14. Satta, R., Fumera, G., Roli, F., Cristani, M., Murino, V.: A multiple component matching framework for person re-identification. ICIAP (2011, in press), available at arXiv:1105.2491 (http://arxiv.org/abs/1105.2491).
15. Truong Cong, D.N., Achard, C., Khoudour, L., Douadi, L.: Video sequences association for people re-identification across multiple non-overlapping cameras. In: ICIAP (2009)
16. Wang, J., Zucker, J.D.: Solving the multiple-instance problem: A lazy learning approach. In: ICML (2000)
17. Yang, J., Jiang, Y.G., Hauptmann, A.G., Ngo, C.W.: Evaluating bag-of-visual-words representations in scene classification. In: Proceedings of the international workshop on Workshop on multimedia information retrieval. pp. 197–206 (2007)