Appearance-based People Recognition by Local Dissimilarity Representations

Riccardo Satta, Giorgio Fumera, 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

Among the possible applications of computer vision to video-surveillance, person re-identification over a network of camera sensors, using cues related to clothing appearance, is gaining much interest. Re-identification techniques can be used for various tasks, e.g., online tracking of a person, and off-line retrieval of all video sequences containing an individual of interest, whose image is given as a query. Recently, some authors proposed to exploit clothing appearance descriptors also to retrieve video sequences of individuals that match a *textual* description of clothing (e.g., "person wearing a black t-shirt and white trousers"), instead of an image. We name this task "appearance-based people search". This functionality can be useful, e.g., in forensics investigations, where a textual description can be provided by a witness. In this paper, we present and experimentally evaluate a general method to perform both person re-identification and people search, using any given descriptor of clothing appearance that exploits widely used multiple part/multiple component representations. It is based on turning the considered appearance descriptor into a dissimilarity-based one, through a framework we previously proposed for speeding up person re-identification methods. Our approach allows one to deploy systems able to perform both tasks with the same pipeline and processing stages for constructing descriptors.

Categories and Subject Descriptors

I.4 [**Computing Methodologies**]: Image Processing and Computer VisionApplications

Keywords

person re-identification, people search, clothing appearance, dissimilarity, video-surveillance

1. INTRODUCTION

Among the possible applications of computer vision to video-surveillance, people recognition among a network of cameras, generally non-overlapping, is gaining much interest. This task is often called "person re-identification" [5], and can enable various useful functionalities. For instance, on-line tracking of an individual over the camera network, and off-line retrieval of all video sequences containing a person of interest, given an image of that person as a query.

In real-world scenarios, strong biometrics like the face cannot be exploited for this task. This is mostly due to unconstrained poses, and to the typically low size of the region of video frames containing a person. Instead, a widely used approach is to exploit *soft* cues, like the gait or clothing appearance. In particular, the latter has proven to be very effective and relatively easy to extract. Therefore, many techniques proposed in the literature exploit clothing appearance descriptors [5], that can be built from one or more video frame containing the individual of interest.

Most recently, in [18, 16] another possible use of clothing appearance descriptors was proposed. It consists of retrieving video sequences showing individuals that match a *textual* description of clothing, like "person wearing a black t-shirt and white trousers". We refer to this task as "appearance-based people search" in the following. This functionality can be very useful, e.g., in forensics investigations, where a textual description can be provided by a witness.

During the past few years, many appearance descriptors have been proposed in the literature. Despite the differences among them, in [15] we showed that most descriptors exploit a multiple component representation (e.g., patches, or interest points), and a body subdivision model (e.g., upper and lower body). Exploiting these commonalities, in [14] we presented the*Multiple Component Dissimilarity* (MCD) framework, to turn any such descriptor into a *dissimilarity-based* appearance descriptor. MCD builds upon the classical dissimilarity-based approach to pattern recognition [12]. However, differently from [12], we compute dissimilarities to prototype representative of *local* components. The resulting descriptor is simply a vector of dissimilarity values. As such, matching two descriptors becomes very fast. In particular, we were able to speed up an existing person re-identification method by four orders of magnitude [14].

In this work, we show that dissimilarity-based descriptors obtained through MCD from any given appearance descriptor, can be used both for person re-identification and for people search. This enables the development of systems that perform both tasks with the same pipeline and processing stages for constructing descriptors. We point out that, differently from the approaches of [18, 16], our approach does not require to develop ad hoc descriptors for people search.

The rest of the paper is organised as follows. First, we briefly review existing appearance descriptors in Sect. 2. We then describe the MCD framework in Sect. 3, and show how dissimilarity-based descriptors obtained through MCD can be used both for person re-identification and people search. We provide an experimental analysis of our approach in Sect. 4, applying MCD to an existing appearance descriptor, and evaluating the use of the resulting dissimilarity-based representation for both tasks. Finally, we draw conclusions and suggest possible directions for future research in Sect. 5.

2. METHODS FOR BUILDING CLOTHING APPEARANCE DESCRIPTORS

Most of the appearance descriptors proposed in the literature so far, are related to the specific task of person re-identification.

The SDALF descriptor of [6] subdivides human body exploiting its symmetry properties: anti-symmetry separates torso and legs, while symmetry divides left and right parts. Three different features are extracted from each body part: colour histograms in the HSV colour space; *maximally stable colour regions* (MSCR); *recurrent high-structured patches* (RHSP). All local features are extracted from torso and legs separately. To obtain MSCR and RHSP, several patches are sampled at random, mainly near symmetry axes, and then clustered to find the most significant ones.

The same body part subdivision is exploited in [15]. Each part is represented with the HSV colour histograms of a bag of randomly extracted image patches. To attain robustness to varying lighting conditions, synthetic patches corresponding to different lighting conditions are added together with the original ones.

In [2], a body part detector is used to find fifteen non-overlapping square cells, that have been proven to be "stable regions" of the silhouette. For each cell, a covariance descriptor based on colour gradients is computed. Colour histogram equalisation was performed to achieve a better robustness to varying lighting conditions. Descriptor generation and matching is performed through a pyramid matching kernel.

In [1] two methods are 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. Inter-camera colour calibration is used to deal with changing lighting conditions.

An approach based on harvesting SIFT-like interest points from different frames of a video sequence is proposed in [10]. Different frames are used also in [7], where two methods are proposed. One is based on interest points, that are selected by the Hessian-Affine interest operator. The other exploits a body part subdivision based on decomposable triangulated graphs. Each part is described by features based on colour and shape.

In [9] several features based on colour (histograms in different colour spaces) and texture (Schmid and Gabor filters), are extracted from randomly taken strips, and their weights are computed by a boosting algorithm.

In [17] global color descriptors (histograms, spatiograms, color/path-length) are computed from the whole body. Changing lighting conditions were addressed by colour histogram normalisation techniques. A graph-based method was used to reduce their dimensionality.

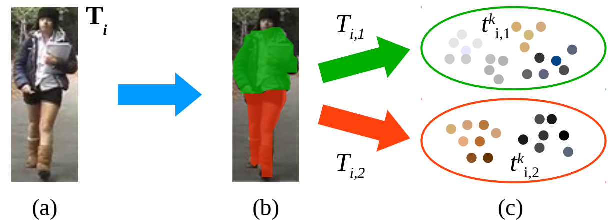

Figure 1: An example of the multiple part-multiple component representation. (a) The image of an individual. (b) The body is subdivided into two parts: upper and lower body, shown respectively in green and in red. (c) A set of components (e.g., image patches) is extracted form each body part. Components are represented here as coloured dots.

In [13] person re-identification is considered as a relative ranking problem, exploiting a discriminative subspace built through an ensemble of SVM-like classifiers specific for ranking tasks. Colour and texture-based features were extracted from six predefined horizontal regions.

In [18, 16] appearance descriptors were used to implement the people search functionality. In [18] the basic idea of building a specific detector for different attributes of interest (e.g., the presence of beard and eyeglasses, the dominant colour of torso and legs, etc.), is proposed, and a specific implementation is developed, mainly for face attributes. The work in [16] focuses on the following attributes: gender, hair/hat colour, clothing colour, and bag (if any) position and colour. A generative model is proposed to build the corresponding descriptors. Both works considered only torso and legs dominant colour as clothing appearance cues.

Despite their differences, most of the above methods use a *multiple instance* representation, obtained by taking several patches, strips, or interest points, and/or exploit some *body part subdivision*. These two observations provide the foundation for a general framework for appearance descriptors proposed in [15]. In this framework the individuals are represented as bags of instances, named "sets of components". Such components can be any kind of local features (e.g., patches, and interest points). If a body part subdivision is used, a different set of components is extracted from each part. The rationale behind this representation is to gain robustness to partial occlusions and pose variations. We can formalise the multiple part/multiple component representation as follows. Let $\mathcal{I} = \{\mathbf{I}_1, \ldots, \mathbf{I}_N\}$ be a *gallery* of images of N individuals. Each image $\mathbf{I}_i$ is represented via an ordered sequence of $M \geq 1$ sets, corresponding to M body parts:

$$\mathbf{I}_i = \{I_{i,1}, \ldots, I_{i,M}\}. \tag{1}$$

According to the multiple component representation, each part $I_{i,m}$ is a set of $n_{i,m}$ of components described by feature vectors $\mathbf{c}^k_{i,m}$:

$$I_{i,m} = \{\mathbf{c}^1_{i,m}, \ldots, \mathbf{c}^{n_{i,m}}_{i,m}\},\ \mathbf{c}^k_{i,m} \in \mathbb{X}, \tag{2}$$

where $\mathbb{X}$ denotes the feature space (assumed to be the same for all the sets, without losing generality). An example of this representation is shown in Fig. 1.

The above framework is the basis of our dissimilarity-based approach, which is described in the next section.

3. DISSIMILARITY-BASED REPRESENTATION OF CLOTHING APPEARANCE

Most appearance-based re-identification methods use descriptors that (1) use a body part subdivision, and (2) represent each body part as a bag of low-level local features. In [14] we proposed a framework, named *Multiple Component Dissimilarity* (MCD), to turn any such descriptor into a dissimilarity-based one, which consists of a vector of dissimilarity values to a predefined set of visual prototypes. The aim of MCD was to reduce processing time, to enable real-time person re-identification systems. In Sect. 3.2 we will show that dissimilarity representations obtained through the MCD framework can be conveniently and naturally exploited also for appearance-based people search. In the following, we summarise the procedure for building MCD descriptors.

Following the representation of Sect. 2, a generic appearance descriptor $\mathbf{I}$ of an individual is a sequence $\{I_m\}_{m=1}^M$ of sets of "components", each one associated to one of the $M \geq 1$ body parts. Each I_m is a bag of local feature vectors $\{\mathbf{c}_m^k\}_{k=1}^{n_m}$.

Let $\mathcal{I}$ be a *gallery* of appearance descriptors (see the example in Fig. 2-a). To represent them in a *dissimilarity space* [12], a set of "visual" prototypes $\mathbf{P}_m = \{P_{m,p}\}_{p=1}^{N_m}$, is first defined for each body part. Prototypes correspond to low-level visual characteristics (e.g., a certain distribution of colours in a body part) that can be shared by several descriptors of $\mathcal{I}$. Then, for each $\mathbf{I} \in \mathcal{I}$, a dissimilarity descriptor $\mathbf{i}^{\mathrm{D}}$ is created, as a vector of dissimilarity values between each $I_m \in \mathbf{I}$, and the corresponding prototypes $\mathbf{P}_m$. Note that, contrary to the original dissimilarity-based approach [12], in MCD prototypes are representative of *local* components of a given body part, instead of the whole part.

Prototypes are created as follows (see Fig. 2-b, c) [14]. For each body part $m = 1, \ldots, M$:

1. Merge the feature vectors of the m-th part of each $\mathbf{I} \in \mathcal{I}$ into a set $X_m = \bigcup_{j=1}^N I_{j,m}$;

2. Cluster the set X_m into a set $\mathbf{P}_m$ of N_m clusters, $\mathbf{P}_m = \{P_{m,1}, \ldots, P_{m,N_m}\}$. Take each cluster as a prototype for the m-th body part.

Each prototype is thus a set of visually similar image components, which can belong to different individuals. In turn, each original descriptor $\mathbf{I}$ consists of a set of components for each body part. Thus, to create a dissimilarity vector from $\mathbf{I}$, dissimilarities can be evaluated via a distance measure between sets. In [14] we used the *k-th Hausdorff Distance*, which is robust to outliers. To reduce computational requirements of dissimilarities computation, in [14] we defined the prototypes as the centroid of the corresponding cluster.

3.1 Dissimilarity-based person re-identification

We illustrate here how the MCD dissimilarity-based descriptors can be exploited for person re-identification, as proposed in [14].

Let $\mathcal{T}$ be a template gallery of N individuals, and $\mathbf{Q}$ be a probe individual. We denote their MCD-based representation respectively as $\mathcal{T}^{\mathrm{D}} = \{\mathbf{T}_1^{\mathrm{D}}, \ldots, \mathbf{T}_N^{\mathrm{D}}\}$ and $\mathbf{Q}^{\mathrm{D}}$. We can

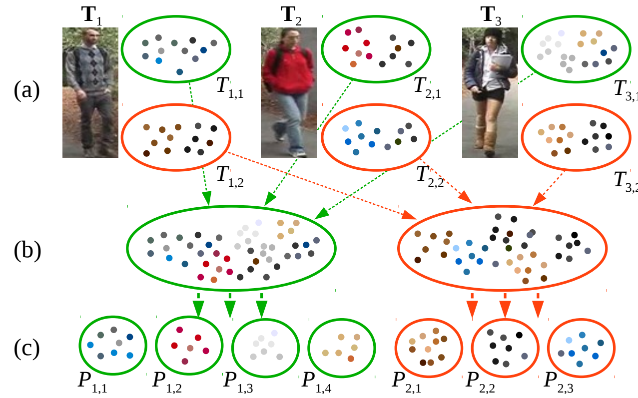

Figure 2: Generation of the prototype gallery in MCD. In this example, the body is subdivided into two parts: upper (in green) and lower body (in red). (a) A template gallery of three individuals, represented with multiple parts and multiple components [15]. (b) All the components of the same part are merged. (c) A clustering algorithm is applied, and a set of prototypes (clusters) is generated for each part.

formulate the problem of re-identifying the individual $\mathbf{Q}^{\mathrm{D}}$ as finding the most similar template:

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

The distance between any template $\mathbf{T}^{\mathrm{D}}$ and $\mathbf{Q}^{\mathrm{D}}$ is computed as:

$$D(\mathbf{T}^{\mathrm{D}}, \mathbf{Q}^{\mathrm{D}}) = f\left(d^{\mathrm{D}}(T_1^{\mathrm{D}}, Q_1^{\mathrm{D}}), \ldots, d^{\mathrm{D}}(T_M^{\mathrm{D}}, Q_M^{\mathrm{D}})\right) \ . \tag{4}$$

where $f : \mathbb{R}^M \to \mathbb{R}$ is a combination of the distances between the corresponding parts of $\mathbf{T}^{\mathrm{D}}$ and $\mathbf{Q}^{\mathrm{D}}$. In this paper, we define such distance measure as the average of the distances $d^{\mathrm{D}}(T_m^{\mathrm{D}}, Q_m^{\mathrm{D}})$. The distance measure $d^{\mathrm{D}}(\cdot, \cdot)$ could be in principle any metric between vectors. In [14], we proposed a weighted Euclidean measure whose weights reflect the degree of relevance of each prototype with respect to the images to be compared (see [14] for further details). The same weighted Euclidean distance is used in this paper.

3.2 Dissimilarity-based people search

In the following we describe a simple and general approach to implement appearance-based people search through MCD descriptors. The approach is independent on the specific body subdivision model and local features used.

Our intuition is that the clothing characteristics that can be detected by a given appearance descriptor, according to its low-level features and body subdivision (e.g., a red shirt, for descriptors based on colour features) may be encoded by one or more visual prototypes. For instance, consider a data set composed by the three images of Fig. 2-a, and suppose that the components of the considered descriptor (represented as coloured dots in Fig. 2) are small image patches described by a RGB colour histogram. After the clustering procedure, it is likely that one prototype will be made up of patches whose dominant color is red, coming from the image $\mathbf{T}_2$. Then, any image of a person wearing a red shirt should exhibit a high similarity to that prototype. Similarly, other prototypes may encode useful information to recognise other clothing characteristics.

Following the above intuition, a possible approach to perform appearance-based people search through an existing appearance descriptor, consists of the following steps: (i) identify a set $\mathcal{Q} = \{\mathbf{Q}_1, \mathbf{Q}_2, \ldots\}$ of clothing characteristics that can be detected by the given descriptor (let we name them *basic queries*); (ii) build a specific detector for each basic query $\mathbf{Q}_i$, using dissimilarity values as *features* of a supervised classification problem.

Clearly, the basic queries of step (i) depend on the features and part subdivision adopted by the original descriptor. For instance, if it separates lower and upper body parts, and uses colour features, one basic query can be "red trousers/skirt". Concerning step (ii), one can define a different supervised binary classification problem for each $\mathbf{Q}_i$, consisting of recognising the presence of the corresponding visual characteristic. A binary classifier (e.g., a Support Vector Machine [4]) can be trained using as features the dissimilarity values of an image descriptor to the prototypes. The set of training samples can be a gallery of images of individuals, labelled according to the presence/absence of the visual characteristic.

Complex queries can be built by connecting basic ones through Boolean operators, e.g., "red shirt AND (blue trousers OR black trousers)". The images relevant to a complex query can be found by simply combining the subsets of images relevant to each basic query, using the set operators corresponding to the Boolean ones (e.g., AND $\rightarrow \bigcap$, OR $\rightarrow \bigcup$).

4. EXPERIMENTAL ANALYSIS

In this section we apply MCD to an existing appearance descriptor, named MCMimpl, previously proposed by the authors in [15]. In MCMimpl, first background and foreground are separated through a STEL generative model [11]. Then the body is divided into torso and legs. From each part, a set of partly overlapping patches is randomly extracted and represented via HSV colour histograms. To obtain robustness to changing lighting conditions, synthetic components corresponding to different lighting conditions are generated from the original ones (see [15] for further details).

To obtain a dissimilarity version of MCMimpl, we followed the procedure described in Sect. 3, using a two-stage clustering scheme to obtain the prototype gallery. In the first stage we used the Mean-Shift algorithm [3] to separately cluster the components of each individual (excluding the simulated patches). The bandwidth parameter of Mean-Shift, which governs the spread of each cluster, was set to $BW = 0.3$. The k-means algorithm was applied at the second stage on the resulting centroids. We set the number of clusters to $k = 150$ for both body parts (torso and legs). Each prototype was finally associated to a bag containing 1) the original patch nearest to each centroid, and 2) the set of synthetic patches created from that patch.

In the following, we evaluate the dissimilarity version of the MCMimpl descriptor, denoted with MCMimpl$^{\mathrm{Dis}}$, for both person re-identification and people search.

4.1 Results on person re-identification

Experiments were carried out on the benchmark VIPeR data set, which was used in many previous works on person re-identification [8]. It is made up of two non-overlapping views of 632 different pedestrians, taken from two different cameras, under different poses, viewpoint and lighting conditions. It is the most challenging dataset currently available for person re-identification.

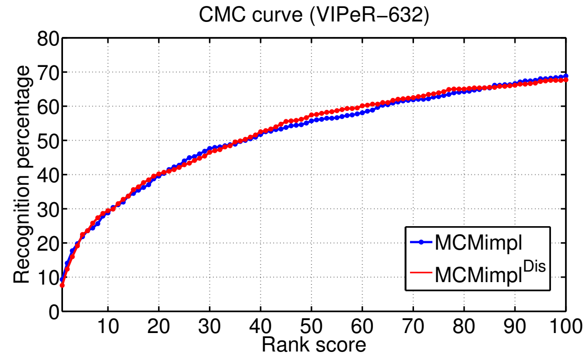

Figure 3: CMC curves attained by MCMimpl and MCMimpl$^{\mathrm{Dis}}$ on the VIPeR dataset.

	MCMimpl	MCMimpl$^{\mathrm{Dis}}$
Avg time for a single matching	28.6 ms	0.004 ms
Avg total time for a single run	11550.6 sec	179.4 sec
Size of the descriptor	96 KB	1.2 KB (1)
Size of the prototype gallery	-	48 KB (1)

Table 1: Processing time and memory requirements of MCMimpl and MCMimpl^{Dis}. (1) 32 bit floating point values.

The first view of each pedestrian is used to build a gallery of 632 templates, while the second view is used as a query. We then ranked the templates in respect of the similarity with each of the 632 the queries. The re-identification accuracy was evaluated using the cumulative matching characteristic (CMC) curve. It is defined as the cumulative distribution of the correct rank R, namely the probability that the template image of the query individual is among the top-r ranked images: $P(R \leq r)$, $r = 1, \ldots, N$. We report the CMC curve attained by MCMimpl and MCMimpl$^{\mathrm{Dis}}$ in Fig. 3.

In terms of pure recognition performance, MCMimpl$^{\mathrm{Dis}}$ equals MCMimpl in the considered dataset. This is reasonable, since the MCD dissimilarity-based descriptor use the same part subdivision and the same local features as the corresponding original descriptor.

Computational requirements of the two methods exhibit strong differences instead. In Table 1 we show the processing time and memory requirements of MCMimpl and MCMimpl$^{\mathrm{Dis}}$, attained on a 2.4 GHz CPU. MCMimpl$^{\mathrm{Dis}}$ clearly outperforms MCMimpl in terms of processing time and memory usage. In particular, a speed-up of four orders of magnitude is attained for matching descriptors. The average overall time required to perform a run of the experiments is much lower as well.

4.2 Results on people search

As explained in Sect. 3.2, the same dissimilarity-based descriptor used for person re-identification can be used for the task of appearance-based people search. We experimentally evaluated the MCMimpl$^{\mathrm{Dis}}$ descriptor on this task, using again the VIPER data set, as it shows a wide range of different clothing characteristics.

We defined 14 different basic queries and labelled a subset of 512 images according to the presence/absence of each of the 14 basic clothing characteristics. These labelled images will be made available in the authors' web site. The basic queries are reported in Table 2. The corresponding number of relevant images is shown between brackets. Since the original descriptor uses a upper/lower body part subdivision, and exploits local features related to colour, we focused on colour clothing characteristics of the upper and lower body parts. Additionally, we considered the presence of short sleeves and short trousers/skirts.

We evaluated the retrieval performance of our approach on each basic query, in terms of the precision-recall (P-R) curve.[1] The P-R curve is widely used in document retrieval, because it is insensitive to the ratio between relevant and non-relevant samples. Typically a higher recall can only be attained at the expense of a lower precision, and vice versa. A useful summary of the whole P-R curve into a scalar value is the break-even point (BEP), namely the point at which precision equals recall.

We first extracted the MCD visual prototypes from the 512 labelled images. Then, for each basic query, we randomly subdivided these images into a training and a testing set of equal size, using a stratified sampling approach to preserve the ratio between relevant and non-relevant images of each class, and trained a Support Vector Machine classifier with linear kernel on training images, to implement the corresponding detector. The P-R curve was evaluated on testing images by varying the SVM decision threshold. We repeated the whole procedure ten times, and averaged the resulting P-R curves.

The performance on each basic query is summarised as the corresponding average BEP in Table 2. In Fig. 4 we report four representative examples of the average P-R curves. An example of the ten top-ranked images for two basic queries is also shown in Fig. 5.

The proposed approach attained a good retrieval performance almost on all basic queries. The best performance was attained on basic queries related to the red, white and black colours (see Table 2). The reason is that such colours are well separated in the HSV space, which is used by all the considered descriptors. A rather good performance has been attained also on basic queries related to the presence of skin on lower arms and legs, namely "short sleeves" and "short trousers/skirt". Even if the considered body subdivision does not separate lower and upper arms and legs, still the descriptor was able to recognise the presence of short sleeves/trousers/skirt by detecting skin-like colour in the whole arms or legs.

5. CONCLUSIONS

We presented a general method to perform both person re-identification and people search using any given descriptor of clothing appearance that exploits a widely used multiple part and/or multiple component representation. It is based on a *local* dissimilarity-based approach that 1) provides very compact descriptors that allow one to considerably speed up the matching phase in a person re-identification task, and 2) can be naturally exploited to train detectors for the task of people search. The proposed method can thus be useful to deploy a system able to perform both tasks with the same acquisition and processing pipeline for building descriptors. Experiments on a well-known benchmark dataset provided evidence of the potential effectiveness of our method, in both tasks above. Interesting directions for further research include the application of MCD to other clothing appearance descriptors, and a proper extension of the framework to deal with video sequences.

[1]Precision is the ratio between the number of images correctly labelled as relevant, and the total number of images labelled as relevant. Recall is the ratio between the number of images correctly labelled as relevant, and the total number of relevant images.

Class (cardinality)	BEP
red shirt (51)	0.845
blue/light blue shirt (34)	0.645
pink shirt (35)	0.534
white/light gray shirt (140)	0.771
black shirt (156)	0.728
orange shirt (10)	0.689
violet shirt (18)	0.422
green shirt (34)	0.687
short sleeves (220)	0.631
red trousers/skirt (16)	0.713
black trousers/skirt (12)	0.683
white/light gray trousers/skirt (81)	0.758
blue/light blue trousers/skirt (175)	0.641
short trousers/skirt (82)	0.416

Table 2: Left: the 14 basic queries considered in the experiments, and the corresponding number of relevant images (between brackets). Right: average break-even point (BEP) attained using the MCMimpl$^{\mathrm{Dis}}$ descriptor.

Figure 4: Average P-R curves of 4 queries.

6. REFERENCES

[1] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Person re-identification using haar-based and dcd-based signature. In *Proc. 7th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8, 2010.

[2] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Person re-identification using spatial covariance regions of human body parts. In *Proc. 7th IEEE Int.*

Figure 5: The top-ten retrieved images for (a) “red shirt“and (b) “black trousers”, sorted from left to right for decreasing relevance score.

Conf. on Advanced Video and Signal Based Surveillance (AVSS), pages 435–440, 2010.

[3] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 24:603–619, 2002.

[4] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[5] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher. Appearance-based person reidentification in camera networks: problem overview and current approaches. *Journal of Ambient Intelligence and Humanized Computing*, 2:127–151, 2011.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Proc. of the 2010 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2360–2367, 2010.

[7] N. Gheissari, T. B. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *Proc. 2006 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1528–1535, 2006.

[8] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proc. 10th IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, pages 41–47, 2007.

[9] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proc. 10th European Conf. on Computer Vision (ECCV)*, pages 262–275, 2008.

[10] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux. Interest points harvesting in video sequences for efficient person identification. In *Proc. 8th Int. Workshop on Visual Surveillance (VS)*, 2008.

[11] N. Jojic, A. Perina, M. Cristani, V. Murino, and B. Frey. Stel component analysis: Modeling spatial correlations in image class structure. *Proc. 2009 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2044–2051, 2009.

[12] E. Pekalska and R. P. W. Duin. *The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence)*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2005.

[13] B. Prosser, W. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *Proc. 2010 British Machine Vision Conf. (BMVC)*, pages 21.1 – 21.10, 2010.

[14] R. Satta, G. Fumera, and F. Roli. Fast person re-identification based on dissimilarity representations. *Pattern Recognition Letters, Special Issue on Novel Pattern Recognition-Based Methods for Reidentification in Biometric Context*, 2012, in press.

[15] R. Satta, G. Fumera, F. Roli, M. Cristani, and V. Murino. A multiple component matching framework for person re-identification. In *Proc. 16th Int. Conf. on Image Analysis and Processing (ICIAP)*, volume 2, pages 140–149, 2011.

[16] J. Thornton, J. Baran-Gale, D. Butler, M. Chan, and H. Zwahlen. Person attribute search for large-area video surveillance. In *Proc. 2011 IEEE Int. Conf. on Technologies for Homeland Security (HST)*, pages 55 –61, 2011.

[17] D. N. Truong Cong, C. Achard, L. Khoudour, and L. Douadi. Video sequences association for people re-identification across multiple non-overlapping cameras. In *Proc. 15th Int. Conf. on Image Analysis and Processing (ICIAP)*, pages 179–189, 2009.

[18] D. Vaquero, R. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk. Attribute-based people search in surveillance environments. In *Proc. IEEE Ws. on Appl. of Computer Vision (WACV)*, 2009.