Person Re-identification by Discriminatively Selecting Parts and Features

Amran Bhuiyan(✉), Alessandro Perina, and Vittorio Murino

Pattern Analysis and Computer Vision (PAVIS),
Istituto Italiano di Tecnologia, Genova, Italy
Amran.Bhuiyan@iit.it

Abstract. This paper presents a novel appearance-based method for person re-identification. The core idea is to rank and select different body parts on the basis of the discriminating power of their characteristic features. In our approach, we first segment the pedestrian images into meaningful parts, then we extract features from such parts as well as from the whole body and finally, we perform a salience analysis based on regression coefficients. Given a set of individuals, our method is able to estimate the different importance (or salience) of each body part automatically. To prove the effectiveness of our approach, we considered two standard datasets and we demonstrated through an exhaustive experimental section how our method improves significantly upon existing approaches, especially in multiple-shot scenarios.

Keywords: Pedestrian re-identification · STEL segmentation · Lasso regression

1 Introduction

Person re-identification is becoming an important topic in Computer Vision, especially in video surveillance scenarios. Its goal is to recognize (indeed, re-identify) an individual captured in diverse locations over different non-overlapping camera views, considering a large set of candidates.

The common assumption in re-identification (re-id) is that individuals do not change their clothing so their appearance in all the views is similar. This is still a complex task due to the nonrigid structure of the human body, the different perspectives with which a pedestrian can be observed, and the highly variable illumination conditions. Re-identification approaches can be divided in two classes of algorithms: learning-based and direct methods. In the former group, a dataset of different individuals is used to learn the features and the metric space where to compare them, in order to guarantee a high re-identification rate (e.g., see [8-14,15]). In contrast, direct methods are mainly devoted to the search of the most discriminant features and their combination so to design a powerful descriptor (or signature) for each individual. Besides, re-identification algorithms can also be categorized in single-shot and multiple-shot classes of methods.

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part III, LNCS 8927, pp. 147–161, 2015.
DOI: 10.1007/978-3-319-16199-0_11

The former focuses on associating pairs of images for each individual, while the latter employs multiple images of the same person as the probe and/or in the gallery set, trying to exploit this additional information.

As for learning-based approaches, in [10], pairwise dissimilarity profiles between individuals are learned and adapted for nearest-neighbor classification. The approach presented in [11] uses boosting to select a combination of spatial and color information for viewpoint invariance. In [8], a high-dimensional signature composed by multiple features is projected into a low-dimensional latent space by a Partial Least Squares reduction method. In [12], contextual visual knowledge is exploited, enriching a bag-of-word-based descriptor by features derived from neighboring people, assuming that groups of persons are invariant across different camera views. Re-identification is cast as a binary classification problem (one vs. all) in [13] adopting Haar-like features and a part-based MPEG7 dominant color descriptor. In [24], re-id is considered as a relative ranking problem in a higher dimensional feature space where true and wrong matches become easily separable. Finally, re-identification is considered as a Multiple Instance Learning problem in [15], where a method of synthetically augmenting the training dataset is also proposed.

Direct methods focus more on designing novel features for capturing the most distinguishing aspects of an individual. In [16], a descriptor is proposed by subdividing the person in horizontal stripes, keeping the median color of each stripe accumulated over different frames. A spatio-temporal local feature grouping and matching is proposed in [17], where a decomposable triangular graph is built in order to capture the spatial distribution of the local descriptor over time. The method proposed in [18] segments a pedestrian image into regions, and stores the spatial relationship of the colors into a co-occurrence matrix. This technique proved to work well when pedestrians are seen from similar perspectives. In [19], SURF interest points are collected over short video sequences and used to characterize human bodies. Symmetry- and asymmetry-driven features are explored on [5,7] based on the idea that features closer to the body symmetry axes are more robust against scene clutter and body extremities. Covariance features, originally employed for pedestrian detection [20], are extracted from coarsely located body parts and tailored for re-id purposes in [21]. This work has been extended in [22] by considering the case where multiple images of the same individual are available. The authors adopt the manifold mean as surrogate of the different covariances coming from the multiple images. Similar features, i.e., MSCR and color histograms, have been also employed in [3,4], and used to match signatures based on Custom Pictorial Structures [1] to finely segment body parts and extract appearance descriptors.

In addition to color based features, there are some other features that have been proved to be promising for the re-id task, such as: textures [8,9,11], edges [8], Haar-like features [13], interest points [19], image patches [11], and segmented regions [18]. All these features can be extracted from horizontal stripes [16], triangular graphs, concentric rings [17], symmetry-driven structures [5,7], and horizontal patches [21]. Another unconventional application of re-id considers

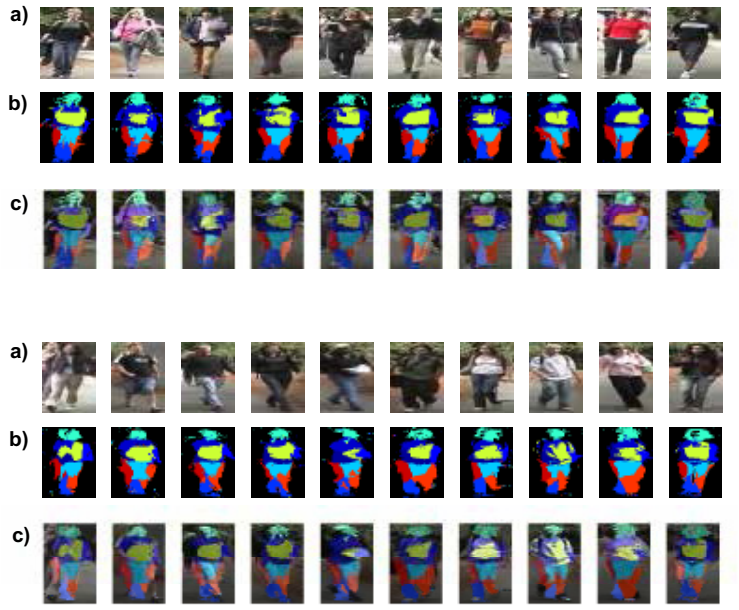

Fig. 1. (a) Lineups of pedestrians; and (b,c) corresponding segmentations (best viewed in color).

Pan-Tilt-Zoom cameras, where distance between signatures are also computed across different scales [6] while estimating the most discriminant part.

Within this context, we propose an approach that focuses on discriminating parts and features, taking inspiration from [6]. The idea is to learn the most discriminant body parts (and associated features) able to separate (or match) at best a given a set of pedestrians. A least shrinkage and selective operator (Lasso) regression method [23] is used for this selection task adopting standard features extracted by segmented body parts, together with the generative model introduced in [2]. Performance has been evaluated by testing our method on two well recognized publicly available datasets, CAVIAR4REID and VIPeR.

Our approach, belongs to the class of the direct methods and it differs from previous works in two important aspects: i) we use Stel Component Analysis (SCA) segmentation technique which is quite effective for pedestrian segmentation, and ii) unlike [3], no manual weighting of individual parts is required for all the pedestrians, instead this is automatically carried out by exploiting regression weights.

The idea of considering parts of the body for re-id is not new. In [3,4], the authors used the pictorial models (see Fig.1c), and in the experimental section we will show that SCA yields to better results.

The rest of the paper has been organized as follows. Section 2 describes of the proposed approach in detail. Sec. 3 reports the experimental results, and concluding remarks are drawn in Sec. 4.

2 The Approach

The pipeline that characterizes our approach consists of the following steps:

Pedestrian segmentation: We segmented each image using SCA [2]. This segmentation method yields to consistent segments across images and it allowed us to discard background regions and focus the analysis on the foreground parts.

Feature extraction: We extracted standard features from each foreground segment.

Rank-to-mask strategy: We applied Lasso regression on every pedestrian in a one-vs-all fashion, and used the regression coefficients to determine the more discriminating parts and/or features.

Signature matching: We evaluated our approach using standard matching approaches.

In the following, each step is described and analyzed focusing on multi-shot re-identification.

2.1 Pedestrian Segmentation

The aim of this phase is to isolate the actual body appearance from the rest of the scene. This allows the proposed approach to focus solely on the individual and its parts, ignoring the context in which it is immersed.

We performed this separation by using Stel Component Analysis (SCA) [2]. This segmentation algorithm is based on the notion of "structure element", or stel, which can be explained as an image portion (often discontinuous) whose topology is consistent over an image class. A stel often represents a meaningful and semantic part for a class of objects, like an arm or the head for pedestrians. For example, in Fig. 1 we show few images of pedestrians and their related segmentation in stels. Each color indexes a particular stel s_i while maintaining consistent the segmentation (same color, same part of the body) across the whole dataset. This is very important as it allows us to compare consistently feature signatures extracted in the various body parts.

More formally, SCA captures the common structure of an image class by blending together multiple stels: it assumes that each pixel measurement x_i, with its 2D coordinate i, has an associated discrete variable s_i, which takes a label from the set $\{1, ..., S\}$. Such a labeling is generated from a stel prior $p(s_i)$, which captures the common structure of the set of images.

The model detects the image self-similarity within a segment: pixels with the same label s are expected to follow a tight distribution over the image measurements. Finally, while the appearance similarity is local (for each image), the model insists on consistent segmentation by means of a stel prior, which is common for all the images.

SCA has been previously considered for re-id, in particular in [5,7], where the authors used it to perform a background-foreground segmentation (S=2): an

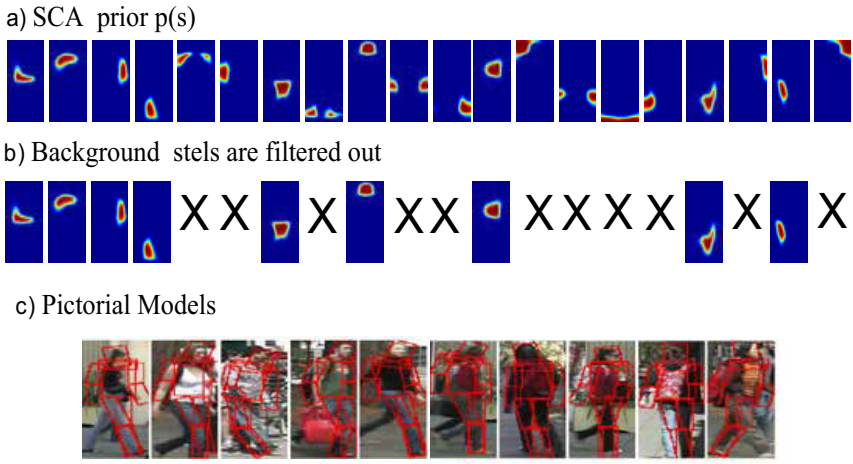

Fig. 2. a) Segmented part-masks for whole image. b) Foreground extraction procedure. c) Lineup of pedestrians and superimposed pictorial structures.

example of the segmentation prior $p(s_i)$ used is shown in Fig. 2a. Here, for the first time, we exploited SCA's segmentation in multiple parts ($S > 2$), discarding background parts and considering the features in each each part separately. An example of a learned SCA segmentation prior is shown in Fig. 2a, where S=20. After learning the stel prior, we manually filtered out the background stels, as shown in Fig. 2b. It is worth to note that the model is learned only once and it is independent on the dataset. Furthermore the background suppression is performed once and not for each individual image, because all the images have consistent segmentation.

In our experiments, we set the number of segment to S=20 and we modeled the distribution over the image measurements as a mixture of Gaussians. To learn the segmentation prior we set the number of iterations to 50. Segmentation of new images (i.e., probe) consists in fast inference and is done in real time.

2.2 Feature Extraction

The feature extraction stage consists in distilling complementary aspects from each body part in order to encode heterogeneous information and to capture distinctive characteristics of the individuals. There are many possible cues useful for a fine visual characterization and we considered two types of features: color histograms and maximally stable color regions (MSCR) [25]. This is the same signature already used in [3–5, 7].

Color histograms represent a good compromise when they encode separately shades of gray from colored pixels. To do so, first, all RGB pixel values are converted to the HSV color space, h,s,v. Then, they are subject to the following operations: all pixels with value $v < \tau_{black}$ are counted in the bin of blacks, all pixels with saturation $s < \tau_{gray}$ are counted in the gray bins according to their

value v, and finally all remaining pixels are counted in the color bins according to their hue-saturation coordinates (h,s). Basically dark and unsaturated pixels are counted separately from the others, and the brightness of the colored pixels are ignored by counting only their chromaticity in a 2D histogram.

This procedure is also tweaked in several ways to improve speed and accuracy: the HSV channels are quantized into [20,10,10] levels, respectively, the votes are (bi-)linearly interpolated into the bins to avoid aliasing and the residual chromaticity of the gray pixels is counted into the color histograms with a weight proportional to their saturation. The image regions of each part are processed separately and provide a combined gray-color histogram (GC histogram in short) which is vectorized and normalized.

The other extracted feature is the Maximally Stable Color Regions (MSCR). Here we used the full body masks independently to constrain the extraction of the MSCR blobs. The MSCR operator detects a set of blob regions by looking at the successive steps of an agglomerative clustering of the pixels. At each step, neighboring pixels with similar color are clustered, employing a threshold that represents the maximal chromatic distance between colors. The threshold is varied at each step and the regions that are stable over a range of steps constitute the maximally stable color regions of the image. As in [3,4], we extracted a signature $MSCR = \{(y_i, c_i) \mid i = 1, ..., N\}$ that contains the height (vertical size) and color of the maximum stable regions, or blobs. The algorithm is set in a way that provides many small blobs and avoids creating big ones. The rationale is to localize details of the pedestrians appearance which is more accurate for small blobs.

2.3 Discriminative Analysis: Rank-to-Mask Strategy

The second novel contribution of our approach is a discriminative analysis of parts and features, which in our case are histogram bins. Our idea is that we can identify a pedestrian by looking only to a subset of the signature features which is different for each pedestrian in the gallery. We accomplish this by ranking features and parts based on the means of a regression approach.

Given a pool of training images for N pedestrians (the gallery) we perform a sparse regression analysis using Lasso [23], which is a general form of regularization in a regression problem. In the simple linear regression problem every training image, described by the proposed feature vector and denoted with $x^{(n)}$, is associated with a target variable $z^{(n)}$ which represents the identity. The target variable can then be expressed as a linear combination of the image features as follows:

$$z^{(n)} = \left(\boldsymbol{\alpha}^{(n)} \right)^{\intercal} \cdot x^{(n)} \tag{1}$$

where $\intercal$ represents the transposed vector.

The standard least square estimate calculates the weight vector $\boldsymbol{\alpha}^{(n)}$ by minimizing the error function.

$$E(\boldsymbol{\alpha}) = \sum_{n=1}^{N} \left(z^{(n)} - \left(\begin{array}{c} \boldsymbol{\alpha}^{(n)} \end{array} \right)^{\intercal} \cdot x^{(n)} \right)^2 \tag{2}$$

Again, in our case N corresponds to the total number of pedestrians we have in the training set. The regularizer in the Lasso estimate is simply expressed as a threshold on the L1-norm of the weight $\boldsymbol{\alpha}$:

$$\sum_{\mathbf{j}} \mid \alpha_{\mathbf{j}} \mid \leq \mathtt{K} \tag{3}$$

This term acts as a constraint that has to be taken into account when minimizing the error function, being $\mathtt{K}$ a constant. After doing this, it has been proven that (depending on the parameter $\mathtt{K}$), many of the coefficients $\alpha_{\mathbf{j}}$ become exactly zero [24].

In our approach, the aim is to determine the visual characteristics of each pedestrian that discriminate him/her from all the others pedestrians present in the gallery. To this end, we performed Lasso regression for each pedestrian separately, considering all its training images has positive samples. In other words, we solved N regression problems, each one returning a pedestrian-specific weight vector $\boldsymbol{\alpha}^{(n)} \quad n = 1, \ldots, N$. Since each dimension $\alpha_{\mathbf{j}}^{(n)}$ weighs a different feature in a different part, $\mathbf{j} = (p, f)$, it is possible to figure out which parts (indexed by p) or signature feature (indexed by f) are the most discriminant for a each pedestrian, and those which can be neglected.

One important consideration is that one cannot weigh the histogram comparisons directly with Lasso outputs as they are not normalized across samples. To solve this issue we used ranking. First, we sorted regression coefficients in decreasing order of their absolute value $|\alpha_{p,f}^{(n)}|$, and second, we assigned a weight to each feature j based on its ranking position $r_{p,f}^{t}$ as follows:

$$R_{p,f}^{(n)} = \begin{cases} 1 & \text{if} \quad r_{p,f}^{t} \leq P \quad \text{and} \quad \alpha_{p,f}^{(n)} > 0 \\ 0 & \text{Otherwise} \end{cases} \tag{4}$$

Where P is a position in the rank; for example if $P = 1$, only the top feature is considered, etc..

We called $R_{p,f}^{(n)}$, the *Rank-to-mask* coefficients: They filter out features which are not important to discriminate the identity of the person n, where the importance is given by the regression coefficients. Furthermore, by summing $\alpha_{p,f}^{(n)}$ over the parts, the individual color bins or/and the pedestrians, we can highlight the *individual* parts or color bins which are more important. To summarize, we introduced the following strategies:

Best parts for pedestrian - BPP. We summed over all the features for each pedestrian $\hat{\alpha}_{p}^{(n)} = \sum_{f} \alpha_{p,f}^{(n)}$ and then we applied the aforementioned procedure to $\hat{\alpha}_{p}^{(n)}$. The resulting $\hat{R}_{p}^{(n)}$ are then simply replicated for each feature f in the part p to reconstruct a mask vector of size $F \times P$ which is subsequently used in the signature matching phase. This strategy aims at highlighting the most discriminative parts for each pedestrians.

Best parts for dataset - BPD. We summed over the features and pedestrians $\hat{\alpha}_p = \sum_{f,n} \alpha_{p,f}^{(n)}$ and we proceeded likewise the previous case. This strategy aims at highlighting the most discriminative parts in the whole dataset.

Best feature for pedestrian - BFP. We summed over the parts for each pedestrian $\hat{\alpha}_f^{(n)} = \sum_{p} \alpha_{p,f}^{(n)}$ and we proceeded likewise the previous cases. This strategy highlights which are the best feature useful to recognize each pedestrian.

Best features for dataset - BFD. We summed over the parts and pedestrians $\hat{\alpha}_f = \sum_{p,n} \alpha_{p,f}^{(n)}$ and we proceeded likewise the previous cases. This strategy aims at highlighting the most discriminative features for each pedestrian.

As illustrative example, Fig. 3a shows the training images of pedestrian. After Lasso-based training, we compute the ranking using the **BPP** strategy resulting in *rank-to-mask* coefficients shown on the right of the figure. The pedestrian presents a white cross in the middle of torso and it is easily understandable that this should be a very discriminant part for this individual, and in fact torso is the most discriminating part. Similar considerations can be drawn by looking at the other two examples in Fig. 3b.

2.4 Feature Matching

Likewise [3], we employed color histogram and the MSCR blobs as image signature. To match two signatures $S_a = (h_a, MSCR_a)$ and $S_b = (h_b, MSCR_b)$ we calculated the distance

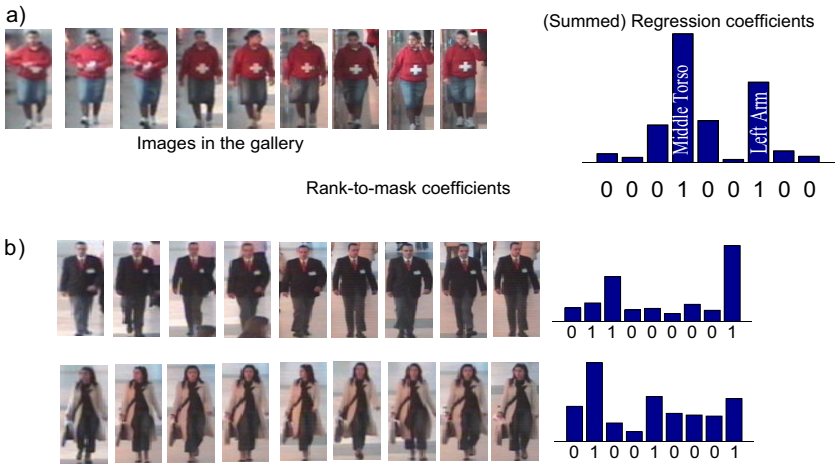

Fig. 3. a) Multiple (gallery) images of a same pedestrian. The regression coefficients summed up across the features highlight which are the most important parts to identify that pedestrian. Coefficients are then ranked and only the top P are retained. In this case $P = 2$. b) Two other examples considering P=3.

$$d\left(S_a, S_b\right) = \beta \cdot d_h\left(R \odot h_a, R \odot h_b\right) + (1-\beta) \cdot d_{MSCR}\left(MSCR_a, MSCR_b\right) \quad (5)$$

where

$$d_h\left(R \odot h_a, R \odot h_b\right) = \log\left(\sqrt{R \odot h_a}^{\mathsf{T}} \cdot \sqrt{R \odot h_b}\right) \quad (6)$$

is the Bhattacharyya distance ($\odot$ represent point-wise multiplications), and

$$d_{MSCR}\left(MSCR_a, MSCR_b\right) = \frac{1}{\mid M_a \cup M_b \mid} \sum_{(i,j) \in M_a \cup M_b} \delta_{ij} \quad (7)$$

is the MSCR distance. Rs are the *rank-to-mask* coefficients introduced in the previous section (at some rank P), and β balances the two distances defined by Eq. 6 and Eq. 7. The latter is obtained by first computing the set of distances between all blobs $(y_i, c_i) \in MSCR_a$ and $(y_j, c_j) \in MSCR_b$:

$$\delta_{ij} = \gamma \cdot d_y\left(y_i, y_j\right) + (1-\gamma) \cdot d_{lab}\left(c_i, c_j\right) \quad (8)$$

where γ balances the height distance $d_y = \frac{|y_i - y_j|}{H}$, and $d_{lab} = \frac{\|lab(c_i) - lab(c_j)\|}{200}$ is the Euclidean distance in the CIELAB color space. Then, the sets $M_a = \{(i,j) \mid \delta_{ij} \leq \delta_{ik}\}$ and $M_b = \{(i,j) \mid \delta_{ij} \leq \delta_{kj}\}$ of minimum distances are calculated from the two point of views, and finally we calculate their average as shown in Eq. 7.

The normalization factor H for the height distance is set to the height of the images in the dataset, while the parameters β and γ are tuned through cross-validation.

3 Experimental Results

The aims of the experimental section are *i)* to compare the SCA segmentation with pictorial models [3,4] and with the BG-FG segmentation (based on SCA) used in [5], and *ii)* to show to which extent the *rank-to-mask* strategy works.

As comparisons, we considered [3–5] because the three methods use exactly the same features, thus making the comparison fair. We compared our results with the best performance obtained by [3].

We considered two public available datasets, CAVIAR4REID [4,28] and VIPeR [26,27]. In particular, CAVIAR4REID covers almost all the challenging aspects of the person re-identification problem, such as shape deformation, illumination changes, occlusions, image blurring, low resolution images, etc.

The most important performance evaluation report tool for re-id is the Cumulative Matching Characteristic (CMC) curve, which is a plot of the recognition performance vs. the re-id ranking score and it represents the expectation of finding the correct match in the top n matches. Higher curves represent better performance. The performance can also be evaluated by computing the ranked matching rates, and these results are shown in the following tables.

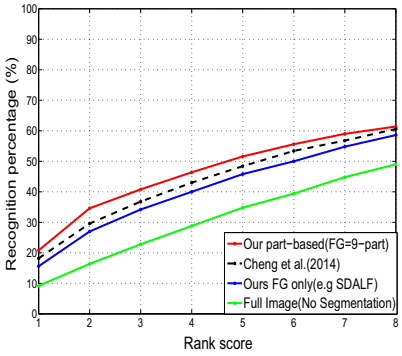

Fig. 4. CMC curves for multiple-shot trials on CAVIAR4REID.

CAVIAR4REID Dataset : It contains images of pedestrians extracted from the CAVIAR repository [4], providing a challenging real world setup. From the 72 identified different individuals (with images varying from 17×39 to 72×144), 50 are captured by two cameras and 22 from only one camera and each pedestrian has 10 images from each camera. Here we restricted to the subjects taken from 2 cameras, and selected M=5 images from the first camera for the probe set and M=5 images from the second camera as gallery set. Then, we performed multi-shot re-id (multi-vs-multi or CMvsM strategy); this is actually the same setup used in [3]. All images are re-sized to 32 × 96 pixels.

From the experimental findings shown in Fig.4-6, it will be evident that our approach outperforms convincingly the methods in the literature at different ranks.

As a first test we evaluated SCA segmentation for S>2. Fig. 4 reports the recognition accuracies: SCA segmentation clearly outperforms pictorial models used in [3,4] and the background-foreground segmentation used in SDALF which correspond to SCA run with S=2. We also reported the re-id performance obtained by considering the full image (without any segmentation): despite in this dataset all the pedestrians appear in the same environment, this actually affects the accuracy.

In the second part of the experiments we evaluated the *rank-to-mask* strategy, varying the rank P (see Eq.4). We set Lasso constant `K` (see Eq. 3) to 50, however the performance of the method has not changed much by varying this parameter.

In Fig. 5 and Tables 1 and 2, we reported the result of the *rank-to-mask* strategy based on the pedestrian specific parts-based (BPP) and the dataset specific parts-based (BPD) ranking scenarios, for different ranking (P) values, respectively.

Although the top-1 part for BPD shows better performance than the same ranking for BPP, still, considering all the instances, it is evident BPP scenario works better than BPD scenario.

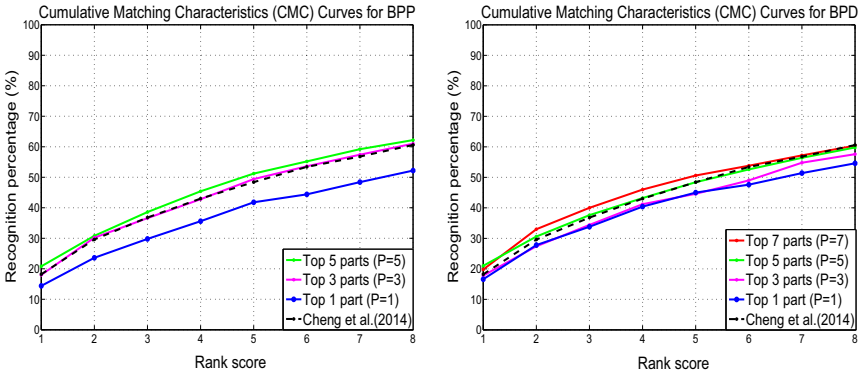

Fig. 5. Comparative plots between best-part for pedestrians (BPP) scenario and best-part for dataset (BPD) scenario on CAVIAR4REID.

Table 1. Comparison with Cheng et al.(2014) [3] methods for top-ranked matching rate (%) on the CAVIAR4REID dataset using best-part for pedestrian (BPP) scenario

Methods	**r=2**	**r=4**	**r=6**	**r=8**
Top 5 parts (P=5)	30.8	45.4	55.2	62.2
Top 3 parts (P=3)	30.2	42.8	53.8	61.0
Top 1 part (P=1)	23.6	35.6	44.4	52.2
Cheng et al.(2014)	29.6	43.0	53.4	60.6

Table 2. Comparison with Cheng et al.(2014) [3] methods by top-ranked matching rate (%) on the CAVIAR4REID dataset using best-part for database (BPD) scenario

Methods	**r=2**	**r=4**	**r=6**	**r=8**
Top 7 parts (P=7)	33.0	46.0	53.8	60.6
Top 5 parts (P=5)	30.6	43.0	52.6	59.8
Top 3 parts (P=3)	27.8	41.4	49.0	57.8
Top 1 parts (P=1)	27.8	40.4	47.6	54.6
Cheng et al.(2014)	29.6	43.0	53.4	60.6

In Fig.6 and Table 3,4, the result of the *rank-to-mask* strategy based on the pedestrian specific feature-based (BFP) and the dataset specific feature-based (BFD) ranking scenarios, for different ranking (P) values have been reported, respectively. The BFP scenario based top-1 feature showed almost equivalent accuracy of BFD based on the top-5 features. In general, we reached the results of [3] by considering only the top-10 features in BFP ranking scenario, while for the BFD ranking scenario it took about 35 top features.

From all the above analysis, we can confirm how the Lasso regression method introduced here is able to extract a significant ranking of features and improve re-id performance. It is also worth mentioning that the performance gets saturated after considering a certain number of top features ranking scores, i.e., after that, it does not show considerable variations.

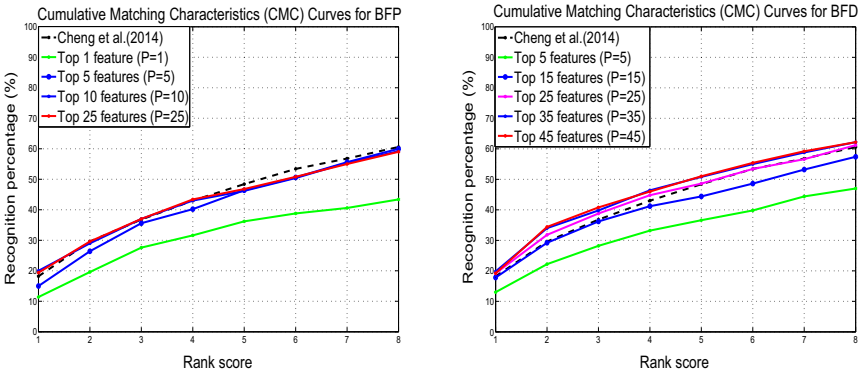

Fig. 6. CMC curves comparing best-feature for pedestrians (BFP) and best-feature for dataset (BFD) scenarios on CAVIAR4REID.

Table 3. Comparison with Cheng et al.(2014) [3] methods by top ranked matching rate (%) on the CAVIAR4REID dataset using best-feature for pedestrians (BFP) scenario

Methods	**r=2**	**r=4**	**r=6**	**r=8**
Top 25 features (P=25)	29.6	43.4	50.8	59.0
Top 10 feature (P=10)	29.0	43.0	50.8	59.2
Top 5 features (P=5)	26.4	40.2	50.8	60.0
Top 1 feature (P=1)	19.6	31.6	38.8	43.4
Cheng et al.(2014)	29.6	43.0	53.4	60.6

Table 4. Comparison with Cheng et al.(2014) [3] methods by top ranked matching rate (%) on the CAVIAR4REID dataset using best-feature for database (BFD) scenario

Methods	**r=2**	**r=4**	**r=6**	**r=8**
Top 45 features (P=45)	34.4	46.0	55.4	62.2
Top 35 features (P=35)	34.2	46.4	55.0	62.2
Top 25 features (P=25)	31.8	44.6	53.8	61.2
Top 15 features (P=15)	29.2	41.2	48.6	57.4
Top 5 features (P=5)	22.2	33.2	39.8	47.0
Cheng et al.(2014)	29.6	43.0	53.4	60.6

VIPeR Dataset : This dataset contains two views of 632 pedestrians. Each pair is made up of image of same pedestrians taken from arbitrary viewpoints under varying illumination conditions. Each image is 128 × 48 pixels and presents a centered unoccluded human figure, although cropped short at the feet in some side views. In the literature, results on VIPeR are typically produced by mediating over ten runs, each consisting in a partition of randomly selected 316 image pairs. For this dataset, we reported the results by computing the normalized area under curve (nAUC) value.

Fig. 7 reports the comparison of the recognition accuracy of our part-based approach with pictorial structure (PS) part-based model of Cheng et al. [3].

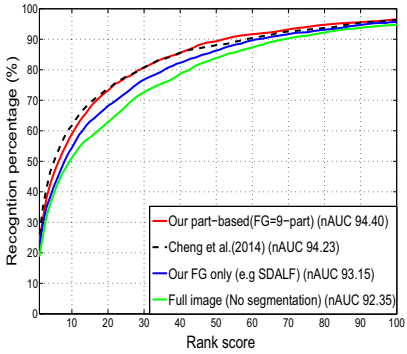

Fig. 7. CMC curves for single-shot trials on VIPeR.

In this dataset, SCA segmentation only works slightly better than pictorial models used in [3,4], but definitely works better than the background-foreground segmentation used in SDALF. Finally, our *rank-to-mask* strategy has not improved the results because the regressors overtrained. Here, in fact, we are in a single shot case and we have only a single positive instance for training.

4 Conclusions

In this work, we have proposed a discriminatively masked part-based re-identification method. For the first time, we exploited the segmentation provided by stel component analysis using a large number of parts. As second contribution, we proposed a method based on Lasso regression to rank the body parts and/or the features which makes our approach quite effective and efficient in multiple-shot scenarios while showing slightly poorer performance in the single-shot case for the datasets analyzed. Empirical results suggest that specific ranking of human body parts and associated features is a promising strategy for re-identification.

References

1. Andriluka, M., Roth, S., Schiele, B.: Pictorial structures revisited: People detection and articulated pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1014–1021 (2009)
2. Jojic, N., Perina, A., Cristani, M., Murino, V., Frey, B.: Stel component analysis: Modeling spatial correlations in image class structure. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2044–2051 (2009)
3. Cheng, D.S., Cristani, M.: Person Re-identification by articulated appearance matching. In: Person Re-Identification. Springer (2014) ISBN 978-1-4471-6295-7
4. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: British Machine Vision Conference (BMVC) (2011)

5. Bazzani, L., Cristani, M., Murino, V.: Symmetry-driven accumulation of local features for human characterization and re-identification. Computer Vision and Image Understanding **117**(2), 130–144 (2013)
6. Salvagnini, P., Bazzani, L., Cristani, M., Murino, V.: Person re-identification with a ptz camera: An introductory study. In: IEEE International Conference on Image Processing (ICIP 2013) (2013)
7. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2010)
8. Schwartz, W., Davis, L.: Learning discriminative appearance-based models using partial least squares. In: SIBGRAPI (2009)
9. Prosser, B., Zheng, W., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: British Machine Vision Conference, pp. 1–11 (2010)
10. Lin, Z., Davis, L.S.: Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Porikli, F., Peters, J., Klosowski, J., Arns, L., Chun, Y.K., Rhyne, T.-M., Monroe, L. (eds.) ISVC 2008, Part I. LNCS, vol. 5358, pp. 23–34. Springer, Heidelberg (2008)
11. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008)
12. Zheng, W., Gong, S., Xiang, T.: Associating groups of people. In: British Machine Vision Conference (2009)
13. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using haarbased and DCD-based signature. In: Workshop on Activity Monitoring by Multi-Camera Surveillance Systems (2010)
14. Sivic, J., Zitnick, C.L., Szeliski, R.: Finding people in repeated shots of the same scene. In: Proceedings of the British Machine Vision Conference (2006)
15. Satta, R., Fumera, G., Roli, F., Cristani, M., Murino, V.: A multiple component matching framework for person re-identification. In: Maino, G., Foresti, G.L. (eds.) ICIAP 2011, Part II. LNCS, vol. 6979, pp. 140–149. Springer, Heidelberg (2011)
16. Bird, N., Masoud, O., Papanikolopoulos, N., Isaacs, A.: Detection of loitering individuals in public transportation areas. IEEE Trans. Intell. Transp. Syst. **6**(2), 167–177 (2005)
17. Gheissari, N., Sebastian, T.B., Tu, P.H., Rittscher, J., Hartley, R.: Person reidentification using spatiotemporal appearance. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 1528–1535 (2006)
18. Wang, X., Doretto, G., Sebastian, T.B., Rittscher, J., Tu, P.H.: Shape and appearance context modeling. In: IEEE Intl. Conf. on Computer Vision (ICCV), pp. 1–8 (2007)
19. Hamdoun, O., Moutarde, F., Stanciulescu, B., Steux, B.: Person re-identification in multicamera system by signature based on interest point descriptors collected on short video sequences. In: ACM/IEEE Intl. Conf. on Distributed Smart Cameras (ICDSC), pp. 1–6 (2008)
20. Tuzel, O., Porikli, F., Meer, P.: Pedestrian detection via classification on riemannian manifolds. IEEE Trans. PAMI, 1713–1727 (2008)
21. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using spatial covariance regions of human body parts. In: AVSS (2010)
22. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Multiple-shot human reidentification by mean riemannian covariance grid. In: Advanced Video and Signal-Based Surveillance, Klagenfurt, Autriche (2011)

23. TibShirani, R.: Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistics Society. Series B(Methodological) **58**(1), 267–288 (1996)
24. Prosser, B., Zheng, W., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: British Machine Vision Conference, pp. 1–11 (2010)
25. Forssén, P.E.: Maximally stable colour regions for recognition and matching. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2007)
26. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition and tracking. In: IEEE Intl. Workshop on Performance Evaluation for Tracking and Surveillance (PETS) (2007)
27. VIPeR Dataset. http://vision.soe.ucsc.edu/?q=node/178
28. Caviar dataset (2004). http://homepages.inf.ed.ac.uk/rbf/CAVIAR/