Open access • Book Chapter • DOI:10.1007/978-3-642-33863-2_39

Person re-identification: what features are important? — Source link

Chunxiao Liu, Shaogang Gong, Chen Change Loy, Xinggang Lin

Institutions: Tsinghua University, Queen Mary University of London

Published on: 07 Oct 2012 - International Conference on Computer Vision

Topics: Feature (computer vision) and Matching (statistics)

Person Re-identification: What Features Are Important?

Chunxiao Liu[1], Shaogang Gong[2], Chen Change Loy[3], and Xinggang Lin[1]

[1] Dept. of Electronic Engineering, Tsinghua University, China
[2] School of EECS, Queen Mary University of London, UK
[3] Vision Semantics Ltd., UK

Abstract. State-of-the-art person re-identification methods seek robust person matching through combining various feature types. Often, these features are implicitly assigned with a single vector of global weights, which are assumed to be universally good for all individuals, independent to their different appearances. In this study, we show that certain features play more important role than others under different circumstances. Consequently, we propose a novel unsupervised approach for learning a bottom-up feature importance, so features extracted from different individuals are weighted adaptively driven by their unique and inherent appearance attributes. Extensive experiments on two public datasets demonstrate that attribute-sensitive feature importance facilitates more accurate person matching when it is fused together with global weights obtained using existing methods.

1 Introduction

Appearance-based person re-identification is a non-trivial problem owing to visual ambiguities and uncertainties caused by illumination changes, viewpoint and pose variations, and inter-object occlusions [1]. Under such stringent constraints, most existing methods [2, 3] combine different appearance features, such as colour and texture, to improve reliability and robustness in person matching. Typically, the feature histograms are concatenated and weighted in accordance to their *importance*, i.e. their discriminative power in distinguishing a target of interest from other individuals.

State-of-the-art approaches [4–7] implicitly assume a feature weighting or selection mechanism that is *global*, by assuming a single weight vector (or a linear weight function) that is globally optimal across all circumstances, e.g. colour is the most important and universally good feature across all individuals. In this study, we term this weight as global feature importance. They can be learned either through boosting [7], rank learning [4], or distance metric learning [5]. Scalability is the main bottleneck of such approaches as the learning process requires exhaustive supervision on pairwise individual correspondence.

We believe that certain appearance features can be more important than others in describing an individual and distinguishing him/her from other people. For instance, colour is more informative to describe and distinguish an individual

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part I, LNCS 7583, pp. 391–401, 2012.

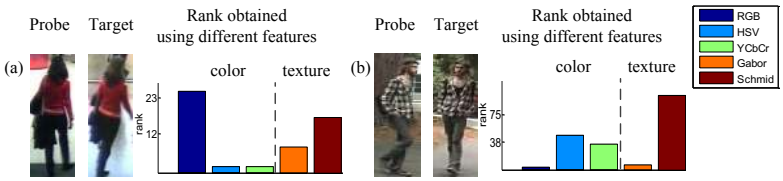

Fig. 1. We show the probe image and the target image, together with the rank of correct matching by using different feature types separately

wearing textureless bright red shirt, but texture information can be equally or more critical for a person wearing plaid shirt (Fig. 1).

Thus, it is desired not to bias all the weights to the features that are universally good for all individuals but also *selectively distribute some weights to informative feature given different appearance attributes*, which refer to appearance characteristics of individuals, e.g. dark shirt, blue jeans. This intuition is well motivated by the study in human visual attention [8], of which results suggest that visual attention is not only governed by top-down global feature importance, but also affected by bottom-up salient features of individual objects as a result of attentional competition between features.

To this end, we first investigate what features are more important under what circumstances. In particular, we show that selecting features specifically for different individuals can yield more robust re-identification performance than feature histogram concatenation with uniform weighting [9, 10]. Motivated by this observation, we propose an effective approach based on the random forest [11] to adaptively determine the feature importance of an individual driven by his/her inherent appearance attributes. Extensive experiments conducted on two challenging person re-identification datasets demonstrate that person matching can benefit from complementing existing 'global weighting' approaches with the proposed attribute-sensitive feature importance.

Related Work - Most existing approaches [4–7] can be considered as 'global weighting' approaches. For example, the RankSVM method in [4] aims to find a linear function to weight the absolute difference of samples via optimisation given pairwise relevance constraints. The Probabilistic Relative Distance Comparison (PRDC) [5] maximises the probability of a pair of true match having a smaller distance than that of a wrong matched pair. The output is an orthogonal matrix that essentially encodes the global importance of each feature.

The method proposed in [12] shares a similar spirit to our work, i.e. it aims to discover what is important given specific appearance. In contrast to [12] that requires labelled gallery images to discover gallery-specific feature importance, our method is fully unsupervised. Importantly, our method is more flexible since the feature importance is attribute-driven, thus it is not limited to specific gallery. A more recent work in [13] starts to explore prototype relevance for improving processing time in re-identification problem. In contrast, we systematically investigate salient feature importance mining for improving matching accuracy.

Contributions - (1) we draw insights into what features are more important under what circumstances. To our best knowledge, this is the first study that systematically investigates the role of different feature types given different appearance attributes; and (2) we formulate a novel unsupervised approach for on-the-fly mining of attribute-sensitive feature importance. Combining it with global feature importance leads to more accurate person re-identification while requiring no more supervision cost than existing learning-based approaches.

2 Attribute-Sensitive Feature Importance

The summary of our approach is depicted in Fig. 2. The three main steps are: (1) discovering prototypes by a clustering forest; (2) attribute-sensitive feature importance mining; (3) determining the feature importance of a probe image on-the-fly.

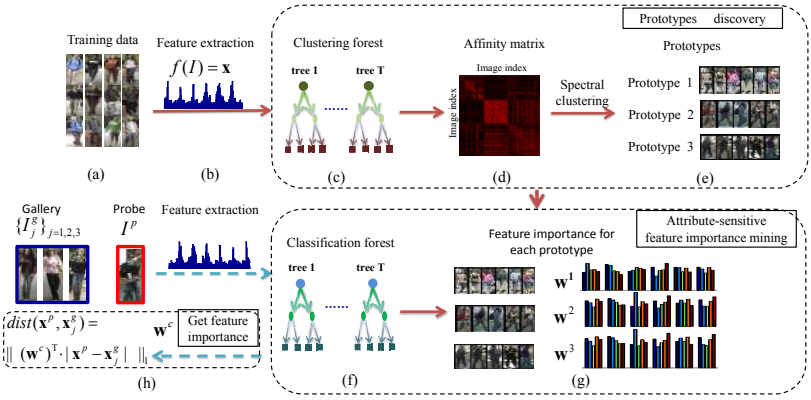

Fig. 2. Overview of attribute-sensitive feature importance mining. Training steps are indicated by red solid arrows and testing steps are denoted by blue slash arrows.

Prototypes Discovery - The first step of our method is to cluster a given set of unlabelled images into several *prototypes*, each of which compose of images that possess similar appearance attributes, e.g. wearing colorful shirt, with backpack, dark jacket (Fig. 2(e)).

Formally, given an input of n unlabelled images $\{I_i\}$, where $i = 1, \ldots, n$, feature extraction $f(\cdot)$ is first performed on every image to extract a D-dimensional feature vector, that is $f(I) = \mathbf{x} = (x_1, \ldots, x_D)^{\mathsf{T}} \in \mathbb{R}^D$ (Fig. 2(b)). We wish to discover a set of prototypes $c \in \mathcal{C} = \{1, \ldots, K\}$, i.e. low-dimensional manifold clusters that group images $\{I\}$ with similar appearance attributes. We treat the prototype discovery problem as a graph partitioning problem, which requires us to first estimate the pairwise similarity between images.

To estimate the similarity between images, we construct a clustering forest[14], an ensemble of T_{cluster} clustering trees (Fig. 2(c)). Each clustering tree t defines a partition of the input samples $\mathbf{x}$ at its leaves, $l(\mathbf{x}) : \mathbb{R}^D \rightarrow \mathcal{L} \subset \mathbb{N}$, where l

represent a leaf index and $\mathcal{L}$ is the set of all leaves in a given tree. For each tree, we compute an $n \times n$ affinity matrix A^t, with each element A^t_{ij} defined as

$$A^t_{ij} = \exp^{-\text{dist}^t(\mathbf{x}_i, \mathbf{x}_j)}, \tag{1}$$

where

$$\text{dist}^t(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 0 & \text{if } l(\mathbf{x}_i) = l(\mathbf{x}_j) \\ \infty & \text{otherwise} \end{cases}. \tag{2}$$

Following the Eqn. (2), we assign closest affinity=1 (distance=0) to samples $\mathbf{x}_i$ and $\mathbf{x}_j$ if they fall into the same leaf node, and affinity=0 (distance=∞) otherwise. To obtain a smooth forest affinity matrix, we compute the final affinity matrix as $A = \frac{1}{T_{\text{cluster}}} \sum_{t=1}^{T_{\text{cluster}}} A^t$. This method offers a few advantages as compared to conventional similarity measuring approaches: (1) avoiding manual definition of distance function since the pairwise affinities are defined by the tree structure itself, and (2) implicit selection of optimal features and corresponding forest parameters via optimisation of the well-defined clustering information gain function [11].

Given the affinity matrix, the normalised cuts algorithm [15] is employed to partition the weighted graph into K prototypes. Thus, each unlabelled probe image $\{I_i\}$ is assigned to a prototype c_i(Fig. 2(e)). In this study, K is predefined but one can estimate the cluster number automatically using alternative methods such as [16].

Attribute-Sensitive Feature Importance - As discussed in Sec. 1, unlike the global weight vector that is assumed to be universally good for all images, attribute-sensitive feature importance is specific to prototype characterised by different appearance characteristics. That is each prototype c has its own attribute-sensitive weighting $\mathbf{w}^c = (w^c_1, \ldots, w^c_D)^\mathsf{T}$, of which high value should be assigned to unique features of that prototype. For example, texture features gain higher weights than others if the images in the prototype have rich textures but less bright colours.

Based on the above intuition, we compute the importance of a feature according to its ability in discriminating different prototypes. Specifically, we train a classification random forest [11] using $\{\mathbf{x}\}$ as inputs and treating the associated prototype labels $\{c\}$ as classification outputs (Fig. 2(f)). For each tree t, we reserve $\frac{1}{3}$ of the original training data as out-of-bag (oob) validation samples. First, we compute the classification error $\epsilon_d^{c,\,t}$ for every dth feature in prototype c. Then we randomly permute the value of the dth feature in the oob samples and compute the $\widetilde{\epsilon}_d^{c,\,t}$ on the perturbed oob samples of prototype c. The importance of the dth feature of prototype c is then computed as the error gain [11]

$$w^c_d = \frac{1}{T_{\text{class}}} \sum_{t=1}^{T_{\text{class}}} (\widetilde{\epsilon}_d^{c,\,t} - \epsilon_d^{c,\,t}), \tag{3}$$

where T_{class} is the total number of trees in the classification forest. Higher value in w^c_d indicates higher importance of the dth feature in prototype c. Intuitively,

the dth feature is important if perturbing its value in the samples causes a drastic increase in classification error, which suggests its critical role in discriminating between different prototypes.

Ranking - Given feature vector of an unseen probe image $\mathbf{x}^p$, our method will determine its feature importance on-the-fly driven by its appearance. First, we classify $\mathbf{x}^p$ using the learned classification forest to obtain its prototype label c (Fig. 2(h)). Then we compute the distance $\mathbf{x}^p$ against a feature vector of a gallery/target image $\mathbf{x}^g$ using the following function

$$\text{dist}(\mathbf{x}^p, \mathbf{x}^g) = \|(\mathbf{w}^c)^\mathsf{T}|\mathbf{x}^p - \mathbf{x}^g|\|_1. \tag{4}$$

The matching ranks of $\mathbf{x}^p$ against a gallery of images can be obtained by sorting the distances computed from Eqn. (4). A smaller distance results in a higher rank.

Fusion with Global Feature Weight Vector - We investigate the fusion between the global feature weight matrix $\mathbf{V}$ obtained from existing methods [4, 5] and our attribute-sensitive feature importance vector $\mathbf{w}$ to gain more accurate person re-identification performance. We adopt a weighted sum scheme as follows

$$\text{dist}_{\text{fusion}}(\mathbf{x}^p, \mathbf{x}^g) = \alpha\|(\mathbf{w}^c)^\mathsf{T}|\mathbf{x}^p - \mathbf{x}^g|\|_1 + (1-\alpha)\|\mathbf{V}^\mathsf{T}|\mathbf{x}^p - \mathbf{x}^g|\|_1, \tag{5}$$

where α is a parameter that controls the weight between global attribute-sensitive importances.

3 Experiments

In Sec. 3.1, we first investigate the re-identification performance of using different features given individuals with different inherent appearance attributes. In Sec. 3.2, the qualitative results of prototype discovery are presented. We then compare feature importances produced by our unsupervised bottom-up solution and two top-down global weighting methods, namely RankSVM [4] and PRDC [5], in Sec. 3.3. Finally, we report the results on combining these two types of feature importance.

Datasets - Two publicly available person re-identification datasets, namely VIPeR [7] and i-LIDS Multiple-Camera Tracking Scenario (MCTS) [17] were used for evaluation. The VIPeR dataset contains 632 persons, each of which has two images captured in outdoor views. The dataset is challenging due to drastic appearance difference between most of the matched image pairs caused by viewpoint variations and large illumination changes at outdoor environment (see Fig. 3). The i-LIDS MCTS dataset was captured in a busy airport arrival hall using multiple cameras. It contains 119 people with a total of 476 images, with an average of four images per person. Apart from the illumination changes and pose variations, many images in this dataset are also subject to severe inter-object occlusions (Fig. 3(f)).

Features - We employed a mixture of colour and texture histograms similar to those employed in [4, 5]. Specifically, we divided an image of a person equally

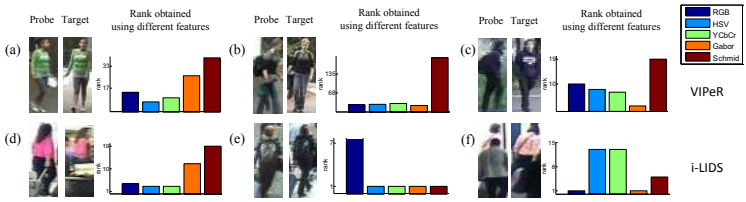

Fig. 3. In each subfigure, we show the probe image and the target image, together with the rank of correct matching by using different feature types separately.

into six horizontal stripes, to roughly capture the head, upper and lower torsos, and leg regions. In each stripe, we considered 8 colour channels (RGB, HSV and YCbCr)[1] and 21 texture filters (8 Gabor filters and 13 Schmid filters) applied to luminance channel [4]. Each channel was represented by a 16-dimensional vector. Concatenating all the feature channels resulted in 2784-dimensional feature vector for each image.

Evaluation - We used the $\ell 1$-norm as the matching distance metric. The matching performance was measured using the averaged cumulative match characteristic (CMC) curve [7] over 10 trials. The CMC curve represents the correct matching rate at the top r ranks. We selected all the images of p person to build the test set. The remaining data was used for training. In the test set of each trial, we randomly chose one image from each person to set up the test gallery set and the remaining images were used as probe images.

3.1 Performance of Using Different Features

We believe that certain features can be more important than others in describing an individual and distinguishing him/her from other people. To validate our hypothesis, we analysed the matching performance of using different features individually.

We first provide a few examples in Fig. 3 (also presented in Fig. 1) to compare the ranks returned by using different feature types. It is observed that no single feature type was able to constantly outperform the others. In the VIPeR dataset, for individuals wearing textureless but colourful and bright clothing (e.g. Fig. 3(a)), the colour features yielded a higher rank. For person wearing clothing with rich texture or with a logo, e.g. Figures 3 (b) and (c), texture features especially the Gabor features tend to dominate. The results suggest that certain features can be more informative than others given different appearance attributes.

The overall matching performance is presented in Fig. 4. In general, HSV and YCbCr features exhibited very close performances, which were much superior

[1] Since HSV and YCbCr share similar luminance/brightness channel, dropping one of them results in a total of 8 channels.

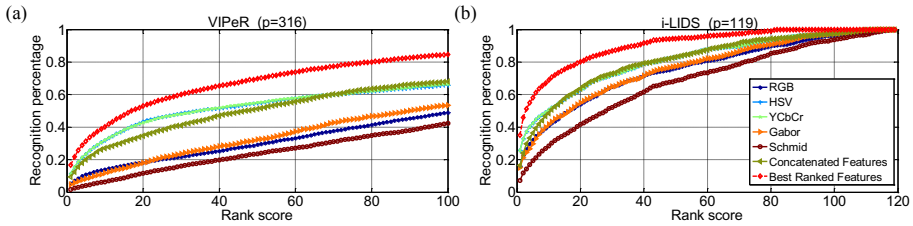

Fig. 4. The CMC performance comparison of using different features on the VIPeR and i-LIDs datasets. 'Concatenated Features' refers to the concatenation of all feature histograms with uniform weighting. In the 'Best Ranked Features' strategy, ranking for each individual was selected based on the best feature that returned the highest rank during matching.

over all other features. This observation of colours being the most informative features agreed with the past studies [7]. Simply concatenating all the feature histograms with uniform weighting did not necessary yield better performance, as can be observed in Fig. 4. The results suggest a more careful feature weighting according to their level of informativeness is necessary. The 'Best Ranked Features' strategy yielded the best performance, i.e. 13.97% and 11.31% improvement of AUC (area under curve) on the VIPeR and i-LIDS datasets, respectively, in comparison to 'Concatenated Features'. In the 'Best Ranked Features' strategy, the final rank was obtained by selecting the best feature that returned the highest rank for each individual, e.g. selecting HSV feature for Fig 3(a) whilst choosing Gabor feature for Fig 3(c). This is a heuristic way. Nevertheless, the results suggest that the overall matching performance can potentially be boosted by weighting features selectively according to the inherent appearance attributes.

3.2 Prototype Discovery

To weigh features in accordance to the inherent appearance attributes, our method first discovers prototypes, i.e. low-dimensional manifold clusters that model similar appearance attributes (see Sec. 2). The number of cluster K is set to 10 and 5 for the VIPeR and i-LIDS datasets, respectively, roughly based on the amount of training samples. We set $T_{\text{cluster}} = T_{\text{class}} = 200$. The minimum forest node size was set to 1.

Some examples of prototype discovered on the VIPeR dataset are depicted in Fig. 5. Each colour-coded row represents a prototype. A short list of possible attributes discovered in each prototype is given next to it. Note that these inherent attributes were neither pre-defined nor pre-labelled, but automatically discovered by the unsupervised clustering forest. As shown by the example members in each prototype, images with similar attributes were likely to be categorised into the same cluster. For instance, a majority of attributes in the second prototype can be characterised with bright and high contrast colour appearance. In the forth prototype, the key attributes are 'carrying backpack' and 'side pose'.

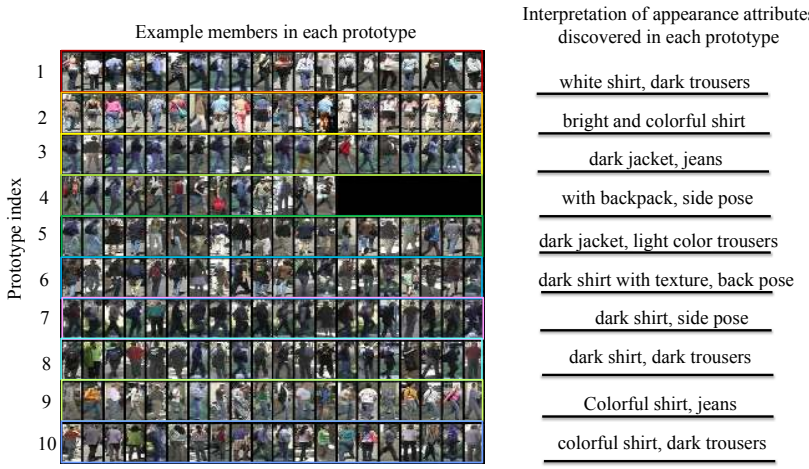

Fig. 5. Example of prototypes discovered on the VIPeR dataset. Each prototype represents a low-dimensional manifold cluster that models similar appearance attributes. Each image row in the figure shows a few examples of images in a particular prototype, with their interpreted unsupervised attributes listed on the right.

The results demonstrate that our method is capable of generating reasonably good clusters of inherent attributes, which can be employed in subsequent step for attribute-sensitive feature importance mining.

3.3 Attribute-Sensitive vs. Global Feature Importance

Comparing Global and Attribute-Sensitive Importance: The aim of this experiment is to compare the feature importances produced by existing approaches [4, 5] and the proposed attribute-sensitive feature importance mining method. Two state-of-the-art methods, i.e. the RankSVM [4] and the PRDC [5] (see Sec. 1), were evaluated using the authors'code. The global feature importances/weights were learned using the labelled images, and averaged over 10-fold cross validation. We set the penalty parameter C in RankSVM to 100 for both datasets and used the default parameter values for PRDC.

The left pane of Fig. 6 shows the feature importance discovered by both RankSVM and PRDC. For PRDC, we only show the first learned orthogonal projection, i.e. feature importance. Each region in the partitioned silhouette images were masked with the labelling colour of the dominant feature. In the feature importance plot, we show in each region the importance of each type of features. The importance of a certain feature type is derived by summing the weight of all the histogram bins belong to this type. The same steps were repeated to depict the attribute-sensitive feature importance on the right pane.

In general, the global feature importance emphasised more on the colour features for all the regions, whereas the texture features were assigned higher

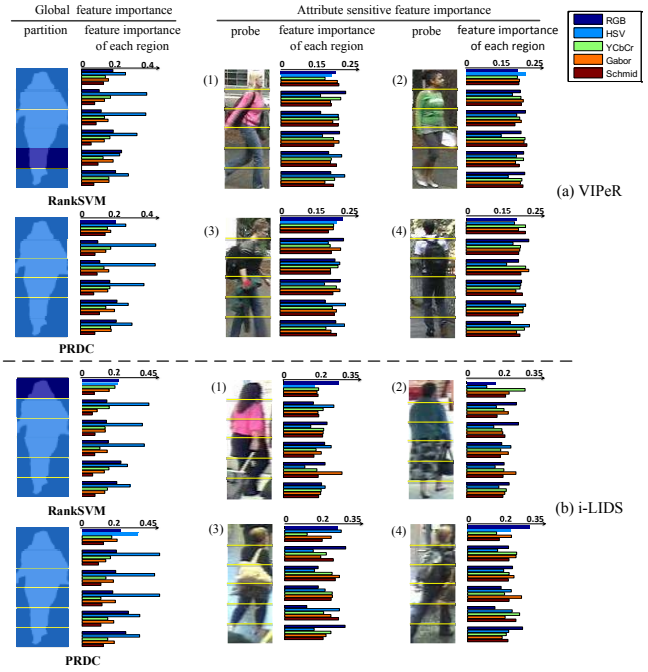

Fig. 6. Comparison of the global feature importance/weights produced by the RankSVM[4], PRDC[5], and the attribute-sensitive feature importance discovered using the proposed method

weights in the leg region than the torso region. This weight assignment or importance was applied universally to all images. In contrast, the attribute-sensitive feature importance are more person-specific. For example, for image regions with colourful appearance, e.g. Fig. 6(a)-1, the colour features in torso region were assigned with higher weights than the texture features. For image regions with rich texture, such as the stripes on the jumper (Figure. 6(a)-3), flower skirt (Figure. 6(b)-2), and bag (Figure. 6(b)-4), the importance of texture features increased. For instance, in Fig. 6(b)-2, the weight of gabor feature in the fifth region was 36.7% higher than that observed in the third region.

Table 1. Comparison of top rank matching rate (%) on the VIPeR and i-LIDS datasets. r is the rank and p is the size of gallery set.

Methods	VIPeR ($p = 316$)				i-LIDS ($p = 50$)			
	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 1$	$r = 5$	$r = 10$	$r = 20$
Uniform weight[9, 10]	9.43	20.03	27.06	34.68	30.40	55.20	67.20	80.80
Our method	9.56	22.44	30.85	42.82	27.60	53.60	66.60	81.00
RankSVM[4]	14.87	37.12	50.19	65.66	29.80	57.60	73.40	84.80
Our method+RankSVM	**15.73**	**37.66**	**51.17**	**66.27**	**33.00**	**58.40**	**73.80**	**86.00**
PRDC[5]	16.01	37.09	**51.27**	**65.95**	32.00	58.00	71.00	83.00
Our method+PRDC	**16.14**	**37.72**	50.98	**65.95**	**34.40**	**59.20**	**71.40**	**84.60**

Integrating Global and Attribute-Sensitive Importance: As shown in Table. 1, in comparison to the baseline uniform feature importance, our method yielded improved matching rate on the VIPeR dataset. No improvement was observed on the i-LIDS dataset. A possible reason is the small training size in the i-LIDS dataset, which leads to suboptimal prototype discovery. This can be resolved by collecting more unsupervised images during prototype discovery. We integrated both global and attribute-sensitive feature importance following the method described in Sec. 2 by setting $\alpha = 0.1$. An improvement as much as 3.2% on rank 1 matching rate can be obtained when we combined our method with RankSVM [4] and PRDC [5] on these two datasets. It is not surprised to observe that the supervised learning-based approaches [4, 5] outperformed our unsupervised approach. Nevertheless, the global approaches benefited from slight bias of feature weights driven by specific appearance attributes of individuals. The results suggest that these two kinds of feature importance are not exclusive, but can complement each other to gain improved matching rate.

4 Conclusion

In this study, we have shown that certain appearance features can be more important than others in describing an individual and distinguishing him/her from other people. The results suggested that instead of biasing all the weights to features that are universally good for all individuals, selectively distributing some weights to informative feature specific to certain appearance attributes can lead to better re-identification result. Future work include the investigation of better integration strategy of both global and attribute-sensitive feature importance, and incremental update of prototypes.

Acknowledgments. Chunxiao Liu was supported by NSF 61132007.

References

1. Doretto, G., Sebastian, T., Tu, P., Rittscher, J.: Appearance-based person reidentification in camera networks: problem overview and current approaches. Journal of Ambient Intelligence and Humanized Computing 2(2), 127–151 (2011)
2. Farenzena, M., Bazzani, L., Perina, A., Cristani, M., Murino, V.: Person reidentification by symmetry-driven accumulation of local features. In: CVPR, pp. 2360–2367 (2010)
3. Bazzani, L., Cristani, M., Perina, A., Murino, V.: Multiple-shot person reidentification by chromatic and epitomic analyses. Pattern Recognition Letters (2011)
4. Prosser, B., Zheng, W., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: BMVC, pp. 21.1–21.11 (2010)
5. Zheng, W., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: CVPR, pp. 649–656 (2011)
6. Mignon, A., Jurie, F.: PCCA: A new approach for distance learning from sparse pairwise constraints. In: CVPR (2012)

7. 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)
8. Yantis, S.: Control of visual attention. Attention 1, 223–256 (1998)
9. Wang, X., Doretto, G., Sebastian, T., Rittscher, J., Tu, P.: Shape and appearance context modeling. In: ICCV, pp. 1–8 (2007)
10. Loy, C.C., Xiang, T., Gong, S.: Time-delayed correlation analysis for multi-camera activity understanding. IJCV 90(1), 106–129 (2010)
11. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
12. Schwartz, W., Davis, L.: Learning discriminative appearance-based models using partial least squares. In: Proc. the 22nd Brazilian Symposium on Computer Graphics and Image Processing, pp. 322–329 (2009)
13. Satta, R., Fumera, G., Roli, F.: Fast person re-identification based on dissimilarity representations. Pattern Recognition Letters (2012)
14. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends in Computer Graphics and Vision 7(2-3), 81–227 (2012)
15. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE TPAMI 22(8), 888–905 (2000)
16. Perona, P., Zelnik-Manor, L.: Self-tuning spectral clustering. In: NIPS, pp. 1601–1608 (2004)
17. Zheng, W., Gong, S., Xiang, T.: Associating groups of people. In: BMVC, pp. 23.1–23.11 (2009)