Person Re-Identification by Efficient Impostor-based Metric Learning*

Martin Hirzer, Peter M. Roth and Horst Bischof
Institute for Computer Graphics and Vision
Graz University of Technology
{hirzer,pmroth,bischof}@icg.tugraz.at

Abstract

Recognizing persons over a system of disjunct cameras is a hard task for human operators and even harder for automated systems. In particular, realistic setups show difficulties such as different camera angles or different camera properties. Additionally, also the appearance of exactly the same person can change dramatically due to different views (e.g., frontal/back) of carried objects. In this paper, we mainly address the first problem by learning the transition from one camera to the other. This is realized by learning a Mahalanobis metric using pairs of labeled samples from different cameras. Building on the ideas of Large Margin Nearest Neighbor classification, we obtain a more efficient solution which additionally provides much better generalization properties. To demonstrate these benefits, we run experiments on three different publicly available datasets, showing state-of-the-art or even better results; however, on much lower computational efforts. This is in particular interesting since we use quite simple color and texture features, whereas other approaches build on rather complex image descriptions!

1. Introduction

The re-identification of individuals across spatially disjoint cameras has recently attracted a lot if interest in the scientific community. Especially, since the topic is practically highly relevant for security applications and statistical analyses, however, still showing a large number of open, unresolved issues. These include but are not limited to (a) the possibly extremely varying appearance of an individual across a network of cameras (changing view points, illumination, different poses, etc.), (b) the potentially high number of "similar" persons (*e.g.*, people wear rather dark clothes in winter), or (c) missing temporal and spatial constraints that could be exploited to ease the task.

*The work was supported by the Austrian Research Promotion Agency (FFG) within the project SHARE in the IV2Splus program.

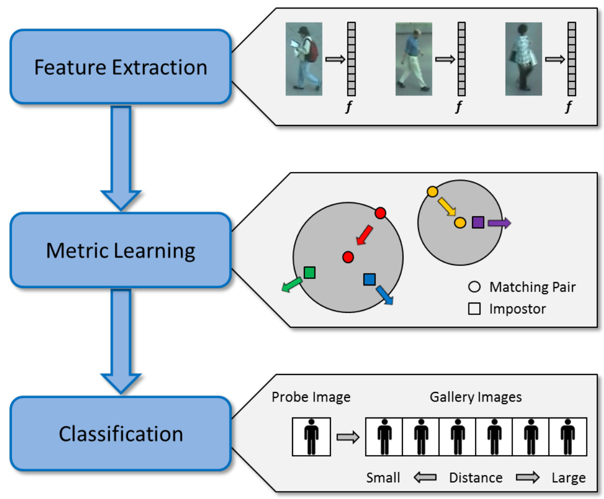


Figure 1. Proposed person re-identification system: (a) feature extraction – dense sampling of color and texture features, (b) metric learning – exploiting the structure of similar and dissimilar pairs, and (c) classification – nearest neighbor search in the projected space.

There are different ways to address these problems. For instance, descriptive methods seek a very distinctive and at the same time stable feature representation for describing a person's appearance. Gheissari *et al.* [9] try to fit a triangulated graph to each person to cope with pose variations. However, their approach works only if people are seen from similar viewpoints, which is not the case in most practical setups. The same restriction applies to the work of Wang *et al.* [21]. The image of a person is divided into regions and their color spatial structure is captured by a co-occurrence matrix. In [8], Farenzena *et al.* try to combine multiple features to describe the appearance of a person by exploiting perceptual principles. Cheng *et al.* [4] use Pictorial Structures for person re-identification. They fit a body configuration composed of chest, head, thighs and legs on pedestrian images and extract per-part color information as well as color displacement within the whole body. The extracted descriptors are then used in a matching step.

As such approaches often are not distinctive enough and often rely on hand-crafted features other methods aim at learning discriminative models. For instance, Bak *et al.* [1] first apply a person detector, and then use AdaBoost to generate a visual signature consisting of Haar-like features. Gray and Tao [11] use AdaBoost to select the most relevant features out of a set of color and texture features. To compare corresponding features they additionally estimate a likelihood ratio test providing a similarity function. Lin and Davis [15] propose to learn pairwise dissimilarities that can be applied to nearest neighbor classification. Schwartz and Davis [20] use Partial Least Squares reduction to project high dimensional signatures onto a low dimensional discriminant space. Another method is presented by Prosser *et al.* [18]. Here, the person re-identification problem is formulated as a ranking problem. The authors introduce Ensemble RankSVM, a method that learns a subspace where the potential true match gets the highest rank.

Hirzer *et al.* [13] combine both of the aforementioned strategies, *i.e.*, they apply a descriptive and a discriminative model in parallel, showing that using the complementary information captured by both models leads to improved performance. Exploiting geometry was proposed by Baltieri *et al.* [2]. They generate a highly sophisticated 3D human body model from foreground segmented person images, which can then be matched using a histogram based distance. Makris *et al.* [16] and Rahimi *et al.* [19] simplify the problem by applying temporal constraints based on the spatial layout of the observed scene. Javed *et al.* [14] try to learn transitions between cameras to cope with illumination changes, and Zheng *et al.* [23] use contextual visual information that comes from surrounding people.

A midway between descriptive and discriminative approaches is metric learning. The idea is similar to descriptive methods, *i.e.*, the data is modeled by descriptive features, however, for comparing the descriptors not the Euclidean distance is used but a metric is learned; thereby learning the camera transitions, making such approaches more suitable for real world scenarios. For instance, Dikmen *et al.* [6] learn a Mahalanobis distance that is optimal for k-nearest neighbor classification using a maximum margin formulation. Similarly, Zheng *et al.* [24] also use metric learning, but formulate it in a probabilistic manner. They seek a distance that maximizes the probability of a matching pair having a smaller distance than a non-matching pair.

In general, a key advantage of metric learning is the computational efficiency during runtime; once learned only a linear projection has to be estimated. However, as a drawback, for calculating the metric typically computationally expensive optimization problems have to be solved (*e.g.*, [22, 5, 12]). Thus, considering a typical practical setup consisting of tens or hundreds of cameras, also the training time is highly relevant!

To overcome these limitations, we propose a more efficient metric learning approach also exploiting the natural constraints given by the person re-identification task. First, since we are given image pairs from different cameras, we formulate the metric learning task as a binary problem. Second, we take into account that already well separable samples have only little influence in the optimization and can be skipped. Third, we relax the hard to fulfill positivity constraint, yielding a closed-form solution via an eigenproblem. The experimental results on three different publicly available datasets show that in this way a sufficient approximation is obtained and that we finally get state-of-the-art or even better results. This is in particular of interest, since the computational effort can be reduced by magnitudes and a rather simple image description extracting color and texture information is applied.

The rest of the paper is organized as follows. First, in Section 2 we introduce our new person re-identification system, which is then demonstrated for several publicly available datasets of different complexity in Section 3. Finally, in Section 4 we summarize and conclude the paper.

2. Person Re-ID System

In this section, we introduce our proposed person re-identification system, consisting of three stages: feature extraction and dimensionality reduction, metric learning, and classification. The overall system is illustrated in Figure 1. Once the embedding and the metric have been learned, the samples in the search space can be projected onto a lower dimensional space. During search time an unknown sample can efficiently be projected onto this space by a linear transformation followed by a ranking step. In the following, these steps are discussed in more detail.

2.1. Representation

Color and texture features have proven to be successful for the task of person re-identification. We use HSV and Lab color channels as well as Local Binary Patterns to create a person image representation. The features are extracted from 8x16 rectangular regions sampled from the image with a grid of 4x8 pixels, *i.e.*, 50% overlap in both directions. In each rectangular patch we calculate mean values per color channel, which are then discretized to the range 0 to 40. Additionally, a histogram of LBP codes is generated from a gray value representation of the patch. These values are then put together to form a feature vector. The vectors from all regions are concatenated to build a representation for the whole image. To reduce the computational effort, we apply dimensionality reduction to the feature vectors. This is done using standard PCA.

2.2. Pairwise Metric Learning

In the following, we introduce a metric learning approach for person re-identification exploiting specific constraints of the given task. Moreover, we introduce an efficient way to approximately solve the arising optimization problem. Given n data points $x_i \in \mathbb{R}^m$, one prominent approach for metric learning is to find a linear projection

$$\hat{x} = \mathbf{L}x \tag{1}$$

obtained from squared distances

$$d_{\mathbf{L}}(x_i, x_j) = ||\mathbf{L}(x_i - x_j)||^2 \ . \tag{2}$$

The matrix $\mathbf{L} \in \mathbb{R}^{m \times m}$ induces a metric if it has full rank and a pseudo-metric otherwise. If additionally class labels are given, not only the generative structure of the data but also the discriminative information can be exploited.

In our case, we are given only a huge amount of image pairs (x_i, x_j) sharing the same label y, making it hard to find a meaningful discriminative projection. Thus, we define the sets $\mathcal{S} = \{(x_i, x_j)|y(x_i) = y(x_j)\}$ and $\mathcal{D} = \{(x_i, x_j)|y(x_i) \neq y(x_j)\}$ of similar and dissimilar points, respectively, allowing us to break down the original multi-class problem to a binary problem. The goal is to minimize the distance between similar points and to maximize the distance between dissimilar points. This can be achieved using the following objective function

$$\mathcal{L}(\mathbf{L}) = \sum_{(i,j)\in\mathcal{S}} ||\mathbf{L}(x_i - x_j)||^2 - \sum_{(i,j)\in\mathcal{D}} ||\mathbf{L}(x_i - x_j)||^2 \ . \tag{3}$$

However, inspired by advanced sampling techniques (*e.g.*, [17]), we propose to ignore those samples which can already be separated well in the original space and focus on the hard samples. As illustrated in Figure 1, similar to Large Margin Nearest Neighbor (LMNN) classification [22], we define a subset $\mathcal{I}_{(i,j)} \subset \mathcal{D}$ over samples x_l (impostors), which invade the perimeter of a given pair (x_i, x_j):

$$\mathcal{I}_{(i,j)} = \left\{(x_i, x_l)| \ ||(x_i - x_l)||^2 \leq ||(x_i - x_j)||^2\right\} \ . \tag{4}$$

The conjunction of all thus obtained sets finally yields the impostor set

$$\mathcal{I} = \bigcup \mathcal{I}_{(i,j)} \ . \tag{5}$$

To take also into account how much an impostor invades the perimeter of a pair, we additionally weight the importance of an impostor:

$$w_{il} = e^{-\frac{||x_i - x_l||}{||x_i - x_j||}} \ . \tag{6}$$

Thus, we can re-write the objective function Eq. (3) to

$$\tilde{\mathcal{L}}(\mathbf{L}) = \sum_{(i,j)\in\mathcal{S}} ||\mathbf{L}(x_i - x_j)||^2 - \sum_{(i,l)\in\mathcal{I}} ||\mathbf{L} \ w_{il} \ (x_i - x_l)||^2 \tag{7}$$

and define the following optimization problem:

$$min \ \tilde{\mathcal{L}}(\mathbf{L}) \tag{8}$$

$$s.t.$$

$$\mathbf{L}\mathbf{L}^\top = \mathbf{I} \ , \tag{9}$$

where the constraint Eq. (9) avoids that the solution collapses to the trivial solution. It can be recognized that Eq. (8) and Eq. (9) define a semi-definite program (SDP) [3], which would typically be solved using an iterative procedure. However, in the following we will show that the actual problem is much simpler and can be solved more efficiently.

Thus, estimating the derivative of the Lagrange function

$$\tilde{\mathcal{L}}(\mathbf{L}) + \lambda\left(\mathbf{L}\mathbf{L}^\top - \mathbf{I}\right) \tag{10}$$

and setting it to zero, we get

$$(\mathbf{\Sigma}_S - \mathbf{\Sigma}_I)\mathbf{L} = \Lambda\mathbf{L} \ , \tag{11}$$

where

$$\mathbf{\Sigma}_S = \sum_{(i,j)\in\mathcal{S}} (x_i - x_j)(x_i - x_j)^\top \tag{12}$$

$$\mathbf{\Sigma}_I = \sum_{(i,l)\in\mathcal{I}} w_{il}(x_i - x_l)(x_i - x_l)^\top \ . \tag{13}$$

It is clear that the solution of Eq. (11) yields the eigenvectors and eigenvalues of $\mathbf{\Sigma}_S - \mathbf{\Sigma}_I$. Finally, the eigenvectors corresponding to the smallest $k << m$ eigenvalues are selected as projection matrix: $\mathbf{L}_k \in \mathbb{R}^{k \times m}$. In this way, the most informative directions according to the objective function Eq. (7) are captured. Contrary, since $\mathbf{L}$ is an orthonormal matrix, keeping all m eigenvectors would rotate the feature space but the original distances would be preserved.

2.3. Classification

In person re-identification we want to recognize a certain person across different camera views. In the work at hand, we assume that we have already detected the persons in all camera views, *i.e.*, we do not tackle the detection problem.

Thus, the goal of person re-identification now is to find a person image that has been selected in one view (*probe image*) in all the images from another view (*gallery images*). This is achieved by calculating the distance between the probe image and all gallery images using the learned metric, and returning those gallery images with the smallest distance as potential matches.

3. Experimental Results

We evaluated our approach on three publicly available datasets, the VIPeR dataset [10], the PRID 2011 dataset (single shot version) [13], and the ETHZ dataset [20]. We chose these datasets because they provide many challenges faced in real world person re-identification applications, *e.g.*, viewpoint, pose and illumination changes, different backgrounds, image resolutions, occlusions, etc. For the VIPeR dataset we reduced the feature vectors described in Section 2.1 to 100 dimensions. Since the other datasets contain significantly less training samples, we used fewer dimensions there in order to avoid overfitting (80 for PRID 2011 and ETHZ SEQ. #1, 40 for ETHZ SEQ. #2 and ETHZ SEQ. #3). Interestingly, on PRID 2011 we achieved better results by using color features alone, *i.e.*, no LBP codes. This has also been reported in [13]. The number of eigenvectors k used for building the projection matrix is set to half of the feature dimension in all experiments.

3.1. VIPeR Dataset

The VIPeR dataset contains 632 person image pairs taken from two different camera views. Changes of viewpoint, illumination and pose are the most prominent sources of appearance variation between the two images of a person. For evaluation we followed the procedure described in [11]. The set of 632 image pairs is randomly split into two sets of 316 image pairs each, one for training and one for testing. In the test case, the two images of an image pair are randomly assigned to a probe and a gallery set. A single image from the probe set is then selected and matched with all images from the gallery set. This process is repeated for all images in the probe set. The whole evaluation procedure is carried out 10 times, and the average result is reported in form of a Cumulative Matching Characteristic (CMC) curve [21], which represents the expectation of finding the true match within the first r ranks.

The thus obtained results are shown in Figure 2a. For comparison we also give results obtained by using all point pairs of the class $\mathcal{D}$ (Diff) as stated in Eq. (3) (*i.e.*, no impostors), Linear Discriminative Analysis (LDA) projection, which could be seen as basic metric learning baseline, and simple feature matching using Euclidean distance. It is obvious that using the proposed method leads to a huge performance gain over LDA projection and simple feature matching. Furthermore, specifically sampling those points that invade the perimeter of a pair leads to a much higher performance than using all point pairs of class $\mathcal{D}$ (Diff). In Table 1 we additionally show a comparison of our to other person re-identification methods. As can be seen, our method outperforms the state-of-the-art, however, at a much reduced computational cost. The results for LMNN [22] and ITML [5] were obtained using exactly the same features and training and test set splits as for our method. The results for the remaining methods originate from the corresponding papers.

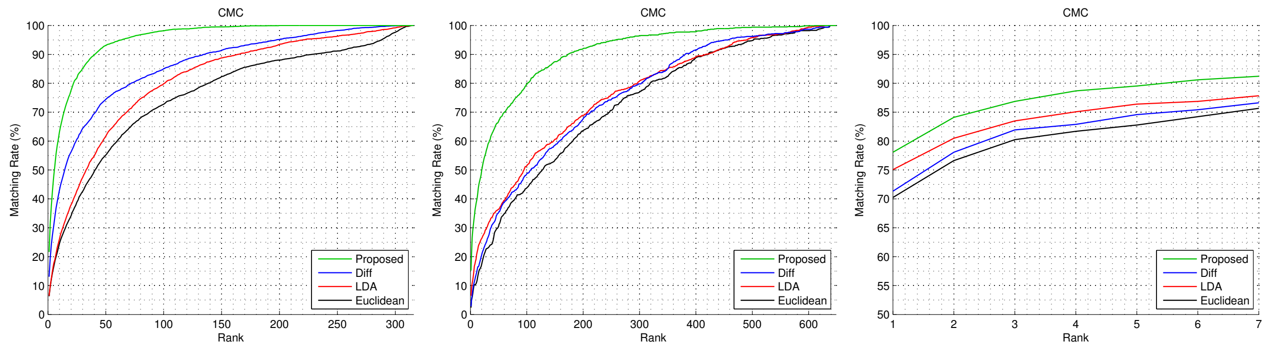
In Table 1 we also analyze the computation time of our method using a Matlab implementation on a 2.83 GHz quad core CPU. The big advantage of our approach compared to others is its training time efficiency, since it does not rely on computationally complex optimization schemes. Using the learned metric in the evaluation step is efficient either, making it suitable for even large scale problems.

Method	r = 1	10	20	50	100	t_{train}
Proposed	**22**	**63**	**78**	**93**	**98**	**0.3 sec**
LMNN [22]	17	54	69	88	96	2 min
ITML [5]	13	53	71	90	97	25 sec
ELF [11]	12	43	60	81	93	5 hrs
SDALF [8]	20	50	65	85	-	-
ERSVM [18]	13	50	67	85	94	13 min
DDC [13]	19	52	65	80	91	-
PS [4]	**22**	57	71	87	-	-
PRDC [24]	16	54	70	87	97	15 min
LMNN-R* [6]	20	68	80	93	99	-

Table 1. Comparison of matching rates in [%] at different ranks r and, if available, average training times per trial on the VIPeR dataset. (* Indicates that the best run was reported, which cannot directly be compared to the other results!)

3.2. PRID 2011 Dataset

The PRID 2011 dataset consists of person images recorded from two different static surveillance cameras. Two scenarios are provided, a multi shot (multiple images per person in each camera view) and a single shot (one image per person in each camera view) scenario. We used the latter one for our evaluation. Typical challenges on this dataset are viewpoint and pose changes as well as significant differences in illumination, background and camera characteristics. Camera view A contains 385 persons, camera view B contains 749 persons, with 200 of them appearing in both views. Hence, there are 200 person image pairs in the dataset. These image pairs are randomly split into a training and a test set of equal size. For evaluation on the test set, we followed the procedure described in [13], *i.e.*, camera A is used for the probe set and camera B is used for the gallery set. Thus, each of the 100 persons in the probe



(a) VIPeR (b) PRID 2011 (c) ETHZ SEQ. #1

Figure 2. Average CMC curves of our approach, the eigenvalue problem according to Eq. (3), LDA, and feature matching using Euclidean distance on (a) the VIPeR, (b) the PRID 2011, and (c) the ETHZ SEQ. #1 dataset.

set is searched in a gallery set of 649 persons (all images of camera view B except the 100 training samples). Again, the whole procedure is repeated 10 times and the result is reported in form of an average CMC curve in Figure 2b. As can be seen, applying the proposed metric leads to superior performance compared to using LDA, Euclidean distance, or directly solving Eq. (3).

In Table 2 we compare our approach to the descriptive model of Hirzer *et al.* [13], which also uses the single shot setup. As can be seen by the numbers, our method clearly outperforms their model at all ranks.

Method	$r = 1$	10	20	50	100
Proposed	**15**	**38**	**50**	**67**	**80**
Descr. Model [13]	4	24	37	56	70

Table 2. Matching rates of our approach and the descriptive model of [13] in [%] at different ranks r on the PRID 2011 dataset.

3.3. ETHZ Dataset

The ETHZ dataset [7] contains video sequences of urban scenes captured from moving cameras. It has originally been proposed for pedestrian detection, but Schwartz and Davis [20] provide a modified version of the dataset for the task of person re-identification. This version consists of person images extracted from three video sequences structured as follows: SEQ. #1 contains 83 persons (4.857 images), SEQ. #2 contains 35 persons (1.961 images), and SEQ. #3 contains 28 persons (1.762 images). All images have been resized to 64x32 pixels. The most challenging aspects of the ETHZ dataset are illumination changes and occlusions. However, since person images are captured from a single moving camera, the dataset does not provide a realistic scenario for person re-identification with multiple disjoint cameras, different viewpoints, different camera characteristics, etc. Despite this limitation it is commonly used for person re-identification.

We use a single shot evaluation strategy, *i.e.*, we randomly sample two images per person to build a training pair, and another two images to build a test pair. The images of the test pairs are then assigned to the probe and the gallery set. After learning a metric, each image in the probe set is matched with all images in the gallery set. The whole procedure is repeated 10 times to generate an average result. The CMC curve for SEQ. #1 is shown in Figure 2c, the results for all three sequences compared to two other methods also using a single shot evaluation, *i.e.*, SDALF [8] (single shot version) and PLS [20], are provided in Table 3.

Method	$r = 1$	2	3	4	5	6	7
	SEQ. #1						
Proposed	78	84	**87**	**89**	**90**	**91**	**91**
SDALF [8]	65	73	77	79	81	82	84
PLS [20]	**79**	**85**	86	87	88	89	90
	SEQ. #2						
Proposed	**74**	**81**	**84**	**87**	**89**	**91**	**92**
SDALF	64	74	79	83	85	87	89
PLS	**74**	79	81	83	84	85	87
	SEQ. #3						
Proposed	**91**	**95**	**97**	**98**	**98**	**98**	**99**
SDALF	76	83	86	88	90	92	93
PLS	77	81	82	84	85	87	89

Table 3. Matching rates of our approach, SDALF (single shot), and PLS in [%] on the ETHZ dataset at the first 7 ranks.

Our results on SEQ. #1 are comparable to PLS [20] and clearly outperform SDALF [8]. Interestingly, even matching image descriptors directly (Euclidean distance) slightly outperforms SDALF. On the other two sequences, our method also delivers state-of-the-art performance (SEQ. #2) or better (SEQ. #3). However, since the number of training samples is very small on these two sequences, learning a metric gets difficult, and compared to LDA or Euclidean distance matching we get only a little performance gain.

4. Conclusion

Recently, metric learning was introduced for the task of person re-identification, which is a considerable tradeoff between descriptive and discriminative modeling. Even showing good results and being effective during evaluation, the computational costs for training are quite high. We target this problem by taking into account constraints given by the task and apply an efficient closed-form solution for the arising optimization problem. Given labeled image pairs, we search for a linear projection that keeps similar pairs together and pushes impostors, *i.e.*, samples that invalidate the perimeter of a similarly labeled image pair, away. In fact, it is well known from sampling theory that already well separable samples provide only little information for a discriminative model. Finally, we obtain an eigenproblem, which can efficiently be solved in closed-form. The experimental results on three different datasets demonstrate the strength of impostor information. We show that using the learned metric state-of-the-art and even better results can be obtained; however, on a much lower computational effort and even using only simple color and texture features. Nevertheless, future work would include a study of more sophisticated features and a generalization of the applied metric learning approach.

References

[1] S. Bak, E. Corvee, F. Brémond, and M. Thonnat. Person re-idendification using Haar-based and DCD-based signature. In *Workshop on Activity Monitoring by Multi-Camera Surveillance Systems*, 2010.

[2] D. Baltieri, R. Vezzani, and R. Cucchiara. Sarc3d: A new 3d body model for people tracking and re-identification. In *Proc. Int'l Conf. on Image Analysis and Processing*, 2011.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[4] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *Proc. British Machine Vision Conf.*, 2011.

[5] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *Proc. Int'l Conf. on Machine Learning*, 2007.

[6] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. In *Proc. Asian Conf. on Computer Vision*, 2010.

[7] A. Ess, B. Leibe, and L. V. Gool. Depth and appearance for mobile scene analysis. In *Proc. IEEE Int'l Conf. on Computer Vision*, 2007.

[8] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2010.

[9] N. Gheissari, T. B. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006.

[10] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proc. IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, 2007.

[11] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proc. European Conf. on Computer Vision*, 2008.

[12] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? Metric learning approaches for face identification. In *Proc. IEEE Int'l Conf. on Computer Vision*, 2009.

[13] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *Proc. Scandinavian Conf. on Image Analysis*, 2011.

[14] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005.

[15] Z. Lin and L. S. Davis. Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance. In *Advances Int'l Visual Computing Symposium*, 2008.

[16] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.

[17] J.-H. Park and Y.-K. Choi. On-line learning for active pattern recognition. *IEEE Signal Processing Letters*, 3(11):301–303, 1996.

[18] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *Proc. British Machine Vision Conf.*, 2010.

[19] A. Rahimi, B. Dunagan, and T. Darrell. Simultaneous calibration and tracking with a network of non-overlapping sensors. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.

[20] W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. In *Proc. Brazilian Symposium on Computer Graphics and Image Processing*, 2009.

[21] X. Wang, G. Doretto, T. B. Sebastian, J. Rittscher, and P. H. Tu. Shape and appearance context modeling. In *Proc. IEEE Int'l Conf. on Computer Vision*, 2007.

[22] K. Q. Weinberger and L. K. Saul. Fast solvers and efficient implementations for distance metric learning. In *Proc. Int'l Conf. on Machine Learning*, 2008.

[23] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *Proc. British Machine Vision Conf.*, 2009.

[24] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2011.