

CAMERA COMPENSATION USING FEATURE PROJECTION MATRIX FOR PERSON RE-IDENTIFICATION

Yimin Wang¹, Ruimin Hu¹, Chao Liang¹, Chunjie Zhang², Qingming Leng¹

¹ National Engineering Research Center for Multimedia Software, School of Computer,
Wuhan University, Wuhan, 430072, China

² School of Computer and Control Engineering, University of Chinese Academy of Sciences,
Beijing, 100190, China

ABSTRACT

Matching individuals across a group of spatially non-overlapping surveillance cameras, also known as person re-identification, has recently attracted a lot of research interests. Current methods mainly focus on feature extraction or metric learning, which directly compare person images captured by different cameras, but seldom consider device differences caused by various surveillance conditions, e.g. view switching, scale zooming and illumination variation. Although brightness transfer function was proposed to address the problem of illumination variation, it could not handle view and scale changes among various cameras. In this paper, we propose an effective data-driven method to conquer device differences in the practical surveillance camera network. More precisely, with the help of a set of labelled pair-wise person images captured by two disjoint cameras, a feature projection matrix can be learned to project the person images of one camera to the feature space of the other camera, and thus images from these two different cameras can be accurately compared in a *common feature space*. Extensive comparative experiments conducted on three standard datasets have shown the promising prospect of our proposed methods.

Index Terms— Person re-identification, non-overlapping camera tracking, feature projection matrix

1. INTRODUCTION

Recently, more and more non-overlapping camera networks have been deployed to monitor pedestrian activities over a large area, such as airports, metro stations and parking lots. Therefore matching individuals across these spatially non-overlapping surveillance cameras, also known as person re-identification, is becoming a hot research spot in the computer vision community [1–4]. However, person re-identification remains an unsolved problem due to the challenges caused by view variation, scale zooming and illumination change (see Fig. 1), which make different persons appear more alike than the same person in various cameras.



Fig. 1. Examples of appearance changes caused by variant views, lighting and scales. Each column shows two images of the same person taken from two different cameras.

Current research on person re-identification generally falls into two categories, *feature-based* and *metric-based* methods [3]. The former aims at finding a set of discriminative and robust features which accurately indicate the person identity in various camera environments. Gheissari et al. used a spatial-temporal segmentation algorithm to generate salient edges and obtained an invariant identity signature by combining normalized color and salient edge histograms [5]. An appearance model using a co-occurrence matrix to capture the spatial distribution of the appearance relative to each of object parts was studied in [6]. Farenzena et al. [1] tried to combine multiple features to describe the appearance image, which was divided into regions by exploiting symmetry and asymmetry perceptual principles. However, under the conditions where view changes can cause significant appearance variations, designing a set of discriminative and robust features is of great difficulty [3].

To solve the above problem, *metric-based* methods were proposed to seek a suitable distance metric function rather than feature representation for person re-identification task. Dikmen et al. [7] improved the Large Margin Nearest Neighbor Classification (LMNN) algorithm for person re-identification by exploiting a fixed bound for neighbors. Zheng et al. [3] proposed a relative distance compari-

son formulation and used logistic function for soft measure. The Gaussian distribution was used to fitting the distribution of pair-wise samples [8], with which the author got a simpler metric function without iterative procedures and called it KISSME. These methods generally adopted a Mahalanobis-like distance. Specially, given two person image feature o_a and o_b , their distance is defined as $D(o_a, o_b) = (o_a - o_b)^\top \mathbf{M} (o_a - o_b)$, where \mathbf{M} is a positive semi-definite matrix for the validity of metric. By performing eigenvalue decomposition on \mathbf{M} with $\mathbf{M} = \mathbf{L}^\top \mathbf{L}$, the above distance can be rewritten as $D(o_a, o_b) = \|\mathbf{L} \cdot (o_a - o_b)\|_2^2 = \|\mathbf{L} \cdot o_a - \mathbf{L} \cdot o_b\|_2^2$. With this definition, it is easy to see that the essence of the metric-based method is to seek a projection matrix that transforms original image features into a new feature space, where feature distance of the same person is smaller than that of different persons. In particular, the above distance definition applies the same feature transformation to both o_a and o_b , which implies a latent assumption that o_a and o_b belong to the same feature space.

However, the above assumption is not always valid, especially in the practical surveillance application where disjoint cameras are installed in various poses, scales and illumination conditions. In this paper, we propose a feature projection matrix (FPM) method to compensate the complex environment difference between two cameras. Specifically, given two person images (or image features) o_a and o_b captured by camera C_a and C_b , their feature distance is defined as $D(o_a, o_b) = \|\mathbf{T}_{a,b} \cdot o_a - o_b\|_2^2$, where $\mathbf{T}_{a,b}$ represents the image feature FPM from C_a to C_b . With the help of labeled images of pedestrians across C_a and C_b , which make the $\mathbf{T}_{a,b}$ can compensate the difference caused by different devices and surveillance environment effectively, we can learn an optimal FPM using the logistic loss function [3, 9]. And then using FPM, the images from different cameras can be directly compared.

A similar idea of the transfer function has been investigated by some early person re-identification methods. In [10], brightness transfer function (BTF) f_{ab} was used to compensate the different illumination conditions of different cameras. It assumed that the percentage of pixels in an observation o_a with the brightness value less than B_a is equal to the percentage of image points seen in o_b of brightness no more than B_b . More specifically, denoting a person image as I , the count of brightness value B in I as $I(B)$, and the cumulative histogram as $H(I) = \sum I(B)$, their assumption can be formulated as $H_a(B_a) = H_b(B_b)$. Thereafter, f_{ab} can be computed by mapping an observed color value in camera C_a to the corresponding observation in camera C_b as follows

$$f_{ab} = H_b^{-1}(H_a(B_a)) \quad (1)$$

where $H^{-1}(\cdot)$ represents the inverted cumulative histogram. Prosser et al. [11] extended the BTF to a cumulative brightness transfer function (CBTF), in which a bi-directional matching was used to avoid false positives. However, the

BTF-based methods have some limitations. (1) It only addresses the problem caused by illumination changes, while cannot handle more complex appearance variation caused by view changes, because they need the assumption for brightness. (2) It just uses positive samples corresponding to image pairs of the same person in BTF learning, while ignores negative samples of different persons.

Compare to BTF and its derivatives, the proposed method has two obvious advantages. (1) It does not need the base assumption of BTF, and more feasible and efficient when the difference caused by not only illumination changes but also variant view and scale. (2) The learning of FPM not only takes into account image pairs from the same person, but also those from different persons, which make the result more accurate.

2. THE APPROACH

This section presents our approach. Firstly, a formulation introduce for person re-identification is given. Then, the FPM is introduced followed by defining a new feature distance function based FPM. Finally, a solution to learn the FPM from the labeled image pairs through optimizing a logistic loss function by the gradient-descent method is given.

2.1. Person re-identification problem

For the convenience of following discussion, we consider a pair of cameras C_a and C_b with non-overlapping field of views, and further assume that $O_a = \{o_a^1, o_a^2, \dots, o_a^m\}$ and $O_b = \{o_b^1, o_b^2, \dots, o_b^n\}$ are two sets of person images captured by C_a and C_b , respectively. The purpose of the person re-identification task is that for each instance o_a^i in O_a , finds the same person images from O_b . This problem is commonly addressed by computing a distance between o_a^i and o_b^j using a distance function with which the distance between images of the same person is relatively small. Usually, the instance is represented by a d -dimensional feature vector, then the Euclidean distance can be formulated as

$$D(o_a^i, o_b^j) = (o_a^i - o_b^j)^\top (o_a^i - o_b^j) = \|o_a^i - o_b^j\|_2^2 \quad (2)$$

where $(\cdot)^\top$ represents the transpose of a vector or matrix.

An underlying assumption of Eq. (2) is that both o_a^i and o_b^j are in the same feature space, otherwise the difference by directly subtract o_b^j from o_a^i does not make sense and inaccurate. However, considering the device difference in a practical surveillance network, person images captured by various cameras hardly locate in a common feature space due to poses variation, scale zooming and illumination change.

2.2. Distance measure based on FPM

To solve the above problem, a feature projection matrix, which maps persons from one camera to the other, is intro-

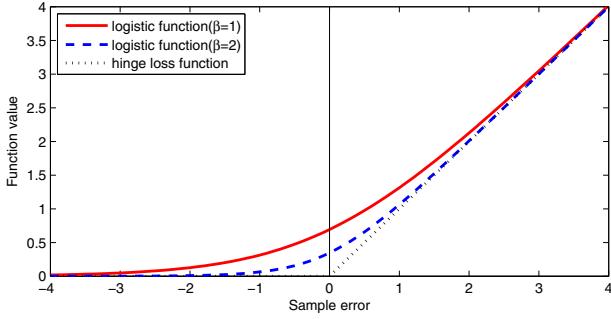


Fig. 2. Logistic loss function.

duced to compensate the discrepancy of the surveillance environment. More especially, assume that $\mathbf{T}_{a,b}$ is the feature projection matrix from C_a to C_b , then $(o_a^i)_b = \mathbf{T}_{a,b} \cdot o_a^i$ is the feature presentation of o_a^i transferred from C_a to C_b , where $\mathbf{T}_{a,b}$ is a $d \times d$ matrix.

Then the new feature distance between o_a^i and o_b^j can be defined as

$$D_{\mathbf{T}_{a,b}}(o_a^i, o_b^j) = \| (o_a^i)_b - o_b^j \|_2^2 = \| \mathbf{T}_{a,b} o_a^i - o_b^j \|_2^2 \quad (3)$$

Obviously, when FPM is a identity matrix, it degenerate into a direct comparison as shown in Eq. (2).

Given the above new distance function, the person re-identification problem is now converted into learning FPM.

2.3. Objective function for FPM learning

The learning process of FPM is similar to that of metric learning [3]. Specifically, we introduce a triple sample $(o_a^i, o_b^{j+}, o_b^{j-})$, where o_a^i is an instance of person captured by C_a while o_b^{j+} and o_b^{j-} are instances of person captured by C_b , and o_a^i and o_b^{j+} are instances coming from the same person, while o_a^i and o_b^{j-} are instances from different persons. Consider matching o_a^i with instances in camera C_b , we call the o_b^{j+} is a positively related image of o_a^i , while o_b^{j-} is a negatively related image. Then, the training set can be denoted as $S = \{(o_a^i, o_b^{j+}, o_b^{j-})_k | k = 1, \dots, s\}$, where s is the size of the training set. We define the error function for one training sample $(o_a^i, o_b^{j+}, o_b^{j-})$ as

$$e(o_a^i, o_b^{j+}, o_b^{j-}) = D_{\mathbf{T}_{a,b}}(o_a^i, o_b^{j+}) - D_{\mathbf{T}_{a,b}}(o_a^i, o_b^{j-}) \quad (4)$$

Seeking the optimal value of $\mathbf{T}_{a,b}$ can be done by minimizing the following objective function

$$\min_{\mathbf{T}_{a,b}} E(\mathbf{T}_{a,b}) = \sum_{k=1}^s \ell_\beta(e((o_a^i, o_b^{j+}, o_b^{j-})_k)) \quad (5)$$

where $\ell_\beta(x) = \frac{1}{\beta} \log(1 + e^{\beta x})$ is the generalized logistic loss function. The objective function thus penalizes the samples of

Algorithm 1 Learning the FPM

Input: The training data $S = \{(o_a^i, o_b^{j+}, o_b^{j-})_k\}$

1: Initialize $\mathbf{T}_{a,b}^0$ as identical matrix;

2: **for** $i = 1$ to $MaxIter$ **do**

3: Compute $\nabla E(\mathbf{T}_{a,b}^i) = \frac{\partial E(\mathbf{T}_{a,b}^i)}{\partial \mathbf{T}_{a,b}^i}$ as Eq.(6)

4: Choose a proper step λ

5: Compute $\mathbf{T}_{a,b}^{i+1} = \mathbf{T}_{a,b}^i - \lambda \nabla E(\mathbf{T}_{a,b}^i)$ as Eq.(7)

6: **if** converge **then**

7: break;

8: **end if**

9: **end for**

Output: The optimal FPM $\mathbf{T}_{a,b}^*$

which the error is bigger than 0. The logistic loss function is a smooth approximation of the hinge loss $h(x) = \max(0, x)$ to which it converges asymptotically as the sharpness parameter β increase [9], i.e. $\lim_{\beta \rightarrow \infty} \ell_\beta(x) = h(x)$ (see Fig.2). In this paper, the β is fixed to 1 with reference to [3].

2.4. Optimization algorithm

Before discussing the detailed learning algorithm, it is worthy noting that the Semi-Definite Programming constraint, which is indispensable in metric learning, is not needed in our solution, for the distance of our definition is default greater than zero.

Since the logistic loss function is convex, Eq. (5) is a convex optimization problem with respect to $\mathbf{T}_{a,b}$. Consequently, the problem in Eq. (5) can be solved using a gradient-descent method. The gradient of the objective function $E(\mathbf{T}_{a,b})$ is given as

$$\frac{\partial E(\mathbf{T}_{a,b})}{\partial \mathbf{T}_{a,b}} = 2 \sum_{k=1}^s g(e(S_k))(o_b^{j-} - o_b^{j+})_k (o_a^i)_k^\top \quad (6)$$

where $g(x) = (1 + e^{-x})^{-1}$ is the derivative of logistic loss function $\ell_\beta(x)$ for $\beta = 1$, S_k is the k -th sample in the training set.

With the gradient, an iterative optimization algorithm can be used to learn the FPM. Starting from an initial identical matrix, which means no projection to instance, the FPM is optimized iteratively with the gradient as follows

$$\mathbf{T}_{a,b}^{i+1} = \mathbf{T}_{a,b}^i - \lambda \cdot \frac{\partial E(\mathbf{T}_{a,b}^i)}{\partial \mathbf{T}_{a,b}^i} \quad (7)$$

where $\lambda > 0$ is a step length automatically determined at each gradient update step using a similar strategy in [3]. The iteration of the algorithm is terminated when the update times are greater than the maximum iterative times (i.e. 1000 in this work) or the following criterion is met

$$|E_{i+1} - E_i| < \varepsilon \quad (8)$$

Table 1. Top ranked matching rate(%) on VIPeR

Methods	rank@1	5	10	20	50
Ours	15.1	41.6	58.2	76.4	91.5
SDALF[1]	20	40	50	65	85
PRDC[3]	15.7	38.4	53.9	70.1	87
RankSVM[2]	16.3	38.2	53.7	69.9	85
ELF[15]	12	28	43	60	81
DDC[13]	19	40	52	65	80
DAM[16]	14	39	53	71	89

where ε is a small positive value set to 10^{-9} in this paper. The complete algorithm flow is showed in Algorithm 1.

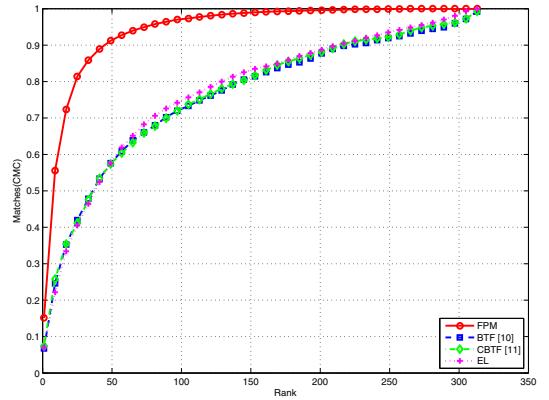
3. EXPERIMENTAL RESULTS

In this section, the proposed approach is validated by comparing with several state-of-the-art person re-identification methods on three publicly available datasets, the VIPeR dataset [12], the PRID 2011 dataset (single shot version) [13] and the 3DPes dataset [14]. The reason of selecting these three datasets are as follows. (1) These datasets cover a wide range of problems faced in the real world person re-identification applications, e.g. viewpoint, pose, and lighting changes. (2) They provide two labeled image sets of persons captured by two cameras with non-overlapping fields of views, in which the images of the same person have the same label, while the images of the different persons have different labels.

3.1. Implementation details

To represent the appearance image of people, a mixture of color and texture features extracted in overlapping blocks was used similar to [8]. Specially, a person image is divided into overlapping blocks of size 16×16 and stride of 8×8 , i.e. 50% overlap in both directions. For each block, the RGB, HSV and Lab color features and LBP texture feature are extracted and represented by histograms. The features are then put together to form a vector. The vectors from all regions are concatenated to build a representation for the whole image. For the computational efficiency, the dimensionality of the concatenated descriptors are reduced to 50 (if no special instructions) by principle component analysis (PCA).

The results are shown in terms of recognition rate by the Cumulative Matching Characteristic (CMC) curve suggested in [6] and exploited by most papers on the person re-identification problem [1–3, 15, 16]. The CMC curve represents the expectation of finding the correct match in the top n matches. Specifically, let $P = \{p_1, \dots, p_{|P|}\}$ be a probe set, where $|P|$ is the size of P , and $G = \{g_1, \dots, g_n\}$ a gallery set. For each probe image $p_i \in P$, all gallery images $g_j \in G$ are ranked by comparing the distance between p_i and g_j in ascending order. The image of the same person p_i in the gallery

**Fig. 3.** Comparative results on VIPeR dataset

set is denoted as g_{p_i} , and the index of which in the sorted gallery is denoted as $r(g_{p_i})$. The CMC value of rank k is defined as

$$CMC_k = \frac{\sum_{i=1}^{|P|} \mathbf{1}(r(g_{p_i}) \leq k)}{|P|} \quad (9)$$

where $\mathbf{1}(\cdot)$ is the indicator function.

3.2. VIPeR dataset

The VIPeR dataset contains two views of 632 persons. Each pair is made up of images of the same person taken from two different cameras, under different viewpoint, pose and light conditions. View changes was the most significant cause of appearance change with most of the matched image pairs containing one front/back view and one side-view. All images are normalized to 128×48 pixels. Most of the examples contain a viewpoint change of 90 degrees.

Comparing to the state-of-the-art and BTF. For comparing our method with the results reported by state-of-art methods in their paper, the parameters were set as $p = 316$ and $q = 20$ following Sec. 3.1. The comparison result with the top 50 ranks is showed in Table 1. As can be seen from the table, the proposed approach obtains competitive results across all ranks even though only the standard Euclidean distance was exploited after projection without studding a more discriminative metric function.

Moreover, a further comparing for the performance of our approach to BTF, CBTF and Euclidean distance (EL) is given in Fig. 3. For BTF and CBTF, the images were first transferred from Camera A to B, then the same feature representation described in Sec. 3.1 was exploited. It is obvious that using proposed method leads to a very large performance gain over BTF and CBTF. BTF and CBTF do not improve and even worse than the Euclidean distance. The reason may be that pose changes are the most significant cause of appearance change with most of the matched image pairs containing one front/back view and one side-view, and only a few image pairs have an distinct light changes.

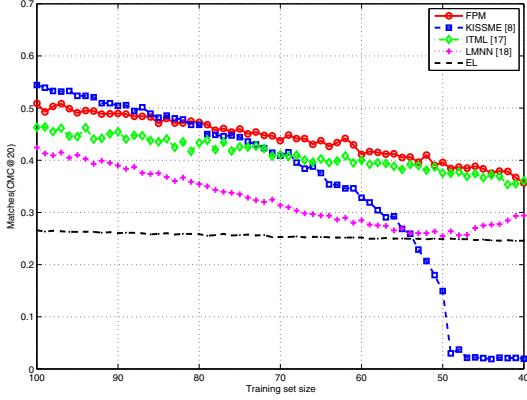


Fig. 4. Matching rate with different training data sizes

Influence of the training data size. In a real person re-identification system, it is usually expensive or difficult to obtain sufficient labelled images for training. Therefore, the property of robust performance under rare training samples is of great importance. Since the state-of-the-art person re-identification methods do not evaluate the efficiency with rare training set, our approach was validated by comparing to the state-of-the-art metric learning methods, i.e. Information Theoretic Metric Learning (ITML) [17], LMNN [18], KISSME [8], which are often used as baselines in [3, 4, 7, 9, 16]. Following the procedure in Sec. 3.1, the parameters were set as $p = 30, 31, \dots, 100$ and $q = 20$. For each p , the procedure was repeated 50 times, and the top ranking 20 results are given in Fig. 4. As can be seen, our method exhibits robust performance under sever decrease of the training set. In particular, the matching rate of KISSME falls sharply as reducing the training set, because it needs enough training data to estimate the distribution of samples.

Influence of the feature dimension. The influence of feature dimension to the algorithm performance was further evaluated. Using PCA, the descriptors was projected into a low dimensional subspace from 25 to 300. Following the procedure of Sec. 3.1, the parameters were set as $p = 316$ and $q = 20$. For each p , the procedure was repeated 50 times, and the top ranking 20 results is shown in Fig. 5. As can be seen, our method is stable at different feature dimensions comparing to KISSME and LMNN, and outperforms ITML.

3.3. PRID 2011 dataset

The PRID 2011 dataset consists of person images from two different static surveillance cameras. Camera A contains 385 persons, and camera B contains 749 persons, with 200 of them appearing in both cameras. Different to the VIPeR dataset, this dataset has significant lighting changes. The efficient for different illumination conditions of our method was validated by comparing with BTF and CBTF. The 200 image pairs were selected for training and testing. Following the procedure in Sec. 3.1, the parameters were set as $p = 100$

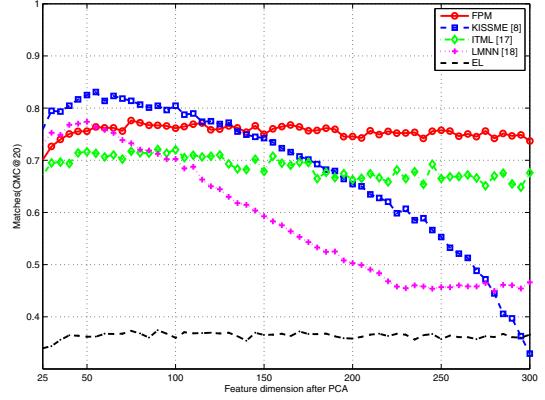


Fig. 5. Matching rate with different feature dimensions

and $q = 20$. This procedure was repeated 50 times and the averaged results are reported in Fig. 6. The BTF and CBTF were processed as the same as in Sec. 3.2.

As can be seen from Fig. 6, applying the proposed method leads to superior performance compared to using BTF and CBTF, while BTF and CBTF perform almost and better than the Euclidean distance. It equal to say that our method is more effective than BTF and CBTF for illumination compensation. The main reason may be that our approach is not need the assumption of brightness which is not accurate in the practical surveillance condition.

3.4. 3DPes dataset

The 3DPes dataset consists of 50 people captured by several cameras. For each person, there are four view images, i.e. front, back, left and right, with the foreground ground truth. In our experiment, the front and left view images were used as taken from two disjoint cameras C_a and C_b . Following the procedure in Sec. 3.1, the parameters were set as $p = 20$ and $q = 10$. This procedure was repeated 50 times and the averaged results are reported in Fig. 7.

The performance of our method outperforms LMNN and standard Euclidean distance slightly, and as the same as ITML with $r > 5$, while KISSME gets a performance equally by the chance. There may be two reasons. First, the dataset is too small which regards the terrible result of KISSME. Second, the differences between images of different persons is distinctive, for that the EL reaches a very well effectiveness.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a feature projection matrix (FPM) to address the person re-identification. In order to matching the person images taken from two non-overlapping cameras, an optimal FPM, which is used to transfer the person images from one camera to the other, was learned by utilizing logistic loss function. Extensive comparative experimental results

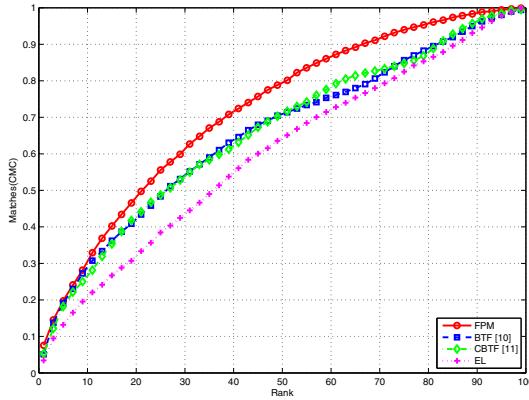


Fig. 6. Comparative results on the PRID 2011 dataset

reported in Sec. 3 show that our method is both effective and robust compared to BTF, CBT and several state-of-the-art person re-identification methods on three challenging public datasets, VIPeR, PRID 2011 and 3DPes. In the future, we plan to further investigate the combination of FPM and metric learning. The FPM is used to project objects from different cameras into the same feature space, while the metric learning is exploited to find the suitable space.

Acknowledgement

The research was supported by the major national science and technology special projects(2010ZX03004-003-03, 2010ZX03004-001-03), the National Basic Research Program of China (973 Program) (2009CB320906),the National Natural Science Foundation of China (60970160, 61070080, 61003184, 61172173, 61170023, 61231015).

References

- [1] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani, “Person re-identification by symmetry-driven accumulation of local features,” in *Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [2] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, “Person re-identification by support vector ranking,” in *British Machine Vision Conference (BMVC)*, 2010.
- [3] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, “Person re-identification by probabilistic relative distance comparison,” in *Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [4] Martin Hirzer, PeterM Roth, Martin Kstinger, and Horst Bischof, “Relaxed pairwise learned metric for person re-identification,” in *European Conference on Computer Vision (ECCV)*, 2012.
- [5] Niloofar Gheissari, Thomas B. Sebastian, and Richard Hartley, “Person reidentification using spatiotemporal appearance,” in *Computer Vision and Pattern Recognition (CVPR)*, 2006.
- [6] Xiaogang Wang, Gianfranco Doretto, Thomas Sebastian, Jens Rittscher, and Peter Tu, “Shape and appearance context modeling,” in *International Conference on Computer Vision (ICCV)*, 2007.
- [7] Mert Dikmen, Emre Akbas, Thomas S. Huang, and Narendra Ahuja, “Pedestrian recognition with a learned metric,” in *Asian Conference on Computer vision (ACCV)*, 2011.
- [8] M. Kostinger, M. Hirzer, P. Wohlhart, P.M. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in *Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [9] A. Mignon and F. Jurie, “Pcca: A new approach for distance learning from sparse pairwise constraints,” in *Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [10] K Shafique and Khurram Shafique, “Appearance modeling for tracking in multiple non-overlapping cameras,” in *Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [11] Bryan Prosser, Shaogang Gong, and Tao Xiang, “Multi-camera matching using bidirectional cumulative brightness transfer functions,” in *British Machine Vision Conference (BMVC)*, 2008.
- [12] S. Brennan D. Gray and H. Tao, “Evaluating appearance models for recognition, reacquisition, and tracking,” in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 2007.
- [13] Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof, “Person re-identification by descriptive and discriminative classification,” in *Scandinavian Conference on Image Analysis*, 2011.
- [14] Davide Baltieri, Roberto Vezzani, and Rita Cucchiara, “3dps: 3d people dataset for surveillance and forensics,” in *Proceedings of the 1st International ACM Workshop on Multimedia access to 3D Human Objects*, 2011.
- [15] Douglas Gray and Hai Tao, “Viewpoint invariant pedestrian recognition with an ensemble of localized features,” in *European Conference on Computer Vision (ECCV)*, 2008.
- [16] Martin Hirzer, Csaba Beleznai, Martin Kstinger, Peter M. Roth, and Horst Bischof, “Dense appearance modeling and efficient learning of camera transitions for person re-identification,” in *ICIP*, 2012.
- [17] Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon, “Information-theoretic metric learning,” in *International Conference on Machine learning (ICML)*, 2007.
- [18] Kilian Q. Weinberger, John Blitzer, and Lawrence K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

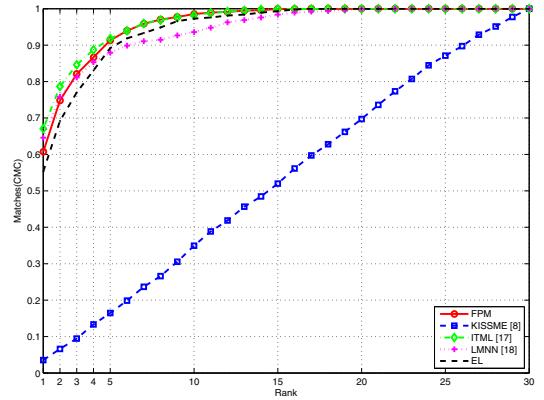


Fig. 7. Comparative results on the 3DPes dataset