This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

Cross-Domain Person Re-Identification Using Domain Adaptation Ranking SVMs

Andy J Ma, Jiawei Li, Pong C Yuen, *Senior Member, IEEE*, and Ping Li

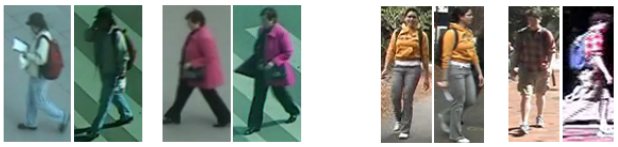
***Abstract*—This paper addresses a new person re-identification problem without label information of persons under non-overlapping target cameras. Given the matched (positive) and unmatched (negative) image pairs from source domain cameras, as well as unmatched (negative) and unlabeled image pairs from target domain cameras, we propose an Adaptive Ranking Support Vector Machines (AdaRSVM) method for re-identification under target domain cameras without person labels. To overcome the problems introduced due to the absence of matched (positive) image pairs in the target domain, we relax the discriminative constraint to a necessary condition only relying on the positive mean in the target domain. To estimate the target positive mean, we make use of all the available data from source and target domains as well as constraints in person re-identification. Inspired by adaptive learning methods, a new discriminative model with high confidence in target positive mean and low confidence in target negative image pairs is developed by refining the distance model learnt from the source domain. Experimental results show that the proposed AdaRSVM outperforms existing supervised or unsupervised, learning or non-learning re-identification methods without using label information in target cameras. Moreover, our method achieves better re-identification performance than existing domain adaptation methods derived under equal conditional probability assumption.**

***Index Terms*—Person Re-Identification, Domain Adaptation, Target Positive Mean, Adaptive Learning, Ranking SVMs.**

I. Introduction

A. Background

In recent years, person re-identification across a camera network comprising multiple cameras with non-overlapping views has become an active research topic due to its importance in many camera-network-based computer vision applications. The goal of person re-identification is to re-identify a person when he/she disappears from the field-of-view of a camera and appears in another. Matching individuals over disjoint cameras can be substantially challenging when variations in illumination condition, background, human pose and scale are significant among those views. Moreover, the temporal transition time between cameras varies greatly for each individual which makes the person re-identification task even harder.



(a) PRID (b) VIPeR

Fig. 1. Comparison between person images from different datasets (better viewed in color): matched image pairs in (a) PRID [27] and (b) VIPeR [28] dataset. There are three major differences between these two image sets: 1) different backgrounds; 2) different viewpoint changes; 3) different illumination conditions.

To address this problem, existing schemes mainly focus on developing either robust feature representations [1]–[17] or discriminative learning models [18]–[26]. For the discriminative learning methods, it is generally assumed that the label information of persons is available for training. With the person labels, matched (positive) and unmatched (negative) image pairs are generated to train the discriminative distance model. While these methods could achieve encouraging re-identification performance, the assumption that label information is available for all the cameras, could only be practically feasible in a small-scale camera network.

Contrarily, in the case of large-scale camera network, collecting the label information of every training subject from every camera in the network can be extremely time-consuming and expensive. Therefore, labels of the training subjects may not be able to be collected from certain cameras. This renders existing approaches inapplicable, since the person labels are not available. Apart from this reason, significant inter-camera variations as exemplified in Fig. 1 would also lead to dramatic performance deterioration, when the distance model learnt from other camera set with label information is directly applied to the cameras missing person labels.

These setbacks pose the need for new methods to handle the afore-described person re-identification issue in the large-scale camera network setting.

B. Motivation

Motivated by domain adaptation approach (see [29] for a review), we consider data from the camera set with label

This project is partially supported by Hong Kong RGC General Research Fund HKBU 212313 and HKBU 12202514. The research of A. J. Ma and P. Li is partially supported by ONR-N00014-13-1-0764 and AFOSR-FA9550-13-1-0137. The authors would like to thank the editor and reviewers for their helpful comments which improve the quality of this paper.

A. J. Ma is with the Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218-2608, USA. This work was mainly done when he was with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. E-mail: jhma@comp.hkbu.edu.hk

J. Li is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. E-mail: jwli@comp.hkbu.edu.hk

P. C. Yuen is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. He is also with the BNU-HKBU United International College, Zhuhai, China. E-mail: pcyuen@comp.hkbu.edu.hk

P. Li is with the Department of Statistics & Biostatistics and Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA. E-mail: pingli@stat.rutgers.edu

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

information as the source domain; while data from the camera set missing label information as the target domain. Here, we denote the source and target domains as s and t, respectively. Due to non-trivial inter-camera variations as indicated in Fig. 1, the source and target joint distributions of the positive or negative tag y and feature vector $\boldsymbol{z}$ for an image pair are supposed to be different, i.e. $\Pr_s(y,\boldsymbol{z}) \neq \Pr_t(y,\boldsymbol{z})$. In order to overcome the problem of the mismatch of marginal distributions, a mapping Φ, s.t. $\Pr_s(\Phi(\boldsymbol{z})) \approx \Pr_t(\Phi(\boldsymbol{z}))$, can be learnt via domain adaptation techniques, e.g. [30] [31]. In general, existing unsupervised domain adaptation schemes assume that, after projection, such Φ also satisfies the equal conditional probability condition, i.e. $\Pr_s(y|\Phi(\boldsymbol{z})) \approx \Pr_t(y|\Phi(\boldsymbol{z}))$. Since the conditional probability $\Pr_s(y|\Phi(\boldsymbol{z}))$ or $\Pr_t(y|\Phi(\boldsymbol{z}))$ can be interpreted as classification score, the condition that $\Pr_s(y|\Phi(\boldsymbol{z})) \approx \Pr_t(y|\Phi(\boldsymbol{z}))$ implies an equivalence of the distance models in the source and target domains. In this case, existing person re-identification algorithms, e.g. [18]–[22], can be employed to learn the distance model in the source domain (consisting of projected data with positive and negative image pairs generated by the label information), which can be applied to the target domain without significant performance degradation.

However, it is almost impossible to verify the validity of the assumption that $\Pr_s(y|\Phi(\boldsymbol{z})) \approx \Pr_t(y|\Phi(\boldsymbol{z}))$ in practice. As a result, there is no way to guarantee that the distance model learnt from the projected data in the source domain is equivalent to the target one. Thus, we propose to learn the target distance model using data from both source and target domains. If a small amount of positive and negative data is available in the target domain, multi-task learning [32] [33] or adaptive learning methods [34] [35] can be employed to learn the distance model for the target cameras without the assumption that the conditional distributions are equal with each other in the source and target domains. However, in large-scale camera networks, it is still time-consuming and expensive to label even a small amount of person images. Due to the absence of label information under target cameras, these domain adaptation techniques [32]–[35] cannot be applied directly. To ensure the equality of the conditional probabilities in the source and target domains, this paper will study on how to estimate the target label information and incorporate the labeled data from source domain with the estimated target label information for discriminative learning.

C. Contributions

The contributions of this paper are two-fold.

- We develop a new method to estimate target positive information based on the labeled data from the source domain, negative data (unmatched image pairs generated from non-overlapping target cameras) and unlabeled data from the target domain. Without positive image pairs generated by the label information of persons, we propose to relax the discriminative constraint into a necessary condition to it, which only relies on the mean of positive pairs. Since source and target domains must be related, we estimate the target positive mean by the labeled data from the source domain. While the estimation based on the source domain data may deviate from the true target positive mean, we propose to estimate it in another way that potential positive data is selected from the unlabeled data in the target domain by maximizing the positive joint distribution with properties and constraints in person re-identification. To further reduce the estimation error, the two estimations of the target positive mean are combined to determine the optimal estimation by the training data.
- We propose a novel Adaptive Ranking Support Vector Machines (AdaRSVM) method to rank the individuals for person re-identification. Inspired by adaptive learning methods [34] [35], RankSVM [19] is employed to learn a distance model by the labeled data from the source domain. After that, the estimated target positive mean and target negative data are used to learn the discriminative model for target domain by adaptively refining the distance model learnt in the source domain.

Although the motivation of this paper is similar to that in our conference version [36], the proposed algorithm is almost different from the previous one. In this paper, we propose a new method (different from that in [36]) to better estimate the target positive mean by all the available information from both source and target domains. Besides, the asymmetric domain adaptation algorithm in this paper is better than the symmetric one in the previous method, since it is more important to train a discriminative model for the target domain (rather than both). Moreover, more experiments have been performed to evaluate the proposed method, e.g. we add more datasets for evaluation and compare with two domain adaptation algorithms and appearance-based methods in this paper.

D. Organization

The rest of this paper is organized as follows. We will first give a brief review on existing person re-identification and domain adaptation methods. Section III will report the proposed method. Experimental results and conclusion are given in Section IV and Section V, respectively.

II. Related Works

Before introducing the proposed method, we give a brief review on person re-identification and domain adaptation in this section.

A. Person Re-Identification

In order to ensure that feature representation of the person image is less sensitive to large inter-camera variations, many existing re-identification methods focus on extracting robust features. Popular ones include SIFT [7] [10], texture [4] [5] [6] [11] [12], color distribution [3] [15], space-time methods [1] [2] and pictorial structures [8].

Besides feature extraction, discriminative distance learning methods are proposed to further improve the re-identificaiton performance. In [19], person re-identification was formulated as a ranking problem and the RankSVM model is learnt by assigning higher confidence to the positive image pairs and vice versa. Denote $\boldsymbol{x}_i$ as the feature vector for image i, $\boldsymbol{x}_{ij}^{+}$ for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

$j = 1, \cdots, n_i^+$ as feature vectors of the images with the same identity, and $\boldsymbol{x}_{il}^-$ for $l = 1, \cdots, n_i^-$ as feature vectors of the images with different identities, where n_i^+ (n_i^-) is the number of the matched (unmatched) observations. And, the absolute difference vector for the positive (resp. negative) image pair of $\boldsymbol{x}_i$ and $\boldsymbol{x}_{ij}^+$ (resp. $\boldsymbol{x}_{il}^-$) is calculated by $\boldsymbol{z}_{ij}^+ = \boldsymbol{d}(\boldsymbol{x}_i - \boldsymbol{x}_{ij}^+)$ (resp. $\boldsymbol{z}_{il}^- = \boldsymbol{d}(\boldsymbol{x}_i - \boldsymbol{x}_{il}^-)$), where $\boldsymbol{d}$ is an entry-wise function of absolute values. The weight vector $\boldsymbol{w}$ in RankSVM is obtained by solving the following optimization problem,

$$\begin{aligned} &\min_{\boldsymbol{w},\xi} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i,j,l}\xi_{ijl} \\ &\text{s.t. } \boldsymbol{w}^T(\boldsymbol{z}_{ij}^+ - \boldsymbol{z}_{il}^-) \geq 1 - \xi_{ijl}, \\ &\quad \xi_{ijl} \geq 0, \forall i, j, l \end{aligned} \tag{1}$$

where ξ_{ijl} is the slack variable and C is a positive parameter. Similar to RankSVM, Zheng *et al.* [22] proposed a Relative Distance Comparison (RDC) method using a second-order distance learning model. This method is able to exploit higher-order correlations among different features, compared with RankSVM. In order to solve the computational complexity issues in RankSVM and RDC, a Relaxed Pairwise Metric Learning (RPML) method [21] was proposed by relaxing the original hard constraints, which leads to a simpler problem that can be solved more efficiently.

Besides the supervised distance learning methods [19] [21] [22], Kuo *et al.* [37] proposed an online-learnt appearance affinity model to decrease the required number of labeled samples under some specific assumptions. On the other hand, an adaptive feature weighting method was proposed in [38] under the observation that the universal model may not be good for all individuals. Different from traditional per-individual identification scheme, Zheng *et al.* [39] addressed a watch list (set) based verification problem and proposed to transfer the information from non-target person data to mine the discriminative information for the target people in the watch list.

B. Domain Adaptation

The main objective of domain adaptation approach is to adapt the classification model learnt from the source domain to target domain without serious deterioration of recognition performance. The target domain refers to data from the target task usually without or with only a small amount of labeled training data, while there are plenty of labeled training data in the source domain. In the last decade, many algorithms (see [29] for a review) have been proposed to solve the joint distribution mismatch problem, i.e. $\Pr_s(y, \boldsymbol{z}) \neq \Pr_t(y, \boldsymbol{z})$.

For unsupervised domain adaptation, the instance re-weighting or covariate shift approach [40] learns the target classification model by re-weighting the labeled samples in the source domain to minimize the approximated empirical classification error in the target domain. To estimate the sample weights calculated by $\Pr_s(\boldsymbol{z})$ dividing $\Pr_t(\boldsymbol{z})$, many density ratio estimation methods [41] have been proposed. Besides instance re-weighting, the feature representation methods [30] [31] [42] [43] [44] [45] construct feature vectors to reduce the difference between features in the source and target domains. Blitzer *et al.* [42] proposed a structural correspondence learning algorithm by selecting pivot features for natural language processing, while other methods [30] [31] [43] [44] [45] try to learn a mapping Φ, s.t. $\Pr_s(\Phi(\boldsymbol{z})) \approx \Pr_t(\Phi(\boldsymbol{z}))$. Without label information in the target domain, these methods assume that the conditional probabilities are equal to each other in the source and target domains. And it was shown in [46] that the empirical classification error can be very small under this assumption. However, this assumption may not be valid, so that the recognition performance may deteriorate.

For supervised domain adaptation with target labeled data, existing methods learn an informative prior using the source domain data and estimate the target model based on such prior [34] [35] [47] [48]. Based on the assumption that the recognition tasks in the source and target domains are related, multi-task learning methods [32] [33] can be employed to discover the task relationship and learn the classification models in the source and target domains simultaneously. Unlike supervised domain adaptation techniques, unlabeled data in the target domain are considered together with the labeled data to learn the target classification model for better performance in [49]–[53]. However, labeling person images for each camera is expensive, especially in large-scale camera networks applications. Thus, existing supervised or semi-supervised domain adaptation algorithms cannot be employed directly.

III. Domain Adaptation Ranking SVMs for Person Re-Identification

To present the algorithm more clearly, let target domain contain images from a pair of (two) cameras a and b. For multiple target cameras, multiple classification models can be trained for each camera pair. Since feature extraction is not the focus of this paper, all general feature representation methods, e.g. color histogram, can be used to extract feature vectors for the person images. As indicated in [22], the absolute difference space shows some advantages over the common difference space, so we follow [22] to use the absolute difference vectors for both positive and negative image pairs. Given two feature vectors $\boldsymbol{x}_i^a$ and $\boldsymbol{x}_j^b$ representing two images under two cameras a and b, the absolute difference vector $\boldsymbol{z}_{ij}$ is defined by

$$\boldsymbol{z}_{ij} = \boldsymbol{d}(\boldsymbol{x}_i^a - \boldsymbol{x}_j^b) = (|\boldsymbol{x}_i^a(1) - \boldsymbol{x}_j^b(1)|, \cdots, |\boldsymbol{x}_i^a(R) - \boldsymbol{x}_j^b(R)|)^T \tag{2}$$

where $\boldsymbol{x}(r)$ is the r-th element of the input vector $\boldsymbol{x}$ and R is the dimension of $\boldsymbol{x}$.

The available training data is introduced as follows. In the source domain, label information is available, so difference vectors of positive and negative image pairs can be generated and denoted as $\boldsymbol{z}_{sij}^+$ and $\boldsymbol{z}_{skl}^-$, respectively. For the target domain, the label information of persons is not available, so positive image pairs cannot be generated. *However, negative image pairs can be easily generated, because same person cannot be presented at the same instant under different non-overlapping cameras.* Denote the difference vectors for the target negative image pairs as $\boldsymbol{z}_{tij}^-$. In addition, unlabeled image pairs in the target domain are also available and the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

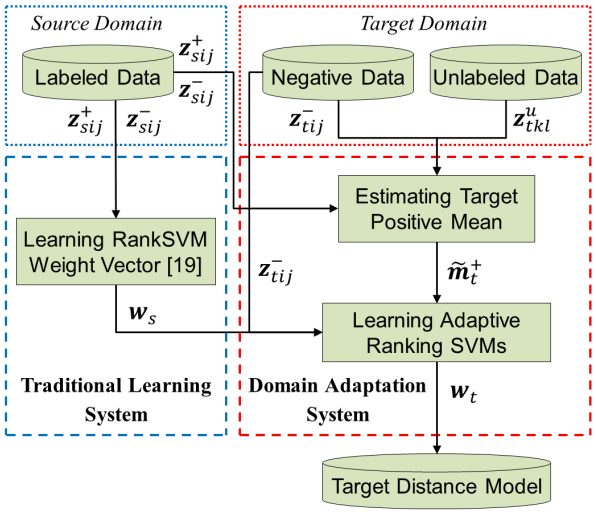


Fig. 2. Proposed domain adaptation framework for person re-identification

unlabeled difference vector is denoted as $\boldsymbol{z}^u_{tkl}$. The block diagram of the proposed method is shown in Fig. 2. With $\boldsymbol{z}^+_{sij}, \boldsymbol{z}^-_{skl}$ in the source domain and $\boldsymbol{z}^-_{tij}, \boldsymbol{z}^u_{tkl}$ in the target domain, we propose a new method to estimate the target positive mean $\widetilde{\boldsymbol{m}}^+_t$, which will be discussed in Section III-A. Then, the labeled data from the source domain are used to train a source distance model $\boldsymbol{w}_s$ by employing RankSVM [19]. At last, the target distance model $\boldsymbol{w}_t$ is learnt by the proposed adaptive ranking SVMs method, which will be presented in Section III-B.

A. Estimating Positive Mean in Target Domain

Since different feature elements in a feature vector give different importance in identifying a person, we follow [19] to use the weighted summation of the absolute difference vector to calculate the confident score for the image pairs in the target domain, i.e. $\boldsymbol{w}_t^T \boldsymbol{z}_{tij}$, where $\boldsymbol{w}_t^T$ denotes the transposition of the target weight vector. If positive image pairs are available, the scores of positive image pairs must be larger than those of the negative ones for a discriminative weight vector $\boldsymbol{w}_t$, i.e.

$$\boldsymbol{w}_t^T \boldsymbol{z}^+_{tij} > \boldsymbol{w}_t^T \boldsymbol{z}^-_{tkl}, \forall i, j, k, l \tag{3}$$

However, positive image pairs are not available in the target domain, so we cannot obtain the absolute difference vectors $\boldsymbol{z}^+_{tij}$ in practice. One way to solve this problem is to determine the weight vector $\boldsymbol{w}_t$ by assigning smaller values to the difference vectors $\boldsymbol{z}^-_{tkl}$ of negative image pairs using one-class SVM [54]. Nevertheless, it is possible that the scores of positive image pairs also decrease when minimizing those of the negative ones. Thus, it cannot be guaranteed that the learnt weight vector $\boldsymbol{w}_t$ satisfies the discriminative constraint (3). In order to deal with this problem, we propose to learn the weight vector $\boldsymbol{w}_t$ by imposing a necessary condition to constraint (3).

Taking the summation of constraint (3) over the difference vectors $\boldsymbol{z}^+_{tij}$ of positive image pairs for all i and j, it has

$$\boldsymbol{w}_t^T \boldsymbol{m}^+_t > \boldsymbol{w}_t^T \boldsymbol{z}^-_{tkl}, \forall k, l \tag{4}$$

where $\boldsymbol{m}^+_t$ denote the mean of positive image pairs in the target domain. Therefore, a necessary condition to constraint (3) is given by equation (4) such that the score of the positive mean is larger than those of the negative image pairs.

While there is no positive data in the target domain, the target positive difference vectors $\boldsymbol{z}^+_{tij}$ are difficult to estimate due to the highly data imbalance problem. According to equation (4), the target positive mean $\boldsymbol{m}^+_t$ can be used to give a necessary condition to the discriminative constraint (3). On the other hand, in domain adaptation, the distance between source and target domain distributions can be measured by the distance between the empirical means of the two domains [31]. Moreover, in one-class classification problems, the positive mean is usually used to represent the positive distribution [55]. Thus, we propose to estimate the target positive mean for domain adaptation in person re-identification. Although it is still challenging to estimate the target positive mean without any positive data, we solve this problem by using both the labeled data in the source domain and the unlabeled data in the target domain.

1) Estimating Target Positive Mean by Labeled Data in Source Domain: To estimate the target positive mean, we propose to make use of the data with label information of persons in the source domain. With the label information, the true means of positive and negative image pairs in the source domain can be calculated and denoted as $\boldsymbol{m}^+_s$ and $\boldsymbol{m}^-_s$, respectively. Since the source and target domains are related, the positive and negative distributions in the source domain must be related to those in the target domain. We suppose the relationship can be modeled in a way that the difference between the positive and negative means in the source domain is close to that in the target domain, i.e.

$$\boldsymbol{m}^+_t - \boldsymbol{m}^-_t \approx \boldsymbol{m}^+_s - \boldsymbol{m}^-_s \tag{5}$$

where $\boldsymbol{m}^+_t$ and $\boldsymbol{m}^-_t$ denote the target genuine positive and negative means, respectively.

Since not all of the negative data in the target domain are available, the true negative mean $\boldsymbol{m}^-_t$ cannot be calculated. Instead, we can estimate the negative mean $\widetilde{\boldsymbol{m}}^-_t$ by the available negative difference vectors $\boldsymbol{z}^-_{tij}$. With equation (5), the positive mean in the target domain can be estimated by the following equation,

$$\widetilde{\boldsymbol{m}}^+_{t1} = \widetilde{\boldsymbol{m}}^-_t + \boldsymbol{m}^+_s - \boldsymbol{m}^-_s \tag{6}$$

The upper bound of the estimation error using equation (6) is given by

$$\begin{aligned} \|\boldsymbol{m}^+_t - \widetilde{\boldsymbol{m}}^+_{t1}\| \leq \|\boldsymbol{m}^-_t - \widetilde{\boldsymbol{m}}^-_t\| \\ + \|(\boldsymbol{m}^+_t - \boldsymbol{m}^-_t) - (\boldsymbol{m}^+_s - \boldsymbol{m}^-_s)\| \end{aligned} \tag{7}$$

Since lots of negative image pairs can be obtained from the non-overlapping target cameras, the estimated mean of negative pairs is close to the true one. Under the assumption given by equation (5), the upper bound of the estimation error for the positive mean in the target domain is small.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

2) Estimating Target Positive Mean by Unlabeled Data in Target Domain: Utilizing the labeled data in the source domain, equation (6) estimates the target positive mean without using the unlabeled data $\boldsymbol{z}_{tkl}^u$ in the target domain. On the other hand, $\widetilde{\boldsymbol{m}}_{t1}^+$ may not be a good estimation, if the assumption given by equation (5) is not valid. Therefore, we propose to estimate the target positive mean in another way based on the target domain data in this subsection.

Denote the mean calculated by the difference vectors of all the image pairs in the target domain as $\boldsymbol{m}_t$. It can be calculated by the following equation,

$$\boldsymbol{m}_t = (N_t^+ \boldsymbol{m}_t^+ + N_t^- \boldsymbol{m}_t^-)/N_t \tag{8}$$

where N_t, N_t^+ and N_t^- denote the number of the overall, positive and negative image pairs, respectively. With equation (8), the mean of positive image pairs can be estimated by the following equation,

$$\widetilde{\boldsymbol{m}}_t^+ = (N_t \boldsymbol{m}_t - N_t^- \widetilde{\boldsymbol{m}}_t^-)/N_t^+ \tag{9}$$

However, N_t^+ and N_t^- are difficult to compute, if target positive samples are not available. On the other hand, the estimation error for the target positive mean with equation (9) can be very large, since the number of negative pairs is much larger than that of positive pairs, i.e. $N_t^- \gg N_t^+$. The estimation error for the positive mean with equation (9) is given by the following equation,

$$\|\boldsymbol{m}_t^+ - \widetilde{\boldsymbol{m}}_t^+\| = \frac{N_t^-}{N_t^+}\|\boldsymbol{m}_t^- - \widetilde{\boldsymbol{m}}_t^-\| \tag{10}$$

Since $N_t^- \gg N_t^+$, the division value of N_t^-/N_t^+ is very large. Therefore, according to equation (10), the estimation error for the positive mean is very large, even though the error for the negative mean is small.

To solve this problem, we propose to calculate the mean by selecting Q potential positive difference vectors from the unlabeled data $\boldsymbol{z}_{tkl}^u$ and maximizing the following joint distribution,

$$\max_{\boldsymbol{z}_{tk_ql_q}^u, q=1,\cdots,Q} \Pr(\boldsymbol{z}_{tk_1l_1}^u, \cdots, \boldsymbol{z}_{tk_Ql_Q}^u|+) \tag{11}$$

If $\boldsymbol{z}_{tk_1l_1}^u, \cdots, \boldsymbol{z}_{tk_Ql_Q}^u$ are independent with each other given the positive tag, the optimization problem (11) can be rewritten as

$$\max_{\boldsymbol{z}_{tk_ql_q}^u, q=1,\cdots,Q} \prod_{q=1}^{Q} \Pr(\boldsymbol{z}_{tk_ql_q}^u|+) \tag{12}$$

Denote $p_{tkl}^u = \Pr(\boldsymbol{z}_{tk_ql_q}^u|+)$. We estimate the positive conditional probability p_{tkl}^u as follows. According to the definition given by equation (2), if the norm $\|\boldsymbol{z}_{tkl}^u\|$ of the absolute difference vector is close to zero, images k and j have a high probability that they represent the same person, and thus they form a positive (matched) image pair, i.e. $\boldsymbol{z}_{tkl}^u$ is likely to be positive. Therefore, we estimate the positive conditional probability by[1]

$$p_{tkl}^u \propto e^{-\|\boldsymbol{z}_{tkl}^u\|} \tag{13}$$

[1]While the norm could be more general, we use the l_1 norm in our experiments, and hence the probability becomes a Laplace distribution.

Algorithm 1 Selecting positive difference vectors

Input: Unlabeled difference vectors $\boldsymbol{z}_{tkl}^u$ in target domain and number of selected positive difference vectors Q;

1: Calculate positive posteriors p_{tkl}^u by equation (13);
2: Group unlabeled data to obtain $G_{k\cdot}$ and $G_{\cdot l}$ by equation (14);
3: Select $\boldsymbol{z}_{tk_ql_q}^u$ with the highest positive probability $p_{tk_ql_q}^u$ from the unlabeled data set;
4: Delete elements in groups $G_{k_q\cdot}$ and $G_{l_q\cdot}$ from the unlabeled data set;
5: Go to step 2 until Q positive difference vectors are selected;

Output: Selected difference vectors $\boldsymbol{z}_{tk_1l_1}^u, \cdots, \boldsymbol{z}_{tk_Ql_Q}^u$.

Under the independent assumption, potential positives can be selected by the unlabeled data with top Q scores $p_{tk_1l_1}^u, \cdots, p_{tk_Ql_Q}^u$. However, the independent assumption is not valid. To explain the reasons, we denote $G_{k\cdot}$ as the set of difference vectors related to image k under camera a and those images under the other camera b, i.e.

$$G_{k\cdot} = \{\boldsymbol{z}_{tkl}^u = \boldsymbol{d}(\boldsymbol{x}_k^a - \boldsymbol{x}_l^b)|\forall \boldsymbol{x}_l^b\} \tag{14}$$

Let us consider two elements $\boldsymbol{z}_{tkl_1}^u$ and $\boldsymbol{z}_{tkl_2}^u$ in $G_{k\cdot}$. Denote the number of positives and the number of all elements in $G_{k\cdot}$ as N_k^+ and N_k, respectively. If $\boldsymbol{z}_{tkl_1}^u$ is positive, the probability that $\boldsymbol{z}_{tkl_2}^u$ is positive is equal to $(N_k^+ - 1)/(N_k - 1)$. If $\boldsymbol{z}_{tkl_1}^u$ is negative, such probability becomes $N_k^+/(N_k - 1)$. Therefore, the independence is not valid for the unlabeled data in the same group $G_{k\cdot}$. And, we propose to add a constraint to the optimization problem (12) to ensure the independence, i.e. $\boldsymbol{z}_{tk_1l_1}^u, \cdots, \boldsymbol{z}_{tk_Ql_Q}^u$ come from different groups $G_{k\cdot}$ and $G_{\cdot l}$[2]. Thus, we have the following optimization problem,

$$\max_{p_{tk_ql_q}^u, \text{ s.t. } k_1 \neq \cdots \neq k_Q, l_1 \neq \cdots \neq l_Q} \prod_{q=1}^{Q} p_{tk_ql_q}^u \tag{15}$$

To solve the optimization problem (15), we propose an efficient greedy method. Once a potential positive difference vector $\boldsymbol{z}_{tk_ql_q}^u$ is selected, the elements in $G_{k_q\cdot}$ and $G_{\cdot l_q}$ are removed for the constraint in the optimization problem (15). The algorithmic procedure is given in Algorithm 1. Since it cannot be guaranteed that the selected unlabeled difference vectors are really generated by the true positive image pairs, we calculate the mean of them to reduce the negative impact for wrongly labeling an unlabeled difference vector as positive, i.e.

$$\widetilde{\boldsymbol{m}}_{t2}^+ = \frac{1}{Q}(\boldsymbol{z}_{tk_1l_1}^u + \cdots + \boldsymbol{z}_{tk_Ql_Q}^u) \tag{16}$$

where $\boldsymbol{z}_{tk_1l_1}^u, \cdots, \boldsymbol{z}_{tk_Ql_Q}^u$ are Q unlabeled difference vectors selected as positive.

3) Combining Estimated Target Positive Means: In the previous discussions, labeled data in the source domain and unlabeled data in the target domain are used to estimate the target positive mean and result in two estimations $\widetilde{\boldsymbol{m}}_{t1}^+$ and

[2]$G_{\cdot l}$ can be defined similarly to $G_{k\cdot}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

$\widetilde{\boldsymbol{m}}_{t2}^{+}$, respectively. Let us consider the target positive mean $\boldsymbol{m}_t^+$ as a random vector. We estimate the conditional distribution of $\boldsymbol{m}_t^+$ given $\widetilde{\boldsymbol{m}}_{t1}^{+}$ and $\widetilde{\boldsymbol{m}}_{t2}^{+}$ as follows. By Bayes' rule [56], the conditional probability $\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+},\widetilde{\boldsymbol{m}}_{t2}^{+})$ can be computed as

$$\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+},\widetilde{\boldsymbol{m}}_{t2}^{+}) = \frac{\Pr(\widetilde{\boldsymbol{m}}_{t1}^{+},\widetilde{\boldsymbol{m}}_{t2}^{+}|\boldsymbol{m}_t^+)\Pr(\boldsymbol{m}_t^+)}{\Pr(\widetilde{\boldsymbol{m}}_{t1}^{+},\widetilde{\boldsymbol{m}}_{t2}^{+})} \tag{17}$$

Since $\widetilde{\boldsymbol{m}}_{t1}^{+}$ and $\widetilde{\boldsymbol{m}}_{t2}^{+}$ are estimated by the source domain and target domain data, respectively, they can be considered to be independent and conditionally independent given the true positive mean $\boldsymbol{m}_t^+$, i.e.

$$\begin{aligned}\Pr(\widetilde{\boldsymbol{m}}_{t1}^{+},\widetilde{\boldsymbol{m}}_{t2}^{+}) &= \Pr(\widetilde{\boldsymbol{m}}_{t1}^{+})\Pr(\widetilde{\boldsymbol{m}}_{t2}^{+})\\ \Pr(\widetilde{\boldsymbol{m}}_{t1}^{+},\widetilde{\boldsymbol{m}}_{t2}^{+}|\boldsymbol{m}_t^+) &= \Pr(\widetilde{\boldsymbol{m}}_{t1}^{+}|\boldsymbol{m}_t^+)\Pr(\widetilde{\boldsymbol{m}}_{t2}^{+}|\boldsymbol{m}_t^+)\end{aligned} \tag{18}$$

Substituting equation (18) into equation (17), we get

$$\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+},\widetilde{\boldsymbol{m}}_{t2}^{+}) = \frac{\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+})\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t2}^{+})}{\Pr(\boldsymbol{m}_t^+)} \tag{19}$$

Without any information about the positive distribution, the prior probability $\Pr(\boldsymbol{m}_t^+)$ can be considered as a constant. Therefore, the key problem is to compute the conditional probabilities $\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+})$ and $\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t2}^{+})$.

To estimate $\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+})$, it is reasonable to assume that the probability of the target positive mean is higher if its distance is closer to $\widetilde{\boldsymbol{m}}_{t1}^{+}$. On the other hand, when the assumption given by equation (5) is satisfied, $\widetilde{\boldsymbol{m}}_{t1}^{+}$ is close to the true target positive mean according to equation (7). Otherwise, $\widetilde{\boldsymbol{m}}_{t1}^{+}$ may not be a good estimation. To measure the uncertainty about the validity of the assumption (5), we define $\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+})$ by Gaussian distribution as follows,

$$\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+}) \propto e^{-\|\boldsymbol{m}_t^+-\widetilde{\boldsymbol{m}}_{t1}^{+}\|_2^2/(2\sigma_1^2)} \tag{20}$$

where $\|\cdot\|_2$ denotes l_2 norm and σ_1^2 is the variance in Gaussian distribution to measure the uncertainty. Similarly, $\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t2}^{+})$ can be defined by

$$\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t2}^{+}) \propto e^{-\|\boldsymbol{m}_t^+-\widetilde{\boldsymbol{m}}_{t2}^{+}\|_2^2/(2\sigma_2^2)} \tag{21}$$

where σ_2^2 is the variance to measure the uncertainty of the positive mean estimated by the potential positive difference vectors selected by Algorithm 1.

With equations (19) (20) (21), the likelihood function of the target positive mean is given by

$$ln(\Pr(\boldsymbol{m}_t^+|\widetilde{\boldsymbol{m}}_{t1}^{+},\widetilde{\boldsymbol{m}}_{t2}^{+})) \propto -\frac{\|\boldsymbol{m}_t^+-\widetilde{\boldsymbol{m}}_{t1}^{+}\|_2^2}{2\sigma_1^2} - \frac{\|\boldsymbol{m}_t^+-\widetilde{\boldsymbol{m}}_{t2}^{+}\|_2^2}{2\sigma_2^2} \tag{22}$$

To compute the maximum value of the likelihood function, we take the first derivative of the right hand side in equation (22) and set it as zero. Then, the optimal estimation of the target positive mean $\widetilde{\boldsymbol{m}}_t^+$ is derived as

$$\widetilde{\boldsymbol{m}}_t^+ = \alpha\widetilde{\boldsymbol{m}}_{t1}^{+} + (1-\alpha)\widetilde{\boldsymbol{m}}_{t2}^{+} \tag{23}$$

where $\alpha = \sigma_2^2/(\sigma_1^2+\sigma_2^2)$. Since $\sigma_1^2 \geq 0$ and $\sigma_2^2 \geq 0$, the range of α is $0 \leq \alpha \leq 1$. And, α is determined by the training data, which will be discussed in the following section.

B. Adaptive Ranking SVMs

With positive and negative image pairs in the source domain, RankSVM [19] is employed to learn a source domain distance model $\boldsymbol{w}_s$. Inspired by the adaptive learning methods [34] [35] for asymmetric domain adaptation, the weight vector $\boldsymbol{w}_t$ for the target domain can be defined as follows,

$$\boldsymbol{w}_t = \theta\boldsymbol{w}_s + \boldsymbol{w} \tag{24}$$

where θ is the coefficient to measure the importance of the classifier $\boldsymbol{w}_s$ trained from the source domain data and $\boldsymbol{w}$ is the perturbation weight vector adapted for the target domain.

Substituting $\boldsymbol{m}_t^+$ and $\boldsymbol{w}_t$ by equations (23) and (24), respectively, the inequality (4) becomes

$$(\theta\boldsymbol{w}_s+\boldsymbol{w})^T(\alpha\widetilde{\boldsymbol{m}}_{t1}^{+} + (1-\alpha)\widetilde{\boldsymbol{m}}_{t2}^{+}) > (\theta\boldsymbol{w}_s+\boldsymbol{w})^T\boldsymbol{z}_{tkl}^{-}, \forall k,l \tag{25}$$

Similar to RankSVM [19], the order relationship given by inequality (25) needs to be preserved for discriminability. Thus, we propose to learn the optimal α, θ and $\boldsymbol{w}$ by solving the following optimization problem,

$$\begin{aligned}&\min_{\boldsymbol{w},\theta,\alpha}\frac{1}{2}(\|\boldsymbol{w}\|_2^2+\mu\theta^2) + C\sum_{k,l}\xi_{kl}\\ &\text{s.t. } (\theta\boldsymbol{w}_s+\boldsymbol{w})^T(\alpha\widetilde{\boldsymbol{m}}_{t1}^{+} + (1-\alpha)\widetilde{\boldsymbol{m}}_{t2}^{+} - \boldsymbol{z}_{tkl}^{-}) \geq 1-\xi_{kl},\\ &\quad \xi_{kl}\geq 0, 0\leq\alpha\leq 1, \forall k,l\end{aligned} \tag{26}$$

where μ is a positive parameter to balance the regularization terms for $\boldsymbol{w}$ and θ.

The optimization problem (26) can be solved by rewriting it as

$$\min_{\alpha} F(\alpha), \text{s.t. } 0\leq\alpha\leq 1 \tag{27}$$

$$\begin{aligned}&F(\alpha) = \min_{\boldsymbol{w},\theta}\frac{1}{2}(\|\boldsymbol{w}\|_2^2+\mu\theta^2) + C\sum_{k,l}\xi_{kl}\\ &\text{s.t. } (\theta\boldsymbol{w}_s+\boldsymbol{w})^T(\alpha\widetilde{\boldsymbol{m}}_{t1}^{+} + (1-\alpha)\widetilde{\boldsymbol{m}}_{t2}^{+} - \boldsymbol{z}_{tkl}^{-}) \geq 1-\xi_{kl},\\ &\quad \xi_{kl}\geq 0, \forall k,l\end{aligned} \tag{28}$$

With this reformulation, we can linearly search α from 0 to 1 with the minimal cost. Fixing α, the optimization problem (26) can be solved efficiently by converting it to the standard RankSVM formulation as in equation (1).

Denote the column concatenation of $\boldsymbol{w}$ and $\sqrt{\mu}\theta$ as $\boldsymbol{v}$, i.e.

$$\boldsymbol{v} = \begin{pmatrix}\boldsymbol{w}\\ \sqrt{\mu}\theta\end{pmatrix} \tag{29}$$

On the other hand, we construct the feature map as

$$\boldsymbol{f}_{kl} = \begin{pmatrix}\alpha\widetilde{\boldsymbol{m}}_{t1}^{+} + (1-\alpha)\widetilde{\boldsymbol{m}}_{t2}^{+} - \boldsymbol{z}_{tkl}^{-}\\ \boldsymbol{w}_s^T(\alpha\widetilde{\boldsymbol{m}}_{t1}^{+} + (1-\alpha)\widetilde{\boldsymbol{m}}_{t2}^{+} - \boldsymbol{z}_{tkl}^{-})/\sqrt{\mu}\end{pmatrix} \tag{30}$$

With the notations in (29) and (30), it has the following equations,

$$\begin{aligned}&\|\boldsymbol{v}\|_2^2 = \|\boldsymbol{w}\|_2^2+\mu\theta^2\\ &\boldsymbol{v}^T\boldsymbol{f}_{kl} = (\theta\boldsymbol{w}_s+\boldsymbol{w})^T(\alpha\widetilde{\boldsymbol{m}}_{t1}^{+} + (1-\alpha)\widetilde{\boldsymbol{m}}_{t2}^{+} - \boldsymbol{z}_{tkl}^{-})\end{aligned} \tag{31}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

Algorithm 2 Training AdaRSVM

Input: Difference vectors $\boldsymbol{z}_{sij}^{+}$ and $\boldsymbol{z}_{skl}^{-}$ in source domain, $\boldsymbol{z}_{tij}^{-}$ and $\boldsymbol{z}_{tkl}^{u}$ in target domain, parameters C and μ;

1: Compute $\boldsymbol{m}_s^{+}$ by $\boldsymbol{z}_{sij}^{+}$, $\boldsymbol{m}_s^{-}$ by $\boldsymbol{z}_{skl}^{-}$, and $\widetilde{\boldsymbol{m}}_t^{-}$ by $\boldsymbol{z}_{tij}^{-}$;
2: Estimate the target positive mean $\widetilde{\boldsymbol{m}}_{t1}^{+}$ by source domain data with equation (6);
3: Select potential positive difference vectors from the unlabeled target data $\boldsymbol{z}_{tkl}^{u}$ by Algorithm 1;
4: Estimate the target positive mean $\widetilde{\boldsymbol{m}}_{t2}^{+}$ by target domain data with equation (16);
5: Set the function cost F^* as infinite;
6: **for** $\alpha \in [0, 1]$ **do**
7: Construct the feature map by equation (30);
8: Solve the optimization problem (33) by the efficient method [57] and obtain the weight vector $\boldsymbol{v}$;
9: **if** $F(\alpha) < F^*$ **then**
10: Set $F^* = F(\alpha)$;
11: Calculate the target domain weight vector $\boldsymbol{w}_t$ by equations (24) and (29);
12: **end if**
13: **end for**

Output: Weight vector $\boldsymbol{w}_t$.

Therefore, the optimization problem (28) is rewritten as

$$\begin{aligned} F(\alpha) &= \min_{\boldsymbol{w},\theta} \frac{1}{2}\|\boldsymbol{v}\|_2^2 + C\sum_{k,l}\xi_{kl} \\ \text{s.t. } & \boldsymbol{v}^T\boldsymbol{f}_{kl} \geq 1 - \xi_{kl}, \xi_{kl} \geq 0, \forall k, l \end{aligned} \tag{32}$$

To solve the optimization problem (32) more efficiently, we reformulate (32) by the square hinge loss as

$$\min_{\boldsymbol{v}} \frac{1}{2}\|\boldsymbol{v}\|^2 + C\sum_{k,l}\max(0, 1 - \boldsymbol{v}^T\boldsymbol{f}_{kl})^2 \tag{33}$$

Employing the efficient algorithm [57] based on primal Newton method to solve the optimization problem (33), the optimal weight vector $\boldsymbol{v}$ can be obtained. Then, the target weight vector $\boldsymbol{w}_t$ is calculated by equations (24) and (29). At last, the algorithmic procedure for training the proposed Adaptive Ranking Support Vector Machines (AdaRSVM) model is presented in Algorithm 2.

IV. Experiments

In this section, we first give an introduction to the datasets and settings used for evaluation. Then, the comparison results are reported in Sections IV-B to IV-E.

A. Datasets and Settings

Four publicly available datasets, namely PRID[3] [27], VIPeR[4] [28], CUHK[5] [23] and i-LIDS[6] [58], are used for evaluating the proposed method. PRID dataset consists of person images from two static surveillance cameras. In total 385 persons were captured by camera A, while 749 persons captured by camera B. The first 200 persons appeared in both cameras, and the remainders only appear in one camera. In our experiments, the single-shot version is used, in which at most one image of each person from each camera is available. VIPeR is a re-identification dataset containing 632 person image pairs captured by two cameras outdoor. CUHK dataset contains five pairs of camera views. Under each camera view, there are two images for each person. Following the single shot setting in [23], images from camera pair one with 971 persons are used for experiments. The i-LIDS Multiple-Camera Tracking (MCT) dataset contains a number of video clips captured by five cameras indoor. In re-identification application, total 476 person images from 119 persons are used for experiments as in [22]. Example images in these four datasets are shown in Fig 1(a), Fig. 1(b), Fig. 3(a) and Fig. 3(b), respectively.

(a) CUHK (b) i-LIDS

Fig. 3. Example matched image pairs in (a) CUHK [23] and (b) i-LIDS [58]

In our experiments, we use PRID, VIPeR or CUHK as the target domain. For the i-LIDS dataset, the camera information is not available. Since the proposed method requires camera information to generate groups defined by equation (14) for positive difference vector selection, we do not use the i-LIDS dataset as the target domain. Without the time acquisition information in the PRID, VIPeR and CUHK datasets, target negative image pairs from non-overlapping cameras are generated by simulating the synchronization using label information. Fixing the target domain dataset as PRID, VIPeR or CUHK, one of the other three datasets is used as the source domain to train the proposed AdaRSVM. When PRID is used as the target dataset, 100 out of the 200 image pairs are randomly selected as the training set, and the others for testing. If VIPeR is used as the target dataset, 632 image pairs are randomly separated into half for training and the other half for testing. For the CUHK dataset, 971 persons are randomly split as 485 for training and 486 for testing. Given a query image in the testing data from one camera view, the evaluation is performed by ranking the person images from another view. For each source dataset, following [22], one positive and one negative image pair for each person are used for training, while the training data in the target domain contains only one negative image pair for each person. Each experiment was repeated ten times and the mean accuracy is reported.

Three state-of-the-art distance learning methods for person re-identification, namely Rank Support Vector Machines (RankSVM) [19], Relative Distance Comparison (RDC) [22], and Relaxed Pairwise Metric Learning (RPML) [21], are used for comparison. Since the label information of persons is supposed to be not available in the target dataset, cross-

[3]https://lrs.icg.tugraz.at/datasets/prid/
[4]http://soe.ucsc.edu/~dgray/VIPeR.v1.0.zip
[5]http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html
[6]http://www.eecs.qmul.ac.uk/~jason/data/i-LIDS_Pedestrian.tgz

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

(a) PRID (b) VIPeR (c) CUHK (d) i-LIDS

Fig. 4. Example masked results in (a) PRID [27], (b) VIPeR [28], (c) CUHK [23] and (d) i-LIDS [58] dataset (better view in color)

validation cannot be performed to select the best parameters. Thus, we empirically set the parameters in existing methods and the proposed AdaRSVM. The PCA dimension in RPML is set as 80. The parameter C in RankSVM and the proposed method is set as 1, while μ in the proposed AdaRSVM is set as 100. For the number of selected positive difference vectors, Q should be less than the number of detected persons in each camera, so we set $Q = 90$ in PRID, 300 in VIPeR and 450 in CUHK dataset.

To evaluate the robustness of the proposed method, we extract two kinds of features for the detected person images. The first feature (Fea1) is constructed by dividing a person image into six horizontal stripes and compute the RGB, YCbCr, HSV color features and two types of texture features extracted by Schmid and Gabor filters on each stripe as reported in [18] [19] [22]. For the second feature (Fea2), we concatenate the first feature with another one with foreground detection. The spatial hierarchy pose estimation method [59] with source code online[7] is employed to detect the local human parts as shown in the middle row of Fig. 4. Though it is reasonable to compute the parts based difference, the pose estimation is not accurate enough and does not consider the pair-wise matching relationship. Instead, we use the estimated human parts to mask the images as shown in the last row of Fig. 4(a)-4(d) for PRID, VIPeR, CUHK and i-LIDS datasets, respectively. Then, the masked person image is divided into 3×1 vertically overlapped boxes. Color histogram and SIFT [16] features are extracted on each box.

B. Comparison with Non-Learning Baselines

To evaluate our method, we compare with two commonly used non-learning based metrics namely L_1 and L_2 norms as baselines. The top r rank matching accuracies (%) are shown in Tables I-VI. From these results, we can see that Fea2 by concatenating Fea1 and another feature with foreground detection is more discriminative than Fea1. Comparing the results on VIPeR dataset in Table III and Table IV for Fea1 and Fea2, respectively, the rank one accuracy of L_1 using Fea2 is over three times higher than that using Fea1. And, the simple non-learning methods L_1 and L_2 can achieve high rank-one accuracies of 37.97% and 33.53%. Such good performance may be due to the reason that the combination of foreground detection and global feature extraction (on a large region of an image) is very effective for VIPeR dataset[8]. On the other hand, our method achieves better performance than L_1 and L_2 on the three target datasets with different source domains and features. Moreover, Table IV shows that when using CUHK as the source domain, the rank one accuracy of our method on VIPeR dataset with Fea2 is 9.50% higher than L_1 and 13.94% higher than L_2. These results convince that the proposed method can learn useful information from the source domain and target domain data to robustly improve the recognition performance with different source domains and features over the non-learning based methods.

TABLE I
TOP r RANK MATCHING ACCURACY (%) ON PRID USING FEA1

Source	Method	1	5	10	20
—	L_1	3.65	9.80	14.25	17.90
—	L_2	1.35	5.05	9.55	14.00
VIPeR	**Ours**	**4.85**	**13.10**	**18.35**	**26.25**
	RankSVM	1.05	5.90	9.70	16.20
	RDC	1.95	5.30	8.05	12.90
	RPML	1.10	7.00	11.85	17.40
CUHK	**Ours**	**4.50**	**11.70**	**16.85**	**24.50**
	RankSVM	1.95	4.40	6.80	12.60
	RDC	1.40	2.95	7.10	10.15
	RPML	0.45	4.75	7.95	12.85
i-LIDS	**Ours**	**4.85**	**12.65**	**18.55**	**27.45**
	RankSVM	2.95	6.75	11.40	19.65
	RDC	2.35	4.75	8.35	13.40
	RPML	0.90	4.20	6.80	12.65

TABLE II
TOP r RANK MATCHING ACCURACY (%) ON PRID DATASET FEA2

Source	Method	1	5	10	20
—	L_1	7.40	17.30	24.70	34.55
—	L_2	2.90	10.00	14.95	23.40
VIPeR	**Ours**	**7.60**	**19.95**	**28.05**	**37.25**
	RankSVM	5.90	14.15	20.25	26.70
	RDC	3.10	7.70	10.45	17.05
	RPML	3.05	8.70	15.65	22.05
CUHK	**Ours**	**10.35**	**22.95**	**30.65**	**40.25**
	RankSVM	5.50	16.50	21.85	29.15
	RDC	4.55	12.20	17.75	23.85
	RPML	3.30	9.75	13.15	19.20
i-LIDS	**Ours**	**9.20**	**21.05**	**28.15**	**39.50**
	RankSVM	5.35	16.60	24.65	34.15
	RDC	3.50	11.30	18.30	25.05
	RPML	2.10	7.30	12.10	17.55

C. Comparison with Supervised Learning Methods

Since label information is assumed to be not available in the target domain, existing supervised distance learning methods for person re-identification, e.g. RankSVM, RDC and RPML, cannot be employed to train a discriminative model for the target domain. For comparison, we use the labeled data from the source domain to train the RankSVM, RDC and RPML. Their results are recorded in Tables I-VI. From these tables,

[7] http://www.cs.cmu.edu/~ILIM/projects/IM/humanpose/humanpose.html

[8] It is interesting to further investigate the deeper reasons for the good performance on VIPeR dataset, but this is not the focus of this paper.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

TABLE III
TOP r RANK MATCHING ACCURACY (%) ON VIPER USING FEA1

Source	Method	1	5	10	20
—	L_1	8.86	18.20	23.80	33.84
—	L_2	9.53	18.80	25.60	34.38
PRID	**Ours**	**10.11**	**21.23**	**27.99**	**38.89**
	RankSVM	10.00	20.92	27.37	35.71
	RDC	9.70	19.21	26.46	35.63
	RPML	5.65	14.03	21.34	30.09
CUHK	**Ours**	**9.75**	**22.25**	**31.09**	**42.33**
	RankSVM	5.00	15.38	22.53	33.10
	RDC	7.30	17.75	25.98	36.72
	RPML	5.00	12.47	18.91	29.05
i-LIDS	**Ours**	**10.87**	**23.70**	**33.12**	**44.49**
	RankSVM	8.94	20.19	29.11	40.60
	RDC	7.23	17.10	24.53	35.41
	RPML	7.14	17.28	25.17	37.71

TABLE IV
TOP r RANK MATCHING ACCURACY (%) ON VIPER USING FEA2

Source	Method	1	5	10	20
—	L_1	37.97	58.26	65.89	75.08
—	L_2	33.53	53.83	62.56	70.65
PRID	**Ours**	**44.94**	**64.15**	**71.33**	**77.48**
	RankSVM	12.72	26.47	35.97	46.03
	RDC	25.89	44.27	53.58	63.40
	RPML	4.54	14.08	20.25	30.40
CUHK	**Ours**	**47.47**	**66.84**	**72.67**	**78.94**
	RankSVM	25.51	44.72	53.84	63.02
	RDC	35.46	54.19	62.15	70.66
	RPML	10.65	26.66	35.85	47.99
i-LIDS	**Ours**	**45.35**	**66.16**	**73.43**	**78.77**
	RankSVM	28.15	46.95	58.13	68.73
	RDC	31.79	52.23	61.16	71.77
	RPML	8.94	20.71	30.41	41.28

TABLE V
TOP r RANK MATCHING ACCURACY (%) ON CUHK USING FEA1

Source	Method	1	5	10	20
—	L_1	3.78	10.31	15.09	22.76
—	L_2	3.41	8.77	13.77	20.07
PRID	**Ours**	**4.85**	**13.30**	**19.71**	**27.85**
	RankSVM	3.38	10.20	15.87	23.13
	RDC	3.73	9.96	15.11	23.23
	RPML	0.94	4.15	7.85	13.94
VIPeR	**Ours**	**5.79**	**15.25**	**22.36**	**31.07**
	RankSVM	3.61	10.53	16.30	24.26
	RDC	2.82	10.20	15.32	21.78
	RPML	1.54	5.82	9.91	16.10
i-LIDS	**Ours**	**5.20**	**13.70**	**19.96**	**28.32**
	RankSVM	3.30	9.48	14.62	21.56
	RDC	1.84	7.12	11.72	18.67
	RPML	1.34	5.35	8.33	13.10

TABLE VI
TOP r RANK MATCHING ACCURACY (%) ON CUHK USING FEA2

Source	Method	1	5	10	20
—	L_1	9.20	20.44	27.24	35.95
—	L_2	6.54	15.59	21.92	29.25
PRID	**Ours**	**9.42**	**22.13**	**29.73**	**39.75**
	RankSVM	7.48	18.33	25.26	34.69
	RDC	9.00	20.63	28.57	37.79
	RPML	1.55	7.14	12.24	19.98
VIPeR	**Ours**	**10.57**	**22.96**	**31.15**	**41.06**
	RankSVM	8.25	20.42	27.28	35.81
	RDC	8.55	19.33	26.55	36.72
	RPML	2.26	8.19	13.42	21.47
i-LIDS	**Ours**	**9.57**	**22.52**	**30.67**	**40.86**
	RankSVM	8.52	18.88	27.07	37.79
	RDC	8.92	21.21	27.97	36.78
	RPML	2.09	8.88	14.68	23.34

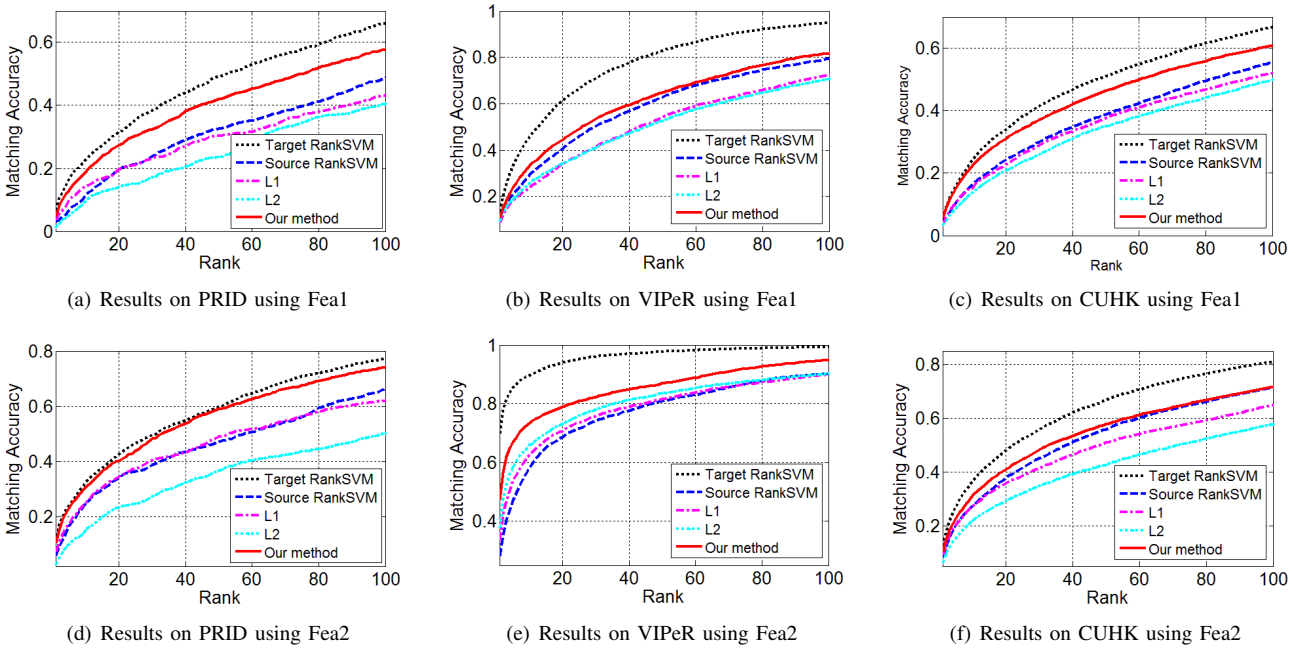
we can see that our method outperforms other supervised distance learning methods when using the source domain data for training. For many domain adaptation scenarios, the improvements by our methods are remarkable. For example, as shown in Table I, when the source domain is VIPeR or CUHK, the rank one accuracy of our method on PRID using Fea1 is higher than twice of others. Moreover, from Table IV, we can see that our method outperforms RankSVM, RDC and RPML by over 10% higher rank one accuracy on VIPeR dataset using Fea2. These results indicate that the proposed method can make good use of negative data (unmatched image pairs generated from non-overlapping cameras) and unlabeled data in the target domain to improve the re-identification performance, when the label information of persons is not available in the target domain.

From Tables II IV VI, we can see that the rank one accuracies of the unsupervised methods L_1 and L_2 with Fea2 are lower than 10% on PRID and CUHK datasets, while on VIPeR they achieve rank one accuracies of 37.97% and 33.53%, respectively. This means Fea2 is a very discriminative feature for VIPeR and less discriminative for PRID and CUHK, which indicates that the same feature can provide different discriminabilities for different datasets/domains. Such differences can cause that the supervised distance learning algorithms, e.g. RankSVM, RDC and RPML, may learn the incorrect information by the labeled data from the source domain. Therefore, in most cases the non-learning methods, e.g. L_1 and L_2, are better than the supervised ones as shown in Tables I-VI. Since our method makes use of the negative and unlabeled data from the target domain to align the distribution mismatch, it can outperform both the non-learning methods and the supervised distance learning algorithms using only labeled data from the source domain.

We further show the CMC curves of the RankSVM, RDC and RPML trained by the labeled data from target or source domain in Figs. 5(a)-(f), Figs. 6(a)-(f) and Figs. 7(a)-(f), respectively. For the results trained by the source domain data, the highest accuracy of each rank is recorded across the three different source datasets. All these figures show that the supervised distance learning methods have a dramatic deterioration of performance, when the classification model is trained with the data from the source domain. From Fig. 5(e), Fig. 6(e) and Fig. 7(e), we can see that the rank one accuracy can be degraded by about 30% with RDC and about 40% with RankSVM and RPML. This means the joint distributions in the source and target domains are different with each other for these datasets. And, the distribution misalignment causes serious performance degradation in person re-identification.

We also plot the CMC curves of our methods in Figs 5(a)-(f), Figs 6(a)-(f) and Figs 7(a)-(f). From these figures, we can see that our method outperforms other learning algorithms using labeled data from source domain, as well as two non-learning based metrics. In some cases, if the target positive mean can be estimated with very small error and represent the target positive data well, our method can achieve very convincing performance which is close to that of the supervised

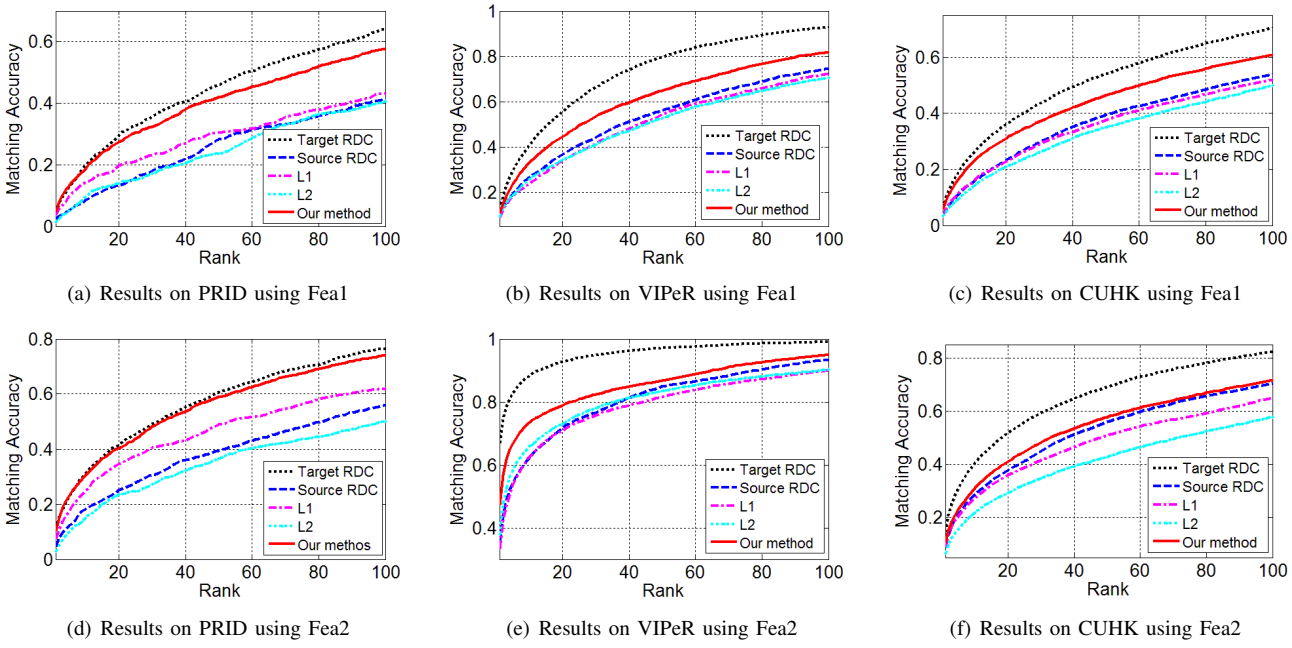
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing



(a) Results on PRID using Fea1 (b) Results on VIPeR using Fea1 (c) Results on CUHK using Fea1

(d) Results on PRID using Fea2 (e) Results on VIPeR using Fea2 (f) Results on CUHK using Fea2

Fig. 5. CMC curve comparison of our method and RankSVM trained by labeled data from target or source domain



(a) Results on PRID using Fea1 (b) Results on VIPeR using Fea1 (c) Results on CUHK using Fea1

(d) Results on PRID using Fea2 (e) Results on VIPeR using Fea2 (f) Results on CUHK using Fea2

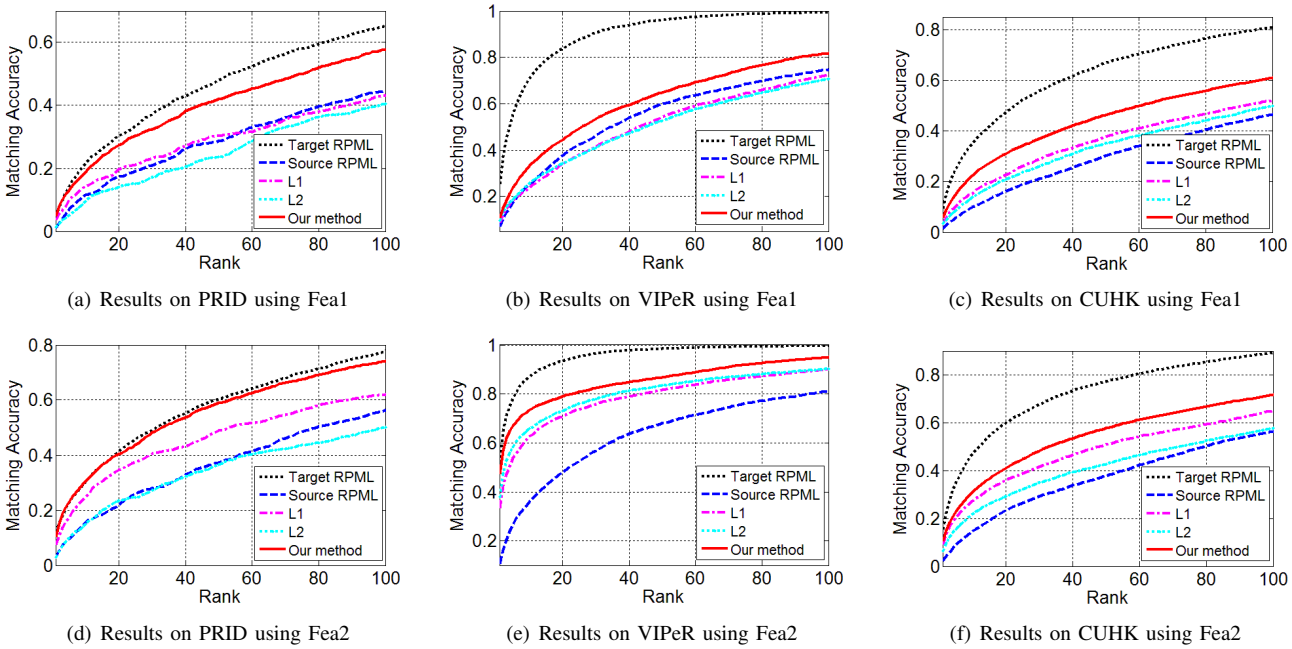
Fig. 6. CMC curve comparison of our method and RDC trained by labeled data from target or source domain

learning method using label information in the target domain. For example, the results in Fig. 5(d), Fig. 6(d) and Fig. 7(d) on PRID dataset using Fea2 show that our CMC curves are very close to the ones using both positive and negative labeled data in target domain for training. On the other hand, it is also possible that the estimated target positive mean contains error or cannot represent the positive data well. Under this situation, the performance of our method is not as good as supervised learning using target labeled data for training as shown in Figs. 5(b)(c)(e)(f), Figs. 6(b)(c)(e)(f) and Figs. 7(b)(c)(e)(f). However, by estimating the target positive mean, our method outperforms the supervised learning methods using source labeled data for training. This convinces that estimating target label information can help to align the joint distributions in source and target domains, so that the recognition performance has been improved.

D. Comparison with Domain Adaptation Algorithms

In this experiment, we would like to evaluate whether the equal conditional probability assumption is satisfied for person re-identification. Two state-of-the-art domain adaptation methods, Geodesic Flow Sampling (GFS) [30] and Transfer

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing



(a) Results on PRID using Fea1 (b) Results on VIPeR using Fea1 (c) Results on CUHK using Fea1

(d) Results on PRID using Fea2 (e) Results on VIPeR using Fea2 (f) Results on CUHK using Fea2

Fig. 7. CMC curve comparison of our method and RPML trained by labeled data from target or source domain

Component Analysis (TCA) [31], are used to learn a projection Φ for the alignment of the marginal distributions between the source and target domains, i.e. $\Pr_s(\Phi(\boldsymbol{z})) \approx \Pr_t(\Phi(\boldsymbol{z}))$. For the implementation, we use the source code provided by the authors for the GFS[9] and re-implement the TCA method. In our experiments, RankSVM is employed to train the discriminative model by the projected labeled data from the source domain. The best PCA dimension and number of intermediate subspace in the GFS are empirically set as 100 and 6, respectively. Linear kernel is employed in the TCA and the reduced dimension is set as 100.

The CMC curves of the domain adaptation methods using Fea2 are shown in Figs. 8(a)-(c), Figs. 9(a)-(c) and Figs. 10(a)-(c). When the source domain is CUHK or i-LIDS and target domain is VIPeR, Figs. 9(b)(c) show that the recognition accuracy by using the domain adaptation methods GFS and TCA are higher than that of the supervised learning method which uses labeled data from the source domain for training. This indicates that the re-identification performance could be improved by aligning the marginal distributions between source and target domain in some domain adaptation settings. Contrarily, simply aligning the marginal distributions may further degrade the classifiers as shown in Figs. 8(a)-(c) and Figs. 10(a)-(c). The reason could be that the source and target classifiers (also known as conditional distributions) may become farther away from each other in the space where the marginal distributions are aligned. On the other hand, all these figures show that the supervised learning method using target label information greatly outperforms GFS and TCA. This means the equal conditional probability assumption is not valid for these domain adaptation scenarios in person re-identification, i.e. $\Pr_s(y|\Phi(\boldsymbol{z})) \neq \Pr_t(y|\Phi(\boldsymbol{z}))$. Therefore, the joint distributions in the source and target domains are not equal with each other ($\Pr_s(y, \Phi(\boldsymbol{z})) \neq \Pr_t(y, \Phi(\boldsymbol{z}))$), though the marginal distributions are equal ($\Pr_s(\Phi(\boldsymbol{z})) \approx \Pr_t(\Phi(\boldsymbol{z}))$) after domain adaptation projection.

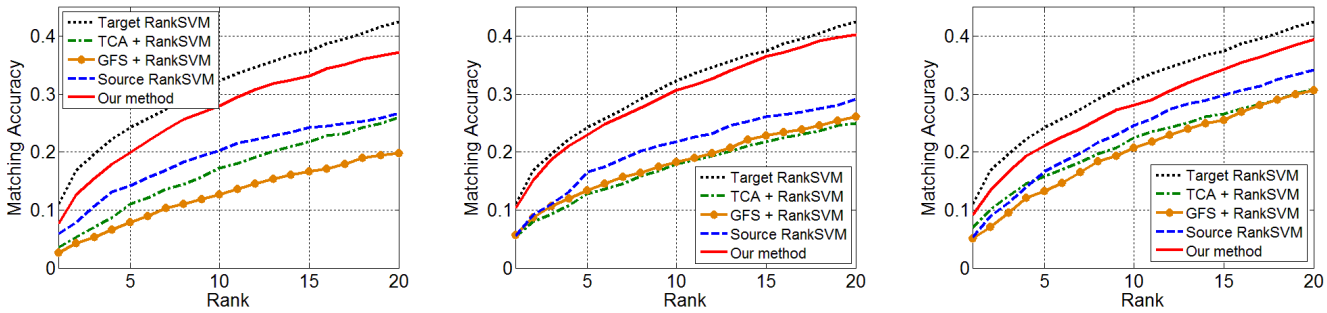
From Figs. 8(a)-(c), Figs. 9(a)-(c) and Figs. 10(a)-(c), we can see that our method clearly outperforms the domain adaptation methods derived based on the equal conditional probability assumption. This convinces that the proposed method can improve the re-identification performance by removing the assumption that the conditional probabilities in the source and target domains are equal. Moreover, Figs. 9(b)(c) show that the rank one accuracies of GFS and TCA are around 30% with CUHK or i-LIDS as the source domain, while the rank one accuracies of them are lower than 20% with PRID as the source domain. This means the domain adaptation methods with the equal conditional probability assumption may have over 10% degradation of rank one accuracy when using a different dataset as the source domain. By estimating the positive information in the target domain to solve the joint distribution mismatch problem, our method achieves much more robust performance (around 45% rank one accuracy) with different datasets as the source domain.

E. Comparing with Appearance-Based Methods

To demonstrate that the proposed domain adaptation approach can outperform unsupervised appearance-based methods for person re-identification, we compare our method with state-of-the-art algorithms, including Symmetry-Driven Accumulation of Local Features (SDALF) [6], Custom Pictorial Structures (CPS) [8], enriched Bio-inspired Covariance (eBiCov) [13], enriched Local Descriptors encoded by Fisher Vectors (eLDFV) [14], Color Invariant Signatures (CIS) [15] and enriched Salience Correspondence (eSDC) [16]. Source

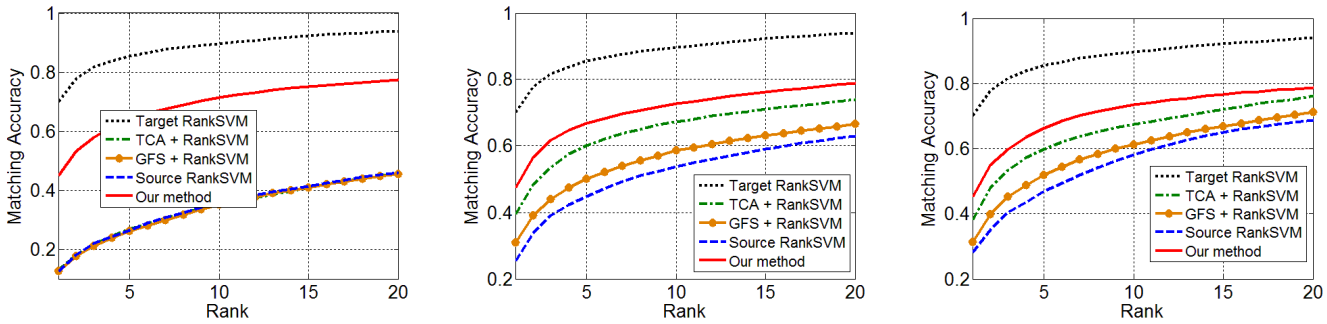
[9]http://www.umiacs.umd.edu/~raghuram/UnsupervisedDA_Grassmann.zip

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing



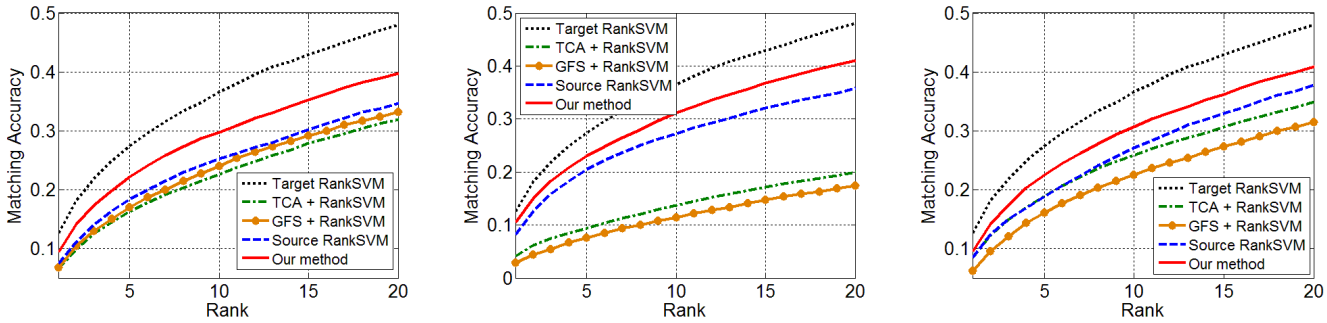
(a) VIPeR as source domain (b) CUHK as source domain (c) i-LIDS as source domain

Fig. 8. CMC curve comparison of our method and Domain Adaptation methods on PRID dataset as the target domain using Fea2



(a) PRID as source domain (b) CUHK as source domain (c) i-LIDS as source domain

Fig. 9. CMC curve comparison of our method and Domain Adaptation methods on VIPeR dataset as the target domain using Fea2



(a) PRID as source domain (b) VIPeR as source domain (c) i-LIDS as source domain

Fig. 10. CMC curve comparison of our method and Domain Adaptation methods on CUHK dataset as the target domain using Fea2

codes of SDALF[10], CIS[11] and eSDC[12] are provided online and used for our experiments, while the results of CPS, eBiCov and eLDFV are copied from their papers. The top r rank accuracy of these methods and the proposed AdaRSVM on VIPeR dataset are recorded in Table VII. From Table VII, we can see that the appearance-based methods outperform ours when using Fea1 to train the domain adaptation model. On the other hand, when Fea2 is used, no matter what the source domain is, our method can remarkably outperforms the appearance-based algorithms. The rank one accuracy of our method with CUHK as the source domain is over 20% higher than that of the best state-of-the-art appearance-based algorithm. Since the proposed method is independent of the input features, these results indicate that domain adaptation approach can improve the performance remarkably over appearance-based methods by employing a discriminative feature.

[10]http://www.lorisbazzani.info/code-datasets/sdalf-descriptor/
[11]http://www.cs.technion.ac.il/~kviat/colorReid.htm
[12]http://mmlab.ie.cuhk.edu.hk/projects/project_salience_reid/index.html

V. Conclusions

In this paper, we propose a novel Adaptive Ranking Support Vector Machines (AdaRSVM) method to deal with the problem that label information of persons is not available under target cameras. Without positive image pairs generated by the label information of persons, we relax the discriminative constraint to a necessary condition, which only relies on the mean of positive pairs. In order to estimate the positive mean in the target domain, we make use of the labeled data from the source domain, the negative and unlabeled data from the target

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

TABLE VII
TOP r RANK MATCHING ACCURACY (%) OF STATE-OF-THE-ART APPEARANCE-BASED ALGORITHMS AND OUR METHOD ON VIPeR

Method	Source	1	5	10	20
Ours+Fea1	PRID	10.11	21.23	27.99	38.89
	CUHK	9.75	22.25	31.09	42.33
	i-LIDS	10.87	23.70	33.12	44.49
Ours+Fea2	PRID	44.94	64.15	71.33	77.48
	CUHK	**47.47**	**66.84**	72.67	**78.94**
	i-LIDS	45.35	66.16	**73.43**	78.77
SDALF	—	19.87	38.89	49.37	65.73
CPS	—	21.84	44.00	57.21	71.00
eBiCov	—	20.66	42.00	56.18	68.00
eLDFV	—	22.34	47.00	60.04	71.00
CIS	—	24.24	44.91	56.55	69.40
eSDC	—	26.74	50.70	62.37	76.36

domain. With two estimations of the target positive mean, the optimal combination is determined by the training data. And, the target distance model is trained by adapting the source domain distance model to target domain.

Extensive experiments show that the proposed method achieve convincing recognition performance for person re-identification. The proposed AdaRSVM not only outperforms non-learning based methods but also is better than state-of-the-art discriminative learning methods using labeled data from the source domain for training. In our experiments, it is shown that the performance deteriorates dramatically when using the learnt model trained on source domain to target domain, which means the joint distributions in source and target domains are not equal to each other. On the other hand, compared with two domain adaptation methods for the alignment of marginal distributions, experimental results demonstrate that the equal conditional probability assumption is not valid for person re-identification. With the help of the negative image pairs generated from non-overlapping target cameras, the proposed AdaRSVM can improve the re-identification performance by estimating the target positive mean for domain adaptation learning. Moreover, it is also shown that the proposed domain adaptation method can remarkably outperform existing appearance-based methods on VIPeR dataset without using target label information for training.

While the proposed method only considers single source domain, multiple source domains are usually available in practice. Therefore, we will further investigate how to select or combine different source domains to train a more discriminative domain adaptation model in the future. On the other hand, our method is developed based on RankSVM. Since RankSVM does not address the problems for large-scale dataset, it may not perform as good as state-of-the-art large-scale ranking methods, e.g. [60], when a large amount of training data is available. Therefore, we are also interested in developing a new domain adaptation method for large-scale learning in person re-identification.

REFERENCES

[1] O. Javed, Z. Rasheed, K. Shafique, and M. Shah, "Tracking across multiple cameras with disjoint views," in *IEEE International Conference on Computer Vision*, vol. 2, 2003, pp. 952–957.

[2] N. Gheissari, T. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1528–1535.

[3] C. Madden, E. Cheng, and M. Piccardi, "Tracking people across disjoint camera views by an illumination-tolerant appearance representation," *Machine Vision and Applications*, vol. 18, no. 3-4, pp. 233–247, 2007.

[4] S. Bąk, E. Corvée, F. Brémond, and M. Thonnat, "Boosted human re-identification using riemannian manifolds," *Image and Vision Computing*, vol. 30, no. 6-7, pp. 443–452, 2010.

[5] S. Bąk, E. Corvee, F. Bremond, and M. Thonnat, "Person re-identification using Haar-based and DCD-based signature," in *IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2010, pp. 1–8.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2360–2367.

[7] M. Bauml and R. Stiefelhagen, "Evaluation of local features for person re-identification in image sequences," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2011, pp. 291–296.

[8] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," in *British Machine Vision Conference*, 2011, pp. 68.1–68.11.

[9] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 2, pp. 127–151, 2011.

[10] K. Jungling and M. Arens, "View-invariant person re-identification with an implicit shape model," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2011, pp. 197–202.

[11] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 898–903, 2012.

[12] S. Bąk, G. Charpiat, E. Corvée, F. Brémond, and M. Thonnat, "Learning to match appearances by correlations in a covariance metric space," in *European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, 2012, vol. 7574, pp. 806–820.

[13] B. Ma, Y. Su, and F. Jurie, "BiCov: a novel image representation for person re-identification and face verification," in *British Machive Vision Conference*, 2012.

[14] ——, "Local descriptors encoded by fisher vectors for person re-identification," in *International Workshop on Re-Identification in conjunction with European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, 2012, vol. 7583, pp. 413–422.

[15] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1622–1634, 2013.

[16] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3586–3593.

[17] Y. Xu, L. Lin, W.-S. Zheng, and X. Liu, "Human re-identification by matching compositional template with cluster sampling," in *IEEE International Conference on Computer Vision*, 2013.

[18] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, 2008, vol. 5302, pp. 262–275.

[19] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *British Machine Vision Conference*, 2010, pp. 21.1–21.11.

[20] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning implicit transfer for person re-identification," in *International Workshop on Re-Identification in conjunction with European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, 2012, vol. 7583, pp. 381–390.

[21] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, 2012, vol. 7577, pp. 780–793.

[22] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 653–668, 2013.

[23] W. Li and X. Wang, "Locally aligned feature transforms across views," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3594–3601.

[24] C. Liu, C. C. Loy, S. Gong, and G. Wang, "POP: Person re-identification post-rank optimisation," in *IEEE International Conference on Computer Vision*, 2013.

[25] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *IEEE International Conference on Computer Vision*, 2013.

[26] L. Ma, X. Yang, and D. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3656–3670, 2014.

[27] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Scandinavian Conference on Image Analysis*, ser. Lecture Notes in Computer Science, 2011, vol. 6688, pp. 91–102.

[28] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, 2007, pp. 41–47.

[29] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[30] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *IEEE International Conference on Computer Vision*, 2011, pp. 999–1006.

[31] S. J. Pan, J. T. K. Ivor W. Tsang, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[32] T. Evgeniou and M. Pontil, "Regularized multi–task learning," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 109–117.

[33] Y. Xue, X. Liao, L. Carin, and B. Krishnapuram, "Multi-task learning for classification with dirichlet process priors," *Journal of Machine Learning Research*, vol. 8, pp. 35–63, May 2007.

[34] J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain video concept detection using adaptive SVMs," in *Proceedings of the ACM International Conference on Multimedia*, 2007, pp. 188–197.

[35] L. Duan, D. Xu, I.-H. Tsang, and J. Luo, "Visual event recognition in videos by learning from web data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1667–1680, 2012.

[36] A. J. Ma, P. C. Yuen, and J. Li, "Domain transfer support vector ranking for person re-identification without target camera label information," in *IEEE International Conference on Computer Vision*, 2013.

[37] C.-H. Kuo, C. Huang, and R. Nevatia, "Inter-camera association of multi-target tracks by on-line learned appearance affinity models," in *European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, 2010, vol. 6311, pp. 383–396.

[38] C. Liu, S. Gong, C. Loy, and X. Lin, "Person re-identification: What features are important?" in *International Workshop on Re-Identification in conjunction with European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, 2012, vol. 7583, pp. 391–401.

[39] W.-S. Zheng, S. Gong, and T. Xiang, "Transfer re-identification: From person to set-based verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2650–2657.

[40] M. Sugiyama and M. Kawanabe, *Machine learning in non-stationary environments: introduction to covariate shift adaptation*. MIT Press, 2012.

[41] T. S. M Sugiyama and T. Kanamori, *Density ratio estimation in machine learning*. Cambridge University Press, 2012.

[42] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Conference on Empirical Methods in Natural Language Processing*, 2006, pp. 120–128.

[43] B. Geng, D. Tao, and C. Xu, "DAML: Domain adaptation metric learning," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2980–2989, 2011.

[44] B. Gong, K. Grauman, and F. Sha, "Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation," in *International Conference on Machine Learning*, 2013, pp. 222–230.

[45] Y.-R. Yeh, C.-H. Huang, and Y.-C. Wang, "Heterogeneous domain adaptation and classification by exploiting the correlation subspace," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2009–2018, 2014.

[46] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, no. 1-2, pp. 151–175, 2010.

[47] H. Daumé III and D. Marcu, "Domain adaptation for statistical classifiers," *Journal of Artificial Intelligence Research*, vol. 26, pp. 101–126, 2006.

[48] R. Raina, A. Y. Ng, and D. Koller, "Constructing informative priors using transfer learning," in *International conference on Machine learning*, 2006, pp. 713–720.

[49] Q. Liu, X. Liao, H. Li, J. R. Stack, and L. Carin, "Semisupervised multitask learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 1074–1086, 2009.

[50] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A dasvm classification technique and a circular validation strategy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 770–787, 2010.

[51] L. Duan, I. Tsang, and D. Xu, "Domain transfer multiple kernel learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 465–479, 2012.

[52] J. Donahue, J. Hoffman, E. Rodner, K. Saenko, and T. Darrell, "Semi-supervised domain adaptation with instance constraints," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 668–675.

[53] Z. Guo and Z. Wang, "Cross-domain object recognition via input-output kernel analysis," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3108–3119, 2013.

[54] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[55] N. Kwak and J. Oh, "Feature extraction for one-class classification problems: Enhancements to biased discriminant analysis," *Pattern Recognition*, vol. 42, no. 1, pp. 17–26, 2009.

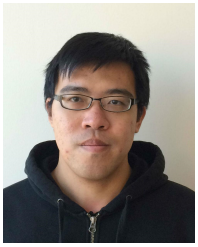
[56] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume I*. Wiley, 1968.

[57] O. Chapelle and S. S. Keerthi, "Efficient algorithms for ranking with SVMs," *Information Retrieval*, vol. 13, no. 3, pp. 201–215, 2010.

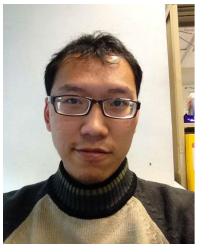
[58] W.-S. Zheng, S. Gong, and T. Xiang, "Associating groups of people," in *British Machine Vision Conference*, 2009, pp. 23.1–23.11.

[59] Y. Tian, C. Zitnick, and S. Narasimhan, "Exploring the spatial hierarchy of mixture models for human pose estimation," in *European Conference on Computer Vision*, ser. Lecture Notes in Computer Science, 2012, vol. 7576, pp. 256–269.

[60] P. Li, Q. Wu, and C. J. Burges, "Mcrank: Learning to rank using multiple classification and gradient boosting," in *Advances in neural information processing systems*, 2007, pp. 897–904.



Andy J Ma received his B.Sc. and M.Sc. degree in applied mathematics from Sun Yat-Sen University in 2007 and 2009, respectively. He obtained his Ph.D. degree in the Department of Computer Science from Hong Kong Baptist University in 2013. He is currently working as a Post-Doctoral Fellow in the Department of Computer Science at Johns Hopkins University. His current research interests focus on developing machine learning algorithms for intelligent video surveillance.



Jiawei Li received the B.Sc. degree in mathematics and applied mathematics, and the M.Sc. degree in mathematics from Sun Yat-sen University, Guangzhou, China, in 2007 and 2011, respectively. He is currently a Ph.D. candidate with the Department of Computer Science, Hong Kong Baptist University, Hong Kong.

His research interests include pattern recognition, computer vision and machine learning. And he is now focusing on transfer learning and person re-identification.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2015.2395715, IEEE Transactions on Image Processing

Pong C Yuen received his B.Sc. degree in Electronic Engineering with First Class Honours in 1989 from City Polytechnic of Hong Kong, and his Ph.D. degree in Electrical and Electronic Engineering in 1993 from The University of Hong Kong. He joined the Hong Kong Baptist University in 1993 and, currently is a Professor and Head of the Department of Computer Science.

Dr. Yuen was a recipient of the University Fellowship to visit The University of Sydney in 1996. He was associated with the Laboratory of Imaging Science and Engineering, Department of Electrical Engineering. In 1998, Dr. Yuen spent a 6-month sabbatical leave in The University of Maryland Institute for Advanced Computer Studies (UMIACS), University of Maryland at college park. From June 2005 to January 2006, he was a visiting professor in GRAVIR laboratory (GRAphics, VIsion and Robotics) of INRIA Rhone Alpes, France. Dr. Yuen was the director of Croucher Advanced Study Institute (ASI) on biometric authentication in 2004 and the director of Croucher ASI on Biometric Security and Privacy in 2007.

Dr. Yuen has been actively involved in many international conferences as an organizing committee and/or technical program committee member. He was the track co-chair of International Conference on Pattern Recognition (ICPR) 2006 and the program co-chair of IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS) 2012. Currently, Dr. Yuen is an Editorial Board Member of Pattern Recognition and Associate Editor of IEEE Transactions on Information Forensics and Security, and SPIE Journal of Electronic Imaging.

Dr. Yuen's current research interests include video surveillance, human face recognition, biometric security and privacy.

Ping Li earned his Ph.D. in Statistics and Masters in EE and CS from Stanford University. He is currently Associate Professor in Statistics and Associate Professor in Computer Science at Rutgers University. His current research focuses on developing hashing algorithms for large-scale search and learning. Ping Li won a prize in the 2010 Yahoo Learning to Rank Grand Challenge. He was the receipt of the Young Investigator Award (YIP) from the AFOSR and the recipient of the Young Investigator Award (YIP) from the ONR.