

Person Re-Identification by Regularized Smoothing KISS Metric Learning

Dapeng Tao, Lianwen Jin, *Member, IEEE*, Yongfei Wang, Yuan Yuan, *Senior Member, IEEE*, and Xuelong Li, *Fellow, IEEE*

Abstract—With the rapid development of the intelligent video surveillance (IVS), person re-identification, which is a difficult yet unavoidable problem in video surveillance, has received increasing attention in recent years. That is because computer capacity has shown remarkable progress and the task of person re-identification plays a critical role in video surveillance systems. In short, person re-identification aims to find an individual again that has been observed over different cameras. It has been reported that KISS metric learning has obtained the state of the art performance for person re-identification on the VIPeR dataset [39]. However, given a small size training set, the estimation to the inverse of a covariance matrix is not stable and thus the resulting performance can be poor. In this paper, we present regularized smoothing KISS metric learning (RS-KISS) by seamlessly integrating smoothing and regularization techniques for robustly estimating covariance matrices. RS-KISS is superior to KISS, because RS-KISS can enlarge the underestimated small eigenvalues and can reduce the overestimated large eigenvalues of the estimated covariance matrix in an effective way. By providing additional data, we can obtain a more robust model by RS-KISS. However, retraining RS-KISS on all the available examples in a straightforward way is time consuming, so we introduce incremental learning to RS-KISS. We thoroughly conduct experiments on the VIPeR dataset and verify that 1) RS-KISS completely beats all available results for person re-identification and 2) incremental RS-KISS performs as well as RS-KISS but reduces the computational cost significantly.

Index Terms—Incremental learning, intelligent video surveillance, metric learning, person re-identification.

Manuscript received September 28, 2012; revised January 3, 2013 and February 21, 2013; accepted February 26, 2013. Date of publication March 28, 2013; date of current version September 28, 2013. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB316400, the National Natural Science Foundation of China under Grants 61125106, 61075021, 61201348, 91120302, and 61072093, the National Science and Technology Support Plan under Grants 2013BAH65F01 and 2013BAH65F04, the Guangdong Natural Science Funds under Grant S2011020000541, the Guangdong Scientific and Technology Research Plan (No. 2012A010701001), the Fundamental Research Funds for the Central Universities of China under Grants 2012ZP0002 and D2116320, and the Shaanxi Key Innovation Team of Science and Technology under Grant 2012KCT-04. This paper was recommended by Associate Editor M. Murshed.

D. Tao, L. Jin, and Y. Wang are with the School of Electronic and Information Engineering, South China University of Technology, GuangZhou 510640, China (e-mail: dapeng.tao@gmail.com; lianwen.jin@gmail.com; yongfei.wang.ee@gmail.com).

Y. Yuan and X. Li are with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mails: yuany@opt.ac.cn; xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2013.2255413

I. INTRODUCTION

PERSON RE-IDENTIFICATION aims to recognize a target of interest over different camera views at different locations and thus, it benefits the booming area of the public security, especially when biometrics, e.g., face [39], gait [61] and fingerprint [7], are not available. In recent years, person re-identification exploiting the body appearance [1], [3], [8], [31], [34], [42], [56] for identification has received increasing attention. This task is challenging, because of small target size, partial occlusion, motion blur, and appearance variation.

In general, person re-identification can be regarded as a special task of the visual retrieval problem [39], i.e., by treating the target of interest as search query, the specific computer system searches the correct match among candidates recorded from different camera views. Typically, it consists of two important stages. First, robust visual features are extracted to represent a person. Second, an effective and efficient matching model is applied to conduct high performance search.

In recent years, a large number of interesting schemes have been proposed to improve the performance of re-identification. We can simply group these schemes into two categories: 1) schemes for extracting new features for robust representation, and 2) schemes for developing new matching models.

Many exciting studies about visual feature extraction and robust representation have been presented and developed over the past years. Most results are readily applicable to the problem of person re-identification. Represented works are briefed below.

Low-level visual features, such as color, texture, and shape are popular in visual recognition. The RGB color histogram and the HSV color histogram are robust against the variability of resolution and perspective [23]. In addition, directly extracting color features from the bounding box areas obtained by detection approaches is not unique. Some research results [12], [52] utilized background subtraction algorithms [2], [59] to minimize impacts arose by the complex background before extracting features. Although the segmentation improves the robustness of the extracted features, the computational cost increases significantly. Considering that the color descriptors are sensitive to light conditions, Gabor filters [15] and Schmid filters [55] have been added to the feature extraction procedures for robust analysis [23], [69].

Local features [6], [16], [43] obtain robust representations over interest points or regions. Scale invariant feature

transform (SIFT) [43] is classic for extracting local features in many computer vision systems [39]. Hamdoun *et al.* [25] improved SIFT by developing SURF to obtain interest points efficiently. Bak *et al.* [1] selected the discriminative Haar like features for matching tasks by utilizing the classical Adaboost approach [18]. Local binary patterns (LBP) [45] are visual feature descriptors that depict the local structures of an interest point. Recently, LBP has been exploited for person re-identification [39]. A comprehensive comparison of different local features is given in [29].

Note that for person re-identification, a combination of selected aforementioned features is usually deployed. Thus, an efficient method of dimension reduction is necessary to retain the most effective features for the subsequent matching [20], [24], [67]. Principal component analysis (PCA) [32], [36], [37] aims to find the principal subspace to achieve the maximized variance of the projected points. It is effective for removing the Gaussian noise. Linear discriminant analysis (LDA) [14] is a linear supervised algorithm and aims to optimally classify samples drawn from Gaussians with different means but an identical covariance. Many manifold learning based dimension reduction techniques, such as locally linear embedding (LLE) [51], Laplacian eigenmaps (LE) [4] and discriminative locality alignment (DLA) [70], are developed to consider the local geometry of a set of high dimensional data. A unifying framework [71] was proposed to properly understand representative manifold learning algorithms and their respective linearizations, such as locality preserving projections [27]. Spatial pyramid matching (SPM) [40] represents the samples by using the local SIFT features extracted from samples. The classical SPM performs better while the classifier is constructed by Mercer kernels.

Schemes for improving matching models are the main focus of this paper and are important for many practical applications, such as image search [41], [46], [47], human action recognition [49], [63], signature verification [13], [33], [65], and face recognition [39], [57], [62]. Although existing approaches achieved top-level performance, they perform poorly for person re-identification. As a metric learning method, KISS Metric Learning is efficient and effective and has obtained top-level performance on various challenging benchmarks, such as face recognition and person re-identification [39]. It is a statistical inference scheme that does not rely on complex iterative optimization. This advantage is important for practical applications and receives intensive attentions. However, KISS has the small sample size problem for calculating the inverse of the covariance matrices and may not perform robustly in practice because the number of the training samples in person re-identification is much smaller than feature dimensionality.

In this paper, we introduce the smoothing and regularization techniques to largely improve KISS for person re-identification. The estimate to a covariance matrix is biased. In particular, the large eigenvalues of the true covariance matrix are biased high in the estimated covariance matrix, while the small eigenvalues of the true covariance matrix are biased low in the estimated covariance matrix. Especially, the lower estimate to small eigenvalues and the higher estimate to large eigenvalues harm the utilization of the estimated covariance

matrix in the subsequent operations, such as classification. To obtain a robust estimate to the covariance matrices for KISS, we first introduce the smoothing technique [38] to improve the estimate to the small eigenvalues of a covariance matrix, and then use the regularization [17] to reduce the effect of the larger estimate to large eigenvalues of the covariance matrix. The improved KISS is termed regularized smoothing KISS or simply RS-KISS.

In real video surveillance systems, the size of the labeled training set is usually increasing along the time. Although we can retrain the whole system, the whole process is very time consuming. Thus, we need to find a solution to incrementally updating the previously trained model given additional training samples. We introduce incremental learning to RS-KISS and further develop incremental RS-KISS (IRS-KISS) for person re-identification.

We summarize the procedure for the RS-KISS/IRS-KISS based person re-identification in the following steps: 1) extracting texture feature and color histogram from each sample; 2) concatenating all the feature descriptors together and conducting PCA to obtain a low-dimension representation for each sample; 3) training RS-KISS or updating the distance metric by using IRS-KISS, and 4) finding the matching rank according the query target. Given the limited page length, we do not detail the other parts, because implementations are easy obtained based upon the references therein.

We organize the rest of the paper as follows. In Section II, we briefly review related works for improving the matching models for person re-identification. We detail the newly proposed RS-KISS and IRS-KISS in Section III. Section IV shows the experimental results on the three representative datasets and Section V concludes the paper.

II. RELATED WORK

In the introduction Section, we briefly reviewed popular features used in person re-identification. In recently, training a robust and efficiency matching scheme has been received increasing attentions [9], [19], [41], [64], [69]. Large margin nearest neighbor metric (LMNN) [64] are proposed to improve the performance of the traditional k NN classification. However, LMNN is time-consuming for using the k closest within-class samples. By minimizing the differential relative entropy between two multivariate Gaussians under constraints on the distance function, information theoretic metric learning (ITML) [9] is built on the Mahalanobis distance metric learned from the information theoretic perspective. Zheng *et al.* [69] proposed a soft discriminative scheme termed relative distance comparison (RDC) by large and small distances corresponding to wrong matches and right matches, respectively. To some extent, the solution of RDC is complicated, and it can be solved by an iterative optimization algorithm. The L2 distance, Mahalanobis Metric, and Bhattacharyya distance are also applied to person re-identification [23]. However, they perform poorly when the view conditions change greatly. As a pairwise method, rank support vector machines (RankSVM) [28] have been extensively used in retrieval related problems. The Ensemble RankSVM is presented by Prosser *et al.* [50]

to handle the scalability issue by using ensemble learning.

Besides distance metric learning based matching schemes, contextual cues are very useful for improving the accuracy and robustness of person re-identification. By utilizing subspace learning, a brightness transfer function [35] is proposed to deal with the illumination changes between different cameras. Makris *et al.* [44] proposed a scheme to determine the topography of cameras by observed location and velocity of moving objects. Hamdoun *et al.* [25] collected interest points from surveillance video shots to estimate the appearance model. The contextual cues can be implemented as a preprocessing step to improve the system of person re-identification.

III. REGULARIZED SMOOTHING KISS AND INCREMENTAL LEARNING

It has been reported that the KISS metric learning (KISS) has obtained the state of the art performance for person re-identification on the VIPeR dataset [39].

Given a feature vector pair \mathbf{x}_i and \mathbf{x}_j referring to two people, respectively, let H_0 denote the hypothesis that the feature vector pair is dissimilar (\mathbf{x}_i and \mathbf{x}_j are different people), and H_1 denote the hypothesis that the feature vector pair is similar (\mathbf{x}_i and \mathbf{x}_j are the same person). The logarithm of the ratio between the two posteriors is

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log \left(\frac{p(H_0|\mathbf{x}_i, \mathbf{x}_j)}{p(H_1|\mathbf{x}_i, \mathbf{x}_j)} \right). \quad (1)$$

For classification, a positive value of $\delta(\mathbf{x}_i, \mathbf{x}_j)$ indicates \mathbf{x}_i and \mathbf{x}_j are different people, while a negative value means \mathbf{x}_i and \mathbf{x}_j are the same person. We denote the difference of the feature vector pair by $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, and thus, we have

$$\delta(\mathbf{x}_{ij}) = \log(p(H_0|\mathbf{x}_{ij}) / p(H_1|\mathbf{x}_{ij})) \quad (2)$$

which can be rewritten as

$$\delta(\mathbf{x}_{ij}) = \log(f(\mathbf{x}_{ij}|H_0) / f(\mathbf{x}_{ij}|H_1)) + \log(p(H_0) / p(H_1)) \quad (3)$$

where $f(\mathbf{x}_{ij}|H_0)$ and $f(\mathbf{x}_{ij}|H_1)$ are the probability density functions of \mathbf{x}_{ij} under the hypothesis of H_0 and H_1 , respectively. That means, $f(\mathbf{x}_{ij}|H_0)$ is the probability density functions of the difference of the similar feature vector pair, while $f(\mathbf{x}_{ij}|H_1)$ is the probability dense functions of the difference of dissimilar feature vector pair. Notice that the mean of \mathbf{x}_{ij} is 0, and it is common to assume that \mathbf{x}_{ij} satisfies the Gaussian distribution. Thus, we have

$$f(\mathbf{x}_{ij}|H_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}_{ij}^T \Sigma_k^{-1} \mathbf{x}_{ij}\right) \quad (4)$$

where $k \in \{0, 1\}$, d is the dimensionality of the feature vector, and Σ_k is the covariance matrix of \mathbf{x}_{ij} .

Given (4), (3) can be simplified as

$$\delta(\mathbf{x}_{ij}) = \frac{1}{2} \mathbf{x}_{ij}^T (\Sigma_1^{-1} - \Sigma_0^{-1}) \mathbf{x}_{ij} + \frac{1}{2} \log \left(\frac{|\Sigma_1|}{|\Sigma_0|} \right) + \log \left(\frac{p(H_0)}{p(H_1)} \right). \quad (5)$$

By dropping the constant terms, we have

$$\delta(\mathbf{x}_{ij}) = \mathbf{x}_{ij}^T (\Sigma_1^{-1} - \Sigma_0^{-1}) \mathbf{x}_{ij}. \quad (6)$$

Define y_{ij} as the indicate variable of \mathbf{x}_i and \mathbf{x}_j : $y_{ij} = 1$ if \mathbf{x}_i and \mathbf{x}_j are the same person, otherwise $y_{ij} = 0$. Let N_0 denote the number of similar feature vector pairs, while N_1 denote the number of dissimilar feature vector pairs. The covariance matrices are estimated as follows:

$$\begin{aligned} \Sigma_0 &= \frac{1}{N_0} \sum_{y_{ij}=0} \mathbf{x}_{ij} \mathbf{x}_{ij}^T = \frac{1}{N_0} \sum_{y_{ij}=0} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T \\ \Sigma_1 &= \frac{1}{N_1} \sum_{y_{ij}=1} \mathbf{x}_{ij} \mathbf{x}_{ij}^T = \frac{1}{N_1} \sum_{y_{ij}=1} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T \end{aligned} \quad (7)$$

Let $M = \Sigma_1^{-1} - \Sigma_0^{-1}$, KISS projects M onto the cone of the positive semi-definite matrices \tilde{M} , i.e.

$$\delta(\mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \tilde{M} \mathbf{x}_{ij}. \quad (8)$$

A. RS-KISS Metric Learning

The estimates to covariance matrices in (6) are critical to obtain robust performance for person re-identification. In practice, it is impossible to collect a large amount of labeled samples to effectively estimate the small eigenvalues of the covariance matrices in person re-identification.

It is known that the estimate to a covariance matrix is always biased. Given a small size training set, the large eigenvalues of the true covariance matrix are biased high in the estimated covariance matrix, while the small eigenvalues of the true covariance matrix are biased low in the estimated covariance matrix. This biased estimation problem is related to ill-posed problem.

In statistics, there are many ways to obtain robust estimations. In this paper, we introduce the smoothing technique [38] and the regularization method [17] to improve estimates to covariance matrices in KISS, because the smoothing technique can enlarge the estimate to the small eigenvalues of a covariance matrix and the regularization technique can reduce the effect of the larger estimate to large eigenvalues of an estimated covariance matrix. By seamlessly integrating them, we can consequently improve the effectiveness of KISS for person re-identification.

We first diagonalize the covariance matrix Σ_i

$$\Sigma_i = \Phi_i \Lambda_i \Phi_i^T \quad (9)$$

where $\Lambda_i = \text{diag}[\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{id}]$ with λ_{ij} being an eigenvalue of Σ_i , and $\Phi_i = [\phi_{i1}, \phi_{i2}, \dots, \phi_{id}]$ with ϕ_{ij} being an eigenvector of Σ_i . Substitute (9) into (6), we have

$$\begin{aligned} \delta(\mathbf{x}_{ij}) &= \mathbf{x}_{ij} (\Sigma_1^{-1} - \Sigma_0^{-1}) \mathbf{x}_{ij}^T \\ &= \mathbf{x}_{ij} (\Phi_1 \Lambda_1^{-1} \Phi_1^T - \Phi_0 \Lambda_0^{-1} \Phi_0^T) \mathbf{x}_{ij}^T \\ &= [\Phi_1^T \mathbf{x}_{ij}]^T \Lambda_1^{-1} [\Phi_1^T \mathbf{x}_{ij}] - [\Phi_0^T \mathbf{x}_{ij}]^T \Lambda_0^{-1} [\Phi_0^T \mathbf{x}_{ij}] \\ &= \sum_{n=1}^d \frac{1}{\lambda_{1n}} (\phi_{1n}^T \mathbf{x}_{ij})^2 - \sum_{n=1}^d \frac{1}{\lambda_{0n}} (\phi_{0n}^T \mathbf{x}_{ij})^2. \end{aligned} \quad (10)$$

According to the smoothing technique, we first replace the small eigenvalues of the covariance matrix with a small constant β_i , and then we have

$$\Lambda_i = \text{diag} \left[\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{ik}, \underbrace{\beta_i, \dots, \beta_i}_{d-k} \right] \quad (11)$$

where the constant β_i takes the value of the average of all the small eigenvalues

$$\beta_i = \frac{1}{d-k} \sum_{n=k+1}^d \lambda_{in}. \quad (12)$$

Thus, (10) can be written as

$$\begin{aligned} \delta(\mathbf{x}_{ij}) &= \sum_{n=1}^d \frac{1}{\lambda_{in}} (\phi_{1n}^T \mathbf{x}_{ij})^2 - \sum_{n=1}^d \frac{1}{\lambda_{0n}} (\phi_{0n}^T \mathbf{x}_{ij})^2 \\ &= \sum_{n=1}^k \frac{1}{\lambda_{in}} (\phi_{1n}^T \mathbf{x}_{ij})^2 + \sum_{n=k+1}^d \frac{1}{\beta_1} (\phi_{1n}^T \mathbf{x}_{ij})^2 \\ &\quad - \sum_{n=1}^k \frac{1}{\lambda_{0n}} (\phi_{0n}^T \mathbf{x}_{ij})^2 - \sum_{n=k+1}^d \frac{1}{\beta_0} (\phi_{0n}^T \mathbf{x}_{ij})^2 \end{aligned} \quad (13)$$

By introducing $\|\mathbf{x}_{ij}\|^2$ to (13), we can avoid calculating $\sum_{n=k+1}^d \frac{1}{\beta_1} (\phi_{1n}^T \mathbf{x}_{ij})^2$ and $\sum_{n=k+1}^d \frac{1}{\beta_0} (\phi_{0n}^T \mathbf{x}_{ij})^2$. Then we have

$$\begin{aligned} \delta(\mathbf{x}_{ij}) &= \sum_{n=1}^k \frac{1}{\lambda_{in}} (\phi_{1n}^T \mathbf{x}_{ij})^2 + \frac{1}{\beta_1} \left(\|\mathbf{x}_{ij}\|^2 - \sum_{n=1}^k (\phi_{1n}^T \mathbf{x}_{ij})^2 \right) \\ &\quad - \sum_{j=1}^k \frac{1}{\lambda_{0n}} (\phi_{0n}^T \mathbf{x}_{ij})^2 - \frac{1}{\beta_0} \left(\|\mathbf{x}_{ij}\|^2 - \sum_{n=1}^k (\phi_{0n}^T \mathbf{x}_{ij})^2 \right) \\ &= \left(\frac{1}{\lambda_{in}} - \frac{1}{\beta_1} \right) \sum_{n=1}^k (\phi_{1n}^T \mathbf{x}_{ij})^2 + \left(\frac{1}{\beta_1} - \frac{1}{\beta_0} \right) \|\mathbf{x}_{ij}\|^2 \\ &\quad - \left(\frac{1}{\lambda_{0n}} - \frac{1}{\beta_0} \right) \sum_{n=1}^k (\phi_{0n}^T \mathbf{x}_{ij})^2 \end{aligned} \quad (14)$$

Equation (14) shows that the score of metric is weighted by the small eigenvalues.

The covariance matrix (9) is further interpolated by an identity matrix according to the regularization technique [17], i.e.

$$\begin{aligned} \tilde{\Sigma}_i &= (1-\gamma) \Sigma_i + \gamma \alpha_i \mathbf{I} \\ &= (1-\gamma) \Phi_i \Lambda_i \Phi_i^T + \gamma \alpha_i \Phi_i \Phi_i^T \\ &= \Phi_i \left[(1-\gamma) \Lambda_i + \gamma \alpha_i \mathbf{I} \right] \Phi_i^T \end{aligned} \quad (15)$$

where $\alpha_i = (1/d) \text{tr}(\Sigma_i)$ and $0 < \gamma < 1$. The parameter γ can shrink $\tilde{\Sigma}_i$ toward identity matrix. A shrunken estimate of the covariance matrix will suppress the larger estimate to the large eigenvalues and thus, improve the prediction performance in practice [17], [26].

By substituting (11) to (15), we have

$$\begin{aligned} \tilde{\Sigma}_i &= \Phi_i \{ \text{diag}[(1-\gamma) \lambda_{i1} + \gamma \alpha_i, \\ &\quad (1-\gamma) \lambda_{i2} + \gamma \alpha_i, \dots, (1-\gamma) \lambda_{ik} + \gamma \alpha_i, \\ &\quad \underbrace{(1-\gamma) \beta_i + \gamma \alpha_i, \dots, (1-\gamma) \beta_i + \gamma \alpha_i}_{d-k}] \} \Phi_i^T. \end{aligned} \quad (16)$$

We define

$$\begin{aligned} \tilde{\Lambda}_i &= \text{diag}[(1-\gamma) \lambda_{i1} + \gamma \alpha_i, \\ &\quad (1-\gamma) \lambda_{i2} + \gamma \alpha_i, \dots, (1-\gamma) \lambda_{ik} + \gamma \alpha_i, \\ &\quad \underbrace{(1-\gamma) \beta_i + \gamma \alpha_i, \dots, (1-\gamma) \beta_i + \gamma \alpha_i}_{d-k}]. \end{aligned} \quad (17)$$

Replace Σ_i with $\tilde{\Sigma}_i$ in equation (10), we obtain

$$\begin{aligned} \delta(\mathbf{x}_{ij}) &= \mathbf{x}_{ij} (\tilde{\Sigma}_1^{-1} - \tilde{\Sigma}_0^{-1}) \mathbf{x}_{ij}^T \\ &= \mathbf{x}_{ij} \left(\Phi_1 \tilde{\Lambda}_1^{-1} \Phi_1^T - \Phi_0 \tilde{\Lambda}_0^{-1} \Phi_0^T \right) \mathbf{x}_{ij}^T \\ &= [\Phi_1^T \mathbf{x}_{ij}]^T \tilde{\Lambda}_1^{-1} [\Phi_1^T \mathbf{x}_{ij}] - [\Phi_0^T \mathbf{x}_{ij}]^T \tilde{\Lambda}_0^{-1} [\Phi_0^T \mathbf{x}_{ij}] \end{aligned} \quad (18)$$

By substituting (17) to (18), we have

$$\begin{aligned} \delta(\mathbf{x}_{ij}) &= \sum_{n=1}^k \frac{1}{(1-\gamma) \lambda_{1n} + \gamma \alpha_1} (\phi_{1n}^T \mathbf{x}_{ij})^2 \\ &\quad + \frac{1}{(1-\gamma) \beta_1 + \gamma \alpha_1} \left(\|\mathbf{x}_{ij}\|^2 - \sum_{n=1}^k (\phi_{1n}^T \mathbf{x}_{ij})^2 \right) \\ &\quad - \sum_{n=1}^k \frac{1}{(1-\gamma) \lambda_{0n} + \gamma \alpha_0} (\phi_{0n}^T \mathbf{x}_{ij})^2 \\ &\quad - \frac{1}{(1-\gamma) \beta_0 + \gamma \alpha_0} \left(\|\mathbf{x}_{ij}\|^2 - \sum_{n=1}^k (\phi_{0n}^T \mathbf{x}_{ij})^2 \right) \\ &= \left(\frac{1}{(1-\gamma) \lambda_{1n} + \gamma \alpha_1} - \frac{1}{(1-\gamma) \beta_1 + \gamma \alpha_1} \right) \sum_{n=1}^k (\phi_{1n}^T \mathbf{x}_{ij})^2 \\ &\quad + \left(\frac{1}{(1-\gamma) \beta_1 + \gamma \alpha_1} - \frac{1}{(1-\gamma) \beta_0 + \gamma \alpha_0} \right) \|\mathbf{x}_{ij}\|^2 \\ &\quad - \left(\frac{1}{(1-\gamma) \lambda_{0n} + \gamma \alpha_0} - \frac{1}{(1-\gamma) \beta_0 + \gamma \alpha_0} \right) \sum_{n=1}^k (\phi_{0n}^T \mathbf{x}_{ij})^2. \end{aligned} \quad (19)$$

Similar to (14), we also introduce $\|\mathbf{x}_{ij}\|^2$ to (19) to reduce the computational cost. Given (19), it is straightforward to conduct matching or retrieval by ranking reference images \mathbf{x}_j according to $\delta(\mathbf{x}_{ij})$ given a query target \mathbf{x}_i . A reference image corresponding to a smaller $\delta(\mathbf{x}_{ij})$ ranks near the top.

B. Incremental RS-KISS Metric Learning

One critical challenge of person re-identification is that the model needs to be updated to incorporate the information carried by the new labeled training samples. We consider two instances of incremental learning in practical scenarios. First, the newly acquired samples are taken from a person that has not been learned by RS-KISS. Second, the newly acquired samples are taken from a person that has been already learned by RS-KISS. We can also understand the new acquired samples are taken from the person's track history. Thus, we propose the incremental learning for RS-KISS.

Given two sets $X = \{\mathbf{x}_i\}$ for $i = 1, 2, \dots, N$ and $Y = \{\mathbf{y}_j\}$ for $j = 1, 2, \dots, L$, which are sets of difference of similar feature vector pairs. Similar to RS-KISS, we assume that the mean vectors of X and Y are $\mathbf{0}$. The covariance matrices of X and Y can be estimated by

$$\begin{aligned} \Sigma_X &= \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \\ \Sigma_Y &= \frac{1}{L} \sum_{i=1}^L \mathbf{y}_i \mathbf{y}_i^T \end{aligned} \quad (20)$$

The whole set $Z = X \cup Y = \{\mathbf{z}_i\}$, $i = 1, 2, \dots, N+L$ thus has a zero mean. The covariance matrix of Z is given by

$$\Sigma_Z = \frac{1}{N+L} \sum_{i=1}^{N+L} \mathbf{z}_i \mathbf{z}_i^T. \quad (21)$$

By substituting (20) to (21), we have

$$\begin{aligned} \Sigma_Z &= \frac{1}{N+L} \sum_{i=1}^{N+L} \mathbf{z}_i \mathbf{z}_i^T \\ &= \frac{1}{N+L} \left(\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T + \sum_{i=1}^L \mathbf{y}_i \mathbf{y}_i^T \right) \\ &= \frac{N}{N+L} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \right) + \frac{L}{N+L} \left(\frac{1}{L} \sum_{i=1}^L \mathbf{y}_i \mathbf{y}_i^T \right) \\ &= \frac{N}{N+L} \Sigma_X + \frac{L}{N+L} \Sigma_Y. \end{aligned} \quad (22)$$

Given (22), it is straightforward to have an incremental learning strategy to update RS-KISS. In particular, we can regard X including the original training examples, while Y including the new training examples. To get the covariance matrix of the whole set $Z = X \cup Y$, we only need to compute the covariance matrix of Y , and then update the covariance

matrix Σ_X by simply adding a re-scaled Σ_Y . This incremental learning strategy saves the computational time.

The covariance matrix of difference of dissimilar feature vector pairs can be obtained in the same way.

After updating the covariance of Σ_0 and Σ_1 , the metric $\delta(\mathbf{x}_{ij}) = \mathbf{x}_{ij}^T (\Sigma_1^{-1} - \Sigma_0^{-1}) \mathbf{x}_{ij}$ can be updated accordingly.

For person re-identification, although the dimensionality of the original feature vector is high, PCA is usually applied before the metric learning to significantly the redundancy. Thus, we can consider incremental PCA [5] to update the projection matrix given a set of additional training examples. Then, we apply the incremental RS-KISS metric learning to update the metric for matching. Note that the matrix inverse operation takes $O(d^{2.4})$ and the calculation of a covariance matrix costs $O(nd^2)$, where the number of examples n is usually much larger than the feature dimensionality d . Thus, incrementally update covariance (saves $O(Nd^2)$) is the key to reduce the time cost for computation. In practice, N is larger than L , so IRS-KISS is preferred.

IV. EXPERIMENTAL RESULTS

In this section, we conducted the experiments of person re-identification on the widely used three datasets, including the VIPeR dataset [22], the ETHZ dataset [11], [56] and the i-LIDS MCTS dataset [68] to validate the effectiveness of the proposed RS-KISS. The VIPeR dataset contains 1264 images collected from 632 individuals. The ETHZ dataset contains 8555 images collected from 146 individuals. The i-LIDS MCTS dataset contains 476 images collected from 119 individuals. Note that these datasets have very different average number samples per person. Thus, they are suitable for evaluating person re-identification schemes under different scenarios. Images are normalized to a standard size of 128×64 and this manipulation causes shape distortion that has limited effect on the human visual systems [22], [58]. For each image, the local binary pattern (LBP) descriptors, HSV histograms, and Lab histograms are used to represent each normalized image [39]. The performance is measured by the average cumulative match characteristic (CMC) curves. To compare RS-KISS against other top-level person re-identification techniques, we have strictly followed the experimental settings of [39], [69]. Details of the experimental setup and baseline models are given below. In addition, we have carefully prepared the incremental learning scheme on the VIPeR dataset and the ETHZ dataset to demonstrate the effectiveness of IRS-KISS.

A. Experimental Settings

There are two different kinds of experimental settings in person re-identification. Some of the previous researches regard person re-identification as a process that template images of each person are selected and matched against probe images by matching models [10], [52]. Satta *et al.* [54] proposed multiple component matching (MCM) to unify the popular person re-identification schemes to better understand the intrinsic differences of these schemes. In this paper, we use another method for performance evaluation. The procedure is



Fig. 1. Some typical samples from the VIPeR dataset. There are same-person paired samples from different camera views in each column. Variations can be observed, such as viewpoint, pose, shooting locations, illumination, and image quality.

summarized as follows: collecting images of several individuals and classifying a new collection of images of different individuals. This method can be deemed as a simple binary classification problem or pair matching technique, and aims to judge whether or not the two images come from a same person. Recently, this pair matching technique has received intensive attentions in person re-identification [39], [69].

In this paper, we present RS-KISS to learn an effective distance metric to measure the distance between two images for person re-identification. It is important to show that the learned metric is robust to different people for re-identification. Thus, we use the second kind of experimental setting, i.e., learning a distance metric on set A while testing the learned metric on set B. This kind of evaluation can demonstrate the robustness of RS-KISS for different datasets.

B. Datasets

The widely used VIPeR Dataset is collected by Gray *et al.* [22] and contains 1264 outdoor images obtained from two views of 632 subjects. Example images are shown in Fig. 1. Each intrapersonal image pair is shown in one column. Intrapersonal image pairs may contain a viewpoint change of 90 degrees. Other variations are also considered, such as light conditions, shooting locations, and the image qualities. Thus, it is challenge to conduct image-based person re-identification on the VIPeR dataset. In our experiments, all samples of p subjects are selected to form the training set, while the rest is used for test. We set $p = 100$ and $p = 316$, respectively, to evaluate the matching performance of different algorithms. In the training stage, we used intrapersonal image pairs as similar pairs and generated interpersonal image pairs (by randomly selecting two images from different subjects) as dissimilar pairs. In the test stage, we randomly divided a test set into two parts, CAM A (a gallery set) and CAM B (a probe set), by randomly putting one image in an intrapersonal image pair in CAM A and another in CAM B. By calculating the CMC curve, we can obtain a ranking for each example in the gallery with respect to the probe. We repeated the above procedure 10 times, and then the average cumulative match characteristic (CMC) curves were depicted.

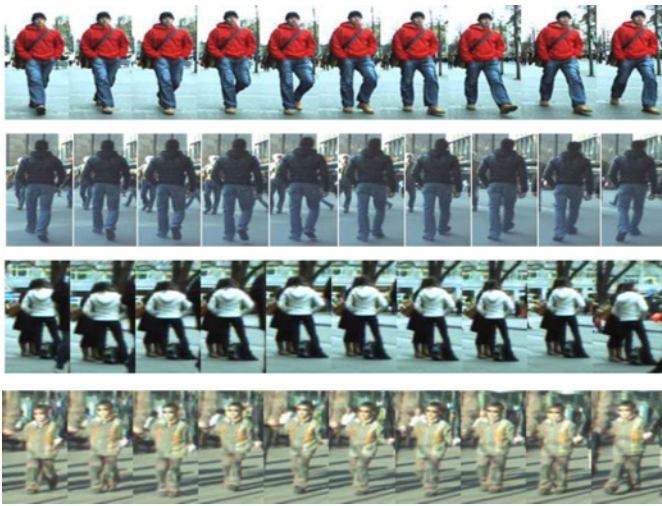


Fig. 2. Some typical samples from the ETHZ dataset. There are same-person samples cropped from the video sequence in each row. We can observe minor variations, including slight changes of viewpoint, pose, shooting locations, illumination, and image quality.

On the VIPeR dataset, we design the incremental learning experiment by considering the first situation. We randomly selected $p = 100$ people to train RS-KISS. Then, we randomly added $ip = 50, 100, 150, 200,$ and 250 people and applied IRS-KISS to update the metric obtained by RS-KISS, respectively. In addition, we randomly selected 270 people, which are blind to RS-KISS and IRS-KISS, for test. The above procedure was repeated 10 times, and then the average CMC curves were depicted.

The ETHZ Dataset is collected by Ess *et al.* [11] and widely used for person detection and tracking. Subsequently, Schwartz *et al.* [56] applied it to person re-identification. There are 8555 images collected from 146 individuals. Some typical example images are given in Fig. 2. Compared with the VIPeR dataset, the ETHZ dataset has more samples collected from a subject. Minor variations include the slight changes of viewpoint, pose, shooting locations, illumination, and image quality. In our experiments, all samples of p subjects are selected to form the training set, while the rest is used for test. We set $p = 76$ and $p = 106$, respectively, to evaluate the matching performance of different algorithms. The above procedure was repeated 10 times, and then the average CMC curves were depicted. In the training stage, we used the above method to generate similar pairs and dissimilar pairs. In the test stage, we randomly selected one sample for each person to build the gallery and rest for the probe.

On the ETHZ dataset, we design the incremental learning experiment considering the second situation. We randomly selected $p = 100$ people to set up the training set and the rest people were used to form the test set. After randomly selecting $b = 2$ images of each person to train RS-KISS, we randomly selected $ib = 1, 2, 3, 4$ images of each person again and applied IRS-KISS to update the metric obtained by RS-KISS, respectively. The process was repeated 10 times, and then the average CMC curves were depicted.

The i-LIDS MCTS dataset [68] was taken at a busy airport



Fig. 3. Some typical samples from the i-LIDS MCTS dataset. There are same-person samples cropped from the video sequence in each row. We can observe minor variations, including slight changes of viewpoint, pose, shooting locations, illumination and image quality. However, many images contain occlusion contaminations.

hall. There are 476 images collected from 119 individuals. We show some typical examples in Fig. 3. Compared with the above two datasets, some images in this dataset have occlusions, because crowds of people always carry big or small baggage. In our experiments, all samples of p subjects are selected to form the training set, while the rest is used for test. We set $p = 89$ to evaluate the matching performance of different algorithms. The process was repeated 10 times, and then the average CMC curves were depicted.

C. Feature Descriptors

It is known that both texture feature and color histogram are useful for person re-identification. According to [39], the LBP descriptors, HSV histogram and Lab histogram are extracted from overlapping blocks of size 8×16 and stride of 8×8 on each image. HSV and Lab histograms encode the different color distribution information in the HSV and Lab color spaces, respectively. LBP descriptors are used to extract texture features. All the feature descriptors are concatenated together. We conducted PCA to obtain a low-dimension representation, to accelerate the learning process, and reduce signal noise. The details can be referred to [39].

D. Baselines and Performance Measures

In this Section, to validate the effectiveness of proposed RS-KISS, we compare five representative metric learning approaches, including Euclidean distance (L2), Mahalanobis metric (MM), information theoretical metric learning (ITML) [9], metric learning for large margin nearest neighbor (LMNN) [64], and KISS [39]. Each of these methods has its own merits. The L2 distance is used to construct a baseline in most existing person re-identification works. Mahalanobis Metric includes the L2 distance as a special case and can perform better than L2 [58], [60]. ITML, LMNN and KISS are the state of the art metric learning algorithms that have shown their effectiveness in many applications.

The average cumulative match characteristic (CMC) curves for illustrating the ranked matching rates are obtained over 10

TABLE I
PERSON RE-IDENTIFICATION TOP MATCHING RATES ON THE VIPeR DATASET: COMPARING WITH THE POPULAR ALGORITHMS

RANK	$p = 100$				$p = 316$			
	1	10	25	50	1	10	25	50
RS-KISS	0.098	0.405	0.608	0.765	0.245	0.666	0.847	0.930
RDC	0.091	0.344	0.535	0.697	0.157	0.539	0.752	0.879
Adaboost	0.042	0.020	0.350	0.503	0.082	0.366	0.582	0.909
Bhat	0.038	0.124	0.203	0.295	0.047	0.166	0.266	0.402
PLS	0.023	0.082	0.142	0.232	0.027	0.109	0.204	0.329
Xing's	0.036	0.121	0.203	0.295	0.047	0.166	0.266	0.415

Some results are directly taken from [69].

TABLE II
PERSON RE-IDENTIFICATION TOP MATCHING RATES ON THE ETHZ DATASET: COMPARING WITH THE POPULAR ALGORITHMS

RANK	$p = 76$				$p = 106$			
	1	5	10	20	1	5	10	20
RS-KISS	0.770	0.921	0.962	0.985	0.835	0.963	0.984	0.996
RDC	0.690	0.858	0.922	0.969	0.727	0.901	0.956	0.988
Adaboost	0.656	0.840	0.905	0.956	0.692	0.878	0.935	0.980
Bhat	0.555	0.761	0.840	0.906	0.610	0.809	0.878	0.941
PLS	0.483	0.694	0.780	0.868	0.546	0.751	0.833	0.924
Xing's	0.544	0.752	0.833	0.904	0.608	0.803	0.874	0.936

Some results are directly taken from [69].

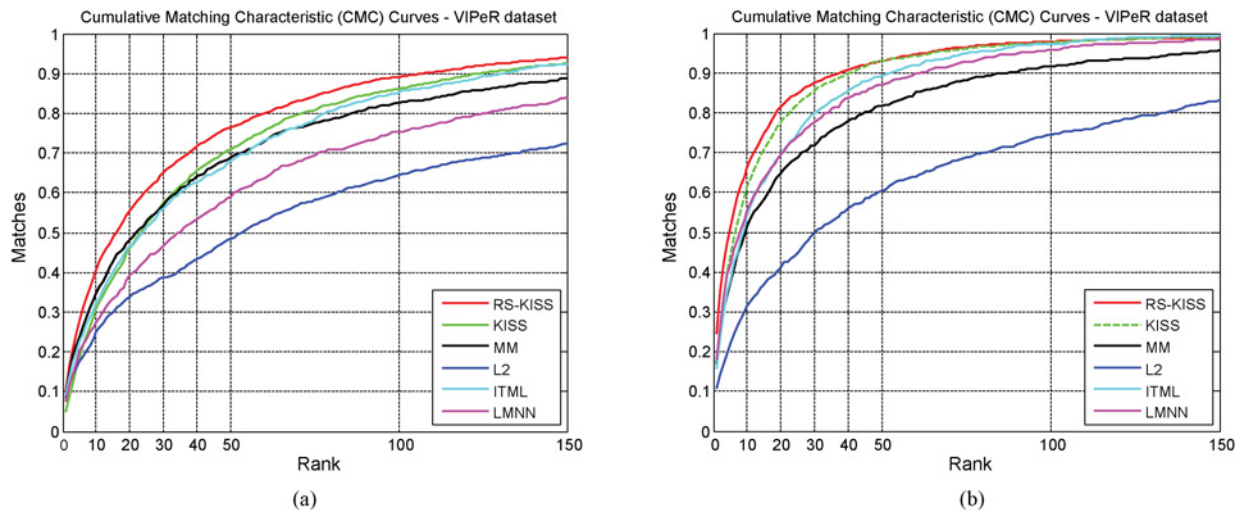


Fig. 4. Performance comparison using CMC curves. In each subfigure, the x-coordinate is the rank score and y-coordinate is the matching rate. We compare RS-KISS with L2, MM, ITML, LMNN and KISS on the VIPeR dataset. Only the top 150 ranking positions are depicted. These subfigures suggest the effectiveness of the proposed RS-KISS.

trials to evaluate the person re-identification performance of a particular learned distance metric. Because the complexity of re-identification problem, the top n ranked matching rate (n is a small value as far as possible) is considered.

E. Experimental Results and Analysis

In Fig. 4, we compare the proposed RS-KISS with KISS, L2, MM, ITML, and LMNN on the VIPeR dataset. In each subfigure, the x-coordinate is the rank score and y-coordinate is the matching rate. Top 150 ranking positions are shown in the figure. In Fig. 5, we compare RS-KISS with KISS, L2, MM, ITML, and LMNN on the ETHZ dataset. In each subfigure, the top 30 ranking positions are shown in the figure. In Fig. 6, RS-KISS is compared with KISS, L2, MM,

ITML, and LMNN on the i-LIDS MCTS dataset. In each subfigure, top 30 ranking positions are shown in the figure. Note that LMNN performs poorly, because LMNN models relative distance that is sensitive to the number of training samples. RS-KISS and KISS perform at the top level.

In Tables I, II, and III, we compared RS-KISS with other popular person re-identification approaches on the three datasets. These approaches include RDC [69], Adaboost [1], Bhat [69], PLS [56] and Xing's [66]. RS-KISS performs best in terms of rank score in most cases.

The main observations from the matching performance comparisons are given below.

- 1) The proposed RS-KISS improves KISS under the situation when the size of the training set is small, because

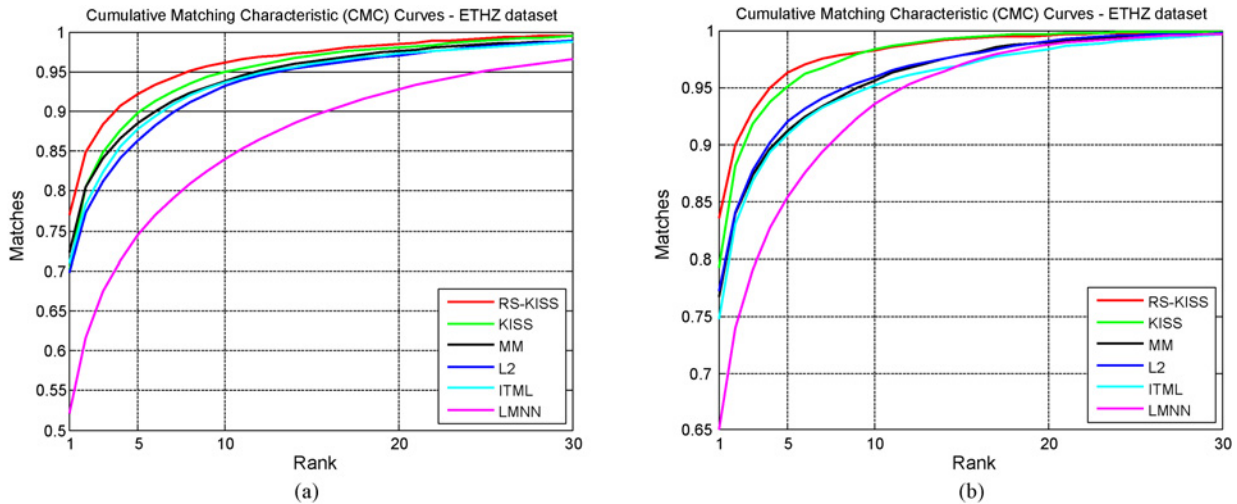


Fig. 5. Performance comparison using CMC curves. In each subfigure, the x-coordinate is the rank score and y-coordinate is the matching rate. We compare RS-KISS with L2, MM, ITML, LMNN and KISS on the ETHZ dataset. Top 30 ranking positions are depicted. These subfigures suggest the effectiveness of the proposed RS-KISS.

TABLE III

PERSON RE-IDENTIFICATION TOP MATCHING RATES ON THE ILIDS MCTS DATASET: COMPARING WITH THE POPULAR ALGORITHMS

RANK	$p = 76$			
	1	5	10	20
RS-KISS	0.437	0.737	0.873	0.960
RDC	0.441	0.727	0.847	0.963
Adaboost	0.356	0.664	0.799	0.932
Bhat	0.318	0.614	0.742	0.895
PLS	0.258	0.574	0.736	0.903
Xing's	0.318	0.626	0.773	0.906

Some results are directly taken from [69].

the covariance matrix in KISS estimated by likelihood maximization is seriously biased given a small size training set.

- 2) When the number of training samples is sufficiently large, RS-KISS performs comparably to KISS."

However, the success of RS-KISS for person re-identification requires a fixed-size feature vector that represents the image of a person [30], [53], [54]. Thus, other combination descriptors concatenated by un-ordered or notfixed size sets of different local features [12], [21], [48], [52] cannot be directly integrated with RS-KISS.

Fig. 7 shows the performance of IRS-KISS. We observed that the newly added samples can improve the retrieval accuracy. Fig. 8 uses CMC curve to compare different metric learning algorithms, including KISS, RS-KISS, and IRS-KISS. We randomly selected $p = 158$ people to form the training set for RS-KISS. On the basis of the RS-KISS model, we randomly selected $ip = 158$ people again to set up the training set for IRS-KISS and KISS. This experiment shows that proposed RS-KISS and IRS-KISS improve KISS. In addition, by using IRS-KISS, we can save the training time, because the computation of the covariance matrix is time consuming. The time cost for matrix inverse operation is acceptable because the matrix size is small in our experiments. In the above

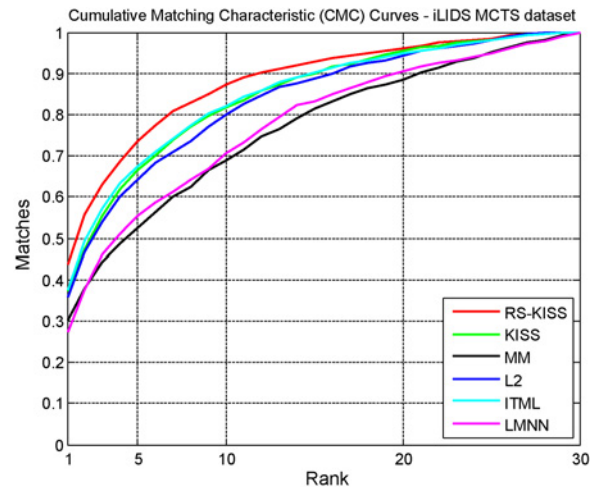


Fig. 6. Performance comparison using CMC curve. In the figure, the x-coordinate is the rank score and y-coordinate is the matching rate. We compare RS-KISS with L2, MM, ITML, LMNN and KISS on the iLIDS MCTS dataset. Top 30 ranking positions are depicted. These subfigures suggest the effectiveness of the proposed RS-KISS.

incremental learning experiment, we utilized IRS-KISS to update the metric that costs 2.7 ms while RS-KISS costs 10.1 ms for completely re-training for $p = 316$. We conduct experiments on an i7-2600 3.4 GHz computer with 24-Gbyte memory.

V. CONCLUSION

Distance metric is critically important for the surveillance task person re-identification. Thus, it is rational to find a proper distance metric learning algorithm to boost the performance of person re-identification. In recent years, many distance metric learning algorithms have been developed, such as information theoretic metric learning (ITML) and metric learning for large margin nearest neighbor (LMNN), but they are not suitable for person re-identification. That is because there are

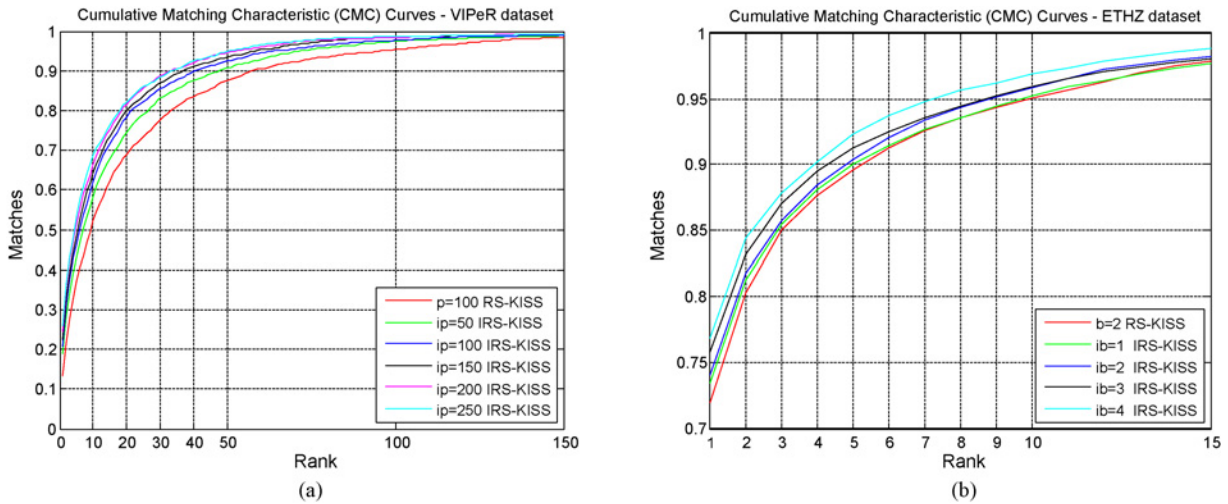


Fig. 7. Experiments for IRS-KISS. (a) We conduct IRS-KISS on the VIPeR dataset. We assumed that the newly acquired samples are taken from a person which has not been learned by RS-KISS. (b) We conduct our IRS-KISS on the ETHZ dataset. We assumed that the newly acquired samples are taken from the people’s track history.

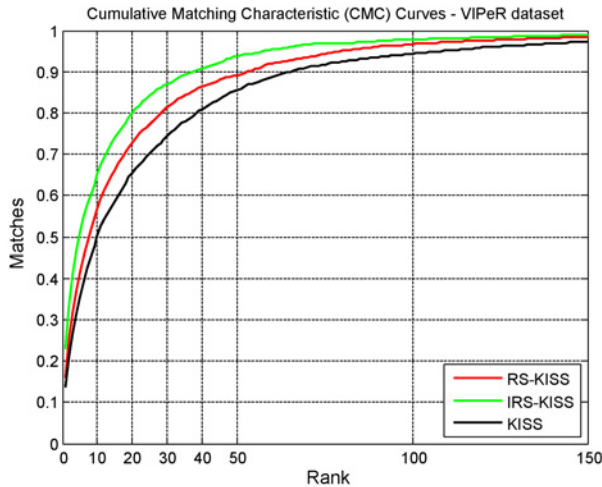


Fig. 8. IRS-KISS versus RS-KISS and KISS. We randomly selected $p = 158$ people to form the training set for RS-KISS. On the basis of RS-KISS model, we randomly selected another $ip = 158$ people to train IRS-KISS. The performance of KISS obtained at the same condition as that for training RS-KISS is shown for comparison. The figure suggests the effectiveness of the proposed RS-KISS and IRS-KISS.

limited training image pairs to learn a metric in person re-identification. Although KISS metric learning obtained the state of the art performance, it shares the same problem. Given a small size training sample, covariance matrices estimated by KISS are highly biased. Therefore, we presented regularized smoothing KISS or RS-KISS for short. The proposed RS-KISS exploited the smoothing technique to enlarge the small eigenvalues of the estimated covariance matrix and the regularization technique to suppress the effect of the large eigenvalues of the estimated covariance matrix. The employed two statistical techniques effectively enlarge the underestimated small eigenvalues and reduce the overestimated large eigenvalues of the estimated covariance matrix. Therefore, RS-KISS significantly improves KISS for person re-identification. Given additional training examples, we advance

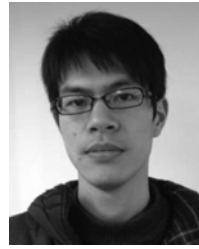
RS-KISS by using incremental learning. Experimental results obtained from the three representative datasets fully demonstrate that the proposed RS-KISS and incremental RS-KISS are very effective in improving the performance of person re-identification.

REFERENCES

- [1] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, “Person re-identification using Haar-based and DCD-based signature,” in *Proc. Workshop Activity Monitor. Multi-Camera Surveillance Syst.*, 2010, pp. 91–102.
- [2] O. Barnich and M. V. Droogenbroeck, “ViBe: A universal background subtraction algorithm for video sequences,” *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.
- [3] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino, “Multiple-shot person re-identification by hpe signature,” in *Proc. Int. Conf. Pattern Recogn.*, 2010, pp. 1413–1416.
- [4] M. Belkin and P. Niyogi, “Laplacian eigenmaps and spectral techniques for embedding and clustering,” *Adv. Neural Inform. Process. Syst.*, vol. 14, pp. 585–591, Dec. 2001.
- [5] M. Brand, “Incremental singular value decomposition of uncertain data with missing values,” in *Proc. Eur. Conf. Comput. Vision*, 2002, pp. 707–720.
- [6] C.-H. Chan, B. Goswami, J. Kittler, and W. J. Christmas, “Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication,” *IEEE Trans. Inform. Forensics Security*, vol. 7, no. 2, pp. 602–612, Apr. 2012.
- [7] X. Chen, J. Tian, X. Yang, and Y. Zhang, “An algorithm for distorted fingerprint matching based on local triangle feature set,” *IEEE Trans. Inform. Forensics Security*, vol. 1, no. 2, pp. 169–177, Jun. 2006.
- [8] D. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, “Custom pictorial structures for re-identification” in *Proc. British Mach. Vision Conf.*, 2011, pp. 68.1–68.11.
- [9] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, “Information-theoretic metric learning,” in *Proc. 24th Int. Conf. Mach. Learn.*, New York, NY, USA, 2007, pp. 209–216.
- [10] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, “Appearance-based person reidentification in camera networks: Problem overview and current approaches,” *J. Ambient Intell. Humanized Comput.*, vol. 2, no. 2, pp. 127–151, 2011.
- [11] A. Ess, B. Leibe, and L. Van Gool, “Depth and appearance for mobile scene analysis,” in *Proc. IEEE 11th Int. Conf. Comput. Vision*, Oct. 2007, pp. 1–8.
- [12] M. Farenzena, L. Bazzani, A. Perina, M. Cristani, and V. Murino, “Person re-identification by symmetry-driven accumulation of local features,” in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, Jun. 2010, pp. 2360–2367.

- [13] M. A. Ferrer, F. Vargas, A. Morales, and A. Ordóñez, "Robustness of offline signature verification based on gray level features," *IEEE Trans. Inform. Forensics Security*, vol. 7, no. 3, pp. 966–977, Jun. 2012.
- [14] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, pp. 179–188, 1936.
- [15] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biol. Cybernet.*, vol. 61, no. 2, pp. 103–113, 1989.
- [16] P.-E. Forssen, "Maximally stable colour regions for recognition and matching," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, Jun. 2007, pp. 1–8.
- [17] J. H. Friedman, "Regularized discriminant analysis," *J. Am. Statist. Ass.*, vol. 84, no. 405, pp. 165–175, 1989.
- [18] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Stat.*, vol. 28, no. 2, pp. 337–407, 2000.
- [19] B. Geng, D. Tao, and C. Xu, "DAML: Domain adaptation metric learning," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2980–2989, Oct. 2011.
- [20] B. Geng, D. Tao, C. Xu, L. Yang, and X. Hua, "Ensemble manifold regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1227–1233, Jun. 2012.
- [21] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, vol. 2, 2006, pp. 1528–1535.
- [22] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE Int. Workshop Performance Eval. Track. Surveillance*, vol. 3, Oct. 2007, p. 5.
- [23] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vision*, Marseille, France, 2008, pp. 262–275.
- [24] N. Guan, D. Tao, Z. Luo, and J. Shawe-Taylor, "MahNMF: Manhattan nonnegative matrix factorization," *ArXiv Preprint*, arXiv:1207.3438, 2012.
- [25] O. Hamdoun, F. Moutarde, B. Stanculescu, and B. Steux, "Person reidentification in multi-camera system by signature based on interest point descriptors collected on short video sequences," in *Proc. ACM/IEEE Int. Conf. Distributed Smart Cameras*, 2008, pp. 1–6.
- [26] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2001.
- [27] X. He and P. Niyogi, "Locality preserving projections," *Adv. Neural Inform. Process. Syst.*, vol. 16, pp. 153–160, Dec. 2003.
- [28] R. Herbrich, T. Graepel, and K. Obermayer, "Large margin rank boundaries for ordinal regression," *Adv. Neural Inform. Process. Syst.*, vol. 12, pp. 115–132, Dec. 1999.
- [29] M. Hirzer, C. Beleznaï, P. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. Scandinavian Conf. Image Anal.*, 2011, pp. 91–102.
- [30] M. Hirzer, P. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *Proc. IEEE Eur. Conf. Comput. Vision*, Oct. 2012, pp. 780–793.
- [31] T. Hospedales, S. Gong, and T. Xiang, "Video behavior mining using a dynamic topic model," *Int. J. Comput. Vision*, vol. 98, no. 3, pp. 303–323, Oct. 2012.
- [32] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 7, pp. 417–441, 1933.
- [33] G.Q. Hu, W. P. Tay, and Y. G. Wen, "Cloud robotics: Architecture, challenges and applications," *IEEE Netw. Special Issue Mach. Robot. Network.*, vol. 26, no. 3, pp. 21–28, May–Jun. 2012.
- [34] A. Iosifidis, A. Tefas, and I. Pitas, "Activity-based person identification using fuzzy representation and discriminant learning," *IEEE Trans. Inform. Forensics Security*, vol. 7, no. 2, pp. 530–542, Apr. 2012.
- [35] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views," *Comput. Vision Image Understanding*, vol. 109, no. 2, pp. 146–162, 2008.
- [36] J. Li and D. Tao, "On preserving original variables in Bayesian PCA with application to image analysis," *IEEE Trans. Image Process.* vol. 21, no. 12, pp. 4830–4843, Dec. 2012.
- [37] J. Li and D. Tao, "Simple exponential family PCA," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 3, pp. 485–497, Mar. 2013.
- [38] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 1, pp. 149–153, Jan. 1987.
- [39] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, Jun. 2012, pp. 2288–2295.
- [40] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recogn.*, Jun. 2006, pp. 2169–2178.
- [41] J.-E. Lee, R. Jin, and A. K. Jain, "Rank-based distance metric learning: An application to image retrieval," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recogn.*, Jun. 2008, pp. 1–8.
- [42] D.-T. Lin and K.-Y. Huang, "Collaborative pedestrian tracking and data fusion with multiple cameras," *IEEE Trans. Inform. Forensics Security*, vol. 6, no. 4, pp. 1432–1444, Dec. 2011.
- [43] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [44] D. Makris, T. J. Ellis, and J. K. Black, "Bridging the gaps between cameras," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, vol. 2, Jun.–Jul. 2004, pp. 205–210.
- [45] T. Ojala, M. Pietikainen, and T. Maenpa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [46] X. Tian, D. Tao, and Y. Rui, "Sparse transfer learning for interactive video search reranking," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 8, no. 3, p. 26, 2012.
- [47] X. Tian, D. Tao, X. Hua, and X. Wu, "Active reranking for Web image Search," *IEEE Trans. Image Process.* vol. 19, no. 3, pp. 805–820, 2010.
- [48] I. d. Oliveira and J. d. S. Pio, "Object reidentification in multiple cameras system," in *Proc. Int. Conf. Embedded Multimedia Comput.*, 2009, pp. 1–8.
- [49] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 34, no. 3, pp. 436–450, Mar. 2012.
- [50] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *Proc. Brit. Mach. Vision Conf.*, 2010, pp. 21.1–21.11.
- [51] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Sci.*, vol. 290, pp. 2323–2326, 2000.
- [52] R. Satta, G. Fumera, F. Roli, M. Cristani, and V. Murino, "A multiple component matching framework for person re-identification," in *Proc. Int. Conf. Image Anal. Process.*, vol. 2, 2011, pp. 140–149.
- [53] R. Satta, G. Fumera, and F. Roli, "Exploiting dissimilarity representations for person re-identification," in *Proc. Int. Workshop Similarity-Based Pattern Anal. Recogn.*, 2011, pp. 275–289.
- [54] R. Satta, G. Fumera, and F. Roli, "Fast person re-identification based on dissimilarity representations," *Pattern Recogn. Lett.*, vol. 33, no. 14, pp. 1838–1848, 2012.
- [55] C. Schmid, "Constructing models for content-based image retrieval," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recogn.*, vol. 2, Dec. 2001, pp. 39–45.
- [56] W. Schwartz and L. Davis, "Learning discriminative appearance-based models using partial least squares," in *Proc. Brazilian Symp. Comput. Graphics Image Process.*, 2009, pp. 322–329.
- [57] S. Si, D. Tao, and B. Geng, "Bregman divergence-based regularization for transfer subspace learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 7, pp. 929–942, Jul. 2010.
- [58] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proc. IEEE*, vol. 94, no. 11, pp. 1948–1962, Nov. 2006.
- [59] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, Jun. 1999, pp. 2246–2252.
- [60] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 39–51, Jan. 1998.
- [61] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
- [62] D. Tao and L. Jin, "Discriminative information preservation for face recognition," *Neurocomput.*, vol. 91, pp. 11–20, 2012.
- [63] L. Wang and C. Leckie, "Encoding actions via the quantized vocabulary of averaged silhouettes," in *Proc. Int. Conf. Pattern Recogn.*, 2010, pp. 3657–3660.
- [64] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.

- [65] Y.G. Wen, W.W. Zhang, and H.Y. Luo, "Energy-optimal mobile application execution: Taming resource-poor mobile devices with cloud clones," in *Proc. 31st IEEE Int. Conf. Comput. Commun.*, 2012, pp. 2716–2720.
- [66] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information," in *Adv. Neural Inform. Process. Syst.*, vol. 15, pp. 505–512, Dec. 2002.
- [67] T. Zhou and D. Tao, "Double shrinking sparse dimension reduction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 244–257, Jan. 2013.
- [68] W. S. Zheng, S. Gong, and T. Xiang, "Associating groups of people," in *Proc. Brit. Mach. Vision Conf.*, vol. 1, Sep. 2009, pp. 1–11.
- [69] W. S. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2012.
- [70] T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in *Proc. IEEE Eur. Conf. Comput. Vision*, Oct. 2008, pp. 725–738.
- [71] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.* vol. 21, no. 9, pp. 1299–1313, Sep. 2009.



Yongfei Wang received the B.S. degree in faculty information engineering from the Guangdong University of Technology, Guangzhou, China. He is currently pursuing the Master's degree in information and communication engineering at the South China University of Technology, Guangzhou, China.

His current research interests include machine learning and computer vision.

Yuan Yuan (M'05–SM'09) is currently a Full Professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi, China.

Xuelong Li (M'02–SM'07–F'12) is currently a Full Professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi, China.



Dapeng Tao received the B.S. degree in electronics and information engineering from Northwestern Polytechnical University, Xi'an, China. He is currently pursuing the Ph.D. degree in information and communication engineering at the South China University of Technology, Guangzhou, China.

His current research interests include machine learning, computer vision, and cloud computing.



Lianwen Jin received the B.S. degree from the University of Science and Technology of China, and the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 1991 and 1996, respectively.

He is currently a Professor with the School of Electronic and Information Engineering, South China University of Technology. He is the author of more than 100 scientific papers. His current research interests include image processing, handwriting analysis and recognition, machine learning,

cloud computing, and intelligent systems.

Dr. Jin is a member of the IEEE Signal Processing Society, the IEEE Communication Society, the IEEE Computer Society, the China Image and Graphics Society, and the Cloud Computing Experts Committee of the China Institute of Communications. He was a recipient of the Award of New Century Excellent Talent Program of MOE in 2006 and the Guangdong Pearl River Distinguished Professor Award in 2011. He served as a Program Committee Member for a number of international conferences, including ICMLC2007~2011, ICFHR2008-2012, ICDAR2009, ICDAR2013, ICPR2010, ICPR2012, ICMLA2012, and so on.