Person Reidentification by Minimum Classification Error-Based KISS Metric Learning

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

***Abstract*—In recent years, person reidentification has received growing attention with the increasing popularity of intelligent video surveillance. This is because person reidentification is critical for human tracking with multiple cameras. Recently, keep it simple and straightforward (KISS) metric learning has been regarded as a top level algorithm for person reidentification. The covariance matrices of KISS are estimated by maximum likelihood (ML) estimation. It is known that discriminative learning based on the minimum classification error (MCE) is more reliable than classical ML estimation with the increasing of the number of training samples. When considering a small sample size problem, direct MCE KISS does not work well, because of the estimate error of small eigenvalues. Therefore, we further introduce the smoothing technique to improve the estimates of the small eigenvalues of a covariance matrix. Our new scheme is termed the minimum classification error-KISS (MCE-KISS). We conduct thorough validation experiments on the VIPeR and ETHZ datasets, which demonstrate the robustness and effectiveness of MCE-KISS for person reidentification.**

***Index Terms*—Intelligent video surveillance, metric learning, minimum classification error, person reidentification.**

I. Introduction

PERSON reidentification is complex but receives intensive attention in the field of intelligent video surveillance (IVS). An aim of person reidentification is to match an instance of a person captured by one camera to another instance of the person captured by different cameras. Traditional biometrics, such as face [1], [2], iris [3], fingerprint [4], and gait [5], are not often available, because images are low-quality, variable, and contain motion blur. However, the use of body appearance is reliable for many person reidentification applications [6]–[8].

As in other visual retrieval applications, such as mobile visual search [9], [10], and face validation [11]–[13], there are two important stages which need to be considered in the process of person reidentification. They are distance learning [12], [14], and visual feature extraction and selection.

Many exciting studies on visual feature extraction and robust representation have been performed which have greatly improved the performance of person reidentification. Here, we briefly review representative works.

Person reidentification is a real-time task. The use of color features (such as RGB, HSV, and YCbCr color histograms), effectively save computational cost. In addition, color features are robust to variability in resolution and perspective [7], [15]. Gabor [16], [17], and Schmid filters [18] are insensitive to light conditions, and have been added to feature extraction procedures [7], [15], [19]. Local image descriptors are widely used to represent interest points or regions within the images. Scale invariant feature transform (SIFT) [20] and speeded up robust features (SURF) have also been used to extract texture features [21], [22]. Haar-like features [23], which are important features in face detection, have been introduced in person reidentification [6]. Finally, local binary patterns (LBP) [24], which were originally proposed for texture classification, have also been exploited for person reidentification [12]. LBP estimates the local geometric structure of an image based on a nonparametric method, and has been widely used in facial image description. A comprehensive comparison of different local image descriptors are given in [25].

Directly choosing the bounding boxes obtained in detection approaches used in person reidentification [6] is not the only method used for feature extraction. Some schemes use stripes which span the whole horizontal dimension [15], [26], while Kostinger *et al*. [12] partitioned the image into a regular grid. From the grid, the color and texture features are extracted from overlapping blocks. In addition, segmentation techniques can be used to obtain a mask that separates the region of the person from the background region [6], [27]. By utilizing spatial and temporal cues, Gheissari *et al*. [28] developed an over-segmentation method to achieve robust performance.

Recent research on image recognition has demonstrated that learning meaningful representations (features) from high-dimensional observations will help to improve the performance of person reidentification. Bak *et al*. [29] utilized spatial pyramid matching (SPM) [30] to represent the instance of a person. The classical SPM scheme, in which the classifier is

Manuscript received March 3, 2013; revised February 23, 2014 and April 30, 2014; accepted May 6, 2014. Date of publication June 3, 2014; date of current version January 13, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61125106, Grant 61075021, and Grant 61201348, in part by the National Science and Technology Support Plan under Grant 2013BAH65F01 and Grant 2013BAH65F04, in part by the Guangdong Natural Science Funds under Grant S2011020000541 and Grant S2012040008016, in part by the Guangdong Scientific and Technology Research Plan 2012A010701001, in part by the Research Fund for the Doctoral Program of Higher Education of China under Grant 20120172110023, and in part by the Shaanxi Key Innovation Team of Science and Technology under Grant 2012KCT-04. This paper was recommended by Associate Editor H. Lu.

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).

X. Li is with the Center for OPTical IMagery Analysis and Learning, 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-mail: 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/TCYB.2014.2323992

constructed by Mercer kernels, provides an effective solution, but it is computationally expensive. Locality-constrained linear coding (LLC) [31] is more suitable than SPM, because LLC is based on local coordinate coding (LCC), and explores the locally linear characteristics of the sample distribution. The effectiveness of LLC is ensured by several attractive properties, namely reconstruction, local smooth sparsity, and an analytical solution.

As well as obtaining robust features for image representation, dimension reduction is necessary to retain the most effective features for subsequent matching. This is because a combination of the aforementioned selected features is usually deployed, and dimension reduction results in a succinct, yet effective representation of a high-dimensional sample. Over the past decade, classical linear dimension reduction algorithms, and the emerging manifold learning algorithms, have enriched our choices for feature selection.

Dimension reduction algorithms have received increasing attentions in recent years. Principal component analysis (PCA) [32] is a representative classical linear algorithm. Laplacian eigenmaps (LE) [33], which are a classical geometrically-motivated algorithm, pay much attention to the nonlinearity of the data distribution. Locally linear embedding (LLE) [34] seeks a low-dimensional, neighborhood-preserving embedding of the high-dimensional data. ISOMap [35] is a variant of multidimensional scaling which considers the geodesic distance between samples. Linear discriminant analysis (LDA) [36], [37] aims to separate samples drawn from different classes. Supervised locality preserving projections (SLPP) [38] and discriminative locality alignment (DLA) [39] consider the local geometry of a set of high-dimensional samples. Zhang *et al.* [40] proposed a framework to unify representative dimension reduction algorithms to better understand their intrinsic differences.

Distance learning can significantly improve the performance of retrieval applications [41], [42]. In this paper, we aim to update the retrieval precision by applying robust distance learning. In earlier studies, several approaches which have achieved top-level performance in image retrieval applications perform poorly for person reidentification. It is worth special mention that KISS metric learning is both efficient and effective [12]. However, it is assumed that pairwise differences are sampled from a Gaussian distribution, has the small sample size problem for estimating the covariance, and therefore results in retrieval precision not always performing robustly in practice.

In this paper, we introduce the minimum classification error (MCE) criterion [43] to improve KISS distance learning for person reidentification. In particular, eigenvalues of the true covariance matrix are biased, which harms the utilization of the estimated covariance matrix in subsequent operations, such as classification. The covariance matrices of KISS are obtained by maximum likelihood (ML) estimation. With increasing the number of training samples, discriminative learning based on MCE is more reliable than classical ML estimation. In addition, the MCE criterion is widely used in the field of machine learning. Many researchers have suggested that automatic speech recognition systems can demonstrate improved performance under the MCE criterion [44]–[46]. Reed and Lee [47] proposed an MCE training algorithm to build a music recommendation tool. However, only introducing the MCE criterion to the training procedure does not work well to estimate the small eigenvalues of the covariance matrices. Therefore, the smoothing technique [48] is required to improve the estimate of the small eigenvalues of a covariance matrix. The improved KISS is termed the minimum classification error-KISS, or simply MCE-KISS.

The procedure for MCE-KISS-based person reidentification can be summarized by the following steps: 1) partitioning the image into a regular grid of size 8×4 and overlapping block of size 8×8, and the color features and texture features are extracted from the overlapping blocks; 2) concatenating all the feature descriptors together and conducting PCA to achieve a robust feature representation for each sample; 3) training MCE-KISS; and 4) finally finding the matching rank according to the query target. Given limited space considerations, we do not describe the other parts in detail, since implementations can be easily found in the references.

The main contribution of this paper include the following.

1) The newly proposed MCE-KISS by seamlessly integrating MCE criterion and smoothing technique to significantly improve the performance of KISS metric learning.
2) We have thoroughly compared MCE-KISS with other the state-of-the-art schemes of person reidentification on two public datasets. Experiment results demonstrate our scheme robust and effectiveness. By contrast to RS-KISS [49] that integrates smoothing and regularization techniques under the frame of KISS metric learning [12], the newly proposed MCE-KISS exploits a discriminative learning procedure to effectively adjust the parameters of Gaussian density model, so MCE-KISS achieves the robust generalization ability on test set.

The rest of the paper is organized as follows: in Section II, we briefly review related works for improving distance learning for person reidentification. We detail the proposed MCE-KISS in Section III. Section IV shows the experiment results on the representative datasets (VIPeR [7] and ETHZ [8]). We conclude the paper in Section V.

II. Related Work

In Section I, we briefly reviewed the techniques used in person reidentification. It is worth noting the importance of distance learning schemes, which have been receiving increasing attention [14], [50] because the retrieval quality is known to be highly dependent on distance metrics. Porikli [51] proposed a new distance learning algorithm to solve the color calibration problem of a multicamera system. Weinberger and Saul [14] proposed a large margin nearest neighbor metric (LMNN) to improve the performance of the classical *k*NN classification. However, the computational processing of *k* closest within-class samples is time-consuming. From the perspective of information theoretic, Davis *et al.* [50] proposed information-theoretic metric learning (ITML), which built on the Mahalanobis distance metric. Recently, Zheng *et al.* [52] proposed a soft discriminative scheme termed relative distance comparison (RDC). In this scheme, it is assumed that wrong and right matches correspond to large and small distances, respectively. In contrast to consider modeling the similarity globally, Yang *et al.* [53] proposed local distance metric to improve the performance of retrieval and classification accuracy. However, most methods may perform poorly when the view conditions change greatly and the training samples are insufficient.

As well as distance metric learning-based matching schemes, researchers have exploited other schemes to improve retrieval precision. By utilizing subspace learning, Javed *et al.* [54] proposed a brightness transfer function to cope with the illumination changes in a multicamera system. Prosser *et al.* [19] introduced Rank support vector machines (RankSVM) to person reidentification and proposed ensemble RankSVM to handle the scalability issue.

III. Minimum Classification Error-Based KISS Metric Learning

A. KISS Metric Learning Review

Recently, Kostinger *et al.* [12] proposed KISS metric learning (KISS) based on an assumption that pairwise differences are Gaussian distributed. This has acquired state-of-the-art retrieval performance for real applications, such as person reidentification and face recognition.

Considering the person reidentification problem, a feature vector pair $\mathbf{x}_i$ and $\mathbf{x}_j$ represents two samples. The hypothesis H_0 can assume that the feature vector pair is dissimilar, i.e., $\mathbf{x}_i$ and $\mathbf{x}_j$ are sampled from different people, and the hypothesis H_1 denotes that the feature vector pair is similar, i.e., $\mathbf{x}_i$ and $\mathbf{x}_j$ are sampled from the same person. Equation (1) defines the logarithm of the ratio between the two posteriors

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

For metric learning, a large $\delta\left(\mathbf{x}_i, \mathbf{x}_j\right)$ indicates $\mathbf{x}_i$ and $\mathbf{x}_j$ represent different people, while a small $\delta\left(\mathbf{x}_i, \mathbf{x}_j\right)$ indicates $\mathbf{x}_i$ and $\mathbf{x}_j$ represent a 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

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

which can be rewritten as

$$\delta\left(\mathbf{x}_{ij}\right) = \log\left(f\left(\mathbf{x}_{ij}|\theta_0\right)/f\left(\mathbf{x}_{ij}|\theta_1\right)\right) \tag{3}$$

where $f\left(\mathbf{x}_{ij}|\theta\right)$ is the probability density functions with parameter θ for hypothesis H.

After assuming the difference space is a Gaussian structure, we have

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

where $k \in \{0, 1\}$, d is the dimensionality of the feature vector, and Σ_k is the covariance matrix of $\mathbf{x}_{ij}$. Note that for specific i and j, since both $\mathbf{x}_{ij}$ and $\mathbf{x}_{ji}$ belong to the pairwise difference set, we have $\sum_{i,j} x_{ij} = 0$, i.e., zero mean and $\theta_1 = (0, \Sigma_1)$ and $\theta_0 = (0, \Sigma_0)$.

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

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

By dropping the constant terms, we have

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

Define y_{ij} as the indicative 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_1 denote the number of similar feature vector pairs, while N_0 denotes the number of dissimilar feature vector pairs. The covariance matrices are estimated as follows:

$$\Sigma_0 = \frac{1}{N_0}\sum_{y_{ij}=0}\mathbf{x}_{ij}\mathbf{x}_{ij}^T = \frac{1}{N_0}\sum_{y_{ij}=0}\left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T$$

$$\Sigma_1 = \frac{1}{N_1}\sum_{y_{ij}=1}\mathbf{x}_{ij}\mathbf{x}_{ij}^T = \frac{1}{N_1}\sum_{y_{ij}=1}\left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T. \tag{7}$$

Equation (7) shows that the eigenvalues of Σ_0 and Σ_1 are positive.

Let KISS project $\Sigma_1^{-1} - \Sigma_0^{-1}$ onto the cone of the positive a semi-definite matrix M, so we have

$$\delta\left(\mathbf{x}_{ij}\right) = \mathbf{x}_{ij}^T M\mathbf{x}_{ij} \tag{8}$$

where M is the KISS metric matrix.

B. MCE-KISS Metric Learning

Although KISS has largely improved the accuracy of person reidentification, there is a lot of room to improve efficiency and stability. It is critical to estimate the covariance matrices in (6) accurately to improve performance for person reidentification. It is known that the model of Gaussian distribution suffers from estimate error given limited training samples. Specifically, it is laborious and tedious to get a large number of labeled samples in real applications, to overcome the estimate error of the small eigenvalues of the covariance matrices which arose through the problem of small sample size.

In statistics, to obtain robust estimations, a large number of techniques have been proposed. In this paper, the smoothing technique [48], and the MCE criterion [43], are introduced to improve the accuracy of estimates of covariance matrices in KISS. By enlarging the estimate to the small eigenvalues of a covariance matrix, the smoothing technique can compensate for the decrease in performance which arose from the estimate errors of small eigenvalues. On the other hand, the covariance matrices of KISS are estimated by ML estimation. It is known that the ML estimation for Gaussian density model is imprecise, the discriminative learning procedure based on MCE aims to adjust the parameters of Gaussian density model and improves the generalization ability by increasing the number of training samples.

The covariance matrix Σ_i in (6) is first diagonalized and can be written as

$$\Sigma_i = \Phi_i\Lambda_i\Phi_i^T \tag{9}$$

where $\Lambda_i = \mathrm{diag}\,[\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{id}]$ with λ_{ij} being an eigenvalue of Σ_i, $\Phi_i = [\phi_{i1}, \phi_{i2}, \ldots, \phi_{id}]$ with ϕ_{ij} being an eigenvector of Σ_i, and eigenvalues in Λ_i are arranged in a descending order.

We substitute (9) into (6) and obtain

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

Through (10), we can explain that the small eigenvalues significantly affect the score of metric.

Next, we replace the small eigenvalues of the covariance matrix with a small constant β_i

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

where d is the dimensionality of training samples. Taking into account the smoothing technique, the constant β_i is set to the value of the average of all the small eigenvalues

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

Thus, (10) can be written as

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

According to the MCE criterion, we need to optimize the parameters of covariance matrices by utilizing the gradient descent method. Then, we have the evaluation of misclassification of a sample $\mathbf{x}$ belonging to class c

$$h_c(\mathbf{x}) = \max_{c}\delta(\mathbf{x}, \mathbf{x}_c) - \min_{r}\delta(\mathbf{x}, \mathbf{x}_r) \qquad (14)$$

where $\mathbf{x}_c$ is a sample of the class c, and $\mathbf{x}_r$ is the closest interclass sample. Equation (14) considers two aspects: 1) the distance between $\mathbf{x}$ and the farthest intraclass sample and 2) the distance between $\mathbf{x}$ and the closest interclass sample. Furthermore, the loss of the misclassification can be written as

$$l_c(\mathbf{x}) = \frac{1}{1 + e^{-\xi h_c(x)}} \qquad (15)$$

where ξ is a trade-off parameter and is selected in the range of $(0, +\infty]$.

Given the training samples $\{\mathbf{x}_n\,|n = 1, 2, \ldots, N\}$, and the label of each sample $\{C_i\,|i = 1, 2, \ldots, M\}$, we can compute the empirical loss by using (16)

$$L = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{M} l_i(\mathbf{x}_n)\, I(\mathbf{x}_n \in C_i) \qquad (16)$$

$$I(\mathrm{x}_n \in C_i) = \begin{cases}1, & \text{if } \mathrm{x}_n \in C_i\\ 0, & \text{if } \mathrm{x}_n \notin C_i\end{cases}. \qquad (17)$$

And (16) can be further deduced to

$$L = \frac{1}{N}\sum_{n=1}^{N} l_c(\mathbf{x}_n) \qquad (18)$$

where c is the class information. According to (18), we expect that the distance between $\mathbf{x}$ and the farthest intraclass sample are as small as possible and the distance between $\mathbf{x}$ and the closest interclass sample are as large as possible.

The parameters in KISS include the eigenvectors and eigenvalues of Σ_0 and Σ_1, i.e., $\lambda_{1n}, \beta_1, \lambda_{0n}, \beta_0, \phi_{1n}$ and ϕ_{0n}. In MCE-KISS, we minimize the empirical loss L by means of adjusting these parameters via gradient descent. Let θ denote the parameters, according to gradient descent, we can get a general update rule of parameters

$$\begin{aligned}\theta(t+1) &= \theta(t) - \varepsilon(t)\frac{\partial L}{\partial\theta}\\ &= \theta(t) - \varepsilon(t)\frac{\partial l_c(\mathbf{x})}{\partial\theta} \qquad (19)\end{aligned}$$

$$\begin{aligned}\frac{\partial l_c(\mathbf{x})}{\partial\theta} &= (-1)\cdot\frac{1}{\left(1 + e^{-\xi h_c(\mathbf{x})}\right)^2}\cdot(-\xi)\,e^{-\xi h_c(\mathbf{x})}\frac{\partial h_c(\mathbf{x})}{\partial\theta}\\ &= \xi l_c^2(\mathbf{x})\cdot\left(\frac{1}{l_c(\mathbf{x})} - 1\right)\frac{\partial h_c(\mathbf{x})}{\partial\theta}\\ &= \xi l_c(\mathbf{x})\,(1 - l_c(\mathbf{x}))\,\frac{\partial h_c(\mathbf{x})}{\partial\theta} \qquad (20)\end{aligned}$$

$$\frac{\partial h_c(\mathbf{x})}{\partial\theta} = \frac{\partial\delta(\mathbf{x}, \mathbf{x}_c)}{\partial\theta} - \frac{\partial\delta(\mathbf{x}, \mathbf{x}_r)}{\partial\theta}. \qquad (21)$$

According (19)–(21), we have

$$\begin{aligned}&\theta(t+1) = \theta(t)\\ &-\varepsilon(t)\,\xi l_c(\mathbf{x})\,(1 - l_c(\mathbf{x}))\left(\frac{\partial\delta(\mathbf{x}, \mathbf{x}_c)}{\partial\theta} - \frac{\partial\delta(\mathbf{x}, \mathbf{x}_r)}{\partial\theta}\right). \qquad (22)\end{aligned}$$

In the learning process, we need to guarantee eigenvalues are positive, so we further define

$$\begin{cases}\lambda_{in} = e^{\sigma_{in}}\\ \beta_i = e^{\tau_i}\end{cases}. \qquad (23)$$

We rewrite (23) to

$$\begin{cases}\sigma_{in} = \ln\lambda_{in}\\ \tau_i = \ln\beta_i\end{cases}. \qquad (24)$$

Based on (22), we convert the parameter updating to the computation of partial derivatives (25)–(28)

$$\begin{cases}\frac{\partial\delta(\mathbf{x},\mathbf{x}_j)}{\partial\tau_1} = -e^{-\tau_1}\left[\left\|\mathbf{x} - \mathbf{x}_j\right\|^2 - \sum_{n=1}^{k}\left[\phi_{1n}^T\left(\mathbf{x} - \mathbf{x}_j\right)\right]^2\right]\\ \\ \frac{\partial\delta(\mathbf{x},\mathbf{x}_j)}{\partial\tau_0} = e^{-\tau_0}\left[\left\|\mathbf{x} - \mathbf{x}_j\right\|^2 - \sum_{n=1}^{k}\left[\phi_{0n}^T\left(\mathbf{x} - \mathbf{x}_j\right)\right]^2\right]\end{cases} \qquad (25)$$

Algorithm 1 Minimum Classification Error-KISS

```
Step 1: The initial $\Sigma_0^{-1}$ and $\Sigma_1^{-1}$ are calculated by using (7).
Step 2: Smooth technique: By using (12) to amend for the estimation errors of small eigenvalues of $\Sigma_0^{-1}$ and $\Sigma_1^{-1}$;
Step 3: MCE technique: By using (25), (26), (27), and (28) to optimize the parameters of $\Sigma_0^{-1}$ and $\Sigma_1^{-1}$;
Step 4: The distance metric is calculated by using (6).
```

$$\begin{cases} \frac{\partial\delta(\mathbf{x},\mathbf{x}_j)}{\partial\sigma_{1n}} = -e^{-\sigma_{1n}}\left[\phi_{1n}^T\left(\mathbf{x}-\mathbf{x}_j\right)\right]^2 \\ \frac{\partial\delta(\mathbf{x},\mathbf{x}_j)}{\partial\sigma_{0n}} = e^{-\sigma_{0n}}\left[\phi_{0n}^T\left(\mathbf{x}-\mathbf{x}_j\right)\right]^2 \end{cases} \tag{26}$$

$$\begin{aligned} \frac{\partial\delta(\mathbf{x},\mathbf{x}_j)}{\partial\phi_{1nl}} &= \frac{\partial}{\partial\phi_{1nl}}\left\{\begin{array}{l} \sum\limits_{n=1}^{k} e^{-\sigma_{1n}}\left[\sum\limits_{l=1}^{d}\phi_{1nl}\left(\mathbf{x}-\mathbf{x}_j\right)_l\right]^2 \\ +e^{-\tau_1}\Bigg[\left\|\mathbf{x}-\mathbf{x}_j\right\|^2 \\ -\sum\limits_{n=1}^{k}\left(\sum\limits_{l=1}^{d}\phi_{1nl}\cdot\left(\mathbf{x}-\mathbf{x}_j\right)_l\right)^2\Bigg] \\ -\sum\limits_{n=1}^{k} e^{-\sigma_{0n}}\left[\sum\limits_{l=1}^{d}\phi_{0nl}\left(\mathbf{x}-\mathbf{x}_j\right)_l\right]^2 \\ -e^{-\tau_0}\Bigg[\left\|\mathbf{x}-\mathbf{x}_j\right\|^2 \\ -\sum\limits_{n=1}^{k}\left(\sum\limits_{l=1}^{d}\phi_{0nl}\cdot\left(\mathbf{x}-\mathbf{x}_j\right)_l\right)^2\Bigg] \end{array}\right\} \\ &= e^{-\sigma_{1n}}\cdot 2\left[\phi_{1n}^T\left(\mathbf{x}-\mathbf{x}_j\right)\right]\cdot\left(\mathbf{x}-\mathbf{x}_j\right)_l \\ &\quad +e^{-\tau_1}.\left(-1\right).2\left[\phi_{1n}^T\left(X-X_j\right)\right].\left(X-X_j\right)_l \\ &= 2\left(e^{-\sigma_{1n}}-e^{-\tau_1}\right)\left[\phi_{1n}^T\left(X-X_j\right)\right]\left(X-X_j\right)_l \end{aligned} \tag{27}$$

$$\frac{\partial\delta\left(\mathbf{x},\mathbf{x}_j\right)}{\partial\phi_{0nl}} = -2\left(e^{-\sigma_{0n}}-e^{-\tau_0}\right)\left[\phi_{0n}^T\left(\mathbf{x}-\mathbf{x}_j\right)\right] \left(\mathbf{x}-\mathbf{x}_j\right)_l. \tag{28}$$

We can optimize the parameters of Σ_0^{-1} and Σ_1^{-1} by using (25)–(28) and the distance metric $\delta\left(\mathbf{x}_{ij}\right)$ satisfied the needs of the MCE.

Based on the above discussions, we summarize MCE-KISS in Algorithm 1.

IV. Experiment Results

In this section, two widely used yet challenging datasets, VIPeR [7] and ETHZ [8], were used to demonstrate the effectiveness of the proposed MCE-KISS method. All images from the two datasets were normalized to a standard size of 128 × 48. In general, this manipulation causes shape distortion which has limited effect on human visual systems. For each image, we concatenated the extracted LBP descriptor [24] and some color features into a high dimensional feature vector.

In our experiments, all samples of p_{ts} subjects were selected to form the test set, while the rest p_{tr} were used for model training. During training, we used intraperson image pairs as similar pairs, and generated interperson image pairs (by randomly selecting two images from different subjects) as dissimilar pairs. The image pairs are used to estimate Σ_0^{-1} and Σ_1^{-1} according Algorithm 1. During testing, the test set were divided into two parts, i.e., a gallery set and a probe set. We randomly chose one sample of each subject to comprise the gallery set. The rest were used for the probe set. Person reidentification aims to identify a person's photo in the probe set by comparing it with images of several individuals stored in the gallery set.

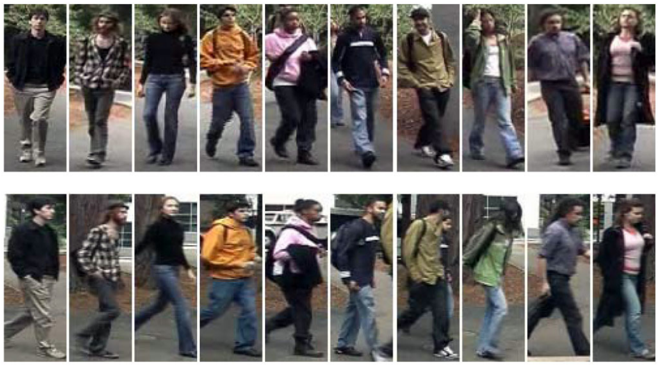


Fig. 1. Some typical samples from the VIPeR dataset. Same-person paired samples from different camera views can be seen in each column, demonstrating the observed variations, such as in viewpoint, pose, shooting locations, illumination, and image quality.

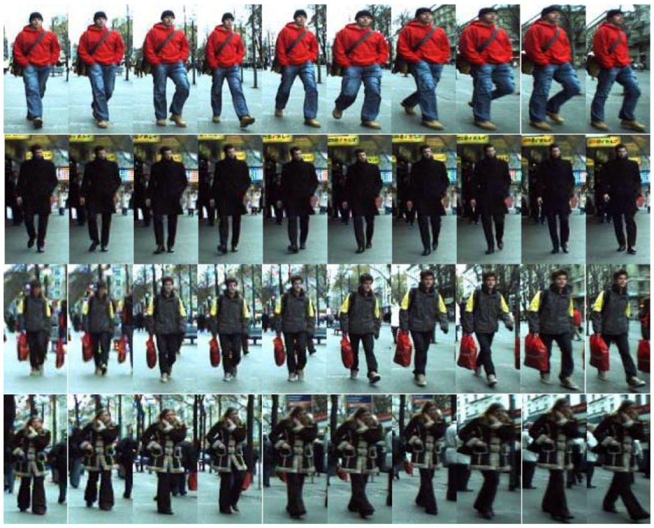


Fig. 2. Some typical samples from the ETHZ dataset. Same-person samples cropped from the video sequence are shown in each row, demonstrating that variations in viewpoint, pose, shooting location, illumination, and image quality, are minor.

By using the average cumulative match characteristic (CMC) curves, we evaluated the performance of the proposed algorithm. Because the complexity of the reidentification problem, the top n-ranked matching rate was considered (n is a small value). We detail of the experiment setup and baseline models as follows.

A. VIPeR Dataset

The VIPeR dataset was collected by Gray *et al.* [7] and contains 1264 outdoor images obtained from two views of 632 subjects. Intraperson image pairs may contain a viewpoint

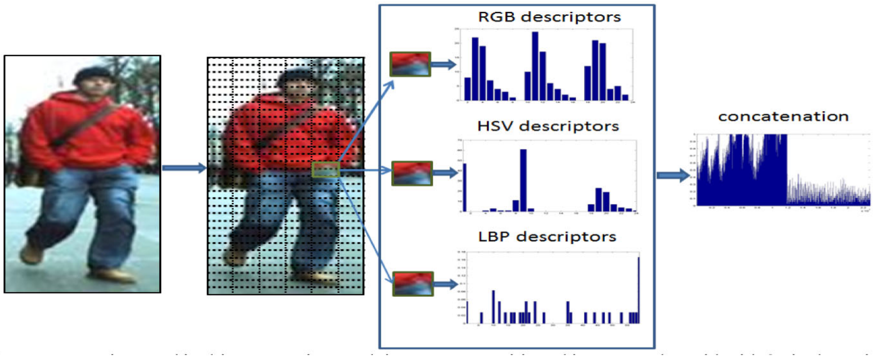
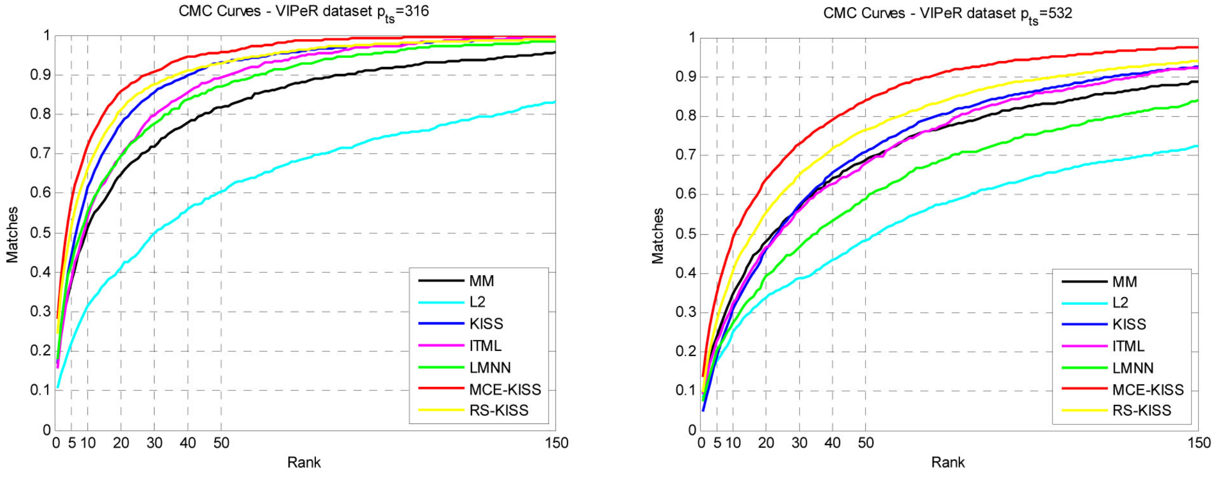


Fig. 3. Process of feature extraction used in this paper. First, each image was partitioned into a regular grid with 8 pixel spacing in the horizontal direction and 4 pixel spacing in the vertical direction. Second, from the grid, the LBP descriptor, HSV histogram, and RGB histogram were extracted from overlapping blocks of size 8 × 8. Third, all the feature descriptors were concatenated together.



(a)

(b)

Fig. 4. Performance comparison using CMC curves. In each subfigure, the *x*-coordinate is the rank score and the *y*-coordinate is the matching rate. We compare MCE-KISS with L2, MM, ITML, LMNN, KISS, and RS-KISS on the VIPeR dataset. (a) $p_{ts} = 316$. (b) $p_{ts} = 532$.

TABLE I
PERSON REIDENTIFICATION TOP RANKED MATCHING RATE ON THE VIPeR DATASET

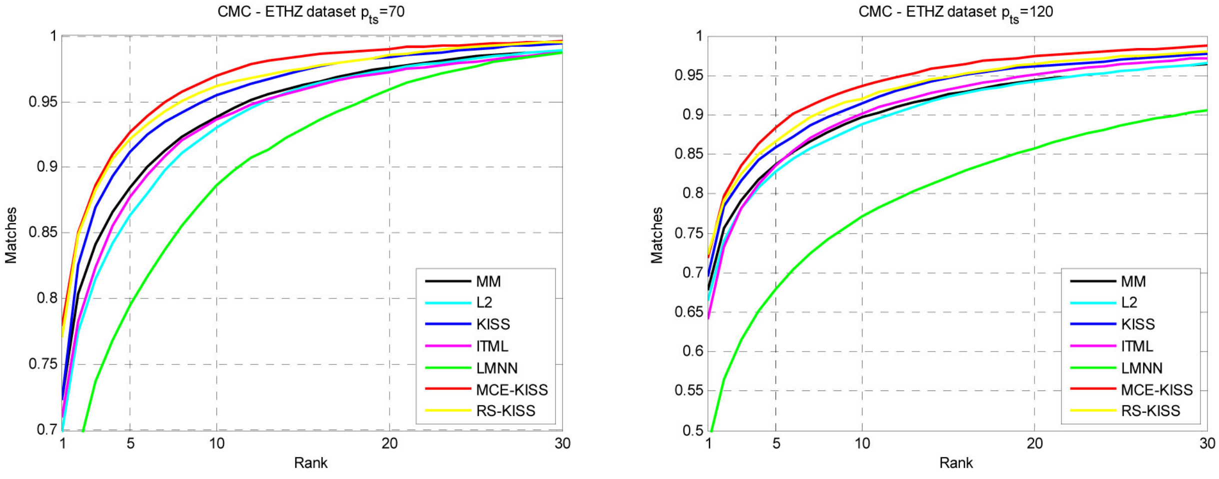
	p_{ts}=316				p_{ts}=532			
RANK	1	10	25	50	1	10	25	50
MM	0.169	0.513	0.688	0.818	0.096	0.347	0.532	0.689
L2	0.108	0.313	0.449	0.604	0.085	0.251	0.363	0.484
KISS	0.170	0.616	0.823	0.931	0.047	0.308	0.524	0.711
ITML	0.155	0.540	0.750	0.892	0.091	0.324	0.515	0.680
LMNN	0.180	0.554	0.741	0.870	0.075	0.274	0.435	0.589
MCE-KISS	**0.282**	**0.721**	**0.889**	**0.956**	**0.136**	**0.490**	**0.688**	**0.840**
RS-KISS	0.244	0.663	0.852	0.930	0.098	0.405	0.608	0.765

change of 90°. Other variations are also considered, such as lighting conditions, shooting locations, and image quality. Thus, it is challenge to conduct image-based person reidentification on the VIPeR dataset. Example images are shown in Fig. 1.

We set $p_{ts} = 316$ and $p_{ts} = 532$, respectively to evaluate the matching performance of different algorithms. We repeated the process 10 times, and the average CMC curves were depicted.

B. ETHZ Dataset

The ETHZ Dataset was collected by Ess *et al*. [55], and is widely used for person detection and tracking. Subsequently, it has been used for the purpose of person reidentification [8].



(a) (b)

Fig. 5. Performance comparison using CMC curves. In each subfigure, the *x*-coordinate is the rank score and the *y*-coordinate is the matching rate. We compare MCE-KISS with L2, MM, ITML, LMNN, KISS, and RS-KISS on the ETHZ dataset. (a) $\mathrm{p_{ts}} = 70$. (b) $\mathrm{p_{ts}} = 120$.

TABLE II
PERSON REIDENTIFICATION TOP RANKED MATCHING RATE ON THE ETHZ DATASET

	p_{ts} =70				p_{ts} =120			
RANK	1	10	20	30	1	10	20	30
MM	0.723	0.938	0.976	0.989	0.678	0.896	0.943	0.965
L2	0.694	0.930	0.975	0.989	0.664	0.888	0.942	0.965
KISS	0.723	0.955	0.984	0.994	0.695	0.915	0.961	0.978
ITML	0.710	0.936	0.973	0.987	0.642	0.902	0.951	0.972
LMNN	0.598	0.886	0.960	0.987	0.475	0.771	0.858	0.906
MCE-KISS	**0.779**	**0.970**	**0.990**	**0.996**	0.719	**0.937**	**0.974**	**0.987**
RS-KISS	0.770	0.962	0.985	0.995	**0.723**	0.921	0.964	0.980

It contains 8555 images collected from 146 individuals. Some typical example images are given in Fig. 2, which show same-person samples cropped from the video sequence in each row. In contrast to VIPeR, ETHZ has more samples collected from a subject. We can see that variations, such as viewpoint, pose, shooting location, illumination, and image quality, are minor.

We set $p_{ts} = 70$ and $p_{ts} = 120$, respectively to evaluate the matching performance of different algorithms. The process was repeated 10 times, and then average CMC curves were depicted.

C. Feature Descriptors

It is known that both texture features and color histograms are useful for person reidentification. In our experiments, each image was partitioned into a regular grid with 8 pixel spacing in the horizontal direction, and 4 pixel spacing in the vertical direction. From the grid, the LBP descriptor, HSV histogram, and RGB histogram were extracted from overlapping blocks of size 8×8. The HSV and RGB histograms encoded the different color distribution information in the HSV and RGB color spaces, respectively. The texture distribution information was modeled effectively by LBP descriptor. All the feature descriptors were concatenated together. Fig. 3 shows the process of feature extraction. We conducted PCA to obtain a 40-dimensional representation, to suppress the Gaussian noise.

D. Baseline Methods

We compared six representative metric learning approaches to validate the effectiveness of our algorithm, including Euclidean distance (L2), Mahalanobis metric (MM), KISS [12], RS-KISS [49], information theoretical metric learning (ITML) [50], and metric learning for LMNN [14]. Each of these methods has its own advantages. The L2 distance has been applied to construct a baseline in most of the existing person reidentification studies. The MM considers the correlations of variables, and can perform better than L2 [56]. KISS, ITML, and LMNN are the state-of-the-art metric learning algorithms that have been applied to many different applications. RS-KISS is superior to KISS through combining the smoothing and regularization techniques.

The impact of the variation of k in (11) has been evaluated on the ETHZ dataset. We set $p_{ts} = 120$ and the number of eigenvalues need to be smoothed $d - k = 1, 4, 7$, respectively. Afterward, MCE-KISS was applied to update the metric. In addition, on the ETHZ dataset, we have thoroughly compared MCE-KISS with KISS on four different size training sets. We selected $p_{tr} = 10, 20, 30, 40$ and applied MCE-KISS and KISS to update the metric, respectively. In addition, the training iterations of MCE-KISS are 1000. We conduct all experiments on an Intel Xeon E5645 2.40GHz computer with a 144-GB memory.

TABLE III
PERSON REIDENTIFICATION TOP MATCHING RATES ON THE VIPeR DATASET (p_{ts} = 316): COMPARING WITH THE POPULAR ALGORITHMS

	p_{ts}=316				
RANK	1	10	20	25	50
MCE-KISS	0.282	**0.721**	**0.857**	**0.889**	0.956
RPML	0.27	0.69	0.83	0.86	0.95
Li's	**0.296**	0.693	-	0.887	**0.968**
RDC	0.157	0.539	0.720	0.752	0.879
PCCA χ^2_{RBF}	0.193	0.649	-	0.83	0.96
Adaboost	0.082	0.366	0.520	0.582	0.909

TABLE IV
PERSON REIDENTIFICATION TOP MATCHING RATES ON THE VIPeR DATASET (p_{ts} = 532): COMPARING WITH THE POPULAR ALGORITHMS

	p_{ts}=532				
RANK	1	10	20	25	50
MCE-KISS	**0.136**	**0.490**	**0.639**	**0.688**	**0.840**
RPML	0.11	0.38	0.52	0.56	0.72
Li's	0.129	0.427	0.580	-	-
RDC	0.091	0.344	0.490	0.535	0.697
PCCA χ^2_{RBF}	0.093	0.374	0.529	-	-
Adaboost	0.042	0.200	0.310	0.350	0.503

E. Experiment Results and Analysis

Fig. 4 shows the comparison of our proposed MCE-KISS metric learning with KISS, RS-KISS, L2, MM, ITML, and LMNN on the VIPeR dataset. In each subfigure, the *x*-coordinate is the rank score, and the *y*-coordinate is the matching rate. Only the top 150 ranking positions are shown in the figure. Table I reports the performance of all the algorithms within the scope of the first 50 ranks.

Fig. 5 compares the proposed MCE-KISS metric learning with KISS, RS-KISS, L2, MM, ITML, and LMNN on the ETHZ dataset. In each subfigure, the *x*-coordinate is the rank score, and the *y*-coordinate is the matching rate. Only the top 30 ranking positions are shown in the figure. Table II reports the performance of all the algorithms in the scope of the first 30 ranks.

In Tables III and IV, we compared MCE-KISS with other popular person reidentification approaches which have different features on the VIPeR dataset. These approaches include RPML [57], Li and Wang's [58], RDC [52], PCCAχ^2_{RBF} [59], and Adaboost [6]. MCE-KISS performs best in terms of rank score in most cases.

Fig. 6 shows the impact of variation of *k* in (11). In the figure, the *x*-coordinate is the rank score and *y*-coordinate is the matching rate. Only the top 10 ranking positions are shown in the figure.

Fig. 7 compares MCE-KISS with KISS on different size training sets on the ETHZ dataset. Only the top 30 ranking positions are shown in the figure. Table V reports mean training time of the experiments.

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

1) MCE-KISS integrates the smoothing technique and the MCE criterion for precise covariance matrix estimation. It thus improves KISS and significantly outperforms L2, MM, ITML, and LMNN.

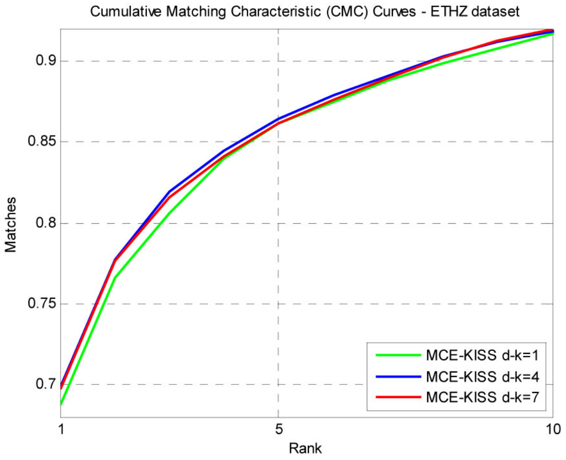


Fig. 6. Performance comparison using the CMC curve. In the figure, the *x*-coordinate is the rank score and *y*-coordinate is the matching rate. Top 10 ranking positions are depicted. This figure suggests the impact of variation of *k* in (11) for MCE-KISS.

TABLE V
AVERAGE TRAINING TIME

Size of training set	KISS	MCE-KISS
p_{tr} = 10	11.8ms	1100s
p_{tr} = 20	23.8ms	1098s
p_{tr} = 30	29.4ms	1084s
p_{tr} = 40	42.6ms	1065s

2) MCE-KISS is superior to RS-KISS, because the discriminative learning procedure for effectively adjusting the parameters of Gaussian density model significantly improves the generalization ability.
3) Fig. 5(a) and (b) shows that LMNN performs poorly, because the variations that cause differences between intraperson images in ETHZ are rather small. It also illustrates that LMNN models relative distance, which is sensitive to training samples.
4) Fig. 6 suggests the smoothing technique is useful for precise estimation of the small eigenvalues.
5) Fig. 7 suggests MCE-KISS can improve the accuracy of the KISS metric and tackle small samples size problem. Table V shows that the training time of MCE-KISS is associated with irritations and requires more training time than that of KISS.
6) Fig. 5(a) and Table II suggest MCE-KISS is more robust than KISS at the top five ranked matching rate. This also illustrates that the proposed algorithm is more reliable.

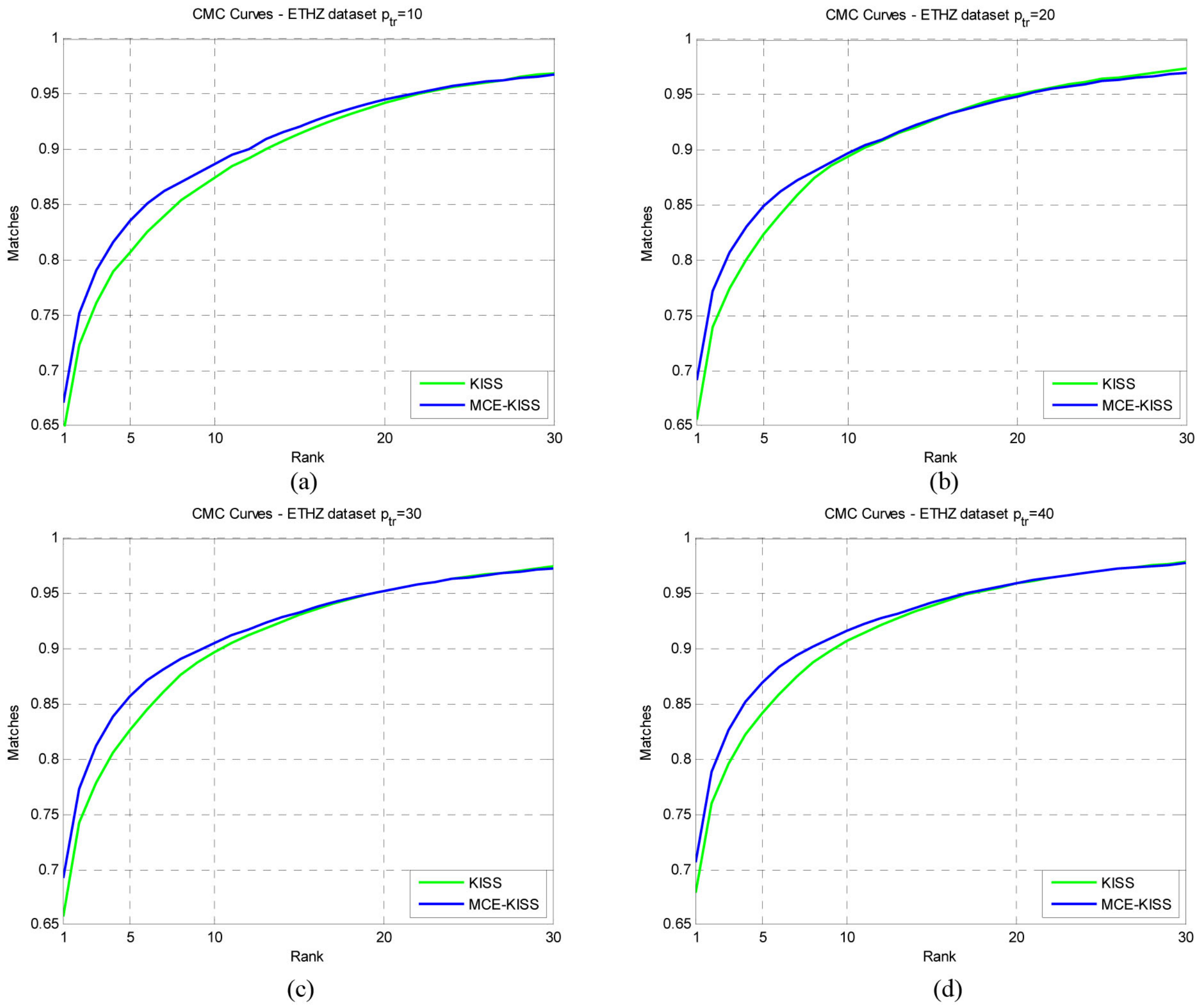


Fig. 7. Performance comparison using the CMC curves. In each subfigure, the *x*-coordinate is the rank score and *y*-coordinate is the matching rate. We compare MCE-KISS with KISS on the ETHZ dataset. Top 30 ranking positions are depicted. These subfigures suggest MCE-KISS can improve the accuracy of the KISS metric and tackle small samples size problem. (a) $p_{\text{tr}} = 10$. (b) $p_{\text{tr}} = 20$. (c) $p_{\text{tr}} = 30$. (d) $p_{\text{tr}} = 40$.

V. Conclusion

The distance metric is critically important for effective person reidentification in surveillance tasks. Thus, it is rational to find a suitable distance metric learning algorithm to boost the performance of person reidentification. In recent years, many distance metric learning algorithms have been developed, such as ITML and metric learning for LMNN. However, these algorithms are not suitable for person reidentification, because there are only limited training image pairs to learn a metric in person reidentification. Although KISS metric learning is considered state-of-the-art, it shares a similar problem.

Given a small number of training samples, we observe that covariance matrices estimated by KISS are biased. Therefore, we present the MCE-KISS. The proposed MCE-KISS algorithm exploits the smoothing technique to enlarge the small eigenvalues of the estimated covariance matrix, and discriminative learning based on MCE which is more reliable than classical ML estimation. The employed two statistical techniques effectively enlarge the underestimated small eigenvalues and better estimate the covariance matrix. Therefore, MCE-KISS significantly improves KISS for person reidentification.

Because the MCE-KISS relies on gradient descent that is an iterative optimization procedure, the training speed of MCE-KISS is much slower than KISS. Thus, we will consider parallelizing MCE-KISS to accelerate the training stage in the future.

References

[1] V. Chatzis, A. G. Bors, and I. Pitas, “Multimodal decision-level fusion for person authentication,” *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 29, no. 6, pp. 674–680, Nov. 1999.

[2] F. Dornaika and A. Bosaghzadeh, “Exponential local discriminant embedding and its application to face recognition,” *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 921–934, Jun. 2013.

[3] R. M. Da Costa and A. Gonzaga, “Dynamic features for iris recognition,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1072–1082, Aug. 2012.

[4] R. Cappelli, M. Ferrara, and D. Maio, “A fast and accurate palmprint recognition system based on minutiae,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 3, pp. 956–962, Jun. 2012.

[5] I. Venkat and P. De Wilde, “Robust gait recognition by learning and exploiting sub-gait characteristics,” *Int. J. Comput. Vis.*, vol. 91, no. 1, pp. 7–23, 2011.

[6] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, “Person re-identification using Haar-based and DCD-based signature,” in *Proc. AVSS*, Boston, MA, USA, 2010, pp. 1–8.

[7] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. 10th PETS*, 2007.

[8] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *Proc. SIBGRAPI*, Rio de Janiero, Brazil, 2009, pp. 322–329.

[9] B. Girod, V. Chandrasekhar, R. Grzeszczuk, and Y. A. Reznik, "Mobile visual search: Architectures, technologies, and the emerging MPEG standard," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 86–94, Mar. 2011.

[10] Y. Wen, W. Zhang, and H. Luo, "Energy-optimal mobile application execution: Taming resource-poor mobile devices with cloud clones," in *Proc. INFOCOM*, Orlando, FL, USA, 2012, pp. 2716–2720.

[11] Y. Gao and Y. Qi, "Robust visual similarity retrieval in single model face databases," *Pattern Recognit.*, vol. 38, no. 7, pp. 1009–1020, 2005.

[12] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE CVPR*, Providence, RI, USA, 2012, pp. 2288–2295.

[13] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face identification using large feature sets," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2245–2255, Apr. 2012.

[14] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.

[15] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. 10th ECCV*, Marseille, France, 2008, pp. 262–275.

[16] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biol. Cybern.*, vol. 61, no. 2, pp. 103–113, 1989.

[17] 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.

[18] C. Schmid, "Constructing models for content-based image retrieval," in *Proc. IEEE CVPR*, 2001, pp. 39–45.

[19] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking," in *Proc. BMVC*, 2010.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[21] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences," in *Proc. ICDSC*, Stanford, CA, USA, 2008, pp. 1–6.

[22] X. Liu *et al.*, "Attribute-restricted latent topic model for person re-identification," *Pattern Recognit.*, vol. 45, no. 12, pp. 4204–4213, 2012.

[23] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. ICIP*, 2002, pp. 900–903.

[24] T. Ojala, M. Pietikainen, and T. Maenpaa, "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.

[25] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. 17th SCIA*, Ystad, Sweden, 2011, pp. 91–102.

[26] N. D. Bird, O. Masoud, N. P. Papanikolopoulos, and A. Isaacs, "Detection of loitering individuals in public transportation areas," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 167–177, Jun. 2005.

[27] S. Bak, S. Suresh, F. Bremond, and M. Thonnat, "Fusion of motion segmentation with online adaptive neural classifier for robust tracking," in *Proc. VISAPP*, 2009, pp. 410–416.

[28] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Proc. CVPR*, 2006, pp. 1528–1535.

[29] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Person re-identification using spatial covariance regions of human body parts," in *Proc. AVSS*, Boston, MA, USA, 2010.

[30] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE CVPR*, Washington, DC, USA, 2006, pp. 2169–2178.

[31] J. Wang *et al.*, "Locality-constrained linear coding for image classification," in *Proc. IEEE CVPR*, San Francisco, CA, USA, Jun. 2010, pp. 3360–3367.

[32] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.

[33] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. NIPS*, 2001, pp. 585–591.

[34] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[35] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[36] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[37] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.

[38] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. ICCV*, Beijing, China, 2005, pp. 1208–1213.

[39] T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in *Proc. ECCV*, Marseille, France, 2008, pp. 725–738.

[40] 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.

[41] J.-E. Lee, R. Jin, and A. K. Jain, "Rank-based distance metric learning: An application to image retrieval," in *Proc. CVPR*, Anchorage, AK, USA, 2008, pp. 1–8.

[42] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.

[43] B.-H. Juang, W. Hou, and C.-H. Lee, "Minimum classification error rate methods for speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 257–265, May 1997.

[44] E. McDermott, T. J. Hazen, J. Le Roux, A. Nakamura, and S. Katagiri, "Discriminative training for large-vocabulary speech recognition using minimum classification error," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 203–223, Jan. 2007.

[45] P. C. Woodland and D. Povey, "Large scale discriminative training of hidden Markov models for speech recognition," *Comput. Lang.*, vol. 16, no. 1, pp. 25–47, 2002.

[46] B. Zamani, A. Akbari, B. Nasersharif, and A. Jalalvand, "Optimized discriminative transformations for speech features based on minimum classification error," *Pattern Recognit. Lett.*, vol. 32, no. 7, pp. 948–955, 2011.

[47] J. Reed and C.-H. Lee, "Preference music ratings prediction using tokenization and minimum classification error training," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 8, pp. 2294–2303, Nov. 2011.

[48] 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.

[49] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing KISS metric learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1675–1685, Oct. 2013.

[50] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. ICML*, Corvallis, OR, USA, 2007, pp. 209–216.

[51] F. Porikli, "Inter-camera color calibration by correlation model function," in *Proc. ICIP*, 2003, pp. 133–136.

[52] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013.

[53] L. Yang, R. Jin, R. Sukthankar, and Y. Liu, "An efficient algorithm for local distance metric learning," in *Proc. AAAI*, 2006, pp. 543–548.

[54] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views," *Comput. Vis. Image Understand.*, vol. 109, no. 2, pp. 146–162, 2008.

[55] A. Ess, B. Leibe, and L. Van Gool, "Depth and appearance for mobile scene analysis," in *Proc. ICCV*, Rio de Janeiro, Brazil, 2007, pp. 1–8.

[56] 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.

[57] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *Proc. ECCV*, Florence, Italy, 2012, pp. 780–793.

[58] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proc. IEEE CVPR*, Portland, OR, USA, 2013, pp. 3594–3601.

[59] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. CVPR*, Providence, RI, USA, 2012, pp. 2666–2672.

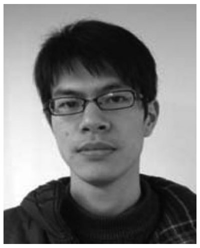
Dapeng Tao received the B.Eng. degree from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 1999 and 2014, respectively.

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

Lianwen Jin (M'98) received the B.S. degree from the University of Science and Technology of China, Anhui, 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. His current research interests include image processing, pattern recognition, machine learning, and intelligent systems. He has authored over 100 scientific papers.

Dr. Jin was the recipient of the New Century Excellent Talent Program Award of MOE, in 2006, and the Guangdong Pearl River Distinguished Professor Award, in 2011. He was the Program Committee Member for several international conferences, including the International Conference on Pattern Recognition, from 2010 to 2014, the International Conference on Machine Learning and Cybernetics, from 2007 to 2011, the International Conference on Frontiers in Handwriting Recognition, from 2008 to 2014, the International Conference on Document Analysis and Recognition, from 2009 to 2013, the International Conference on Multimedia and Expo, in 2014, and the International Conference on Image Processing in 2014.



Yongfei Wang received the B.S. degree in faculty information engineering from the Guangdong University of Technology, Guangzhou, China, in 2011. He is currently pursuing the master's degree in information and communication engineering from South China University of Technology, Guangzhou.

His current research interests include machine learning and computer vision.

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