Mahalanobis Distance Learning for Person Re-Identification

Peter M. Roth[1], Martin Hirzer[1], Martin Köstinger[1], Csaba Beleznai[2], and Horst Bischof[1]

[1] Graz University of Technology, Austria
[2] Austrian Institute of Technology, Austria

Abstract Recently, Mahalanobis metric learning has gained a considerable interest for single-shot person re-identification. The main idea is to build on an existing image representation and to learn a metric that reflects the visual camera-to-camera transitions, allowing for a more powerful classification. The goal of this chapter is twofold. We first review the main ideas of Mahalanobis metric learning in general and then give a detailed study on different approaches for the task of single-shot person re-identification, also comparing to the state-of-the-art. In particular, for our experiments we used Linear Discriminant Metric Learning (*LDML*), Information Theoretic Metric Learning (*ITML*), Large Margin Nearest Neighbor (*LMNN*), Large Margin Nearest Neighbor with Rejection (*LMNN-R*), Efficient Impostor-based Metric Learning (*EIML*), and *KISSME*. For our evaluations we used four different publicly available datasets (i.e., *VIPeR*, *ETHZ*, *PRID 2011*, and *CAVIAR4REID*). Additionally, we generated the new, more realistic *PRID 450S* dataset, where we also provide detailed segmentations. For the latter one, we also evaluated the influence of using well segmented foreground and background regions. Finally, the corresponding results are presented and discussed.

1 Introduction

Person re-identification has become one of the major challenges in visual surveillance, showing a rather wide range of applications such as searching for criminals or tracking and analyzing individuals or crowds. In general, there are two main strategies: single-shot and multi-shot recognition. For the first one, an image pair is matched: one image given as input and one stored in a database. In contrast, for multi-shot scenarios multiple images (i.e., trajectories) are available. In this chapter, we mainly focus on the single-shot case, even though the ideas can simply be extended to the multi-shot scenario.

Corresponding author: Peter M. Roth, e-mail: pmroth@icg.tugraz.at

Even for humans, person re-identification is very challenging for several reasons. First, the appearance of an individual can vary extremely across a network of cameras due to changing view points, illumination, different poses, etc. Second, there is a potentially high number of "similar" persons (e.g., people wear rather dark clothes in winter). Third, in contrast to similar large scale search problems typically no accurate temporal and spatial constraints can be exploited to ease the task. Having these problems in mind and motivated by the high number of practical applications there has been a significant scientific interest during the last years (e.g., [3, 7, 11, 14, 16, 22, 26, 28, 29]), and also various benchmark datasets (e.g., [13, 16, 28]) have been published.

In general, the main idea is to find a suitable image description and then to perform a matching step using a standard distance. For describing images there exist two different strategies: (a) invariant and (b) discriminative description. The goal of invariant methods (e.g., [9, 11, 16, 27, 29]) is to extract visual features that are both, distinctive and stable under changing viewing conditions between different cameras. The large intra-class appearance variations, however, make the computation of distinctive and stable features often impossible under realistic conditions. To overcome this limitation, discriminative methods (e.g., [3, 14, 16, 28] on the other hand take advantage of class information to exploit the discriminative information to find a more distinctive representation. However, as a drawback such methods tend to overfit to the training data. Moreover, they are often based on local image descriptors, which might be a severe disadvantage. For instance, a red bag visible in on view would be very discriminative, however, if it is not visible in the other view it becomes impossible to re-identify a specific person.

An alternative to these two approaches, also incorporating label information, is to adopt metric learning for the given task (e.g., [7, 17, 18, 20, 21, 31]). Similar to the idea of inter-camera color calibration (e.g., [25]), using labeled samples transitions in feature space between two camera views can be modeled. Hence, using a non-Euclidean distance even less distinctive features, which have not to capture the visual invariances between different cameras, are sufficient for getting considerable matching results. However, to estimate such a metric, a training stage is necessary, but once learned, metric learning approaches are very efficient during evaluation, since additionally to the feature extraction and the matching only a linear projection has to be computed.

When dealing with person re-identification we have to cope with three main problems. First, to capture all relevant information often complex, high dimensional feature representations are required. Thus, widely used metric learners such as Large Margin Nearest Neighbor (*LMNN*) [30], Information Theoretic Metric Learning (*ITML*) [6], and Logistic Discriminant Metric Learning (*LDML*) [15] building on complex optimization schemes run into high computational costs and memory requirements, making them infeasible in practice. Second, these methods typically assume a multi-class classification problem, which is not the case for person re-identification. In fact, we are typically given image pairs, so existing methods have to be adapted. There are only a few methods such as [1, 12] which directly intend learning a metric from data pairs. Third, we have to deal with a partially ill-posed

problem. In fact, two images showing the same person might not be similar (e.g., due to camera noise, geometry, or different viewpoints: frontal vs. back). On the other hand, images not showing the same person can be very similar (e.g., in winter many people wear black/dark gray coats). Thus, for standard methods there is a high tendency to overfit to the training data yielding insufficient results during testing.

The goal of this chapter is to analyze the applicability of metric learning for the task of single-shot person re-identification from a more general point of view. Thus, we first review the main idea of Mahalanobis distance metric learning and give an overview of selected approaches targeting at the problem of discriminative metric learning via different strategies. In particular, we selected established methods applied to diverse visual classification tasks, (i.e., Logistic Discriminant Metric Learning (*LDML*) [15], Information Theoretic Metric Learning (*ITML*) [6], and Large Margin Nearest Neighbor (*LMNN*) [30]), as well as approaches that have been developed in particular for person re-identification (i.e., Large Margin Nearest Neighbor with Rejection (*LMNN-R*) [7], Efficient Impostor-based Metric Learning (*EIML*) [17], and *KISSME* [20]).

To show that metric learning is widely applicable, we run experiments on five different datasets showing different characteristics. Four of them, namely *VIPeR*, *ETHZ*, *PRID 2011*, and *CAVIAR4REID*, are publicly available and widely used. For a more thorough evaluation and as additional contribution we created a new, more realistic dataset, *PRID 450S*, where we also provide detailed foreground/background segmentations. The results are summarized and compared to state-of-the-art results for the specific datasets. In addition, to have a generative and discriminative baseline the same experiments were also run using the standard Mahalanobis distance and a slightly adapted version of Linear Discriminant Analysis (*LDA*) [10].

The rest of the chapter is organized as follows. First, in Sec. 2 Mahalanobis metric learning in general is introduced and the approaches used in the study are summarized. Then, in Sec. 3, our specific person re-identification framework consisting of three stages is presented. In Sec. 4 and 5 we first review the five datasets used for our study and then present the obtained results. Finally, in Sec. 6 we summarize and conclude the chapter.

2 Mahalanobis Distance Metric Learning

In this section, we first introduce the general idea of Mahalanobis metric learning and then give an overview of the approaches used in this study. We selected generic methods that have shown good performance for diverse visual classification tasks as well as specific methods that have been developed for the task of person re-identification. Moreover, to give a more generic analysis, we tried to select methods tackling the same problem from different points of view: generative data analysis, statistical inference, information theoretic aspect, and discriminative learning. Additionally, we consider Linear Discriminant Analysis (*LDA*) and standard Mahalanobis metric learning, which can be considered simple baselines. For all methods

the implementations are publicly available, thus allowing (a) for a fair comparison, and (b) for easily exchanging the used representation.

2.1 Mahalanobis Metric

Mahalanobis distance learning is a prominent and widely used approach for improving classification results by exploiting the structure of the data. Given n data points $\mathbf{x}_i \in \mathbb{R}^m$, the goal is to estimate a matrix $\mathbf{M}$ such that

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) \tag{1}$$

describes a pseudo-metric. In fact, this is assured if $\mathbf{M}$ is positive semi-definite, i.e., $\mathbf{M} \succeq 0$. If $\mathbf{M} = \Sigma^{-1}$ (i.e., the inverse of the sample covariance matrix), the distance defined by Eq. (1) is referred to as the *Mahalanobis distance*. An alternative formulation for Eq. (1), which is more intuitive, is given via

$$d_{\mathbf{L}}(\mathbf{x}_i, \mathbf{x}_j) = ||\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)||^2 \,, \tag{2}$$

which is easily obtained from

$$(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \underbrace{\mathbf{L}^\top \mathbf{L}}_{\mathbf{M}} (\mathbf{x}_i - \mathbf{x}_j) = ||\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)||^2 \,. \tag{3}$$

Hence, either directly the metric matrix $\mathbf{M}$ or the factor matrix $\mathbf{L}$ can be estimated from the data. A discussion on factorization and the corresponding optimality criteria can be found in, e.g., [4, 19].

If additionally for a sample $\mathbf{x}$ its class label $y(\mathbf{x})$ is given, not only the generative structure of the data but also discriminative information can be exploited. For many problems (including person re-identification), however, we are lacking class labels. Thus, given a pair of samples $(\mathbf{x}_i, \mathbf{x}_j)$ we break down the original multi-class problem into a two-class problem in two steps. First, we transform the samples from the data space to the label agnostic difference space $\mathcal{X} = \{\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j\}$, which is inherently given by the metric definitions in Eqs. (1) and (2). Moreover, $\mathcal{X}$ is invariant to the actual locality of the samples in the feature space. Second, the original class labels are discarded and the samples are arranged using pairwise equality and inequality constraints, where we obtain the classes *same* $\mathcal{S}$ and *different* $\mathcal{D}$:

$$\mathcal{S} = \{(\mathbf{x}_i, \mathbf{x}_j) | y(\mathbf{x}_i) = y(\mathbf{x}_j)\} \tag{4}$$

$$\mathcal{D} = \{(\mathbf{x}_i, \mathbf{x}_j) | y(\mathbf{x}_i) \neq y(\mathbf{x}_j)\} \,. \tag{5}$$

In our particular case the pair $(\mathbf{x}_i, \mathbf{x}_j)$ consists of images showing persons in different camera views, and sharing a label means that the samples $\mathbf{x}_i$ and $\mathbf{x}_j$ describe the same person. In the following, we exemplary discuss different approaches dealing with the problem described above.

To increase readability we introduce the notation $\mathbf{C}_{ij} = (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top$ and the similarity variable

$$y_{ij} = \begin{cases} 1 & y(\mathbf{x}_i) = y(\mathbf{x}_j) \\ 0 & y(\mathbf{x}_i) \neq y(\mathbf{x}_j) \end{cases}. \tag{6}$$

2.2 Linear Discriminant Analysis

Let $\mathbf{x}_i \in \mathbb{R}^m$ be a sample and c its corresponding class label. Then, the goal of Linear Discriminant Analysis (*LDA*) [10] is to compute a classification function $g(\mathbf{x}) = \mathbf{L}^\top \mathbf{x}$ such that the Fisher-criterion

$$\mathbf{L}_{opt} = \arg\max_{\mathbf{L}} \frac{\left|\mathbf{L}^\top \mathbf{S}_b \mathbf{L}\right|}{\left|\mathbf{L}^\top \mathbf{S}_w \mathbf{L}\right|}, \tag{7}$$

where $\mathbf{S}_w$ and $\mathbf{S}_b$ are the within-class scatter and between-class scatter matrices, is optimized. This is typically realized via solving the generalized eigenvalue problem

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} \tag{8}$$

or directly by computing the eigenvectors for $\mathbf{S}_W^{-1}\mathbf{S}_B$.

However, it is known that the Fisher-criterion given by Eq. (7) is only optimal in Bayes' sense for two classes (see, e.g., [23]). Thus, if the number of classes (image pairs in our case) is increasing LDA is going to fail. To overcome this problem, we can re-formulate the original multi-class objective Eq. (7) to a binary formulation by using the two classes defined in Eqs. (4) and (5). In other words, Eq. (7) tries to minimize the distance between similar pairs and to maximize the distance between dissimilar pairs.

2.3 Logistic Discriminant Metric Learning

A similar idea is followed by Logistic Discriminant Metric Learning (*LDML*) of Guillaumin et al. [15], however, from a probabilistic point of view. Thus, to estimate the Mahalanobis distance the probability p_{ij} that a pair $(\mathbf{x}_i, \mathbf{x}_j)$ is similar is modeled as

$$p_{ij} = p(y_{ij} = 1 | \mathbf{x}_i, \mathbf{x}_j; \mathbf{M}, b) = \sigma(b - d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j)), \tag{9}$$

where $\sigma(z) = (1 + \exp(-z))^{-1}$ is a sigmoid function and b is a bias term. As Eq. (9) is a standard linear logistic model, $\mathbf{M}$ can be optimized by maximizing the log-likelihood

$$\mathcal{L}(\mathbf{M}) = \sum_{ij} y_{ij} \ln(p_{ij}) + (1 - y_{ij}) \ln(1 - p_{ij}). \tag{10}$$

The optimal solution is then obtained by gradient ascent in direction

$$\frac{\partial \mathcal{L}(\mathbf{M})}{\partial \mathbf{M}} = \sum_{ij} (y_{ij} - p_{ij}) \mathbf{C}_{ij} \ , \tag{11}$$

where the influence of each pair on the gradient direction is controlled over the probability. No further constraints, in particular no positive semi-definiteness on $\mathbf{M}$, are imposed on the problem!

2.4 Information Theoretic Metric Learning

Similarly, Information Theoretic Metric Learning (*ITML*) was presented by Davis et al. [6], who regularize the estimated metric $\mathbf{M}$ by minimizing the distance to a predefined metric $\mathbf{M}_0$ via an information-theoretic approach. In particular, they exploit the existence of a bijection between the set of Mahalanobis distances and the set of equal-mean multivariate Gaussian distributions. Let $d_{\mathbf{M}}$ be a Mahalanobis distance, then its corresponding multivariate Gaussian is given by

$$g(\mathbf{x}, \mathbf{M}) = \frac{1}{Z} exp\left(-\frac{1}{2} d_{\mathbf{M}}(\mathbf{x}, \mu) \right) \ , \tag{12}$$

where Z is a normalizing factor, μ is the mean, and the covariance is given by $\mathbf{M}^{-1}$.

Thus, the goal is to minimize the relative entropy between $\mathbf{M}$ and $\mathbf{M}_0$ arising the following optimization problem:

$$min\ KL(g(\mathbf{x}, \mathbf{M}_0) || g(\mathbf{x}, \mathbf{M})) \tag{13}$$

$$s.t.$$

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \leq u \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S} \tag{14}$$

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \geq l \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D} \ , \tag{15}$$

where KL is the Kullback-Leibler divergence, and the constraints in Eqs. (14) and (15) enforce that the distances between similar pairs are small while they are large for dissimilar pairs.

As the optimization problem Eqs. (13)–(15) can be expressed via Bregman divergence, starting from $\mathbf{M}_0$ the Mahalanobis distance matrix $\mathbf{M}$ can be obtained by the following update rule:

$$\mathbf{M}_{t+1} = \mathbf{M}_t + \beta \mathbf{M}_t \mathbf{C}_{ij} \mathbf{M}_t \ , \tag{16}$$

where β encodes both, the pair label and the step size.

2.5 Large Margin Nearest Neighbor

In contrast, Large Margin Nearest Neighbor (*LMNN*) metric learning, introduced by Weinberger and Saul [30], additionally exploits the local structure of the data. For each instance a local perimeter surrounding the k nearest neighbors sharing the same label (target neighbors) is established. Samples having a different label that invade this perimeter (impostors) are penalized. More technically, for a target pair $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}$, i.e, $y_{ij} = 1$, any sample $\mathbf{x}_l$ with $y_{il} = 0$ is an impostor if

$$||\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)||^2 \leq ||\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)||^2 + 1 \; . \tag{17}$$

Thus, the objective is to pull target pairs together and to penalize the occurrence of impostors. This is realized via the following objective function:

$$\mathcal{L}(\mathbf{M}) = \sum_{j \rightsquigarrow i} \left[d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) + \beta \sum_l (1 - y_{il}) \xi_{ijl}(\mathbf{M}) \right] \tag{18}$$

with

$$\xi_{ijl}(\mathbf{M}) = 1 + d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) - d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_l) \tag{19}$$

and β to be a weighting factor. The first term of Eq. (18) minimizes the distance between target neighbors $\mathbf{x}_i$ and $\mathbf{x}_j$, indicated by $j \rightsquigarrow i$, and the second one denotes the amount by which impostors invade the perimeter of $\mathbf{x}_i$ and $\mathbf{x}_j$. To estimate the metric $\mathbf{M}$, gradient descent is performed on the objective function Eq. (18):

$$\frac{\partial \mathcal{L}(\mathbf{M})}{\partial \mathbf{M}} = \sum_{j \rightsquigarrow i} \mathbf{C}_{ij} + \beta \sum_{(i,j,l) \in \mathcal{N}} (\mathbf{C}_{ij} - \mathbf{C}_{il}) \tag{20}$$

where $\mathcal{N}$ describes the set of triplets indices corresponding to a positive slack.

LMNN was later adopted for person re-identification by Dikmen et al. [7], who introduced a rejection scheme not returning a match if all neighbors are beyond a certain threshold: Large Margin Nearest Neighbor with Rejection (*LMNN-R*).

2.6 Efficient Impostor-based Metric Learning

Since both approaches described in Sec. 2.5, *LMNN* and *LMNN-R*, rely on complex optimization schemes, in [17] Efficient Impostor-based Metric Learning (*EIML*) was proposed that allows for exploiting the information provided by impostors more efficiently. In particular, Eq. (17) is relaxed to the original difference space. Thus, given a target pair $(\mathbf{x}_i, \mathbf{x}_j)$, a sample $\mathbf{x}_l$ is an impostor if

$$||(\mathbf{x}_i - \mathbf{x}_l)||^2 \leq ||(\mathbf{x}_i - \mathbf{x}_j)||^2 \; . \tag{21}$$

To estimate the metric $\mathbf{M} = \mathbf{L}^\top \mathbf{L}$ the following objective function has to be minimized:

$$\mathcal{L}(\mathbf{L}) = \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} ||\mathbf{L}(x_i - x_j)||^2 - \sum_{(\mathbf{x}_i,\mathbf{x}_l)\in\mathcal{I}} ||\mathbf{L}\, w_{il}\, (x_i - x_l)||^2 \,, \tag{22}$$

where $\mathcal{I}$ is the set of all impostor pairs and

$$w_{il} = e^{-\frac{||x_i - x_l||}{||x_i - x_j||}} \tag{23}$$

is a weighting factor also taking into account how much an impostor invades the perimeter of a target pair. By adding the orthogonality constraint $\mathbf{L}\mathbf{L}^\top = \mathbf{I}$, Eq. (22) can be re-formulated to an eigenvalue problem:

$$(\Sigma_\mathcal{S} - \Sigma_\mathcal{I})\mathbf{L} = \Lambda\mathbf{L} \,, \tag{24}$$

where

$$\Sigma_\mathcal{S} = \frac{1}{|\mathcal{S}|} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} \mathbf{C}_{ij} \quad \text{and} \quad \Sigma_\mathcal{I} = \frac{1}{|\mathcal{I}|} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{I}} \mathbf{C}_{ij} \tag{25}$$

are the covariance matrices for $\mathcal{S}$ and $\mathcal{I}$, respectively. Hence, the problem is much simpler and can be solved efficiently.

2.7 KISSME

The goal of the Keep It Simple and Straightforward MEtric (*KISSME*) [20] is to address the metric learning approach form a statistical inference point of view. Therefore, we test the hypothesis H_0 that a pair $(\mathbf{x}_i, \mathbf{x}_j)$ is dissimilar against H_1 that it is similar using a likelihood ratio test:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log\left(\frac{p(\mathbf{x}_i, \mathbf{x}_j|H_0)}{p(\mathbf{x}_i, \mathbf{x}_j|H_1)}\right) = \log\left(\frac{f(\mathbf{x}_i, \mathbf{x}_j, \theta_0)}{f(\mathbf{x}_i, \mathbf{x}_j, \theta_1)}\right) \,, \tag{26}$$

where δ is the log-likelihood ratio, and $f(\mathbf{x}_i, \mathbf{x}_j, \theta)$ is a PDF with the parameter set θ. Assuming zero-mean Gaussian distributions Eq. (26) can be re-written to

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log\left(\frac{\frac{1}{\sqrt{2\pi|\Sigma_\mathcal{D}|}} \exp(-1/2\, (\mathbf{x}_i - \mathbf{x}_j)^\top\, \Sigma_\mathcal{D}^{-1}\, (\mathbf{x}_i - \mathbf{x}_j))}{\frac{1}{\sqrt{2\pi|\Sigma_\mathcal{S}|}} \exp(-1/2\, (\mathbf{x}_i - \mathbf{x}_j)^\top\, \Sigma_\mathcal{S}^{-1}\, (\mathbf{x}_i - \mathbf{x}_j))}\right) , \tag{27}$$

where $\Sigma_\mathcal{S}$ and $\Sigma_\mathcal{D}$ are the covariance matrices of $\mathcal{S}$ and $\mathcal{D}$ according to Eq. (25).

The maximum likelihood estimate of the Gaussian is equivalent to minimizing the distances from the mean in a least squares manner. This allows *KISSME* to find respective relevant directions for $\mathcal{S}$ and $\mathcal{D}$. By taking the log and discarding the constant terms we can simplify Eq. (27) to

$$\begin{aligned}\delta(\mathbf{x}_i,\mathbf{x}_i) &= (\mathbf{x}_i-\mathbf{x}_j)^\top \Sigma_\mathcal{S}^{-1} (\mathbf{x}_i-\mathbf{x}_j) - (\mathbf{x}_i-\mathbf{x}_j)^\top \Sigma_\mathcal{D}^{-1} (\mathbf{x}_i-\mathbf{x}_j) \\ &= (\mathbf{x}_i-\mathbf{x}_j)^\top (\Sigma_\mathcal{S}^{-1}-\Sigma_\mathcal{D}^{-1})(\mathbf{x}_i-\mathbf{x}_j)\ . \end{aligned} \tag{28}$$

Hence, the Mahalanobis distance matrix $\mathbf{M}$ is defined by

$$\mathbf{M} = \left(\Sigma_\mathcal{S}^{-1}-\Sigma_\mathcal{D}^{-1}\right)\ . \tag{29}$$

3 Person Re-Identification System

In the following, we introduce the person re-identification system used for our study consisting of three stages: (1) feature extraction, (2) metric learning, and (3) classification. The overall system is illustrated in Fig. 1. During training the metric between two cameras is estimated, which is then used for calculating the distances between an unknown sample and the samples given in the database. The three steps are discussed in more detail in the next sections.

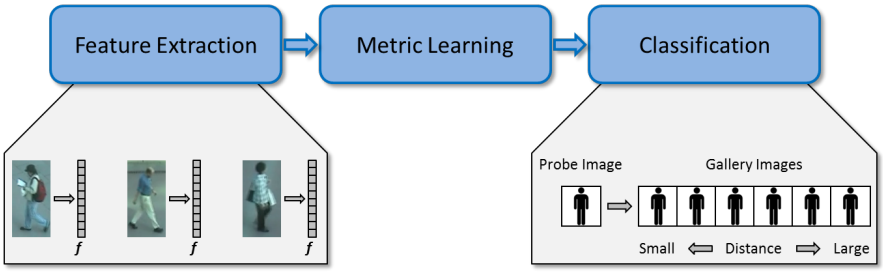


Fig. 1 Person re-identification system consisting of three stages: (1) feature extraction – dense sampling of color and texture features, (2) metric learning – exploiting the structure of similar and dissimilar pairs, (3) classification – nearest neighbor search under the learned metric.

3.1 Representation

Color and texture features have proven to be successful for the task of person re-identification. We use HSV and Lab color channels as well as Local Binary Patterns

to create a person image representation. The features are extracted from 8x16 rectangular regions sampled from the image with a grid of 4x8 pixels, i.e., 50% overlap in both directions, which is illustrated in Fig. 2. In each rectangular patch we calculate the mean values per color channel, which are then discretized to the range 0 to 40. Additionally, a histogram of LBP codes is generated from a gray value representation of the patch. These values are then put together to form a feature vector. Finally, the vectors from all regions are concatenated to generate a representation for the whole image.

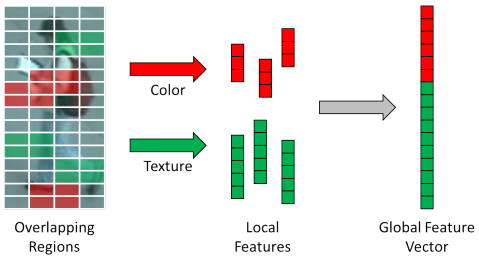


Fig. 2 Global image descriptor: different local features (HSV, Lab, LBP) are extracted from overlapping regions and are then concatenated to a single feature vector.

3.2 Metric Learning

First of all, we run a PCA step to reduce the dimensionality and for noise removal. In general, this step is not critical (the particular settings are given in Sec. 5), however, we recognized that for smaller datasets also a lower dimensional representation is sufficient. During training we learn a Mahalanobis metric $\mathbf{M}$ according to Eq. (1). Once $\mathbf{M}$ has been estimated, during evaluation the distance between two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ is calculated via Eq. (1). Hence, additionally to the actual classification effort only linear projections are required.

3.3 Classification

In person re-identification we want to recognize a certain person across different, non-overlapping camera views. In our setup, we assume that we have already detected the persons in all camera views, i.e., we do not tackle the detection problem. The goal of person re-identification now is to find a person image that has been selected in one view (*probe image*) in all the images from another view (*gallery images*). This is achieved by calculating the distances between the probe image and

all gallery images using the learned metric, and returning those gallery images with the smallest distances as potential matches.

4 Re-Identification Datasets

In the following, we give an overview of the datasets used in our evaluations and explain the corresponding setups. In particular, these are *VIPeR* [13], *PRID 2011* [16], *ETHZ* [28], *CAVIAR4REID* [5], and *PRID 450S*. The first four (see Fig. 3) are publicly available and widely used for benchmarking person re-identification methods; the latter one was newly generated for this study.

Although there are other datasets like *iLIDS* we abstained from using them in this study. "The" *iLIDS* dataset was not used since there are at least four different datasets available that arbitrarily cropped patches from the huge (publicly not available!) *iLIDS* dataset, making it difficult to give fair comparisons.

(a) (b) (c) (d)

Fig. 3 Example image pairs from (a) the *VIPeR*, (b) the *PRID 2011*, (c) the *ETHZ*, and (d) the *CAVIAR4REID* dataset. The upper and lower row correspond to different appearances of the same person, respectively.

4.1 VIPeR Dataset

The *VIPeR* dataset contains 632 person image pairs taken from two different camera views. Changes of viewpoint, illumination and pose are the most prominent sources of appearance variation between the two images of a person. For evaluation we followed the procedure described in [14]. The set of 632 image pairs is randomly split into two sets of 316 image pairs each, one for training and one for testing. In the test case, the two images of an image pair are randomly assigned to a probe and a gallery set. A single image from the probe set is then selected and matched with all images from the gallery set. This process is repeated for all images in the probe set.

4.2 ETHZ Dataset

The *ETHZ* dataset [28], originally proposed for pedestrian detection [8] and later modified for benchmarking person re-identification approaches, consists of three video sequences: SEQ. #1 containing 83 persons (4.857 images), SEQ. #2 containing 35 persons (1.961 images), and SEQ. #3 containing 28 persons (1.762 images). All images have been re-sized to 64×32 pixels. The most challenging aspects of this dataset are illumination changes and occlusions. However, as the person images are captured from a single moving camera, the dataset does not provide a realistic scenario for person re-identification (i.e., no disjoint cameras, different viewpoints, different camera characteristics, etc.). Despite this limitation it is commonly used for person re-identification. We use a single-shot evaluation strategy, i.e., we randomly sample two images per person to build a training pair; and another pair for testing. The images of the test pairs are then assigned to the probe and the gallery set.

4.3 PRID 2011 Dataset[1]

The *PRID 2011* dataset consists of person images recorded from two different static cameras. Two scenarios are provided: multi-shot and single-shot. Since we are focusing on single-shot methods in this work, we use only the latter one. Typical challenges on this dataset are viewpoint and pose changes as well as significant differences in illumination, background and camera characteristics. Camera view A contains 385 persons, camera view B contains 749 persons, with 200 of them appearing in both views. Hence, there are 200 person image pairs in the dataset. These image pairs are randomly split into a training and a test set of equal size. For evaluation on the test set, we followed the procedure described in [16], i.e., camera A is used for the probe set and camera B is used for the gallery set. Thus, each of the 100 persons in the probe set is searched in a gallery set of 649 persons (all images of camera view B except the 100 training samples).

4.4 CAVIAR4REID Dataset

The *CAVIAR4REID* dataset [5] contains images of 72 individuals captured from two different cameras in a shopping center, where the original images have been resized to 128×64. 50 of them appear in both camera views, the remaining 22 only in one view. Since we are interested in person re-identification in different cameras, we only use individuals appearing in both views in our experiments. Each person is represented by 10 appearances per camera view. Typical challenges on

[1] The dataset is publicly available under https://lrs.icg.tugraz.at/download.php.

this dataset are viewpoint and pose changes, different light conditions, occlusions, and low resolution. To compare the different methods we use a multi-shot evaluation strategy similar to [2]. The set of 50 persons is randomly split into a training set of 42 persons, and a test set of 8 persons. Since every person is represented by 10 images per camera view, we can generate 100 different image pairs between the views of two individuals. During training, we use all possible combinations of positive pairs showing the same person, and negative pairs showing different persons. When comparing two individuals in the evaluation stage, we again use all possible combinations in order to calculate the mean distance between the two persons.

4.5 PRID 450S Dataset[2]

The *PRID 450S* dataset builds on *PRID 2011*, however, is arranged according to *VIPeR* by image pairs and contains more linked samples than *PRID 2011*. In particular, the dataset contains 450 single-shot image pairs depicting walking humans captured in two spatially disjoint camera views. From the original images with resolution of 720×576 pixels, person patches were annotated manually by bounding boxes with a vertical resolution of 100-150 pixels. To form the ground truth for re-identification, persons with the same identity seen in the different views were associated. In addition, for each image instance we generated binary segmentation masks separating the foreground from the background. Moreover, we further provide a part-level segmentation[3] describing the following regions: head, torso, legs, carried object at torso level (if any) and carried object below torso (if any). The union of these part segmentations is equivalent to the foreground segment. Exemplary images and corresponding segmentations for both cameras are illustrated in Fig. 4.

Fig. 4 *PRID 450S* dataset: Original images (top) and multi-label segmentations (bottom) for both camera views.

[2] The dataset is publicly available under https://lrs.icg.tugraz.at/download.php.

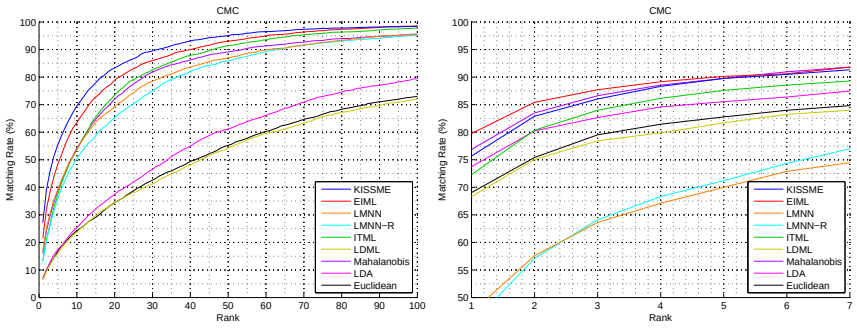
[3] The more detailed segmentations were actually not used for this study, but as they could be beneficial for others they are also provided.

5 Experimental Results

In the following, we give a detailed study on metric learning for person re-identification using the framework introduced in Sec. 4. In particular, we compare the methods discussed in Sec. 2 using the datasets presented in Sec. 4, where all methods get exactly the same data (training/test splits, representation). The results are presented in form of CMC scores [29], representing the expectation of finding the true match within the first r ranks. In particular, we plot the CMC scores for the different metric learning approaches and additionally provide tables for the first ranks, where the best scores are given in boldface, respectively. If available, also comparisons to state-of-the-art methods are given. The reported results are averaged over 10 random runs. Regarding the number of PCA dimensions, we use 100 dimensions for *VIPeR* and *CAVIAR4REID*, 40 for *PRID 2011*, *PRID 450S* and *ETHZ* SEQ. #1, and 20 for *ETHZ* SEQ. #2 and SEQ. #3.

5.1 Dataset Evaluations

The first experiment was carried out on the *VIPeR* dataset, which can be considered the standard benchmark for single-shot re-identification scenarios. The CMC curves for the different metric learning approaches are shown in Fig. 5(a). It can be seen that besides *LDA* and *LDML*, which either have too weak discriminative power or are overfitting to the training data, all approaches significantly improve the classification results over all rank levels. In addition, we provide these results compared to state-of-the-art methods (i.e., *ELF* [14], *SDALF* [9], *ERSVM* [26], *DDC* [16], *PS* [5], *PRDC* [31], and *PCCA* [24]) in Table 1. As for many methods timings are available, these are also included in the table. The results show that metric learning boosts the performance of the originally quite simple representation and finally yields competitive results; however, at dramatically reduced computational complexity.



(a) *VIPeR*. (b) *ETHZ* (SEQ. #1).

Fig. 5 CMC curves for (a) *VIPeR* and (b) *ETHZ* SEQ. #1.

Method	$r = 1$	10	20	50	100	t_{train}
KISSME [20]	**27**	**70**	**83**	**95**	**99**	0.1 sec
EIML [17]	22	63	79	93	**99**	0.3 sec
LMNN [30]	17	54	69	87	96	2 min
LMNN-R [7]	13	50	65	86	95	45 min
ITML [6]	13	54	73	91	98	25 sec
LDML [15]	6	24	35	54	72	0.8 sec
Mahalanobis	16	54	72	89	96	**0.001 sec**
LDA	7	25	37	61	79	0.1 sec
Euclidean	7	24	34	55	73	–
ELF [14]	12	43	60	81	93	5 hours
SDALF [9]	20	50	65	85	–	–
ERSVM [26]	13	50	67	85	94	13 min
DDC [16]	19	52	65	80	91	–
PS [5]	22	57	71	87	–	–
PRDC [31]	16	54	70	87	97	15 min
PCCA [24]	19	65	80	–	–	–

Table 1 CMC scores (in [%]) and average training times per trial for *VIPeR*.

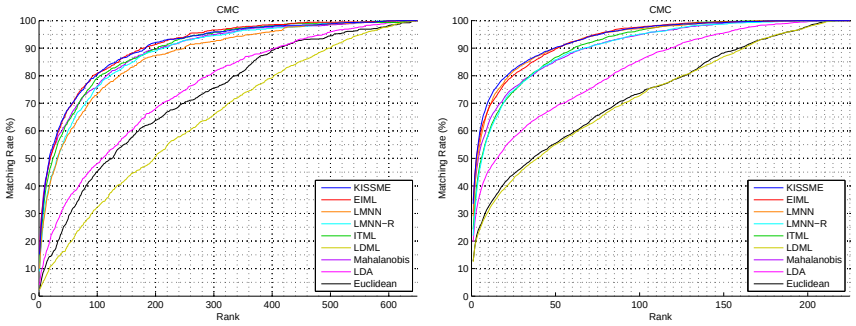
Next, we show results for *ETHZ*, another widely used benchmark, containing trajectories of persons captured from a single camera. Thus, the image pairs show the same characteristics and metric learning has only little influence. Nevertheless, the CMC curves in Fig. 5(b) for SEQ. #1, where metric learning has the largest impact, reveal that a performance gain of more than 5% can be obtained over all ranks. The decrease of *LMNN* can be explained by the evaluation protocol, which generates impostors resulting in an overfitting model.

	SEQ. #1							SEQ. #2							SEQ. #3						
Method	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
KISSME [20]	76	83	86	88	**90**	90	91	69	79	83	86	89	90	91	83	91	93	95	96	98	98
EIML [17]	**80**	**85**	**88**	**89**	**90**	**91**	**92**	**74**	**83**	**87**	**90**	**91**	**92**	**93**	**90**	**94**	95	**96**	**98**	**99**	**99**
LMNN [6]	47	58	64	67	70	73	74	40	51	59	66	70	75	79	34	51	61	66	72	77	79
LMNN-R [7]	45	57	64	68	71	74	77	47	56	65	72	76	79	83	49	64	73	79	83	86	89
ITML [30]	72	80	84	86	88	89	89	70	81	85	87	89	90	91	88	93	**96**	**96**	**98**	98	**99**
LDML [15]	68	75	78	80	82	83	84	64	74	78	81	84	85	86	81	88	91	95	96	96	96
Mahalanobis	77	83	87	**89**	**90**	**91**	**92**	70	81	85	89	89	91	91	84	91	93	95	96	98	98
LDA	74	80	83	85	86	86	87	70	81	85	87	90	91	92	88	**94**	**96**	**96**	**98**	98	98
Euclidean	69	75	80	81	83	84	85	68	77	81	83	85	87	89	85	91	94	95	96	96	97

Table 2 CMC scores (in [%]) for *ETHZ* for the first 7 ranks.

In contrast, *PRID 2011* defines a more realistic setup. In fact, the images stem from multiple cameras and especially the number of gallery images is much higher. Again from the CMC curves in Fig. 6(a) it can be seen that for all methods besides *LDA* and *LDML* a significant improvement can be obtained, especially for the first ranks. The results in Table 3 reveal that in this case also using a standard Mahalanobis distance yields competitive results. Moreover, it can be seen that the

descriptive approach [16], which uses a much more complex representation, can clearly be outperformed.



(a) *PRID 2011*. (b) *PRID 450S* (with segmentation).

Fig. 6 CMC curves for (a) *PRID 2011* and (b) *PRID 450S*.

Method	$r = 1$	10	20	50	100
KISSME [20]	15	39	**52**	**68**	80
EIML [17]	**16**	39	51	**68**	**81**
LMNN [30]	10	30	42	59	73
LMNN-R [7]	9	32	43	60	76
ITML [6]	12	36	47	64	79
LDML [15]	2	6	11	19	32
Mahalanobis	**16**	**41**	51	64	76
LDA	4	14	21	35	48
Euclidean	3	10	14	28	45
Descr. M. [16]	4	24	37	56	70

Table 3 CMC scores (in [%]) for *PRID 2011*.

As the newly created *PRID 450S* dataset builds on *PRID 2011*, it has similar characteristics, however, provides much more linked samples. In addition, we also generated detailed foreground/background masks, allowing us to analyze the effect of using an exact foreground/background segmentation. The CMC curves exploiting the given segmentations are shown in Fig. 6(b). Again it can be seen that using *LDML* has no and using *LDA* has only a little influence on the classification results, whereas for all other approaches a significant improvement can be obtained. The impact of segmentation is analyzed in Table 4, where both, the results with and without segmentation, are compared. It can be recognized that using the foreground information is beneficial for all approaches increasing the performance by up to 5%.

Finally, we show the results for the *CAVIAR4REID* dataset for two reasons. First, to demonstrate that metric learning can also be applied if the number of training samples is small, and, second, to show that the single-shot setup can easily be ex-

Method	$r = 1$	10	20	50	100
KISSME [20]	28	**65**	**76**	**88**	**96**
EIML [17]	**29**	62	73	86	**96**
LMNN [30]	24	62	73	87	**96**
LMNN-R [7]	19	54	66	81	92
ITML [6]	21	53	67	84	95
LDML [15]	4	16	23	40	53
Mahalanobis	27	60	70	82	93
LDA	19	38	46	62	81
Euclidean	4	15	22	40	53

(a)

Method	$r = 1$	10	20	50	100
KISSME [20]	33	**71**	**79**	**90**	97
EIML [17]	**35**	68	77	**90**	**98**
LMNN [30]	29	68	78	**90**	97
LMNN-R [7]	22	59	71	86	95
ITML [6]	24	59	71	87	97
LDML [15]	12	31	39	55	73
Mahalanobis	31	62	73	85	95
LDA	20	46	54	69	86
Euclidean	13	32	41	55	74

(b)

Table 4 CMC scores (in [%]) for *PRID 450S*: (a) without segmentation and (b) with segmentation.

Method	$r = 1$	2	3	4	5	6	7	8
KISSME [20]	**70**	**88**	**95**	**98**	**99**	99	99	**100**
EIML [17]	67	86	92	95	98	99	**100**	**100**
LMNN [30]	43	60	70	81	88	94	98	**100**
ITML [6]	56	76	86	93	97	98	**100**	**100**
LDML [15]	27	46	59	71	81	88	94	**100**
Mahalanobis	55	77	90	95	98	99	**100**	**100**
LDA	37	60	73	83	91	94	98	**100**
Euclidean	28	46	62	71	81	88	94	**100**
ICT [2]	62	81	**95**	97	97	**100**	**100**	**100**

Table 5 CMC scores (in [%]) for *CAVIAR4REID*.

tended to multi-shot. The corresponding CMC scores (due to the small number of samples averaged over 100 runs) are shown in Fig. 7 and Table 5, where we also compare to [2]. Again for all approaches except *LDML* an improvement can be obtained. The higher variance in performance can be explained by the smaller number of training samples, resulting in a higher overfitting tendency.

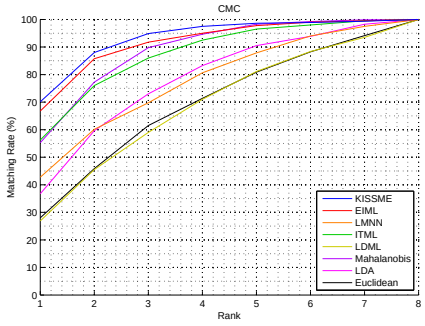


Fig. 7 CMC curves for *CAVIAR4REID*.

5.2 Discussion

The results shown above clearly indicate that metric learning, in general, can drastically boost the performance for single-shot (and even for multi-shot) person re-identification. In fact, by learning a metric we can describe the visual transition from one camera to the other, thus, the applied features do not have to cope with all variabilities, allowing for more meaningful feature matchings. Hence, even if rather simple features are used competitive results can be obtained.

In particular, we used only block-based color and texture descriptions for two reasons. On the one hand side since they are easy and fast to compute and on the other hand side to demonstrate that using even such simple features state-of-the-art or better results can be obtained. However, it is clear that better features, e.g., also exploiting temporal information in a multi-shot scenario will further improve the results.

Surprisingly, even using the standard Mahalanobis distance allows for improving the results and finally yields considerable results. Nevertheless, also incorporating discriminative information yields a further performance gain. However, we have to consider the specific constraints given by the task: (a) images showing the same person might not have a similar visual description whereas (b) images not showing the same person could be very close in the original feature space. Thus, the problem is somehow ill-posed and highly prone to overfitting. This can for instance be recognized for *LDML*, *LMNN*, and *ITML*.

As *LDML* does not use any regularization, it is totally overfitting to the training data and thus yields rather weak results (comparable to the Euclidean distance). The results of *LMNN* are typically better, however, since the impostor handling is not robust against outliers, the problems described above cannot be handled sufficiently. The same applies for *ITML*, which often yields similar results as the original Mahalanobis distance, clearly showing that given somehow "ambiguously labeled" samples no additional discriminative information can be gained. In contrast, *KISSME* and *EIML*, following different strategies, provide some regularization by relaxing the original problem, which seems to be better suited for the given task. Moreover, the metric estimation is computationally much more efficient.

Results on five different datasets showing totally different characteristics clearly demonstrate that metric learning is a general purpose strategy. In fact, the same features were used, only the parameter for PCA was adjusted, which has only a little influence on the results. However, we recognized that for smaller datasets less PCA dimensions are sufficient. The results also indicate the characteristics of the datasets. For *VIPeR* and *CAVIAR4REID* showing a larger variety in the appearance the discriminative power can fully be exploited. For *PRID 2011* and *PRID 450S* containing a larger amount of "similar" instances the improvement from generative to discriminative metric is less significant. Finally, for the *ETHZ* dataset, where the images are taken from the same camera view, metric learning has, as expected, only a little influence.

Thus, if we are given enough data to learn a meaningful metric, metric learning could be highly beneficial in the context of person re-identification. However,

more important than much data is good data. Hence, it would be more meaningful to use temporal information to select good candidates for learning than just using larger amounts of data. Similarly, is was also revealed by the improved results for the *PRID 450S* dataset that using better data (i.e., estimating the metric on the foreground regions only) is beneficial.

6 Conclusion

The goal of this chapter was to analyze the applicability of Mahalanobis metric learning in the context of single-shot person re-identification. We first introduced the main ideas of metric learning and gave an overview on specific approaches addressing the same problem following different paradigms. These were evaluated within a fixed framework on five different benchmark datasets (where one was newly generated). If applicable, we also gave a comparison to the state-of-the-art. Even though some approaches tend to overfit to the training data, we can conclude that metric learning can dramatically boost the classification performance and that even less complex (non-handcrafted) representations could be sufficient for the given task. Moreover, one interesting result is that even a standard Mahalanobis metric not using any discriminative information yields quite good classification results. We also showed that having a perfect segmentation further improves the classification and that it is straight forward to extend the current framework toward multi-shot scenarios. In a similar way also temporal information or a better image representation can be used.

References

1. B. Alipanahi, M. Biggs, and A. Ghodsi. Distance metric learning vs. fisher discriminant analysis. In *Proc. AAAI Conf. on Artificial Intelligence*, 2008.
2. T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch. Learning implicit transfer for person re-identification. In *Proc. ECCV Workshop on Re-Identification*, 2012.
3. S. Bak, E. Corvee, F. Brémond, and M. Thonnat. Person re-idendification using Haar-based and DCD-based signature. In *Workshop on Activity Monitoring by Multi-Camera Surveillance Systems*, 2010.
4. S. Burer and R. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.
5. D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *Proc. British Machine Vision Conf.*, 2011.
6. J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *Proc. Int'l Conf. on Machine Learning*, 2007.
7. M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. In *Proc. Asian Conf. on Computer Vision*, 2010.
8. A. Ess, B. Leibe, and L. V. Gool. Depth and appearance for mobile scene analysis. In *Proc. IEEE Int'l Conf. on Computer Vision*, 2007.

9. M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2010.
10. R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.
11. N. Gheissari, T. B. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006.
12. A. Ghodsi, D. F. Wilkinson, and F. Southey. Improving embeddings by flexible exploitation of side information. In *Proc. Int'l Joint Conf. on Artificial Intelligence*, 2007.
13. D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proc. IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, 2007.
14. D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proc. European Conf. on Computer Vision*, 2008.
15. M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? Metric learning approaches for face identification. In *Proc. IEEE Int'l Conf. on Computer Vision*, 2009.
16. M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *Proc. Scandinavian Conf. on Image Analysis*, 2011.
17. M. Hirzer, P. M. Roth, and H. Bischof. Person re-identification by efficient metric learning. In *Proc. IEEE Int'l Conf. on Advanced Video and Signal-Based Surveillance*, 2012.
18. M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Efficient learning of camera transitions for person re-identification. In *Proc. European Conf. on Computer Vision*, 2012.
19. M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization of the cone of positive semidefinite matrices. *SIAM Journal of Optimization*, 20(5):2327–2351, 2010.
20. M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2012.
21. W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *Proc. Asian Conf. on Computer Vision*, 2012.
22. Z. Lin and L. S. Davis. Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance. In *Advances Int'l Visual Computing Symposium*, 2008.
23. M. Loog, R. P. W. Duin, and R. Haeb-Umbach. Multiclass linear dimension reduction by weighted pairwise fisher criteria. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(7):762–766, 2001.
24. A. Mignon and F. Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2012.
25. F. Porikli. Inter-camera color calibration by correlation model function. In *Proc. Int'l Conf. on Image Processing*, 2003.
26. B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *Proc. British Machine Vision Conf.*, 2010.
27. A. Rahimi, B. Dunagan, and T. Darrell. Simultaneous calibration and tracking with a network of non-overlapping sensors. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.
28. W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. In *Proc. Brazilian Symposium on Computer Graphics and Image Processing*, 2009.
29. X. Wang, G. Doretto, T. B. Sebastian, J. Rittscher, and P. H. Tu. Shape and appearance context modeling. In *Proc. IEEE Int'l Conf. on Computer Vision*, 2007.
30. K. Q. Weinberger and L. K. Saul. Fast solvers and efficient implementations for distance metric learning. In *Proc. Int'l Conf. on Machine Learning*, 2008.
31. W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(3):653–668, 2013.

Index