Sample-Specific SVM Learning for Person Re-identification

Ying Zhang[1], Baohua Li[1], Huchuan Lu[1], Atshushi Irie[2] and Xiang Ruan[2]
[1] Dalian University of Technology, Dalian, 116023, China
[2] OMRON Corporation, Kusatsu, Japan

Abstract

Person re-identification addresses the problem of matching people across disjoint camera views and extensive efforts have been made to seek either the robust feature representation or the discriminative matching metrics. However, most existing approaches focus on learning a fixed distance metric for all instance pairs, while ignoring the individuality of each person. In this paper, we formulate the person re-identification problem as an imbalanced classification problem and learn a classifier specifically for each pedestrian such that the matching model is highly tuned to the individual's appearance. To establish correspondence between feature space and classifier space, we propose a Least Square Semi-Coupled Dictionary Learning (LSSCDL) algorithm to learn a pair of dictionaries and a mapping function efficiently. Extensive experiments on a series of challenging databases demonstrate that the proposed algorithm performs favorably against the state-of-the-art approaches, especially on the rank-1 recognition rate.

1. Introduction

As a fundamental task in multi-camera surveillance system, person re-identification aims to match people observed from different cameras or across different time with a single camera. Although has gained much attention among researchers [11, 8, 33, 44, 16, 28, 19, 32, 42, 29, 40, 2, 20, 15, 24] in recent years, person re-identification remains a challenging problem since a person's appearance can change significantly when large variations in view angle, illumination, background clutter and occlusion are involved.

To address these challenges, a lot of approaches have been proposed to develop robust feature representations which are discriminative for identity, such as Ensemble of Localized Features (ELF) [11], Symmetry-Driven Accumulation of Local Features (SDALF) [8], Covariance descriptor based on Bio-inspired Features (gBiCov) [29] and Local Maximal Occurrence (LOMO) [20].

On the other hand, there are many efforts attempting to learn optimal matching metrics under which instances belonging to the same person are closer than different persons, like Probabilistic Relative Distance Comparison (PRDC) [44], Keep It Simple and Straightforward Metric Learning (KISSME) [16], Local Fisher Discriminant Analysis (LFDA) [32], Cross-view Quadratic Discriminant Analysis (XQDA) [20], etc. Taking the person re-identification as a relative ranking problem, some researchers [33, 24] employ the Support Vector Machine (SVM) model to learn a ranking function such that the scores of matched image pairs are larger than unmatched ones.

However, most of the existing methods focus on learning a fixed distance metric or ranking function to measure the similarity between all images, while ignoring the fact that different instances have different feature representations, and the metric or function derived for matching all objects may not be optimal for every single person. Since the primary goal of person re-identification is to seek the optimal match for each pedestrian, a sample-specific distance metric or ranking function should be more investigated. Zheng *et al*. [43] utilized the initial rank scores of each sample to compute the adaptive weights for feature fusion. The Query based Adaptive Re-Ranking (QARR) algorithm [27] was developed to learn a weighted combination of a base score function and a perturbation linear function for each query. Nevertheless, both of the query-adaptive methods learn the new weighting scheme in the re-ranking stage, and the effectiveness of the model may be easily affected by the initial matching results.

In this paper, we propose a novel framework for person re-identification, where a sample-specific SVM is learned for each pedestrian to seek the optimal match. The matching function parameterized by the classifier weight vector is highly tuned to the individual's appearance, which can provide more discriminative measurements for finding the best candidate. To investigate the intrinsic relationship between the feature space and weight space, we propose a Least Square Semi-Coupled Dictionary Learning (LSSCDL) algorithm to learn a dictionary pair and a mapping function simultaneously, through which the weight parameters of a new sample can be easily inferred by its feature patterns. Figure 1 shows the overview of our approach.

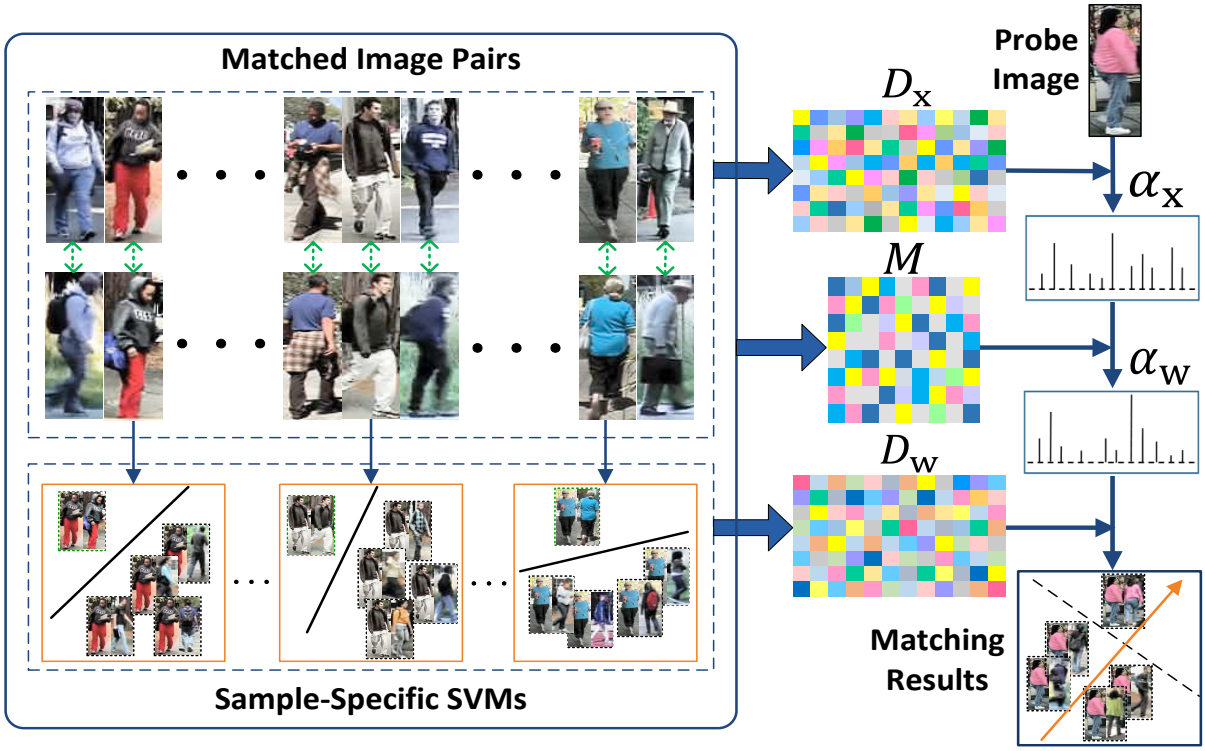


Figure 1. Overview of the proposed approach for person re-identification. We first learn matching classifiers for every training individuals by sample-specific SVMs. The classifier weight vectors are used to further learn a pair of dictionary and a mapping matrix, by which a weight vector of a test probe image can be easily inferred from its feature representation. The re-identification is then performed based on our proposed matching criterion with the learned weight vector.

2. Related Work

2.1. Mainstream Methods for Re-identification

Most existing approaches to tackle the person re-identification problem are mainly carried on from two aspects: developing distinctive feature representations and seeking discriminative distance metrics. Both of them aim to compute the matching distances (or scores) which are optimal for matched image pairs from the gallery and probe set respectively.

For feature representation, a number of approaches [11, 8, 29, 40, 24, 20, 4, 35, 17] have been proposed to design robust descriptors against background and illumination variations. For instance, Gray *et al*. [11] presented the ELF by fusing 8 color channels with 19 texture channels, while Farenzena *et al*. [8] employed the weighted color histograms, Maximally Stable Color Regions (MSCR), Recurrent High-Structured Patches (RHSP) to capture different image properties. Ma *et al*. [29] proposed an image representation based on the combination of Biologically Inspired Features (BIF) and Covariance descriptors. LOMO [20] extracted the maximal pattern of joint HSV color histogram coupled with Scale Invariant Local Ternary Pattern (SILTP) [22], and it is worth mentioning that LOMO based pedestrian representation has shown impressive robustness against viewpoint changes. Chen *et al*. [4] proposed a zero-padding based feature transformation strategy to enables alignment of the feature distributions across disjoint views, which can significantly enhance the performance of existing matching models. Shi *et al*. [35] learned mid-level semantic attributes such as hair-style, shoe-type or clothing-style to achieve more powerful representation.

For metric learning, numerous research works [6, 12, 7, 44, 31, 16, 14, 19, 32, 20] aim to learn a metric matrix under which the distance between images of the same pedestrian is smaller than different ones. Zheng *et al*. [44] proposed the PRDC based method where the probability of a pair of true match having a smaller distance than that of a wrong match pair is maximized. Pedagadi *et al*. [32] employed the LFDA algorithm to maximize the inter-class separability while preserving the multiclass modality. Li *et al*. [19] developed the Locally-Adaptive Decision Functions (LADF), which combines the distance metric with a locally adaptive thresholding rule for each pair of sample images. KISSME [16] derived a Mahalanobis metric by computing the difference between the intra-class and inter-class covariance matrix. As an improvement, XQDA [20] learned a more discriminative distance metric and a low-dimensional subspace simultaneously. Although has achieved inspiring re-identification

reults, these methods did not give sufficient consideration to the individuality of each pedestrian when learning a generic distance metric for all instances. In this paper we propose to learn a matching metric specifically for every person, such that each individual could have the matching model that best suits his or her appearance.

2.2. SVM learning for Re-identification

Some researchers formulated the person re-identification problem as a ranking problem and managed to learn a ranking function parameterized by a weight vector to order relevant images pairs before irrelevant ones. Prosser *et al*. [33] developed the Ensemble RankSVM where a set of weak RankSVMs were learned on different subsets and then combined into a stronger ranker using a boosting principle. In [24], the structural SVM was employed to score correct matches higher than all incorrect ones by a margin.

Although based on SVM learning, our approach differs substantially from these two methods. Instead of computing a fixed weight vector for all pedestrians, we learn specific weight parameters such that the ranking function is highly tuned to the individual's appearance. Furthermore, the previous methods employ the relative ranking relationships for SVM learning, while the proposed algorithm tackles person re-identification as an absolute classification problem, which greatly enlarges the gap between matched and unmatched image pairs.

2.3. Dictionary Learning Methods

SCDL The Semi-Coupled Dictionary Learning (SCDL) algorithm [38] was originally intended to solve the cross-style image synthesis problems, and it assumed that there exists a dictionary pair over which the coding coefficients of the two representations have a stable mapping.

SSCDL To bridge the appearance variations across cameras, Liu *et al*. [23] developed the Semi-Supervised Coupled Dictionary Learning (SSCDL) algorithm where the coupled dictionaries for gallery and probe images are learned jointly.

$\mathbf{SLD^2L}$ Based on the observation that gallery images are high-resolution (HR) while probe images are low-resolution (LR), Jing *et al*. [15] proposed the Semi-coupled Low-rank Discriminant Dictionary Learning (SLD^2L) algorithm to learn a pair of HR and LR dictionaries and a matrix to map the feature from LR to HR.

However, the SCDL model is designed for photo-sketch synthesis and requires large time consumption to solve the sparse coding problem, therefore we propose the LSSCDL algorithm to solve the cross-modal problem with higher efficiency. In contrast with SSCDL and SLD^2L which directly learn a dictionary pair for two camera views or two resolutions, the LSSCDL model attempts to investigate the intrinsic relationship between feature patterns and ranking parameters.

3. The Proposed Algorithm

3.1. Sample-Specific SVM Learning

Given the probe set $\mathcal{D}^p = \{\mathbf{x}_i^p\}_{i=1}^N$ and the gallery set $\mathcal{D}^g = \{\mathbf{x}_j^g\}_{j=1}^N$, we respectively denote i and j to be the identity label of pedestrians from the two groups. A probe-gallery image pair with a matching label is constructed as $\{(\mathbf{x}_i^p, \mathbf{x}_j^g), y_j^p\}$, where $y_j^p = +1$ represents that $(\mathbf{x}_i^p, \mathbf{x}_j^g)$ is a correct matching pair, while $y_j^p = -1$ indicates the incorrect matches.

We explicitly consider the problem of person re-identification as a binary classification problem. Given a probe image $\mathbf{x}_i^p$, we attempt to learn a sample-specific classifier $\mathcal{F}_i^p$ on the probe-gallery set $\{(\mathbf{x}_i^p, \mathbf{x}_1^g), ..., (\mathbf{x}_i^p, \mathbf{x}_N^g)\}$ such that

$$\mathcal{F}_i(\mathbf{x}_i^p, \mathbf{x}_j^g) = \begin{cases} \geq 0, \ y_j^p = +1 \\ < 0, \ y_j^p = -1 \end{cases} \quad j = 1, ..., N \tag{1}$$

We define the classification function with the following form

$$\mathcal{F}_i(\mathbf{x}_i^p, \mathbf{x}_j^g) = \mathbf{w}_i^p \cdot \phi(\mathbf{x}_i^p, \mathbf{x}_j^g) + b_i \tag{2}$$

where $\mathbf{w}_i^p$ denotes the weight vector of $\mathbf{x}_i^p$ and b_i is the bias. $\phi(\mathbf{x}_i^p, \mathbf{x}_j^g)$ is defined as a feature map of the image pair with the following form

$$\phi(\mathbf{x}_i^p, \mathbf{x}_j^g) = [(\mathbf{x}_i^p)^\top, |\mathbf{x}_i^p - \mathbf{x}_j^g|^\top, (\mathbf{x}_j^g)^\top]^\top \tag{3}$$

where $\left|\mathbf{x}_i^p - \mathbf{x}_j^g\right| = (|\mathbf{x}_i^p(1) - \mathbf{x}_j^g(1)|, ..., |\mathbf{x}_i^p(d) - \mathbf{x}_j^g(d)|)^\top$ is the absolute difference vector [44] and d is the feature dimension. This feature map not only takes the image difference into consideration, but also exploits the nature of image itself to enhance the distinctiveness of an image pair.

The traditional SVM model [5] is employed to solve the binary classification problem, where the relevant image pairs are separated from all the irrelevant ones by the largest possible margin. Consider that the number of correct matches (positive set) is much smaller than incorrect ones (negative set), we impose different penalty parameters [36] to handle the imbalance. Learning the sample-specific classifier is equivalent to optimizing the following objective function:

$$\begin{aligned} \min_{\mathbf{w}_i^p} \frac{1}{2} \parallel \mathbf{w}_i^p \parallel^2 + C^+ \sum_{y_j^p = +1} \xi_j + C^- \sum_{y_j^p = -1} \xi_j \\ s.t. \ y_j^p(\mathbf{w}_i^p \cdot \phi(\mathbf{x}_i^p, \mathbf{x}_j^g) + b_i) \geq 1 - \xi_j, \\ \xi_j \geq 0, \ j = 1, ..., N \end{aligned} \tag{4}$$

where $C^+ > 0$ and $C^- > 0$ are regularization parameters for positive and negative classes, respectively, and ξ_j is the slack variable.

Generally, the linear SVM model has good interpretability in the sense that it seeks a direction that can explain the biggest difference between the two classes. Therefore, the binary classification strategy for pedestrian matching actually attempts to find the weight vector maximally enlarging

the gap between matched and unmatched pairs to improve the rank-1 recognition rate [42], which is consistent with the primary goal of person re-identification. Moreover, the positive image pair is separated from its corresponding negative pairs in each sample-specific classifier, under which the instance-level information is effectively utilized to make the matching model highly tuned to the individual's appearance, leading to more powerful discrimination.

Note that the weight vector $\mathbf{w}_i^p$ plays a key role in ordering the relevant image pairs before irrelevant ones, here we define a score function without changing the ranks of matching candidates by omitting the bias in (2).

$$f_i(\mathbf{x}_i^p, \mathbf{x}_j^g) = \mathbf{w}_i^p \cdot \phi(\mathbf{x}_i^p, \mathbf{x}_j^g) \tag{5}$$

where $f_i(\mathbf{x}_i^p, \mathbf{x}_j^g)$ denotes the matching score of $\mathbf{x}_i^p$ and $\mathbf{x}_j^g$. The higher the score is, the more likely that the two images represent the same person. In the following sections we only focus on the discussion of weight vectors.

3.2. Least Square Semi-Coupled Dictionary Learning

Since there is a one-to-one correspondence between the feature vector and weight vector, it is reasonable to assume that the feature patterns and weight parameters of a specific person should have similar intrinsic structures, and there exists a mapping function through which one type of the representation can be inferred by the other. It is hard to define the mapping function between the two styles of representation directly, while the linear reconstruction pattern of each pair of samples in their respective space can be related to some extent. As suggested in [38], for two different styles of representations indicating the same scene, there exist coupled dictionaries over which the coding coefficients of two styles have a stable mapping.

In this paper, we propose a Least Square Semi-Coupled Dictionary Learning (LSSCDL) algorithm to learn a pair of dictionaries and a mapping function efficiently, where the two dictionaries respectively depict the intrinsic structures of the feature space and weight space, and the mapping function characterizes the relationship between the two spaces.

Given the training probe set $\mathbf{X}^p = (\mathbf{x}_1^p, \mathbf{x}_2^p, ..., \mathbf{x}_N^p) \in \mathbb{R}^{d\times N}$ with each column representing a probe image, and the corresponding learned weight set $\mathbf{W}^p = (\mathbf{w}_1^p, \mathbf{w}_2^p, ..., \mathbf{w}_N^p) \in \mathbb{R}^{3d\times N}$, we denote $\mathbf{D}_\mathrm{x} \in \mathbb{R}^{d\times k}$, $\mathbf{D}_\mathrm{w} \in \mathbb{R}^{3d\times k}$ and $\mathbf{M} \in \mathbb{R}^{k\times k}$ to be the feature dictionary, the weight dictionary, and the mapping matrix, respectively. Here k indicates the dictionary size. Then the problem of jointly optimizing the dictionaries and mapping function can be formulated as follows.

$$\min_{\{\mathbf{D}_\mathrm{x}, \mathbf{D}_\mathrm{w}, \mathbf{M}\}} \Phi(\mathbf{D}_\mathrm{x}, \mathbf{D}_\mathrm{w}, \mathbf{M}, \mathbf{\Lambda}_\mathrm{x}, \mathbf{\Lambda}_\mathrm{w}) \tag{6}$$

with

$$\begin{aligned}\Phi = &E_{data}(\mathbf{D}_\mathrm{x}, \mathbf{X}^p) + E_{data}(\mathbf{D}_\mathrm{w}, \mathbf{W}^p) \\ &+ E_{map}(\mathbf{M}) + E_{reg}(\mathbf{\Lambda}_\mathrm{x}, \mathbf{\Lambda}_\mathrm{w}, \mathbf{M}, \mathbf{D}_\mathrm{x}, \mathbf{D}_\mathrm{w})\end{aligned} \tag{7}$$

where $\mathbf{\Lambda}_\mathrm{x}$ and $\mathbf{\Lambda}_\mathrm{w}$ denote the coding coefficients. $E_{data}(\cdot,\cdot)$ is the representation fidelity term indicating the reconstruction error, $E_{map}(\cdot)$ is the mapping fidelity term to represent the mapping error between the coding coefficients, E_{reg} is the regularization term to regularize the coding coefficients, mapping matrix and dictionaries.

Following the assumption that dictionaries are overcomplete, many previous methods [38, 15] impose ℓ_1-norm regularization on coding coefficients to select a few atoms of the learned dictionary to describe a sample. In person re-identification, however, feature dimension is usually much larger than the number of samples, and the sparse representation may be insufficient to capture the correlation structure of data with large variations. Furthermore, it is not efficient to solve the ℓ_1-minimization problem for each instance. To address these issues, we present the Least Square Semi-Coupled Dictionary Learning (LSSCDL) algorithm with the following form.

$$\begin{aligned}\min_{\{\mathbf{D}_\mathrm{x}, \mathbf{D}_\mathrm{w}, \mathbf{M}\}} &\parallel \mathbf{X}^p - \mathbf{D}_\mathrm{x}\mathbf{\Lambda}_\mathrm{x} \parallel_F^2 + \parallel \mathbf{W}^p - \mathbf{D}_\mathrm{w}\mathbf{\Lambda}_\mathrm{w} \parallel_F^2 \\ &+\lambda \parallel \mathbf{\Lambda}_\mathrm{w} - \mathbf{M}\mathbf{\Lambda}_\mathrm{x} \parallel_F^2 + \lambda_\Lambda \parallel \mathbf{\Lambda}_\mathrm{x} \parallel_F^2 + \lambda_\Lambda \parallel \mathbf{\Lambda}_\mathrm{w} \parallel_F^2 \\ &+\lambda_\mathrm{M} \parallel \mathbf{M} \parallel_F^2 + \lambda_\mathrm{D} \parallel \mathbf{D}_\mathrm{x} \parallel_F^2 + \lambda_\mathrm{D} \parallel \mathbf{D}_\mathrm{w} \parallel_F^2\end{aligned} \tag{8}$$

where $\|\cdot\|_F$ indicates the Frobenius norm, λ, λ_Λ, λ_D, and λ_M are regularization parameters to balance the terms in the objective function.

Note that (8) is a non-convex optimization problem with all the matrices $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$, $\mathbf{M}$, $\mathbf{\Lambda}_\mathrm{x}$, $\mathbf{\Lambda}_\mathrm{w}$, but it is convex with respect to each of the variables when the others are fixed. Therefore the optimization problem can be solved by conducting the following steps iteratively until convergency.

1. Fix $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$, $\mathbf{M}$, $\mathbf{\Lambda}_\mathrm{w}$, let $\frac{\partial\Phi}{\partial\mathbf{\Lambda}_\mathrm{x}} = 0$, we have

$$\mathbf{\Lambda}_\mathrm{x} = (\mathbf{D}_\mathrm{x}^\top\mathbf{D}_\mathrm{x} + \lambda\mathbf{M}^\top\mathbf{M} + \lambda_\Lambda\mathbf{I})^{-1}(\mathbf{D}_\mathrm{x}^\top\mathbf{X}^p + \lambda\mathbf{M}^\top\mathbf{\Lambda}_\mathrm{w}) \tag{9}$$

2. Fix $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$, $\mathbf{M}$, $\mathbf{\Lambda}_\mathrm{x}$, let $\frac{\partial\Phi}{\partial\mathbf{\Lambda}_\mathrm{w}} = 0$, we obtain

$$\mathbf{\Lambda}_\mathrm{w} = (\mathbf{D}_\mathrm{w}^\top\mathbf{D}_\mathrm{w} + (\lambda + \lambda_\Lambda)\mathbf{I})^{-1}(\mathbf{D}_\mathrm{w}^\top\mathbf{W}^p + \lambda\mathbf{M}\mathbf{\Lambda}_\mathrm{x}) \tag{10}$$

3. Fix $\mathbf{\Lambda}_\mathrm{x}$, $\mathbf{\Lambda}_\mathrm{w}$, $\mathbf{M}$, let $\frac{\partial\Phi}{\partial\mathbf{D}_\mathrm{x}} = 0$ and $\frac{\partial\Phi}{\partial\mathbf{D}_\mathrm{w}} = 0$, we get

$$\mathbf{D}_\mathrm{x} = \mathbf{X}^p\mathbf{\Lambda}_\mathrm{x}^\top(\mathbf{\Lambda}_\mathrm{x}\mathbf{\Lambda}_\mathrm{x}^\top + \lambda_\mathrm{D}\mathbf{I})^{-1} \tag{11}$$

$$\mathbf{D}_\mathrm{w} = \mathbf{W}^p\mathbf{\Lambda}_\mathrm{w}^\top(\mathbf{\Lambda}_\mathrm{w}\mathbf{\Lambda}_\mathrm{w}^\top + \lambda_\mathrm{D}\mathbf{I})^{-1} \tag{12}$$

4. Fix $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$, $\mathbf{\Lambda}_\mathrm{x}$, $\mathbf{\Lambda}_\mathrm{w}$, let $\frac{\partial\Phi}{\partial\mathbf{M}} = 0$, we have

$$\mathbf{M} = \mathbf{\Lambda}_\mathrm{w}\mathbf{\Lambda}_\mathrm{x}^\top(\mathbf{\Lambda}_\mathrm{x}\mathbf{\Lambda}_\mathrm{x}^\top + \frac{\lambda_\mathrm{M}}{\lambda}\mathbf{I})^{-1} \tag{13}$$

The optimization algorithm for solving (8) is summarized in Algorithm 1.

We can see from (8) that the learned dictionaries $\mathbf{D}_\mathrm{x}$ and $\mathbf{D}_\mathrm{w}$ transfer the representations from two different spaces into a common coding space, and the coding coefficients of $\mathbf{X}^p$ and $\mathbf{W}^p$ can be related by the mapping matrix $\mathbf{M}$. Therefore, the weight vector for a new sample can be easily inferred from its feature pattern with the learned dictionary pair and mapping function.

Algorithm 1: The Optimization of LSSCDL

Input: probe image matrix $\mathbf{X}^p$, weight matrix $\mathbf{W}^p$, parameters λ, λ_Λ, λ_D, and λ_M.
Output: feature dictionary $\mathbf{D}_\mathrm{x}$, weight dictionary $\mathbf{D}_\mathrm{w}$, mapping matrix $\mathbf{M}$.
Initialize: $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$, $\mathbf{M}$, $\mathbf{\Lambda}_\mathrm{x}$ and $\mathbf{\Lambda}_\mathrm{w}$.
Repeat:
1: Fix $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$, $\mathbf{M}$, $\mathbf{\Lambda}_\mathrm{w}$, update $\mathbf{\Lambda}_\mathrm{x}$ by (9).
2: Fix $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$, $\mathbf{M}$, $\mathbf{\Lambda}_\mathrm{x}$, update $\mathbf{\Lambda}_\mathrm{w}$ by (10)
3: Fix $\mathbf{\Lambda}_\mathrm{x}$, $\mathbf{\Lambda}_\mathrm{w}$, $\mathbf{M}$, update $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$ by (11) and (12).
4: Fix $\mathbf{D}_\mathrm{x}$, $\mathbf{D}_\mathrm{w}$, $\mathbf{\Lambda}_\mathrm{x}$, $\mathbf{\Lambda}_\mathrm{w}$, update $\mathbf{M}$ by (13).
Until: convergency.

3.3. Pedestrian Matching

Assuming that we have the test probe set and test gallery set denoted as $\mathcal{T}^p = \{\mathbf{x}_t^p\}_{t=1}^M$ and $\mathcal{T}^g = \{\mathbf{x}_{t'}^g\}_{t'=1}^{M'}$, where t and t' indicate the identity label, respectively. Given a test probe image $\mathbf{x}_t^p$, the corresponding weight vector $\mathbf{w}_t^p$ can be derived with the learned dictionary pair $\mathbf{D}_\mathrm{w}$, $\mathbf{D}_\mathrm{x}$ and mapping matrix $\mathbf{M}$.

We first compute the coding coefficients $\boldsymbol{\alpha}_\mathrm{x}$ of $\mathbf{x}_t^p$ by solving the following problem.

$$\min_{\{\boldsymbol{\alpha}_\mathrm{x}\}} \parallel \mathbf{x}_t^p - \mathbf{D}_\mathrm{x}\boldsymbol{\alpha}_\mathrm{x} \parallel_F^2 + \lambda_\Lambda \parallel \boldsymbol{\alpha}_\mathrm{x} \parallel_F^2 \tag{14}$$

then the coding vector $\boldsymbol{\alpha}_\mathrm{w}$ of $\mathbf{w}_t^p$ is derived by

$$\boldsymbol{\alpha}_\mathrm{w} = \mathbf{M}\boldsymbol{\alpha}_\mathrm{x} \tag{15}$$

and the weight vector $\mathbf{w}_t^p$ of $\mathbf{x}^p$ can be reconstructed by

$$\mathbf{w}_t^p = \mathbf{D}_\mathrm{w}\boldsymbol{\alpha}_\mathrm{w} \tag{16}$$

Finally, we compute the matching score of a test prob-gallery image pair $(\mathbf{x}_t^p, \mathbf{x}_{t'}^g)$ with (5).

4. Experimental Results

In this section, we show the performance of the proposed person re-identification algorithm on the VIPER datatset [10], the QMUL GRID dataset [26], the PRID 450S dataset [34], the CUHK01 dataset [42], the CUHK03 dataset [18] and the OpeRID dataset [21]. Comparisons of the Cumulative Matching Characteristic (CMC) [10] results demonstrate that our approach performs favorably against other state-of-the-art methods, especially on the rank-1 recognition rate.

4.1. Feature Extraction and Parameter Settings

Feature Extraction For each image, we extract the LOMO descriptors to represent the human appearance. The LOMO extractor has shown impressive robustness against viewpoint changes and illumination variations by concatenating the maximal pattern of joint HSV histogram and SILTP descriptor. Consider that the dimensionality of LOMO is extremely high, we employ [20] for dimensionality reduction, which can greatly save the time and memory.

Parameter Settings There are 7 parameters in our approach, C^+, C^-, k, λ, λ_Λ, λ_M and λ_D. In our experiments, we set $C^+ = 300$, $C^- = 0.1C^+$, $k = N$, $\lambda = 0.1$, $\lambda_\Lambda = 0.01$, $\lambda_\mathrm{M} = 0.01$, $\lambda_\mathrm{D} = 0.01$ for all the databases. We find that our experimental results are not sensitive to parameter changes, and please refer to the supplementary material for more details.

4.2. VIPeR Dataset

The VIPER datatset [10] contains 632 pairs of pedestrian images captured from two different cameras in outdoor academic environment, with only one image per person in each view and all the images normalized to 128×48 pixels. Suffering from significant viewpoint changes, pose variation, and illumination difference across cameras, it is one of the most challenging database for person re-identification.

Following the experimental protocol of [20], we randomly divide all the pedestrian pairs into two equal parts, with one part for training and the other for testing. These procedures are repeated for 10 trails and the average matching rates are summarized in Table 1. We can see that the proposed approach achieves comparable performance with other methods. Although the matching rates are a little inferior to MirrorRep [4] and Semantic [35] , the proposed approach achieves the second best rank-1 recognition rate of 42.66%.

Table 1. Comparison of state-of-the-art results on the VIPeR dataset (P=316). The cumulative matching scores (%) at rank 1, 10, and 20 are listed.

Method	rank=1	rank=10	rank =20
Ours	42.66	84.27	91.93
MirrorRep [4]	**42.97**	**87.28**	94.84
Semantic [35]	41.60	86.20	**95.10**
LOMO+XQDA [20]	40.00	80.51	91.08
IDLA [1]	34.81	75.63	84.49
PolyMap [2]	36.80	83.70	91.70
$\mathrm{SLD}^2\mathrm{L}$ [15]	16.86	58.06	79.00
ECM [24]	38.90	78.40	88.90
QALF [43]	30.17	62.44	73.81
QARR-RSVM [27]	22.53	62.20	75.82
SCNCD [40]	37.80	81.20	90.40
gBiCov [29]	31.11	70.71	82.45
Mid-level Filter [42]	29.11	65.95	79.87
MtMCML [30]	28.83	75.82	88.51
SSCDL [23]	25.60	68.10	83.60
LADF [19]	30.22	78.92	90.44
SalMatch [41]	30.16	65.54	79.15
KISSME [16]	19.60	62.20	77.00
PCCA [31]	19.27	64.91	80.28
PRDC [44]	15.66	53.86	70.09
SDALF [8]	19.87	49.37	65.73
RankSVM [33]	14.00	51.00	67.00
ELF [11]	12.00	44.00	61.00

4.3. QMUL GRID Dataset

The QMUL GRID dataset [26] consists of person images recorded from 8 disjoint cameras installed in an underground station. The probe set contains 250 pedestrians, with each one having a matching image in the gallery set. Besides, there are 775 additional images in the gallery set that do not match any person in the probe set, which increases the difficulty of seeking the optimal match for each probe image. We normalize all the images to 128×48 pixels, and adopt the experimental setting of 10 random trials for this dataset. In each trial, we randomly select 125 image pairs for training and use the remaining 125 pairs coupled with 775 irrelevant images for testing.

Table 2 compares the matching rates of our approach with previous methods. The comparison shows that the proposed algorithm improves the rank-1 recognition rate from 16.56% to 22.40% and produces about 10 percents improvement on rank-10 and rank-20 matching rate, showing significant advantage in person re-identification.

Table 2. Comparison of state-of-the-art results on the QMUL GRID database (P=900). The cumulative matching scores (%) at rank 1, 10, and 20 are listed.

Method	rank=1	rank=10	rank =20
Ours	**22.40**	**51.28**	**61.20**
LOMO+XQDA [20]	16.56	41.84	52.40
PolyMap [2]	16.30	46.00	57.60
MtMCML [30]	14.08	45.84	59.84
LCRML [3]	10.68	35.04	46.48
MRank-PRDC [25]	11.12	35.76	46.56
MRank-RankSVM [25]	12.24	36.32	46.56
PRDC [44]	9.68	32.96	44.32
RankSVM [33]	10.24	33.28	43.68

4.4. PRID 450S Dataset

The PRID 450S dataset [34] includes 450 single-shot pedestrian image pairs captured from two disjoint camera views. It is also a challenging person re-identification dataset due to the background interference, partial occlusion and viewpoint changes. In our experiments, all the images are normalized to the size of 128×64 pixels. We randomly select half of the dataset for training and the remaining for testing, and repeat the procedures for 10 times to report the average performance.

We also implement the LOMO+XQDA [20] algorithm on the PRID 450S dataset under the same protocol, and the results comparison are summarized in Table 3, from which one can see that the proposed approach performs well against the existing methods and achieves the second best one of 60.49% on the rank-1 recognition rate, showing competitive performance on this dataset.

Table 3. Comparison of state-of-the-art results on the PRID 450S database (P=225). The cumulative matching scores (%) at rank 1, 10, and 20 are listed.

Method	rank=1	rank=10	rank =20
Ours	60.49	88.58	93.60
LOMO+XQDA [20]	**61.42**	**90.84**	**95.33**
MirrorRep [4]	55.42	87.82	93.87
Semantic [35]	44.90	77.50	86.70
ECM [24]	41.90	76.90	84.90
SCNCD [40]	26.90	64.20	74.90
KISSME [16]	33.00	71.00	79.00
EIML [13]	35.00	68.00	77.00
ELF [11]	30.60	73.60	84.20

4.5. CUHK01 Dataset

The CUHK01 Dataset [42] is captured in a campus environment with two camera views. It contains 971 individuals and each of them has two images in every camera view. Taking the evaluation method in [42], we normalized all the images to 160×60 pixels, and conduct the experiments over 10 random partitions for this dataset, where 485 persons are randomly sampled for training and the rest are utilized for testing.

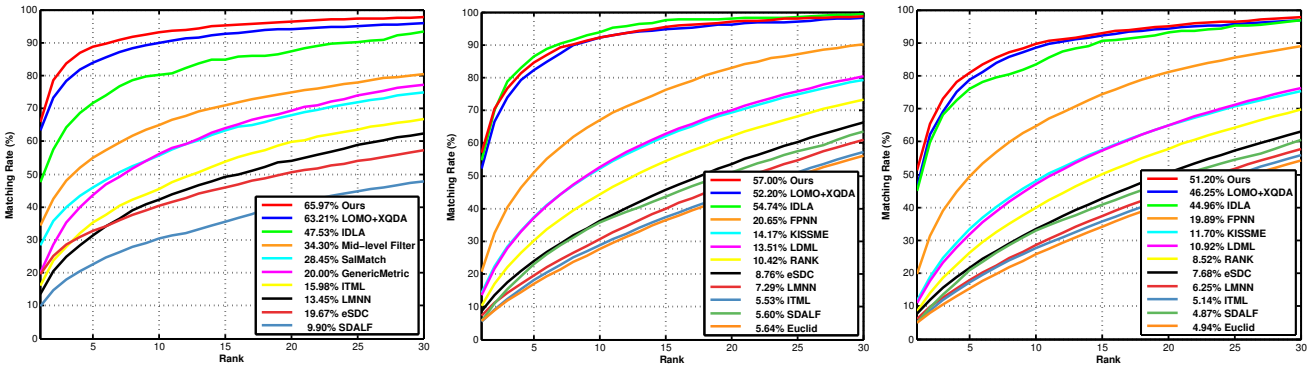
Figure 2 (a) plots the CMC curves of our method and existing state-of-the-art algorithms. Our approach reports the best rank-1 recognition rate of 65.97%, with an improvement of 2.76% over LOMO+XQDA [20].

4.6. CUHK03 Dataset

The CUHK03 dataset [18] consists of 13,164 images of 1,360 pedestrians captured with six surveillance cameras. Each individual is observed by two disjoint camera views, and there are 4.8 images on average for each identity in each view. Apart from the manually labeled pedestrian bounding boxes, this database also provides the samples detected with a pedestrian detector [9], which causes some misalignments and body part missing for a more realistic setting.

Following the experimental settings in [18], we partition the dataset into a training set of 1,160 and a test set of 100 persons. All the experiments are conducted with 20 random splits and the average results are presented.

Figure 2 (b) plots the CMC curves of all the methods on the CUHK03 dataset with the labeled bounding boxes. we can see that proposed algorithm achieves comparable results with XQDA [20], and IDLA [1] reports the best performance from rank-2 to rank-30. The best rank-1 recognition rate reported to date is 54.74%, while we have achieved 57.00% with an improvement of 2.26%. Figure 2 (c) compares the performance of our approach with other state-of-the art methods using automatically detected bounding boxes. Although the performance on detected CUHK03 is inferior to labeled CUHK03 due to the misalignment and in-



(a) CUHK01 (b) Labeled CUHK03 (c) Detected CUHK03

Figure 2. CMC curves and rank-1 identification rates on different datasets.

completeness caused by the detector, the proposed algorithm still shows great advantages over previous method, with a rank-1 recognition rate of 51.20% compared with the second best one of 46.25%.

4.7. OPeRID Dataset

The OpeRID dataset [21] contains 7,413 images of 200 persons collected from a real outdoor surveillance scenario with a network setting of 6 cameras. Each of the person has at least 2 associated camera views and a person may have up to 5 camera views. Lighting changes, viewpoint variations, low resolution and blur can be observed from the images, which are segmented from a interactive labeling software [37]. According to the experimental settings and evaluation protocols in [21], we scale all the images to 128×48 pixels, calculate the detection and identification rates (DIR) and false accept rate (FAR) for evaluation. All the procedures are repeated for 10 trials, and the average performance are reported.

Table 4 summarizes the DIR values at rank 1, 10 under FAR=1% and FAR=10%. We can see that the proposed algorithm is a little inferior to the rate of 3.99% and 4.35% of RRDA [21] at FAR=1%, which achieves the best performance of 15.08% and 18.17% at FAR=10%. The DIR values on open-set dataset are pretty low for all the methods, and there are still much work to do for real applications.

Table 4. Comparison of Detection and identification rates (%) on the OPeRID dataset. The DIR values at rank 1, 10 under FAR=1% and FAR=10% are listed.

	FAR=1%		FAR=10%	
	rank=1	rank=10	rank =1	rank =10
Ours	3.15	3.76	**15.08**	**18.17**
RRDA [21]	**3.99**	**4.35**	14.51	16.72
LADF [19]	1.53	1.74	9.11	10.82
KISSME [16]	1.82	1.92	9.99	11.46
MAHAL [16]	1.89	1.99	10.50	11.97
ITML [6]	1.18	1.21	8.39	9.27
LMNN [39]	0.41	0.41	3.97	4.58

5. Comparative Experiments

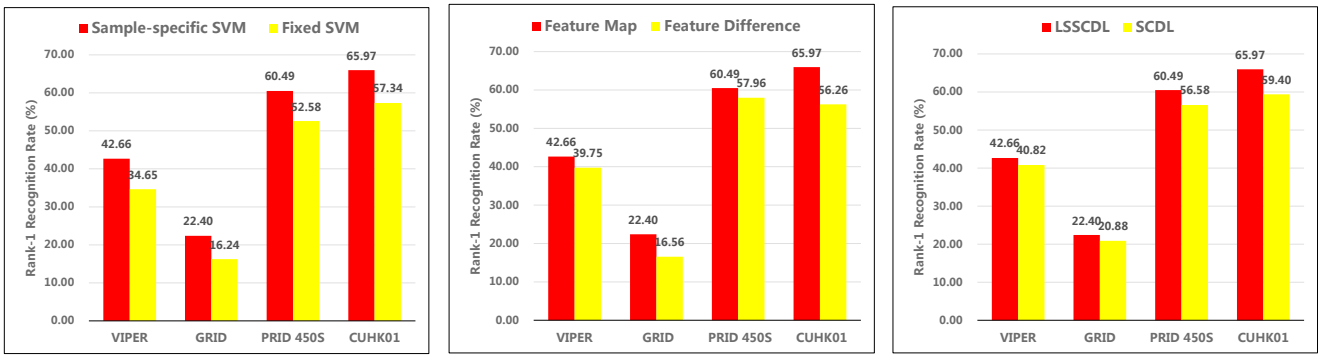
To better understand the function of each part in the proposed algorithm, we conducted further comparative experiments in two aspects: sample-specific SVM vs fixed SVM, feature map vs feature difference and LSSCDL vs SCDL. The complexity analysis in terms of running time and convergence performance are also presented in this section.

5.1. Sample-specific SVM vs Fixed SVM

We compare the results of our approach with the method of learning a fixed SVM. Specifically, the second method employs all matched and unmatched pairs as positive and negative class respectively, and learns a common weight vector for all pedestrians. Figure 3 (a) shows the rank-1 recognition rate comparisons of the two methods. From the results one can easily confirm that learning specific weight vector for each sample significantly outperforms the method of learning fixed weight parameters, by improving the rank-1 accuracy from 34.64% to 42.66% on the VIPER database, 16.24% to 22.40% on the GRID database, 52.58% to 60.49% on the PRID 450S database, and 57.34% to 65.97% on the CUHK01 database. This comparison demonstrates that by taking the appearance individuality into consideration, the proposed algorithm can learn more optimal ranking function for each pedestrian.

5.2. Feature Map vs Feature Difference

To demonstrate the effectiveness of the proposed feature map $\phi(\mathbf{x}_i^p, \mathbf{x}_j^g) = [(\mathbf{x}_i^p)^\top, |\mathbf{x}_i^p - \mathbf{x}_j^g|^\top, (\mathbf{x}_j^g)^\top]^\top$, we compare the experimental results with feature map $\phi(\mathbf{x}_i^p, \mathbf{x}_j^g)$ and feature difference $\phi'(\mathbf{x}_i^p, \mathbf{x}_j^g) = |\mathbf{x}_i^p - \mathbf{x}_j^g|$. From Figure 3 (b) we can see that, the re-identification performance can be obviously improved by the proposed feature map, especially with a great improvement of 9.71% on the CUHK01 dataset. This indicates that the proposed feature map is able to learn more distinctive weight parameters by exploiting both the difference information and natural characters of the pairwise feature representation.



(a) (b) (c)

Figure 3. Rank-1 recognition rate comparison of (a)Sample-specific SVM vs Fixed SVM; (b) Feature Map vs Feature Difference; (c) LSSCDL vs SCDL.

5.3. LSSCDL vs SCDL

In our framework, the LSSCDL algorithm aims to capture the intrinsic relationship between feature space and ranking function space. It can be seen as an improved version of SCDL [38] in person re-identification tasks for higher efficiency. The rank-1 recognition rate comparisons of the two dictionary learning strategies are shown in Figure 3 (c), from which one can see that SCDL achieves a rank-1 rate of 40.82% on VIPER, 20.88% on GRID, 56.58% on PRID 450S and 59.40% on CUHK01 dataset. In contrast, the LSSCDL improves the performance by 1.84%, 1.52%, 3.91% and 6.57% on the four datasets, respectively. The performance of SCDL on small datasets is a slightly lower than LSSCDL, while the running time of SCDL is about five times slower than LSSCDL, which will be discussed in the next section.

5.4. Complexity Analysis

We conduct the proposed approach with Matlab implementation on a desktop PC with Intel i7-4790K @4.00GHz CPU and 32GB RAM, and report the running time of each stage averaged over 10 random trials on the VIPER dataset.

The computation time of learning sample-specific SVMs for all training images is 4.52 seconds, and it should be noted that learning a fixed weight vector costs 41.79 seconds. This demonstrates that solving multiple small SVM learning problems is actually more efficient than solving a large scale classification problem. The optimization time of LSSCDL is about 2.89 seconds, showing notable acceleration compared to 11.88 seconds of the SCDL algorithm. However, the testing time for one probe image is only 0.001 seconds, which indicates good applicability of the proposed approach in real applications.

To investigate the convergence effect of the proposed LSSCDL algorithm, we visualize the change of objective function value during optimization on the VIPER dataset in Figure 4. We initialize $\mathbf{D}_x$, $\mathbf{D}_w$, $\mathbf{\Lambda}_x$, $\mathbf{\Lambda}_w$ to be random matrices, and $\mathbf{M}$ to be the unit matrix in all experiments. From the figure one can see that the objective function value decreases quickly at first and then reaches a minimal, which demonstrates the feasibility of the proposed algorithm.

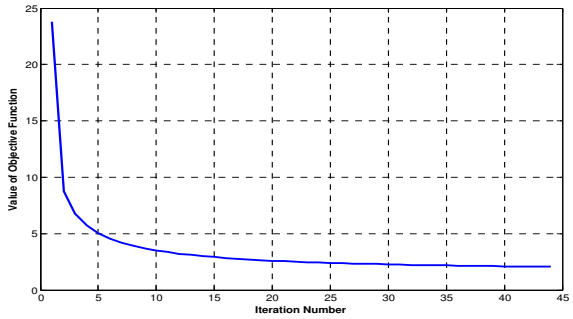


Figure 4. Change of objective function value during LSSCDL optimization on the VIPER dataset.

6. Conclusion

In this paper, we propose a novel and effective method for person re-identification. Motivated by the insight that different matching functions should be designed for different individuals, we formulate the person re-identification problem into a binary classification problem and learn a classifier specifically for each pedestrian. To capture the intrinsic relationship between feature patterns and ranking parameters, we propose an efficient LSSCDL algorithm to learn a pair of dictionary and a mapping function simultaneously. Experimental results on five challenging person re-identification datasets demonstrate the superiority of the proposed algorithm over state-of-the-art methods. In the future work, we will focus on applying the proposed algorithm to more matching applications.

Acknowledgements. Y. Zhang, B. Li, and H. Lu are supported by the Natural Science Foundation of China #61528101 and #61472060.

References

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2015.

[2] D. Chen, Z. Yuan, G. Hua, N. Zheng, and J. Wang. Similarity learning on an explicit polynomial kernel feature map for person re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1565–1573. IEEE, 2015.

[3] J. Chen, Z. Zhang, and Y. Wang. Relevance metric learning for person re-identification by exploiting global similarities. In *Proceedings of IEEE International Conference on Pattern Recognition*, pages 1657–1662. IEEE, 2014.

[4] Y. Chen, W. Zheng, and J. Lai. Mirror representation for modeling view-specific transform in person re-identification. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 3402–3408. AAAI Press, 2015.

[5] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[6] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *Proceedings of International Conference on Machine Learning*, pages 209–216. ACM, 2007.

[7] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. pages 501–512. Springer, 2011.

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

[9] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.

[10] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, volume 3. Citeseer, 2007.

[11] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proceedings of European Conference on Computer Vision*, pages 262–275. Springer, 2008.

[12] M. Guillaumin, J. J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In *Proceedings of IEEE International Conference on Computer Vision*, pages 498–505. IEEE, 2009.

[13] M. Hirzer, P. M. Roth, and H. Bischof. Person re-identification by efficient impostor-based metric learning. In *Proceedings of IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pages 203–208. IEEE Computer Society, 2012.

[14] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *Proceedings of European Conference on Computer Vision*, pages 780–793. Springer, 2012.

[15] X.-Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, and B. Xu. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 695–704. IEEE, 2015.

[16] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2288–2295. IEEE Computer Society, 2012.

[17] R. Layne, T. M. Hospedales, and S. Gong. Person re-identification by attributes. In *Proceedings of British Machine Vision Conference*, pages 1–11. BMVA Press, 2012.

[18] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 152–159. IEEE, 2014.

[19] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3610–3617. IEEE, 2013.

[20] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2197–2206. IEEE, 2015.

[21] S. Liao, Z. Mo, Y. Hu, and S. Z. Li. Open-set person re-identification. arXiv:1408.0872, 2014.

[22] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikäinen, and S. Z. Li. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1301–1306. IEEE Computer Society, 2010.

[23] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu. Semi-supervised coupled dictionary learning for person re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3550–3557. IEEE, 2014.

[24] X. Liu, H. Wang, Y. Wu, J. Yang, and M. Yang. An ensemble color model for human re-identification. In *IEEE Winter Conference on Applications of Computer Vision*, pages 868–875. IEEE Computer Society, 2015.

[25] C. C. Loy, C. Liu, and S. Gong. Person re-identification by manifold ranking. In *Proceedings of IEEE International Conference on Image Processing*, pages 3567–3571. IEEE, 2013.

[26] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1988–1995. IEEE Computer Society, 2009.

[27] A. J. Ma and P. Li. Query based adaptive re-ranking for person re-identification. pages 397–412. Springer, 2014.

[28] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by fisher vectors for person re-identification. In *Proceedings of*

European Conference on Computer Vision Workshops and Demonstrations, pages 413–422. Springer, 2012.

[29] B. Ma, Y. Su, and F. Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379–390, 2014.

[30] L. Ma, X. Yang, and D. Tao. Person re-identification over camera networks using multi-task distance metric learning. *IEEE Transactions on Image Processing*, 23(8):3656–3670, 2014.

[31] A. Mignon and F. Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2666–2672. IEEE Computer Society, 2012.

[32] S. Pedagadi, J. Orwell, S. A. Velastin, and B. A. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3318–3325. IEEE, 2013.

[33] B. Prosser, W. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *Proceedings of British Machine Vision Conference*, pages 1–11. British Machine Vision Association, 2010.

[34] P. M. Roth, M. Hirzer, M. Köstinger, C. Beleznai, and H. Bischof. Mahalanobis distance learning for person re-identification. In *Person Re-Identification*, pages 247–267. Springer, 2014.

[35] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 4184–4193. IEEE, 2015.

[36] V. N. Vapnik and V. Vapnik. *Statistical learning theory*, volume 1. Wiley New York, 1998.

[37] C. Vondrick, D. J. Patterson, and D. Ramanan. Efficiently scaling up crowdsourced video annotation - A set of best practices for high quality, economical video labeling. *International Journal of Computer Vision*, 101(1):184–204, 2013.

[38] S. Wang, L. Zhang, Y. Liang, and Q. Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2216–2223. IEEE Computer Society, 2012.

[39] K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In *Advances in Neural Information Processing Systems*, pages 1473–1480, 2005.

[40] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *Proceedings of European Conference on Computer Vision*, pages 536–551. Springer, 2014.

[41] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *Proceedings of IEEE International Conference on Computer Vision*, pages 2528–2535. IEEE, 2013.

[42] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 144–151. IEEE, 2014.

[43] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian. Query-adaptive late fusion for image search and person re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.

[44] W. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 649–656. IEEE Computer Society, 2011.