Information Sciences 355–356 (2016) 74–89

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Sparse representation matching for person re-identification

Le An[a], Xiaojing Chen[b,*], Songfan Yang[c], Bir Bhanu[d]

[a] National Key Laboratory of Science and Technology on Multi-spectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA
[c] College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China
[d] Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 28 September 2015
Revised 15 January 2016
Accepted 25 February 2016
Available online 22 March 2016

Keywords:
Person re-identification
Surveillance
Subspace learning
Sparse representation

ABSTRACT

The need for recognizing people across distributed surveillance cameras leads to the growth of recent research interest in person re-identification. Person re-identification aims at matching people in non-overlapping cameras at different time and locations. It is a difficult pattern matching task due to significant appearance variations in pose, illumination, or occlusion in different camera views. To address this multi-view matching problem, we first learn a subspace using canonical correlation analysis (CCA) in which the goal is to maximize the correlation between data from different cameras but corresponding to the same people. Given a probe from one camera view, we represent it using a sparse representation from a jointly learned coupled dictionary in the CCA subspace. The ℓ_1 induced sparse representation are regularized by an ℓ_2 regularization term. The introduction of ℓ_2 regularization allows learning a sparse representation while maintaining the stability of the sparse coefficients. To compute the matching scores between probe and gallery, their ℓ_2 regularized sparse representations are matched using a modified cosine similarity measure. Experimental results with extensive comparisons on challenging datasets demonstrate that the proposed method outperforms the state-of-the-art methods and using ℓ_2 regularized sparse representation ($\ell_1 + \ell_2$) is more accurate compared to use a single ℓ_1 or ℓ_2 regularization term.

1. Introduction

The vast deployment of video surveillance cameras in public venues drives the needs for automated surveillance applications such as people tracking [21], anomaly detection in crowd [31], *etc*. Many of these applications require the ability to determine the ID of the subject in different camera views, which is a problem referred as person re-identification that is gaining more attentions in the literature recently [2,12,14,19,32,56,59,67,74]. Specifically, the goal of person re-identification is to accurately match individuals across non-overlapping cameras at different time and locations. The results of person re-identification can be readily used in further processing tasks such as tracklet association for multi-camera people tracking [7].

Despite of the plethora of advanced pattern recognition techniques developed in the past few years, the performance of person re-identification is still not robust enough to warrant high accuracy in practice. The difficulties for person re-identification involve the following aspects:

[*] Corresponding author. Tel.: +1 9516404692.
E-mail address: xchen010@ucr.edu (X. Chen).

http://dx.doi.org/10.1016/j.ins.2016.02.055
0020-0255/© 2016 Elsevier Inc. All rights reserved.

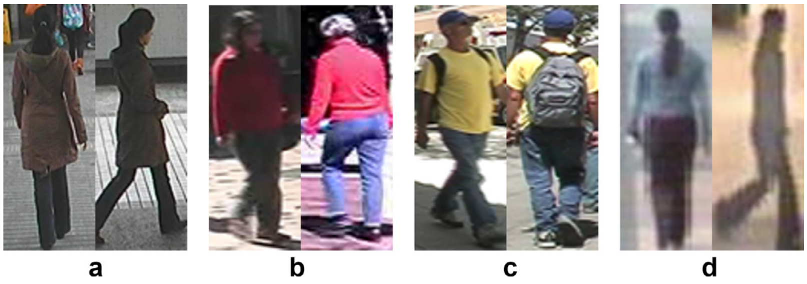

Fig. 1. Samples of image pairs of the same person in different camera views, showing (a) pose variation, (b) illumination change, (c) occlusion, and (d) low image quality, which make re-identification of people in different cameras a challenging problem.

1. *Pose variation*. In different camera views, a subject may have arbitrary poses (Fig. 1(a)),
2. *Illumination change*. The lighting condition is usually not constant in different camera views. As a result, the appearance of the same subject may vary significantly due to changing illumination (Fig. 1(b)),
3. *Occlusion*. A subject in one camera view may be fully or partially occluded by other subject or carrying items such as a backpack (Fig. 1(c)),
4. *Low image quality*. The captured image of a subject may suffer from low resolution, noise, or blur due to limited imaging quality of surveillance cameras (Fig. 1(d)).

In a person re-identification system usually two steps are involved: (1) extracting feature representations from person detections, and (2) establishing the correspondence between feature representations of probe and gallery. A gallery is a dataset composed of images of people with known IDs. A probe is the detection of a person from a different camera. Although other forms of biometrics such as face and gait [17,76] can be used to recognize people, however, acquiring such biometrics is difficult in uncontrolled low-resolution videos. For person re-identification, most of the existing approaches are appearance-based.

With the availability of tools for person detections, most of the previous work on person re-identification can be categorized into two groups:

1. Extracting feature representations which are robust against pose or illumination change, *e.g.*, [2,12,14,59].
2. Developing new matching methods using metric learning or ranking classifiers, *e.g.*, [19,26,53,74].

For the first group, discriminative appearance features are desired. Normally color and texture based features are widely used [19,28]. However, color or texture feature representations are sensitive to pose and illumination change, which may result in larger intra-person variation (difference between features of same person) than inter-person variation (difference between features of different persons). Besides low level image features, attribute or shape information has been applied in conjunction with color or texture features to improve the recognition accuracy [59].

To pursue more reliable matching, feature transformations or distance metrics are learned such that the distance between feature representations of the same person from different cameras is reduced while the distance between feature representations from different persons is increased [9,16,26,61]. SVM with ranking [53] and transfer learning [73] have also been proposed to obtain better matching correspondence.

In this paper, we propose a novel feature representation for person re-identification based on sparse coding. Inspired by coherent subspace learning to handle cross-type image synthesis [58] and face image super-resolution [1], we first learn a transformation to project the original image features into a subspace using canonical correlation analysis (CCA). In this learned subspace, the correlation between the features of the same people from different camera views is maximized. Then, two dictionaries for two camera views are jointly learned using training data in the coherent subspace. Given an image in the gallery, its image features are first projected into the CCA subspace and the sparse coefficients of this gallery subject are obtained using the learned dictionary with ℓ_2 regularization. These coefficients become the new feature representation for this gallery instance. During re-identification, given a probe, its sparse representation is obtained in the same way using the corresponding dictionary. Fig. 2 illustrates the outline of the proposed method for generating the sparse representation. The matching is then performed by computing the similarity between the sparse representations of the probe and gallery.

A related work for person re-identification was introduced in which a sparse representation was directly learned using a dictionary [25]. The dictionary was composed of existing data without any learning and the identity of the probe was determined through the non-zero coefficients by majority voting rule. In contrast to our approach, the sparse representations in [25] were used for determining the identity of the probe, while in our method the sparse representations are used as new feature representations for matching. Another related method was proposed in [41], in which coupled dictionary learning was used. Our method is different from [41] in the following aspects: (1) the learning methodology is different. In [41], both labeled and unlabeled data are required to learn the coupled dictionaries, which is a semi-supervised framework. The unlabeled data are used to exploit the geometry of the data distribution. On the other hand, our framework is supervised and we do not require extra unlabeled training data to carry out learning; (2) The dictionaries are learned in different

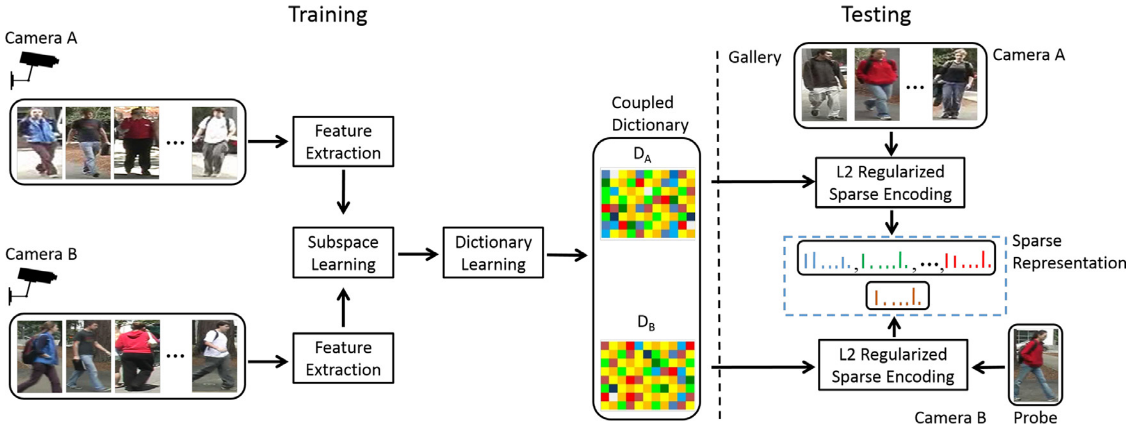

Fig. 2. Outline of the proposed sparse representation for person re-identification. In the training phase, the appearance features are extracted from images captured by two different cameras. A subspace is learned using CCA to project the features into a coherent subspace with maximized correlation between data in two views. Two dictionaries are learned jointly for each camera. In the testing phase, the features of gallery and probe are first projected into the learned subspace. Then, their sparse representations with ℓ_2 regularization are obtained using the coupled dictionaries. The sparse representations are then used as a new representation for probe or galley for matching.

feature spaces. In [41], the coupled dictionary learning uses the original image features, while in our method, a subspace learning and projection are performed first. This creates a better foundation for dictionary learning because features from different cameras are projected into a coherent subspace; (3) The dictionary learning technique is different. In [41], Local Coordinate Coding (LCC) is used. In our method, L2 regularized sparse representation, *i.e.*, elastic net regularization, is used; (4) The matching scheme is different. In [41], a probe is first sparsely represented by the dictionary, and then the sparse representation is used to recover the feature through the gallery dictionary, and then matching is performed by ranking the feature similarities. In our method, both the probe and gallery are represented by the corresponding dictionaries, and then the matching is performed using the sparse representations as features. Moreover, as later validated in the experiments, our method achieves notably better results than those in [41].

The rest of this paper is organized as follows. Section 2 reviews related work and summarizes our contributions. Details of the proposed ℓ_2 regularized sparse representation based method for person re-identification are presented in Section 3. Section 4 provides the experimental results and finally Section 5 concludes this paper.

2. Related work and contributions

2.1. Feature representation

Color information has been widely used for person re-identification. Kviatkovsky et al. [28] discovered an intra-distribution color structure and this structure was found invariant under a wide range of imaging conditions while being discriminative. When used together with a covariance descriptor, state-of-the-art performance was achieved on different re-identification datasets. Yang et al. [66] proposed a salient color names based color descriptor which can guarantee that a higher probability will be assigned to the color name which is nearer to the color. This color descriptor can be computed efficiently in advance. Farenzena et al. [12] combined features including the overall chromatic content, the spatial arrangement and the presence of recurrent local motifs to handle individual matching with appearance variation. Instead of designing handcrafted features, Gray and Tao [15] used Adaboost to select the most discriminative features for person re-identification. Bak et al. [5] learned a model in a covariance metric space to select features based on the idea that different regions of each subject should be matched specifically. Martinel and Micheloni [47] proposed a novel discriminative signature from multiple local features and designed a signature distance measure by exploiting different body parts. Cheng et al. [8] adopted pictorial structures to localize human parts and searched part-to-part correspondences to match subjects across different cameras. Ma et al. [45] used both biologically inspired features and covariance descriptors to handle background and illumination variations. To handle the appearance change in different cameras, Javed et al. [23] showed that brightness transfer functions from one camera to another lie in a subspace which can be learned by probabilistic PCA.

To explore higher level image features, Kuo et al. [27] applied semantic color names to describe an image of a person by probabilities of the presence of the predefined colors and a better performance was achieved than using only the color histograms. An et al. [2] proposed the use of reference descriptors instead of using image features directly. A reference descriptor for a probe or gallery was generated by computing the similarity scores between this probe or gallery to a reference set. This similarity-based representation achieved state-of-the-art performance on a widely used dataset. Layne et al. [29] used mid-level semantic attributes such as gender for improved re-identification results compared to the existing approaches using low-level image features. Zhao et al. [68,69] proposed unsupervised salient features, and it was shown that attributes can be combined with the existing methods to improve the recognition performance. In a recent work, Zhao et al. [70] learned discriminative mid-level filters from automatically discovered patch clusters and those filters were able to identify specific visual patterns in order to distinguish different persons. Li et al. [30] proposed to use clothing attributes to assist person re-identification. Shi et al. [55] invented a semantic attribute model based on fashion photographs and then transferred this model for person re-identification tasks.

Liu et al. [39] studied feature importance for person re-identification and a method for on-the-fly mining of feature was proposed. Martinel et al. [48] extracted multiple features from image pairs and obtained a so-called distance feature vector and re-identification was achieved by classifying the distance feature vector using a binary classifier. Xu *and Zheng* [64] proposed to use multi-resolution color histograms and local structural sparse coding for re-identification. In this method, the query image was sparsely represented by the gallery image and then the sparse codes were used as a complementary feature descriptor to the global appearance features. Liao et al. [35] proposed a Local Maximal Occurrence (LOMO) feature, in which the horizontal occurrence of local features is used to achieve a stable representation against view changes.

With the help of human feedback, Hirzer et al. [18] proposed a two-step method by first using a descriptive model to obtain an initial ranking, which was refined in the second step by human using a discriminative model. Liu et al. [40] proposed a post-rank optimization method in which human was allowed to select negative samples, achieving over 30% performance gain and significantly reduced search time as compared to the exhaustive search.

2.2. Matching methods

For person re-identification, it is important to find an appropriate measure for calculating the distance between images from different cameras. However, the visual discrepancy by multiple views or the so-called semantic gap is a common issue and has attracted much attention in recent years, and it is still a challenging task to bridge this gap. To tackle this problem, different from existing methods which mainly focus on generating the correlation between a subject and its corresponding images, Gao et al. [13] proposed a probabilistic semantic mapping scheme to build the connection between subject level relevance and image level pairwise matching. This method has shown satisfactory performance on the task of view-based object retrieval and introduced a possible direction to bridge the semantic gap. For the task of person re-identification, standard metric learning techniques such as Large Margin Nearest Neighbor (LMNN) [61], Information Theoretic Metric Learning (ITML) [9], and Logistic Discriminant Metric Learning (LDML) [16] have been applied. A variant of LMNN (LMNN-R) was developed in [10] that introduced a reject option to the unfamiliar matches, which improved re-identification results. A metric learning method using the idea “keep it simple and straightforward” (KISSME) was proposed in [26] from a statistical inference point of view such that a decision of dissimilarity can be obtained by a likelihood ratio test. Tao et al. [56] extended the KISSME formulation by introducing a regularization term and a smoothing term to robustly estimate covariance matrices. In this way the instability in calculating the inverse of a covariance matrix from a small sized training set was alleviated. Zheng et al. [74] formulated re-identification as a relative distance comparison problem in which the likelihood of the distance between a pair of images of the same person being smaller than a pair of images of different people was maximized. This framework was improved by Liu et al. [38] with the incorporation of attribute information and feature weighting. A relaxed pairwise learned metric (RPLM) based on Mahalanobis distance learning was proposed in [19] which took advantages of the structure of the data with reduced computational cost and achieved good results with simple image features. Li et al. [34] proposed a filter pairing deep neural network in which misalignment, pose difference, occlusions and background clutter were jointed handled. However, training such a deep network requires a significant amount of data. Xiong et al. [63] applied multiple kernel-based metrics in conjunction with histogram-based features and showed improvement over state-of-the-art on several datasets. Wang et al. [60] proposed a data-specific adaptive metric method by applying the cross-view support and projection consistencies.

Instead of re-identifying people with absolute scores, Prosser et al. [53] formulated a relative ranking problem and an ensemble RankSVM was proposed to overcome the limited scalability in existing SVM-based ranking methods. Re-identification was reformulated as the problem of matching a probe to a watch list of people instead of matching to an individual in [73]. To solve this problem, a transfer learning framework was proposed and different types of variations among a target set and non-target data were defined. A manifold ranking approach was developed by Loy et al. [42] such that the probe information was propagated along the data manifold in an unsupervised manner. It showed that the performance of the existing metric learning based methods could be significantly improved by integrating the manifold ranking. In [36], soft and hard re-weighting were applied to redistribute energy among the sparse coefficients. This iterative process ensures that the best candidates are ranked at each iteration.

Subspace learning has also been explored for person re-identification. The image spaces of two camera views were jointly partitioned into different configurations based on the similarity of cross-view transforms in [32] and image pairs with similar transforms were projected into a common feature space for matching. A pairwise constrained component analysis (PCCA) was proposed in [49] to learn a low-dimensional mapping by complying with a set of sparse training pairwise constraints. Pedagadi et al. [52] used local Fisher discriminant analysis (LFDA) to reduce feature dimensionality for person re-identification and experimental results suggested that LFDA outperformed other metric learning-based methods.

From a different perspective, Jing et al. [24] introduced the problem of super-resolution person re-identification, in which low-resolution probe images are matched with high-resolution images. Since the subject may also be occluded in certain view, Zheng et al. [75] proposed a patch-level matching model to match subjects whose bodies are only partially

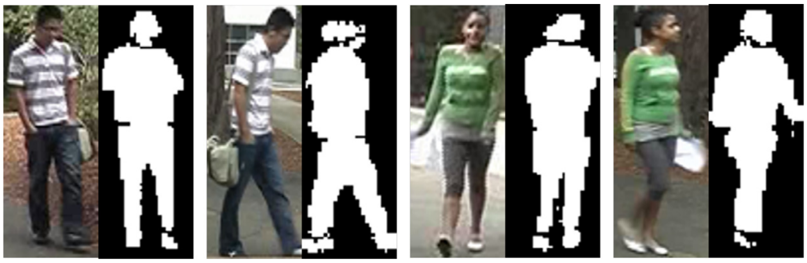

Fig. 3. Sample segmentation results using the method in [43] to separate the foreground subject from the background. The appearance features are extracted from the foreground to mitigate the impact by the cluttered background.

visible. In a recent work [71], a large scale person re-identification dataset was introduced. This dataset is significantly larger than the existing ones and is more practical towards real-world application. Ma et al. [44] tackled a new person re-identification problem without label information. Given the matched and unmatched image pairs from source domain cameras, as well as unmatched and unlabeled image pairs from target domain cameras, an adaptive ranking support vector machines (AdaRSVMs) method was developed for matching under target domain cameras without person labels. For more detailed discussion on person re-identification, comprehensive surveys are available in [57] and [6].

2.3. Contributions of this paper

The main contributions of this paper are:

1. To mitigate the disparity between image data from different views, we propose the use of a coherent subspace learned by CCA and project the multi-view data (images of the same person in different camera views) into this coherent subspace such that the correlation between two views of the same data are maximized. This subspace projection provides a foundation for robust matching across cameras.
2. We propose a novel framework for generating sparse representations of probe and gallery data in the coherent subspace. The generated sparse representations are used for person re-identification. Compared with matching using features extracted directly from images, using the learned sparse representation achieves the state-of-the-art results on different publicly available datasets.
3. We learn the sparse representation with a coupled dictionary sets. The dictionaries are jointly learned using training data from different camera views. In addition, the sparse representation is regularized with an additional ℓ_2 regularization term to ensure the stability of the learned coefficients while preserving sparsity. Experimental results show that ℓ_2 regularized sparse representation outperforms standard sparse representation.

3. Sparse representation for person re-identification

The goal is to re-identify people in non-overlapping cameras. To mitigate significant disparity in appearance feature space for the same subject in different views, the proposed algorithm first finds projection matrices for features from each view such that after projection features of the same person are maximally correlated. To learn this subspace projection, labeled training image pairs are used in CCA. After the projection matrices are obtained, training data are projected into this coherent subspace. The projected training data are then used to jointly learn coupled dictionaries for each camera view. In the re-identification process given probe and gallery, appearance features are first extracted. To minimize the impact of the cluttered background, we use a deep decompositional network based pedestrian parsing method [43] to segment the foreground subject from the background before the appearance features are extracted. Fig. 3 shows some segmentation results. These features are then projected into the learned coherent subspace, in which their sparse representations with ℓ_2 regularization are obtained using the coupled dictionaries. The calculated sparse coefficients are used as a new feature representation for probe and gallery in the matching process which is based on a modified cosine similarity measure. The pipeline for generating sparse representations is illustrated in Fig. 2.

3.1. Coherent subspace learning

Canonical Correlation Analysis (CCA) was first introduced in [20] and it is a multivariate statistical analysis technique. CCA finds projection matrices for two sets of random variables such that the correlation between the projected random variables is maximized in the correlated or coherent subspace. CCA has been applied to problems involving multi-view or multi-modality data such as image super-resolution [1,22] and face recognition under pose variation [54].

For person re-identification, given N image pairs from two cameras A and B, appearance features with dimension m are first extracted from the images. These feature vectors are organized into two data matrices $X_A = \{x_A^i \in \mathbb{R}^m, i = 1, 2, \dots, N\}$ and $X_B = \{x_B^i \in \mathbb{R}^m, i = 1, 2, \dots, N\}$, in which x_A^i and x_B^i correspond to the same person in different views. The goal of CCA

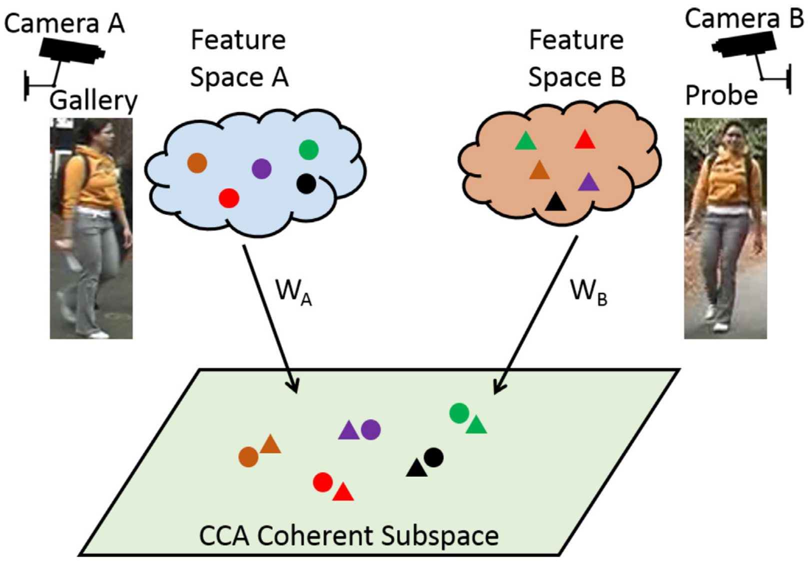

Fig. 4. Illustration of CCA projection. A pair of symbols with the same color but different shapes indicates features of the same person. The projection matrices W_A and W_B transform the data from the original feature space to a coherent subspace in which the data correlation is maximized. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

is to find a pair of projection vectors $w_A \in \mathbb{R}^m$ and $w_B \in \mathbb{R}^m$ such that the correlation coefficient ρ of $w_A^T X_A$ and $w_B^T X_B$ is maximized. Mathematically, the objective function to be maximized is given by

$$\begin{aligned}\rho &= \frac{cov(w_A^T X_A, w_B^T X_B)}{\sqrt{var(w_A^T X_A) var(w_B^T X_B)}} \\ &= \frac{w_A^T C_{AB} w_B}{\sqrt{w_A^T C_{AA} w_A w_B^T C_{BB} w_B}},\end{aligned} \tag{1}$$

where cov is short for covariance and var computes data variance. $C_{AA} = E[X_A X_A^T]$ and $C_{BB} = E[X_B X_B^T]$ are the covariance matrices of X_A and X_B. $C_{AB} = E[X_A X_B^T]$ is the covariance matrix of X_A and X_B.

Eq. (1) can be reformulated as a constrained optimization problem as follows

$$\begin{aligned}&\text{maximize} \quad w_A^T C_{AB} w_B, \\ &\text{subject to} \quad w_A^T C_{AA} w_A = 1, \quad w_B^T C_{BB} w_B = 1.\end{aligned} \tag{2}$$

Equivalently, w_A and w_B can be solved through the following generalized eigenvalue problem

$$\begin{bmatrix} 0 & C_{AB} \\ C_{BA} & 0 \end{bmatrix} \begin{bmatrix} w_A \\ w_B \end{bmatrix} = \lambda \begin{bmatrix} C_{AA} & 0 \\ 0 & C_{BB} \end{bmatrix} \begin{bmatrix} w_A \\ w_B \end{bmatrix}, \tag{3}$$

where $C_{BA} = E[X_B X_A^T] = C_{AB}^T$.

The projection matrices $W_A \in \mathbb{R}^{m \times d}$ and $W_B \in \mathbb{R}^{m \times d}$ are composed of d pairs of projection vectors w_A and w_B corresponding to d largest eigenvalues. In this way, W_A and W_B project the original features from $\mathbb{R}^m$ to a subspace of $\mathbb{R}^d$, where the correlation between the projected features of X_A and X_B is maximized.

Fig. 4 demonstrates the CCA principle that projects the data from different views into a coherent subspace in which the data pair of the same person are maximally correlated. To validate the ability of CCA to find a coherent subspace, we use half of the data from the VIPeR dataset [14] to learn the PCA and CCA projections, respectively and project the other half of the data into the PCA and CCA subspaces with learned projection matrices. As shown in Fig. 5, the CCA projected data from two cameras exhibit more similarity in terms of their manifold structures compared to the structures of the same data in the PCA subspace. This indicates that CCA is able to correlate data from different views through subspace projection.

3.2. Coupled dictionary learning

Given N training data pairs $x_A^i \in \mathbb{R}^m$ and $x_B^i \in \mathbb{R}^m$ consisting of image pairs from cameras A and B, the projected image features in the CCA subspace are denoted by $p_A^i \in \mathbb{R}^d$ and $p_B^i \in \mathbb{R}^d$, respectively. The goal is to jointly learn the dictionaries $D_A \in \mathbb{R}^{d \times k}$ and $D_B \in \mathbb{R}^{d \times k}$ of size k, such that the sparse representations for p_A^i and p_B^i should be as similar as possible. In

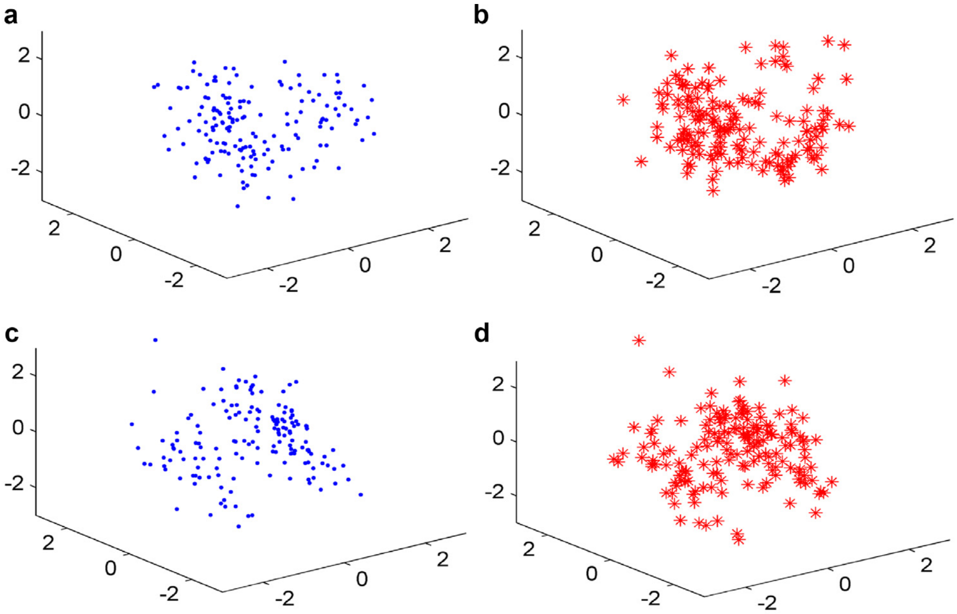

Fig. 5. Features in subspace using different projections. (a) Embedded manifold for features in PCA subspace of testing data in one camera view (in dots). (b) Embedded manifold for features in PCA subspace of testing data in the other camera view (in asterisks). (c) Embedded manifold for features in CCA subspace of testing data in one camera view (in dots). (d) Embedded manifold for features in CCA subspace of testing data in the other camera view (in asterisks). Different from PCA, CCA projects multi-view data into coherent subspaces and the manifold of data in each view is more similar, *i.e.*, the shapes of data in (c) and (d) by CCA projection are more similar than the shapes of data in (a) and (b) by PCA projection. The training data for PCA and CCA are from the VIPeR dataset [14] and testing data are the rest of the same dataset. The numbers on the axes denote normalized feature values.

other words, the idea is that the sparse representations corresponding to the same person but in different camera views should be almost the same. The energy function to be optimized is

$$\min_{D_A,D_B} \frac{1}{N}\left(\sum_{i=1}^{N} \left\| p_A^i - D_A\alpha_i \right\|_2^2 + \left\| p_B^i - D_B\alpha_i \right\|_2^2 + \gamma_1 \|\alpha_i\|_1 + \gamma_2 \|\alpha_i\|_2^2\right), \tag{4}$$

where γ_1 and γ_2 are regularization parameters for ℓ_1 and ℓ_2 regularization terms respectively.

The ℓ_1 regularization term ensures that the coefficients α_i are sparse. Previous study suggested that most images (*e.g.*, faces) can be approximated as a linear combination of the base elements in a dictionary and this representation is naturally sparse [62]. The sparsity resembles human perception system in which the activation of neurons to an image is typically sparse [51]. The ℓ_2 regularization term has the properties as in the Ridge Regression to stabilize the coefficients. Sparse representation with ℓ_2 regularization is also referred as Elastic Net in statistics [77].

Similar to the formulation in [65], Eq. (4) can be written as

$$\min_{D} \frac{1}{N}\left(\sum_{i=1}^{N} \left\| p^i - D\alpha_i \right\|_2^2 + \gamma_1 \|\alpha_i\|_1 + \gamma_2 \|\alpha_i\|_2^2\right), \tag{5}$$

where p^i and D are constructed by

$$p^i = \begin{bmatrix} p_A^i \\ p_B^i \end{bmatrix}, \quad D = \begin{bmatrix} D_A \\ D_B \end{bmatrix}. \tag{6}$$

To obtain the dictionary in Eq. (5), an online optimization algorithm based on stochastic approximations is used [46].

3.3. Sparse representation with ℓ_2 regularization

With the learned coupled dictionaries D_A and D_B, the sparse representations for probe or gallery data can be generated. Assuming images from camera A serve as the gallery, for each image j in the gallery, the appearance features are first extracted and then projected into the CCA subspace with the projected features denoted as g^j. Its sparse representation α_{g^j} is obtained by

$$\operatorname*{argmin}_{\alpha_{g^j}} \left\| g^j - D_A\alpha_{g^j} \right\|_2^2 + \gamma_1 \left\| \alpha_{g^j} \right\|_1 + \gamma_2 \left\| \alpha_{g^j} \right\|_2^2, \tag{7}$$

The ℓ_2 regularized sparse coefficients α_{gj} are used as a new representation for gallery image j. Similarly, given a probe with its appearance feature projected into CCA subspace as p, the sparse representation α_p is learned by solving

$$\underset{\alpha_p}{\operatorname{argmin}} \|p - D_B\alpha_p\|_2^2 + \gamma_1\|\alpha_p\|_1 + \gamma_2\|\alpha_p\|_2^2. \tag{8}$$

Eqs. (7) and (8) can be solved efficiently with stability using least angle regression (LARS) algorithm [11].

3.4. Identity matching

With the sparse representations of probe and gallery ready, the matching is based on the similarity between the sparse coefficients α_p of a probe and the sparse coefficients α_g^j of a gallery. We adopt a modified cosine similarity measure [37] defined by

$$\mathrm{sim}(\alpha_p, \alpha_{g^j}) = \frac{|(\alpha_p)^T \cdot \alpha_{g^j}|}{\|\alpha_p\| \cdot \|\alpha_{g^j}\| \cdot (\|\alpha_p - \alpha_{g^j}\|_{\mathrm{P}} + \epsilon)}, \tag{9}$$

where $\|\cdot\|_{\mathrm{P}}$ is the L_{P} norm and ϵ is a small positive number to prevent division by zero. The reason to apply the modified cosine similarity is that the standard cosine similarity does not take into account the actual distance between two vectors, while the modified cosine similarity is able to address both the distance measure and angular measure and has shown improved performance in recognition tasks [37]. The modified cosine similarity has shown to be useful for person re-identification in [3]. We use the same features and evaluation protocols as those in [3], and we have empirically found out that the modified cosine similarity is better than the standard cosine similarity in our case.

3.5. Summary of the algorithm

The proposed algorithm for person re-identification consists of training and testing phases. In the training phase, CCA projection matrices are learned as well as the coupled dictionaries for different camera views. Given probe and gallery in the testing phase, the appearance features are extracted and projected into the CCA subspace. Then the ℓ_2 regularized sparse representations for probe and gallery are obtained using jointly learned dictionaries. The similarity scores between probe and gallery are computed using their sparse representations. The algorithms for training and testing algorithms are summarized in Algorithms 1 and 2, respectively.

Algorithm 1 CCA subspace and coupled dictionary learning.

```
Input:
    Training image pairs from cameras $A$ and $B$
    $d$: dimension of the CCA subspace
    $k$: size of the dictionaries
1: Extract appearance features from images and obtain data matrices $X_A$ and $X_B$.
2: Solve the generalized eigenvalue problem in Eq. (3).
3: Project $X_A$ and $X_B$ into the learned CCA subspace.
4: Solve the optimization problem in Eq. (5).
Output:
    CCA subspace projection matrices $W_A$ and $W_B$
    Coupled dictionaries $D_A$ and $D_B$
```

Algorithm 2 Re-identification with sparse representation.

```
Input:
    Probe and gallery from cameras $A$ and $B$
    $W_A$ and $W_B$: CCA projection matrices
    $D_A$ and $D_B$: coupled dictionaries
1: Extract appearance features from probe and gallery.
2: Project appearance features of probe and gallery into the learned CCA subspace using $W_A$ and $W_B$.
3: Obtain $\ell_2$ regularized sparse representations for probe and gallery using Eqs. (7) and (8) with $D_A$ and $D_B$.
4: Calculate similarity scores between probe and gallery using Eq. (9).
5: Rank the similarity scores from high to low.
Output:
    Re-identification results
```

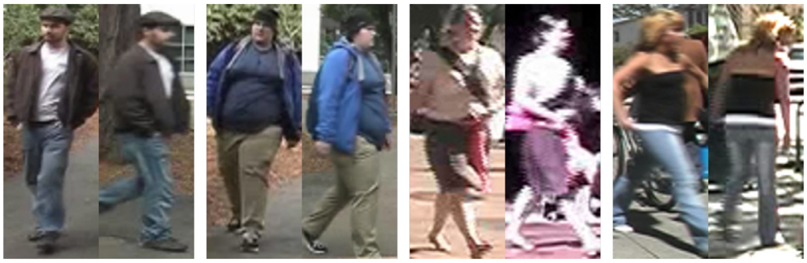

Fig. 6. Sample image pairs from the VIPeR dataset [14].

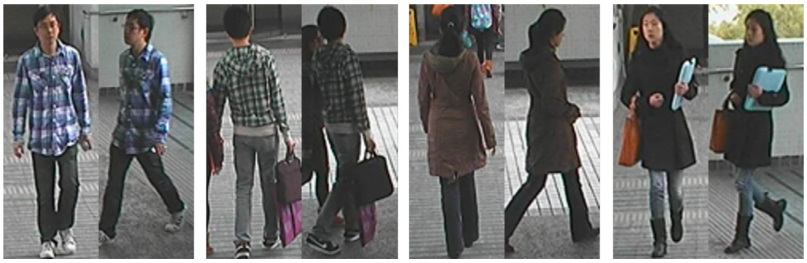

Fig. 7. Sample image pairs from the CUHK Campus dataset [33].

4. Experiments

4.1. Datasets

We evaluate our method on three different publicly available datasets for two typical re-identification scenarios, namely image-based case (single-shot) and video-based case (multi-shot). For single-shot scenario, the VIPeR dataset and the CUHK Campus dataset are used. The VIPeR dataset[1] enrolled 632 persons and is one of the most challenging and widely evaluated benchmark datasets for person re-identification [14]. Some sample image pairs from this dataset are shown in Fig. 6. For each person, one detection is available in each of the two non-overlapping cameras with varying illumination and cluttered background. For most of the subjects the view change is more than 90 degrees. In addition, partial occlusion is frequent due to the subjects carrying items such as backpack or handbag.

The CUHK Campus dataset[2] is a recently released dataset which contains images of 971 subjects from two non-overlapping camera views [33]. One camera captures the frontal or rear view of a person and the other camera captures the profile view of a person. Each person has two detections in each camera view. Some sample image pairs from this dataset are shown in Fig. 7.

For multi-shot video-based re-identification, we use the Person Re-ID 2011 (PRID) dataset[3]. The PRID dataset consists of multiple person trajectories recorded from two surveillance cameras. Camera *A* contains 385 persons and camera *B* shows 749 persons. The first 200 persons appear in both camera views. Each trajectory contains approximately 100 to 150 images depending on the walking speed of a person. Two segments of trajectories of the same person in two cameras are shown in Fig. 8.

4.2. Feature extraction

All of the images in these three datasets are normalized to 128 × 48 in the experiments. For feature extraction, Each image is divided into blocks of size 8 × 16. The blocks are overlapped by 50% in both horizontal and vertical directions. The appearance features extracted from the images include both color and texture features as in [19]. For each block, the color features consist of the quantized mean values of the HSV and Lab color channels. In addition, we include semantic color names [27] as an additional color representation. The texture features are represented by the 8-bit Local Binary Patterns (LBP) [50]. The final appearance features are the concatenation of both color and texture features.

The choices of HSV and Lab to encode color and LBP to encode texture have shown to be effective for the task of person re-identification [2,4,19]. In our case, we have also found out that these features are discriminative. It is worthwhile to mention that our method is compatible with any feature extraction method, which is beyond the scope of this paper. Since,

[1] The VIPeR dataset is available at http://vision.soe.ucsc.edu/?q=node/178

[2] The CUHK Campus dataset is available at http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html

[3] The PRID dataset is available at http://lrs.icg.tugraz.at/datasets/prid

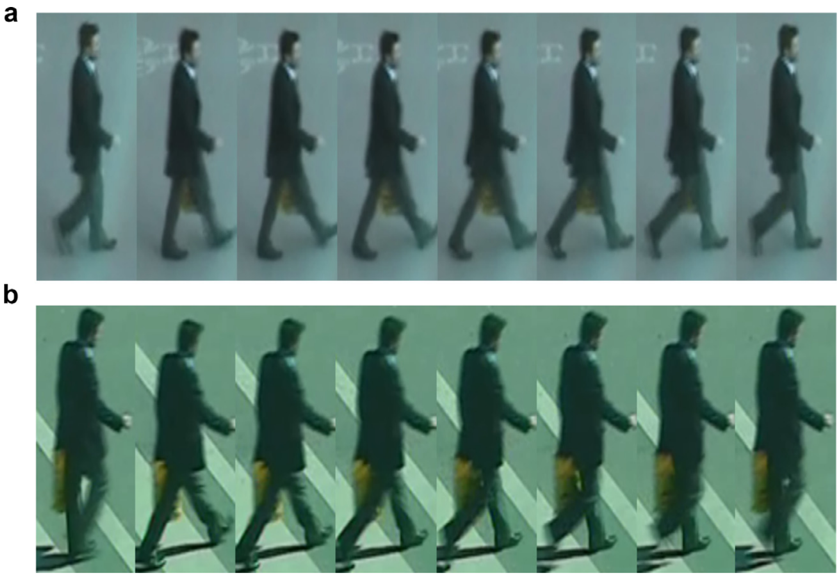

Fig. 8. Sample image pairs from the PRID dataset [18]. (a) Trajectory of a person in camera *A*. (b) Corresponding trajectory in camera *B*.

the feature dimensions significantly exceeds the number of samples, we first use PCA to reduce the feature dimension with 99% of the data variance retained. This can effectively offset the small sample size problem in the subsequent CCA subspace learning. The dimension of the CCA subspace projection matrices W_A and W_B are set to $d = 50$. The ℓ_1 and ℓ_2 regularization parameters to learn the coupled dictionaries and sparse representations are set to 0.01 and 0.02 respectively. The size of the dictionary is set to $k = 100$. These parameters are determined by cross-validation on the training data.

4.3. Evaluation method

For the experiments on the VIPeR and the CUHK Campus datasets, we follow the experimental protocols in the previous work (*e.g.*, [12,19,26] and [68]) for fair comparison. We randomly partition each dataset into two subsets of equal size. Half of the data are used for training and the other half are used for testing. Gallery consists of images from one camera and images from the other camera are used as probes. For the CUHK Campus dataset, since each person has two images in one view, we follow a single-shot approach as in [70], meaning that each probe is matched with every image in the gallery. For the PRID dataset, the data of the common 200 subjects in two camera views are used. The gallery set is constructed by extracting five evenly sampled images per trajectory as done in [18]. A probe of one subject consists of all the detections in the trajectory and a majority voting rule is applied to determine the identity of each probe.

To evaluate the re-identification performance, matching rates at selected ranks and the Cumulative Match Characteristic (CMC) curves are reported for comparison. The CMC curve represents the expectation of finding the correct match in the top *r* matches. In other words, a rank-*r* matching rate shows the percentage of the probes that are correctly recognized from the top *r* matches in the gallery. The experiments are conducted 10 times and the average results are listed.

To ensure fair comparison, the experiments are conducted using the same evaluation protocols as in the other methods being compared with. To accurately report the performance of the other methods, their results are either obtained by using the implementation provided by the authors, or directly cited from the corresponding papers.

4.4. Experimental results

4.4.1. The VIPeR dataset

We first conduct experiments on the VIPeR dataset, the results on which have been reported in most of the recent work on person re-identification. We compare our approach with 19 state-of-the-art alternatives.

We first conduct experiments on the VIPeR dataset, the results on which have been reported in most of the recent work on person re-identification. We compare our approach with the following 19 state-of-the-art alternatives, which are reference-based approach (RD) [2], saliency matching (SalMatch) [68], relaxed pairwise learned metric (RPLM) [19], Semi-supervised coupled dictionary (SSCDL) [41], regularized smoothing KISS metric learning (RS-KISS) [56], custom pictorial structures (CPS) [8], biologically inspired features and covariance descriptors (BiCov) [45], KISS metric (KISSME) [26], large margin nearest neighbor with rejection (LMNN-R) [10], symmetry-driven accumulation of local features (SDALF) [12], manifold ranking (MRank) [42], pairwise constrained component analysis (PCCA) [49], descriptive and discriminative classification (DDC) [18], large margin nearest neighbor (LMNN) [61], attributed-based relative distance comparison (aPRDC) [38], relative

Table 1
Matching rates (in %) on the VIPeR dataset at different ranks.

Rank	$r=1$	10	20	50	100
Proposed	**32.91**	**75.93**	**89.24**	**96.84**	**99.73**
RD [2]	30.25	74.68	86.82	95.70	99.24
SalMatch [68]	30.16	65.54	79.15	91.49	98.10
RPLM [19]	27.34	69.02	82.69	94.56	98.54
SSCDL [41]	25.60	68.10	83.60	95.20	–
RS-KISS [56]	24.50	66.60	81.70	93.50	98.00
CPS [8]	21.84	57.21	71.00	87.00	91.77
BiCov [45]	20.66	56.18	68.00	81.56	88.66
KISSME [26]	20.03	62.39	77.46	92.81	98.19
LMNN-R [10]	20.00	66.00	79.00	92.50	95.18
SDALF [12]	19.87	49.37	65.73	84.84	90.43
MRank [42]	19.34	55.51	70.44	87.69	96.90
PCCA [49]	19.27	64.91	80.28	95.00	97.01
DDC [18]	19.00	52.00	65.00	80.00	91.00
LMNN [61]	17.41	53.86	67.88	88.13	96.23
aPRDC [38]	16.14	50.98	65.95	88.00	93.00
PRDC [74]	15.66	53.86	70.09	87.79	92.84
ITML [9]	15.54	53.13	69.05	88.54	96.93
RankSVM [53]	14.00	51.00	67.00	85.00	94.00
ELF [15]	12.00	43.00	60.00	81.00	93.00

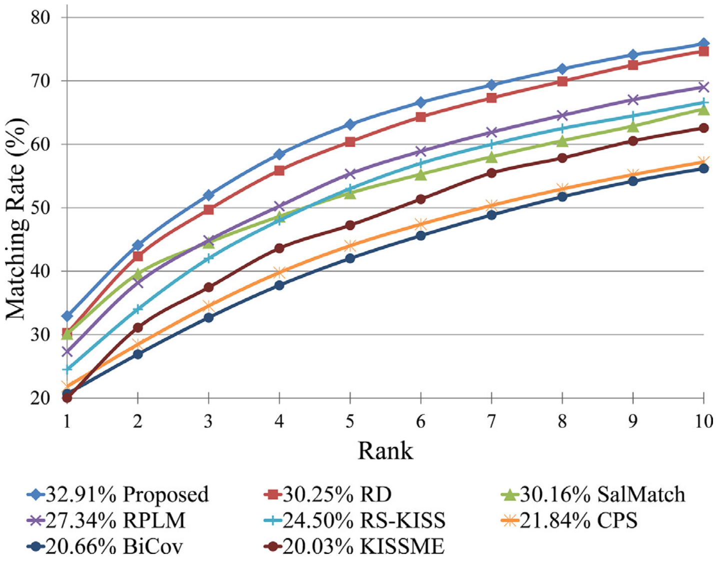

Fig. 9. CMC curves on the VIPeR dataset for the proposed method and the other methods.

distance comparison (PRDC) [74], information-theoretic metric learning (ITML) [9], support vector ranking (RankSVM) [53], and ensemble of localized features (ELF) [15].

The re-identification accuracy of different methods at rank 1, 10, 20, 50 and 100 are reported in Table 1. The proposed method achieves a matching rate of 32.91% at rank 1, which is about 9% relative improvement over the second best result of 30.25% by RD [2]. Furthermore, at all of the other ranks, the proposed method consistently outperforms the competing methods. At rank 100, almost 100% matching rate is reached. The CMC curves are compared in Fig. 9 between our method and the other top performers in Table 1. Similar to the observations from Table 1, the proposed method achieves higher matching rates compared to the other methods at different ranks.

To study the impact of reduced training data size and to make comparison with other methods, we report in Table 2 the re-identification results with different training data sizes. In this case, all the data from the VIPeR dataset are used. As the size of the training set decreases, the number of subjects in the gallery and probe data increases, which makes the re-identification more difficult. The same experiment protocol was used in RD [2], RPLM [19], and PRDC [74], the results of which are included in Table 2 for comparison. The results shown in Table 2 suggest that with a smaller training set the proposed method is still able to perform better than the competing methods at different ranks.

Table 2
Matching rates (in %) on the VIPeR dataset at different ranks with reduced training data size.

Training size	N=200			N=100		
Rank	$r = 1$	10	20	$r = 1$	10	20
Proposed	**23.34**	**60.07**	**75.26**	**16.82**	**49.14**	**62.97**
RD [2]	21.94	59.26	74.58	15.11	47.14	60.30
RPLM [19]	19.51	56.44	71.09	10.88	37.69	51.64
PRDC [74]	12.64	44.28	59.95	9.12	34.40	48.55

Table 3
Matching rates (in %) on the CUHK Campus dataset at different ranks.

Rank	$r = 1$	10	20	50	100
Proposed	**31.34**	**68.39**	**78.14**	**87.63**	**95.26**
RD [2]	29.97	67.78	77.04	87.24	94.21
SalMatch [68]	28.45	55.68	67.95	83.53	92.10
ITML [9]	15.98	45.60	59.81	76.61	88.32
LMNN [61]	13.45	42.25	54.11	73.29	86.65
SDALF [12]	9.90	30.33	41.03	55.99	67.39
ℓ_2 distance	9.84	26.42	33.13	46.98	63.48
ℓ_1 distance	10.33	26.34	33.52	45.62	61.95

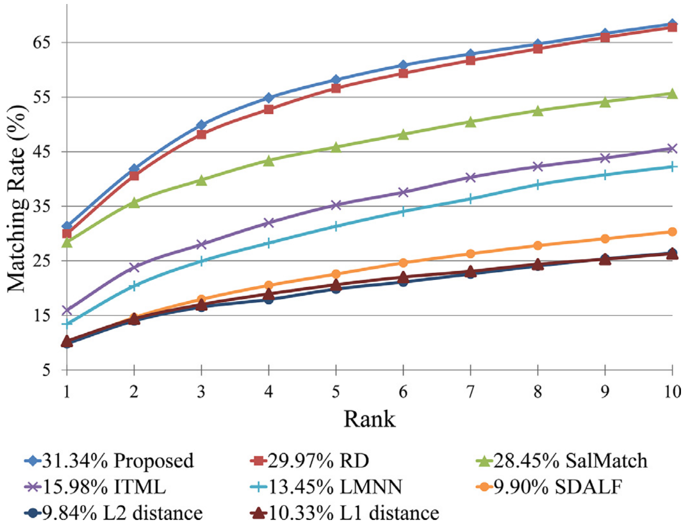

Fig. 10. CMC curves on the CUHK Campus dataset for the proposed method and the other methods.

4.4.2. The CUHK Campus dataset

For the CUHK Campus dataset, we compare the proposed approach with the following methods: RD [2], SalMatch [68], SDALF [12], LMNN [61], ITML [9], as well as direct matching using ℓ_2 and ℓ_1 distance. Table 3 reports the matching rates at rank 1, 10, 20, 50, and 100. As compared to the other methods with all the matching rates below 30%, the proposed method achieves a rank-1 matching rate of 31.34%. Fig. 10 shows the CMC curves of our method and the other methods. The proposed method achieves a higher rate at each rank compared to the second best method (RD), and outperforms the rest of the methods by a large margin.

4.4.3. The PRID dataset

For the PRID dataset, we compare the proposed approach with the following methods: RD [2], KISSME [26], as well as two baseline methods using ℓ_1 distance and ℓ_2 distance. Table 4 reports the matching rates at rank 1, 5, 10, 20, and 50. Fig. 11 compares the CMC curve of our method and the other methods. The performance comparison from Table 4 and Fig. 11 suggest that compared to the other methods, our method has over 10% improvement in terms of matching accuracy at different ranks. Compared to the baseline methods using ℓ_1-norm distance, ℓ_2-norm-distance, and KISSME metric [26], in which the low-level appearance features are used for matching, the performance of the proposed method and RD [2] shows significant advantage due to the adoption of new feature representation instead of using low-level appearance features directly for matching.

Table 4
Matching rates (in %) on the PRID dataset at different ranks.

Rank	$r = 1$	5	10	20	50
Proposed	**27**	**45**	**56**	**69**	**93**
RD [2]	24	41	53	67	92
KISSME [26]	16	38	49	60	92
ℓ_1 distance	13	35	47	58	89
ℓ_2 distance	11	33	42	57	87

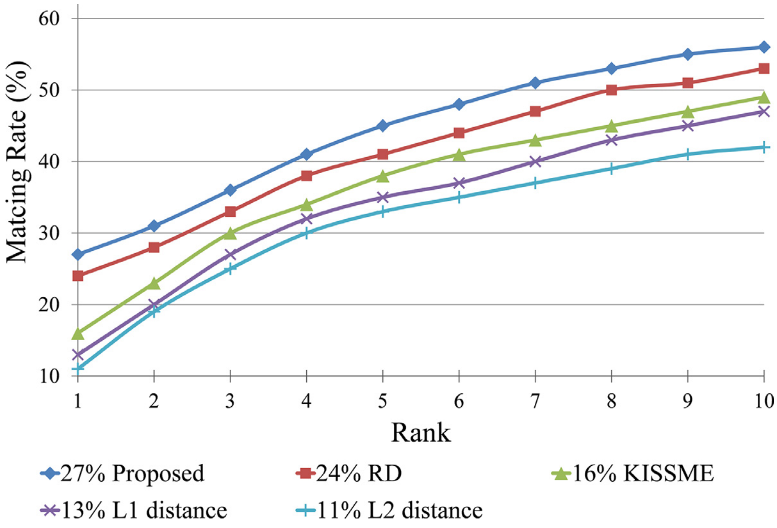

Fig. 11. CMC curves on the PRID dataset for the proposed method and the other methods.

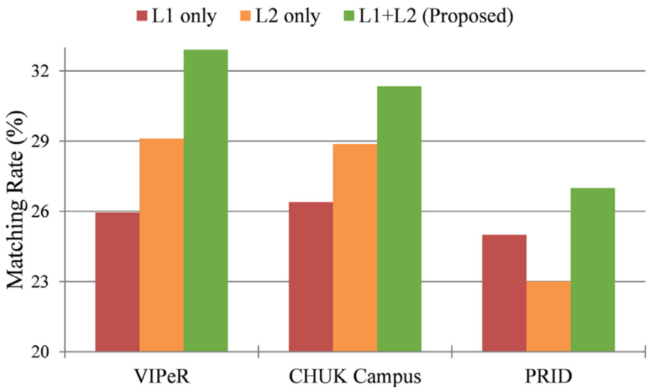

Fig. 12. Comparisons of the rank-1 matching rates on different datasets using the proposed method (ℓ_1 + ℓ_2), sparse representation only (ℓ_1), and ℓ_2 regularization only. Best viewed in color.

4.4.4. Effects of ℓ_2 regularization

To evaluate the effectiveness of the proposed sparse representation with ℓ_2 regularization. Fig. 12 shows the re-identification performance in terms of rank-1 matching rates with different regularization terms. On the VIPeR dataset, when sparse representation is used with ℓ_1 term ($\gamma_2 = 0$ in Eqs. (7) and (8)), the rank-1 matching rate is 25.95%. When the ℓ_1 regularization term is dropped while keeping the ℓ_2 regularization term ($\gamma_1 = 0$ in Eqs. (7) and (8)), a matching rate of 29.11% is achieved. The combination of ℓ_1 and ℓ_2, referred as the ℓ_2 regularized sparse representation, improves over the results using a single regularization term and brings up the rank-1 matching rate to 32.91%. This indicates that joint ℓ_1 and ℓ_2 regularization is effective for the proposed person re-identification approach. On the CUHK Campus dataset, the rank-1 matching rate is the highest (31.34%) by using both ℓ_1 and ℓ_2 terms together. The use of a single regularization term (ℓ_1 or ℓ_2) leads to less accurate matching rates of 26.39% and 28.87%, respectively. Similar observations hold for the PRID dataset for which ℓ_1 + ℓ_2 produces a better performance (27%) compared with using each of the regularization terms alone.

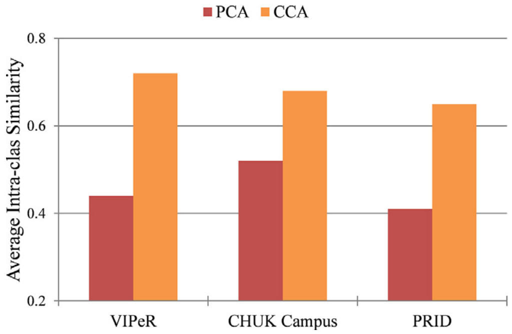

Fig. 13. Comparisons of the average intra-class similarity on different datasets after PCA and CCA projections.

4.4.5. *Effects of CCA projection*

As shown in Fig. 5, the manifold structures of the two views, after CCA projection, are more similar to each other than using a PCA projection. To quantitatively verify the effectiveness of CCA projection, we computed the average intra-class similarity, *i.e.*, the average similarity between projected features of the same person from different cameras. The similarity is computed by Eq. (9). The average intra-class similarities after PCA and CCA projection, for different datasets, are reported in Fig. 13. The similarities are computed from the testing data, while the PCA and CCA projections are learned from the training data. As observed, after CCA projection, the intra-class similarity is notably increased. In other words, CCA makes the features of the same subject in different views much closer in the transformed feature space, and this is very important in subsequent matching.

4.4.6. *Computational cost*

Regarding computational cost, we implement our method in Matlab on a computer with 2.4 GHz CPU and 8GB memory. On average, extracting feature from one image takes about 0.36 s. For dictionary learning, it takes about 10.23, 10.52, and 10.04 s on the VIPeR, CHUK Campus, and PRID dataset, respectively. The dictionary learning is based on the implementation from [46]. For matching one probe, less than 0.05 second is required on all the datasets. Further speedup can be achieved, for example, through parallel feature extraction.

4.4.7. *Discussion*

In our case, it is assumed that a probe is included in the gallery, which is also referred to as closed-set re-identification. This is also the assumption in most previous work. In case a probe is not included in the gallery, or commonly referred to as open-set re-identification, different strategies can be applied. For example, a simple way is to set up a threshold. Once the distance is above this threshold, it may be concluded that the probe does not have a match in the gallery. More advanced techniques, such as transfer local relative distance comparison (t-LRDC) model [72], have also been developed for open-set re-identification. In this work, we focus on closed-set re-identification problem, and open-set re-identification is our ongoing work.

Currently, our method is designed for a two-camera setting. For the case involving more than two cameras, coupled dictionary learning has to be performed on every camera pair, which would increase the computational cost. However, dictionary learning is an offline process and shall be performed only once.

5. Conclusions

In this paper, we proposed a person re-identification method using a sparse representation with ℓ_2 regularization. The ℓ_2 regularized sparse representations learned in a coherent subspace was used as a new feature representation instead of the appearance features for identity matching. Experiments were conducted on three publicly available datasets to evaluate the performance in single-shot and multi-shot re-identification settings. Compared to the state-of-the-art approaches, the proposed method achieved the highest matching rates in different scenarios. In addition, the experimental results suggested that sparse representation with ℓ_2 regularization has superior performance compared to the baseline methods with a single ℓ_1 or ℓ_2 regularization term.

Acknowledgment

This work was supported in part by the National Science Foundation under Grant 1330110 and in part by the Office of Naval Research under Grant N00014-12-1-1026.

References

[1] L. An, B. Bhanu, Face image super-resolution using 2D CCA, Signal Process. 103 (0) (2014) 184–194.

[2] L. An, M. Kafai, S. Yang, B. Bhanu, Reference-based person re-identification, in: Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2013, pp. 244–249.

[3] L. An, M. Kafai, S. Yang, B. Bhanu, Person re-identification with reference descriptor, IEEE Trans. Circuits Syst. Video Technol. PP (99) (2015) 1, doi:10.1109/TCSVT.2015.2416561.

[4] L. An, S. Yang, B. Bhanu, Person re-identification by robust canonical correlation analysis, IEEE Signal Process. Lett. 22 (8) (2015) 1103–1107.

[5] S. Bak, G. Charpiat, E. Corvee, F. Bremond, M. Thonnat, Learning to match appearances by correlations in a covariance metric space, in: Proceedings of the 2012 European Conference on Computer Vision (ECCV), 2012, pp. 806–820.

[6] A. Bedagkar-Gala, S.K. Shah, A survey of approaches and trends in person re-identification, Image Vis. Comput. 32 (4) (2014) 270–286.

[7] X. Chen, L. An, B. Bhanu, Reference set based appearance model for tracking across non-overlapping cameras, in: Proceedings of the Seventh International Conference on Distributed Smart Cameras (ICDSC), 2013, pp. 1–6.

[8] D.S. Cheng, M. Cristan, M. Stoppa, L. Bazzani, V. Murino, Custom pictorial structures for re-identification, in: Proceedings of the 2011 British Machine Vision Conference (BMVC), 2011, pp. 68.1–68.11.

[9] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: Proceedings of the 2007 International Conference on Machine learning (ICML), 2007, pp. 209–216.

[10] M. Dikmen, E. Akbas, T.S. Huang, N. Ahuja, Pedestrian recognition with a learned metric, in: Proceedings of the 2011 Asian Conference on Computer Vision (ACCV), 2011, pp. 501–512.

[11] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2004) 407–499.

[12] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2360–2367.

[13] Y. Gao, M. Wang, R. Ji, X. Wu, Q. Dai, 3-D object retrieval with Hausdorff distance learning, IEEE Trans. Ind. Electron. 61 (4) (2014) 2088–2098.

[14] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, in: Proceedings of the 2007 IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), 2007, pp. 1–7.

[15] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: Proceedings of the 2008 European Conference on Computer Vision (ECCV), 2008, pp. 262–275.

[16] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? Metric learning approaches for face identification, in: Proceedings of the 2009 IEEE International Conference on Computer Vision (ICCV), 2009, pp. 498–505.

[17] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2) (2006) 316–322.

[18] M. Hirzer, C. Beleznai, P.M. Roth, H. Bischof, Person re-identification by descriptive and discriminative classification, in: Proceedings of the 2011 Scandinavian Conference on Image analysis (SCIA), 2011, pp. 91–102.

[19] M. Hirzer, P.M. Roth, M. Köstinger, H. Bischof, Relaxed pairwise learned metric for person re-identification, in: Proceedings of the 2012 European Conference on Computer Vision (ECCV), 2012, pp. 780–793.

[20] H. Hotelling, Relations between two sets of variates, Biometrika 28 (3–4) (1936) pp.321–377.

[21] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou, S. Maybank, Principal axis-based correspondence between multiple cameras for people tracking, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) (2006) 663–671.

[22] H. Huang, H. He, X. Fan, J. Zhang, Super-resolution of human face image using canonical correlation analysis, Pattern Recognit. 43 (7) (2010) 2532–2543.

[23] O. Javed, K. Shafique, Z. Rasheed, M. Shah, Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views, Comput. Vis. Image Underst. 109 (2) (2008) 146–162.

[24] X.-Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, B. Xu, Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning, in: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 695–704.

[25] M. Khedher, M. El Yacoubi, B. Dorizzi, Multi-shot SURF-based person re-identification via sparse representation, in: Proceedings of the 2013 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2013, pp. 159–164.

[26] M. Köstinger, M. Hirzer, P. Wohlhart, P. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2288–2295.

[27] C.-H. Kuo, S. Khamis, V. Shet, Person re-identification using semantic color names and rankboost, in: Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), 2013, pp. 281–287.

[28] I. Kviatkovsky, A. Adam, E. Rivlin, Color invariants for person reidentification, IEEE Trans. Pattern Anal. Mach. Intell. 35 (7) (2013) 1622–1634.

[29] R. Layne, T. Hospedales, S. Gong, Person re-identification by attributes, in: Proceedings of the 2012 British Machine Vision Conference, 2012, pp. 24.1–24.11.

[30] A. Li, L. Liu, K. Wang, S. Liu, S. Yan, Clothing attributes assisted person reidentification, IEEE Trans. Circuits Syst. Video Technol. 25 (5) (2015) 869–878.

[31] W. Li, V. Mahadevan, N. Vasconcelos, Anomaly detection and localization in crowded scenes, IEEE Trans. Pattern Anal. Mach. Intell. 36 (1) (2014) 18–32.

[32] W. Li, X. Wang, Locally aligned feature transforms across views, in: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3594–3601.

[33] W. Li, R. Zhao, X. Wang, Human re-identification with transferred metric learning, in: Proceedings of the 2012 Asian Conference on Computer Vision (ACCV), 2012, pp. 31–44.

[34] W. Li, R. Zhao, T. Xiao, X. Wang, DeepReID: Deep filter pairing neural network for person re-identification, in: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 152–159.

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

[36] G. Lisanti, I. Masi, A. Bagdanov, A. Del Bimbo, Person re-identification by iterative re-weighted sparse ranking, IEEE Trans. Pattern Anal. Mach. Intell. 37 (8) (2015) 1629–1642.

[37] C. Liu, Discriminant analysis and similarity measure, Pattern Recognit. 47 (1) (2014) 359–367.

[38] C. Liu, S. Gong, C. Loy, X. Lin, Person re-identification: What features are important? in: Proceedings of the 2012 European Conference on Computer Vision Workshops and Demonstrations, 2012, pp. 391–401.

[39] C. Liu, S. Gong, C.C. Loy, On-the-fly feature importance mining for person re-identification, Pattern Recognit. 47 (4) (2014) 1602–1615.

[40] C. Liu, C.C. Loy, S. Gong, G. Wang, POP: Person re-identification post-rank optimisation, in: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), 2013, pp. 441–448.

[41] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, J. Bu, Semi-supervised coupled dictionary learning for person re-identification, in: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3550–3557.

[42] C.C. Loy, C. Liu, S. Gong, Person re-identification by manifold ranking, in: Proceedings of the 2013 IEEE International Conference on Image Processing (ICIP), 2013, pp. 3567–3571.

[43] P. Luo, X. Wang, X. Tang, Pedestrian parsing via deep decompositional network, in: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), 2013, pp. 2648–2655.

[44] A. Ma, J. Li, P. Yuen, P. Li, Cross-domain person reidentification using domain adaptation ranking svms, IEEE Trans. Image Process. 24 (5) (2015) 1599–1613.

[45] B. Ma, Y. Su, F. Jurie, BiCov: A novel image representation for person re-identification and face verification, in: Proceedings of the 2012 British Machine Vision Conference (BMVC), 2012, pp. 57.1–57.11.

[46] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: Proceedings of the 2009 International Conference on Machine Learning (ICML), 2009, pp. 689–696.

[47] N. Martinel, C. Micheloni, Re-identify people in wide area camera network, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2012, pp. 31–36.

[48] N. Martinel, C. Micheloni, C. Piciarelli, Learning pairwise feature dissimilarities for person re-identification, in: Proceedings of the Seventh International Conference on Distributed Smart Cameras (ICDSC), 2013, pp. 1–6.

[49] A. Mignon, F. Jurie, PCCA: A new approach for distance learning from sparse pairwise constraints, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2666–2672.

[50] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.

[51] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: A strategy employed by V1? Vis. Res. 37 (23) (1997) 3311–3325.

[52] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3318–3325.

[53] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by support vector ranking, in: Proceedings of the 2010 British Machine Vision Conference (BMVC), 2010, pp. 21.1–21.11.

[54] A. Sharma, M.A. Haj, J. Choi, L.S. Davis, D.W. Jacobs, Robust pose invariant face recognition using coupled latent space discriminant analysis, Comput. Vis. Image Underst. 116 (11) (2012) 1095–1110.

[55] Z. Shi, T.M. Hospedales, T. Xiang, Transferring a semantic representation for person re-identification and search, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4184–4193.

[56] D. Tao, L. Jin, Y. Wang, Y. Yuan, X. Li, Person re-identification by regularized smoothing kiss metric learning, IEEE Trans. Circuits Syst. Video Technol. 23 (10) (2013) 1675–1685.

[57] R. Vezzani, D. Baltieri, R. Cucchiara, People re-identification in surveillance and forensics: a survey, ACM Comput. Surv. 46 (2) (2013) 29:1–29:3.

[58] S. Wang, D. Zhang, Y. Liang, Q. Pan, Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2216–2223.

[59] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, Shape and appearance context modeling, in: Proceedings of the 2007 IEEE International Conference on Computer Vision (ICCV), 2007, pp. 1–8.

[60] Z. Wang, R. Hu, C. Liang, Y. Yu, J. Jiang, M. Ye, J. Chen, Q. Leng, Zero-shot person re-identification via cross-view consistency, IEEE Trans. Multimed. 18 (2) (2016) 260–272, doi:10.1109/TMM.2015.2505083.

[61] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (2009) 207–244.

[62] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.

[63] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person re-identification using kernel-based metric learning methods, in: Proceedings of the 2014 European Conference on Computer Vision (ECCV), 2014, pp. 1–16.

[64] D. Xu, H. Zheng, Person re-identification by multi-resolution saliency-weighted color histograms and local structural sparse coding, in: Proceedings of the Seventh International Conference on Image and Graphics (ICIG), 2013, pp. 477–482.

[65] J. Yang, Z. Wang, Z. Lin, S. Cohen, T. Huang, Coupled dictionary training for image super-resolution, IEEE Trans. Image Process. 21 (8) (2012) 3467–3478.

[66] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, S. Li, Salient color names for person re-identification, in: Proceedings of the 2014 European Conference on Computer Vision (ECCV), 2014, pp. 536–551.

[67] L. Zhang, D.V. Kalashnikov, S. Mehrotra, R. Vaisenberg, Context-based person identification framework for smart video surveillance, Mach. Vis. Appl. 25 (7) (2013) 1711–1725.

[68] R. Zhao, W. Ouyang, X. Wang, Person re-identification by salience matching, in: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), 2013, pp. 2528–2535.

[69] R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for person re-identification, in: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3586–3593.

[70] R. Zhao, W. Ouyang, X. Wang, Learning mid-level filters for person re-identification, in: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 144–151.

[71] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1116–1124.

[72] W. Zheng, S. Gong, T. Xiang, Towards open-world person re-identification by one-shot group-based verification, IEEE Trans. Pattern Anal. Mach. Intell. 38 (3) (2016) 591–606.

[73] W.-S. Zheng, S. Gong, T. Xiang, Transfer re-identification: From person to set-based verification, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2650–2657.

[74] W.-S. Zheng, S. Gong, T. Xiang, Reidentification by relative distance comparison, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 653–668.

[75] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, S. Gong, Partial person re-identification, in: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4678–4686.

[76] X. Zhou, B. Bhanu, Integrating face and gait for human recognition at a distance in video, IEEE Trans. Syst. Man Cybern. Part B Cybern. 37 (5) (2007) 1119–1137.

[77] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. Ser. B 67 (2005) 301–320.