Sparse Representation Matching for Person Re-Identification

Le An^a, Yue Gao^b, Xiaojing Chen^c, Songfan Yang^d, Bir Bhanu^e

^a*Department of Electrical Engineering, University of California, Riverside, USA*

^b*Department of Automation, Tsinghua University, China*

^c*Department of Computer Science, University of California, Riverside, USA*

^d*College of Electronics and Information Engineering, Sichuan University, China*

^e*Center for Research in Intelligent Systems, University of California, Riverside, USA*

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.

Keywords: Person re-identification, surveillance, subspace learning, sparse representation

Email address: lan004@ucr.edu (Le An)

Preprint submitted to Information Sciences

January 14, 2016

1. Introduction

The vast deployment of video surveillance cameras in public venues drives the needs for automated surveillance applications such as people tracking [20], anomaly detection in crowd [30], 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 [13, 12, 58, 31, 18, 74, 55, 67, 2]. 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:

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 [16, 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.

Figure 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.

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, 13, 12, 58].
2. Developing new matching methods using metric learning or ranking classifiers, e.g., [18, 74, 25, 52].

For the first group, discriminative appearance features are desired. Normally color and texture based features are widely used [27, 18]. 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 [61, 15, 9, 25]. SVM with ranking [52] 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 [57] and face image super-resolution [1], we first learn a transformation to project the original image features into

Figure 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 gallery for matching.

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 [24]. 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 [24] 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 [40], in which coupled dictionary learning was used. Our method is different from [40] in the following aspects: 1) the learning methodology is different. In [40], 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 feature spaces. In [40], 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 [40], 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 [40], 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 [40].

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.* [27] discovered an intra-distribution color structure and this structure was found invariant under a wide range of imaging conditions while

111 being discriminative. When used together with a covariance descriptor, state-
 112 of-the-art performance was achieved on different re-identification datasets.
 113 Yang *et al.* [66] proposed a salient color names based color descriptor which
 114 can guarantee that a higher probability will be assigned to the color name
 115 which is nearer to the color. This color descriptor can be computed efficiently
 116 in advance. Farenzena *et al.* [12] combined features including the overall chro-
 117 matic content, the spatial arrangement and the presence of recurrent local
 118 motifs to handle individual matching with appearance variation. Instead of
 119 designing handcrafted features, Gray *et al.* [14] used Adaboost to select the
 120 most discriminative features for person re-identification. Bak *et al.* [5] learned
 121 a model in a covariance metric space to select features based on the idea that
 122 different regions of each subject should be matched specifically. Martinel
 123 *et al.* [46] proposed a novel discriminative signature from multiple local fea-
 124 tures and designed a signature distance measure by exploiting different body
 125 parts. Cheng *et al.* [8] adopted pictorial structures to localize human parts
 126 and searched part-to-part correspondences to match subjects across different
 127 cameras. Ma *et al.* [44] used both biologically inspired features and covariance
 128 descriptors to handle background and illumination variations. To handle the
 129 appearance change in different cameras, Javed *et al.* [22] showed that bright-
 130 ness transfer functions from one camera to another lie in a subspace which
 131 can be learned by probabilistic PCA.

132 To explore higher level image features, Kuo *et al.* [26] applied semantic
 133 color names to describe an image of a person by probabilities of the presence
 134 of the predefined colors and a better performance was achieved than using
 135 only the color histograms. An *et al.* [2] proposed the use of reference de-
 136 scriptors instead of using image features directly. A reference descriptor for
 137 a probe or gallery was generated by computing the similarity scores between
 138 this probe or gallery to a reference set. This similarity-based representa-
 139 tion achieved state-of-the-art performance on a widely used dataset. Layne
 140 *et al.* [28] used mid-level semantic attributes such as gender for improved
 141 re-identification results compared to the existing approaches using low-level
 142 image features. Zhao *et al.* [69, 68] proposed unsupervised salient features,
 143 and it was shown that attributes can be combined with the existing methods
 144 to improve the recognition performance. In a recent work, Zhao *et al.* [70]
 145 learned discriminative mid-level filters from automatically discovered patch
 146 clusters and those filters were able to identify specific visual patterns in or-
 147 der to distinguish different persons. Li *et al.* [29] proposed to use clothing
 148 attributes to assist person re-identification. Shi *et al.* [54] invented a seman-

149 tic attribute model based on fashion photographs and then transferred this
150 model for person re-identification tasks.

151 Liu *et al.* [38] studied feature importance for person re-identification and
152 a method for on-the-fly mining of feature was proposed. Martinel *et al.* [47]
153 extracted multiple features from image pairs and obtained a so-called dis-
154 tance feature vector and re-identification was achieved by classifying the dis-
155 tance feature vector using a binary classifier. Xu *et al.* [64] proposed to use
156 multi-resolution color histograms and local structural sparse coding for re-
157 identification. In this method, the query image was sparsely represented by
158 the gallery image and then the sparse codes were used as a complementary
159 feature descriptor to the global appearance features. Liao *et al.* [34] proposed
160 a Local Maximal Occurrence (LOMO) feature, in which the horizontal oc-
161 currence of local features is used to achieve a stable representation against
162 view changes.

163 With the help of human feedback, Hirzer *et al.* [17] proposed a two-step
164 method by first using a descriptive model to obtain an initial ranking, which
165 was refined in the second step by human using a discriminative model. Liu
166 *et al.* [39] proposed a post-rank optimization method in which human was
167 allowed to select negative samples, achieving over 30% performance gain and
168 significantly reduced search time as compared to the exhaustive search.

169 2.2. Matching Methods

170 Standard metric learning techniques such as Large Margin Nearest Neighbor
171 (LMNN), Information Theoretic Metric Learning (ITML), and Logistic
172 Discriminant Metric Learning (LDML) have been applied to person re-
173 identification [61, 9, 15]. A variant of LMNN (LMNN-R) was developed
174 in [10] that introduced a reject option to the unfamiliar matches, which im-
175 proved re-identification results. A metric learning method using the idea
176 “keep it simple and straightforward” (KISSME) was proposed in [25] from
177 a statistical inference point of view such that a decision of dissimilarity can
178 be obtained by a likelihood ratio test. Tao *et al.* [55] extended the KISSME
179 formulation by introducing a regularization term and a smoothing term to
180 robustly estimate covariance matrices. In this way the instability in calcu-
181 lating the inverse of a covariance matrix from a small sized training set was
182 alleviated. Zheng *et al.* [74] formulated re-identification as a relative dis-
183 tance comparison problem in which the likelihood of the distance between
184 a pair of images of the same person being smaller than a pair of images
185 of different people was maximized. This framework was improved by Liu *et*

186 *al.* [37] with the incorporation of attribute information and feature weighting.
 187 A relaxed pairwise learned metric (RPLM) based on Mahalanobis distance
 188 learning was proposed in [18] which took advantages of the structure of the
 189 data with reduced computational cost and achieved good results with simple
 190 image features. Li *et al.* [33] proposed a filter pairing deep neural network in
 191 which misalignment, pose difference, occlusions and background clutter were
 192 jointed handled. However, training such a deep network requires a signifi-
 193 cant amount of data. Xiong *et al.* [63] applied multiple kernel-based metrics
 194 in conjunction with histogram-based features and showed improvement over
 195 state-of-the-art on several datasets. Wang *et al.* [60] proposed a data-specific
 196 adaptive metric method by applying the cross-view support and projection
 197 consistencies.

198 Instead of re-identifying people with absolute scores, Prosser *et al.* [52]
 199 formulated a relative ranking problem and an ensemble RankSVM was pro-
 200 posed to overcome the limited scalability in existing SVM-based ranking
 201 methods. Re-identification was reformulated as the problem of matching a
 202 probe to a watch list of people instead of matching to an individual in [73].
 203 To solve this problem, a transfer learning framework was proposed and differ-
 204 ent types of variations among a target set and non-target data were defined.
 205 A manifold ranking approach was developed by Loy *et al.* [41] such that the
 206 probe information was propagated along the data manifold in an unsuper-
 207 vised manner. It showed that the performance of the existing metric learning
 208 based methods could be significantly improved by integrating the manifold
 209 ranking. In [35], soft and hard re-weighting were applied to redistribute en-
 210 ergy among the sparse coefficients. This iterative process ensures that the
 211 best candidates are ranked at each iteration.

212 Subspace learning has also been explored for person re-identification. The
 213 image spaces of two camera views were jointly partitioned into different con-
 214 figurations based on the similarity of cross-view transforms in [31] and image
 215 pairs with similar transforms were projected into a common feature space
 216 for matching. A pairwise constrained component analysis (PCCA) was pro-
 217 posed in [48] to learn a low-dimensional mapping by complying with a set
 218 of sparse training pairwise constraints. Pedagadi *et al.* [51] used local Fisher
 219 discriminant analysis (LFDA) to reduce feature dimensionality for person re-
 220 identification and experimental results suggested that LFDA outperformed
 221 other metric learning-based methods.

222 From a different perspective, Jing *et al.* [23] introduced the problem of
 223 super-resolution person re-identification, in which low-resolution probe im-

224 ages are matched with high-resolution images. Since the subject may also be
 225 occluded in certain view, Zheng *et al.* [75] proposed a patch-level matching
 226 model to match subjects whose bodies are only partially visible. In a re-
 227 cent work [71], a large scale person re-identification dataset was introduced.
 228 This dataset is significantly larger than the existing ones and is more prac-
 229 tical towards real-world application. Ma *et al.* [43] tackled a new person
 230 re-identification problem without label information. Given the matched and
 231 unmatched image pairs from source domain cameras, as well as unmatched
 232 and unlabeled image pairs from target domain cameras, an adaptive ranking
 233 support vector machines (AdaRSVMs) method was developed for match-
 234 ing under target domain cameras without person labels. For more detailed
 235 discussion on person re-identification, comprehensive surveys are available
 236 in [56] and [6].

237 2.3. Contributions of This Paper

238 The main contributions of this paper are:

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

Figure 3: Sample segmentation results using the method in [42] to separate the foreground subject from the background. The appearance features are extracted from the foreground to mitigate the impact by the cluttered background.

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 compositional network based pedestrian parsing method [42] 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 [19] 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.

283 CCA has been applied to problems involving multi-view or multi-modality
 284 data such as image super-resolution [1, 21] and face recognition under pose
 285 variation [53].

286 For person re-identification, given N image pairs from two cameras A
 287 and B , appearance features with dimension m are first extracted from the
 288 images. These feature vectors are organized into two data matrices $X_A =$
 289 $\{x_A^i \in \mathbb{R}^m, i = 1, 2, \dots, N\}$ and $X_B = \{x_B^i \in \mathbb{R}^m, i = 1, 2, \dots, N\}$, in which x_A^i
 290 and x_B^i correspond to the same person in different views. The goal of CCA
 291 is to find a pair of projection vectors $w_A \in \mathbb{R}^m$ and $w_B \in \mathbb{R}^m$ such that the
 292 correlation coefficient ρ of $w_A^T X_A$ and $w_B^T X_B$ is maximized. Mathematically,
 293 the objective function to be maximized is given by

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

294 where cov is short for covariance and var computes data variance. $C_{AA} =$
 295 $E[X_A X_A^T]$ and $C_{BB} = E[X_B X_B^T]$ are the covariance matrices of X_A and X_B .
 296 $C_{AB} = E[X_A X_B^T]$ is the covariance matrix of X_A and X_B .

297 Eq. 1 can be reformulated as a constrained optimization problem as fol-
 298 lows

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

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

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

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

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

307 Fig. 4 demonstrates the CCA principle that projects the data from differ-
 308 ent views into a coherent subspace in which the data pair of the same person

Figure 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.

are maximally correlated. To validate the ability of CCA to find a coherent subspace, we use half of the data from the VIPeR dataset [13] 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 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

Figure 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 [13] and testing data are the rest of the same dataset. The numbers on the axes denote normalized feature values.

energy function to be optimized is

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

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

330 a linear combination of the base elements in a dictionary and this repre-
 331 sentation is naturally sparse [62]. The sparsity resembles human perception
 332 system in which the activation of neurons to an image is typically sparse [50].
 333 The ℓ_2 regularization term has the properties as in the Ridge Regression to
 334 stabilize the coefficients. Sparse representation with ℓ_2 regularization is also
 335 referred as Elastic Net in statistics [77].

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

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

337 where p^i and D are constructed by

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

338 To obtain the dictionary in Eq. 5, an online optimization algorithm based
 339 on stochastic approximations is used [45].

340 3.3. Sparse Representation with ℓ_2 Regularization

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

$$\operatorname{argmin}_{\alpha_{g^j}} \|g^j - D_A \alpha_{g^j}\|_2^2 + \gamma_1 \|\alpha_{g^j}\|_1 + \gamma_2 \|\alpha_{g^j}\|_2^2, \quad (7)$$

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

$$\operatorname{argmin}_{\alpha_p} \|p - D_B \alpha_p\|_2^2 + \gamma_1 \|\alpha_p\|_1 + \gamma_2 \|\alpha_p\|_2^2. \quad (8)$$

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

353 3.4. Identity Matching

354 With the sparse representations of probe and gallery ready, the matching
 355 is based on the similarity between the sparse coefficients α_p of a probe and
 356 the sparse coefficients α_g^j of a gallery. We adopt a modified cosine similarity
 357 measure [36] defined by

$$\text{sim}(\alpha_p, \alpha_{g^j}) = \frac{|(\alpha_p)^T \cdot \alpha_{g^j}|}{\|\alpha_p\| \cdot \|\alpha_{g^j}\| \cdot (\|\alpha_p - \alpha_{g^j}\|_P + \epsilon)}, \quad (9)$$

358 where $\|\cdot\|_P$ is the L_P norm and ϵ is a small positive number to prevent di-
 359 vision by zero. The reason to apply the modified cosine similarity is that
 360 the standard cosine similarity does not take into account the actual distance
 361 between two vectors, while the modified cosine similarity is able to address
 362 both the distance measure and angular measure and has shown improved per-
 363 formance in recognition tasks [36]. **The modified cosine similarity has shown**
 364 **to be useful for person re-identification in [3]. We use the same features and**
 365 **evaluation protocols as those in [3], and we have empirically found out that**
 366 **the modified cosine similarity is better than the standard cosine similarity in**
 367 **our case.**

368 3.5. Summary of the Algorithm

369 The proposed algorithm for person re-identification consists of training
 370 and testing phases. In the training phase, CCA projection matrices are
 371 learned as well as the coupled dictionaries for different camera views. Given
 372 probe and gallery in the testing phase, the appearance features are extracted
 373 and projected into the CCA subspace. Then the ℓ_2 regularized sparse repre-
 374 sentations for probe and gallery are obtained using jointly learned dictionar-
 375 ies. The similarity scores between probe and gallery are computed using their
 376 sparse representations. The algorithms for training and testing algorithms
 377 are summarized in Algorithm 1 and Algorithm 2, respectively.

378 4. Experiments

379 4.1. Datasets

380 We evaluate our method on three different publicly available datasets
 381 for two typical re-identification scenarios, namely image-based case (single-
 382 shot) and video-based case (multi-shot). For single-shot scenario, the VIPeR

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 ℓ_2 regularized sparse representations for probe and gallery using Eq. 7 and Eq. 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

383 dataset and the CUHK Campus dataset are used. The VIPeR dataset¹ en-
384 rolled 632 persons and is one of the most challenging and widely evaluated
385 benchmark datasets for person re-identification [13]. Some sample image
386 pairs from this dataset are shown in Fig. 6. For each person, one detection

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

Figure 6: Sample image pairs from the VIPeR dataset [13].

Figure 7: Sample image pairs from the CUHK Campus dataset [32].

387 is available in each of the two non-overlapping cameras with varying illumi-
 388 nation and cluttered background. For most of the subjects the view change
 389 is more than 90 degrees. In addition, partial occlusion is frequent due to the
 390 subjects carrying items such as backpack or handbag.

391 The CUHK Campus dataset² is a recently released dataset which contains
 392 images of 971 subjects from two non-overlapping camera views [32]. One
 393 camera captures the frontal or rear view of a person and the other camera
 394 captures the profile view of a person. Each person has two detections in each
 395 camera view. Some sample image pairs from this dataset are shown in Fig. 7.

396 For multi-shot video-based re-identification, we use the Person Re-ID 2011
 397 (PRID) dataset³. The PRID dataset consists of multiple person trajectories
 398 recorded from two surveillance cameras. Camera *A* contains 385 persons and
 399 camera *B* shows 749 persons. The first 200 persons appear in both camera
 400 views. Each trajectory contains approximately 100 to 150 images depending
 401 on the walking speed of a person. Two segments of trajectories of the same
 402 person in two cameras are shown in Fig. 8.

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

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

(a)

(b)

Figure 8: Sample image pairs from the PRID dataset [17]. (a) Trajectory of a person in camera *A*. (b) Corresponding trajectory in camera *B*.

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 [18]. 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 [26] as an additional color representation. The texture features are represented by the 8-bit Local Binary Patterns (LBP) [49]. 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, 18, 4]. 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 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] [25] [18] 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 [17]. 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, recognition 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 recognition 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.

457 We compare our approach with the following 18 state-of-the-art alternatives,
 458 which are reference-based approach (RD) [2], saliency matching (SalMatch) [68],
 459 relaxed pairwise learned metric (RPLM) [18], **Semi-supervised coupled dic-**
 460 **tionary (SSCDL)** [40], regularized smoothing KISS metric learning (RS-
 461 KISS) [55], custom pictorial structures (CPS) [8], biologically inspired fea-
 462 tures and covariance descriptors (BiCov) [44], KISS metric (KISSME) [25],
 463 large margin nearest neighbor with rejection (LMNN-R) [10], symmetry-
 464 driven accumulation of local features (SDALF) [12], manifold ranking (MRank) [41],
 465 pairwise constrained component analysis (PCCA) [48], descriptive and dis-
 466 criminative classification (DDC) [17], large margin nearest neighbor (LMNN) [61],
 467 attributed-based relative distance comparison (aPRDC) [37], relative dis-
 468 tance comparison (PRDC) [74], information-theoretic metric learning (ITML) [9],
 469 support vector ranking (RankSVM) [52], and ensemble of localized features
 470 (ELF) [14].

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

481 To study the impact of reduced training data size and to make comparison
 482 with other methods, we report in Table 2 the re-identification results with
 483 different training data sizes. In this case, all the data from the VIPeR dataset
 484 are used. As the size of the training set decreases, the number of subjects in
 485 the gallery and probe data increases, which makes the re-identification more
 486 difficult. The same experiment protocol was used in RD [2], RPLM [18],
 487 and PRDC [74], the results of which are included in Table 2 for comparison.
 488 The results shown in Table 2 suggest that with a smaller training set the
 489 proposed method is still able to perform better than the competing methods
 490 at different ranks.

491 4.4.2. The CUHK Campus Dataset

492 For the CUHK Campus dataset, we compare the proposed approach with
 493 the following methods: RD [2], SalMatch [68], SDALF [12], LMNN [61],

Table 1: Recognition 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 [18]	27.34	69.02	82.69	94.56	98.54
SSCDL [40]	25.60	68.10	83.60	95.20	—
RS-KISS [55]	24.50	66.60	81.70	93.50	98.00
CPS [8]	21.84	57.21	71.00	87.00	91.77
BiCov [44]	20.66	56.18	68.00	81.56	88.66
KISSME [25]	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 [41]	19.34	55.51	70.44	87.69	96.90
PCCA [48]	19.27	64.91	80.28	95.00	97.01
DDC [17]	19.00	52.00	65.00	80.00	91.00
LMNN [61]	17.41	53.86	67.88	88.13	96.23
aPRDC [37]	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 [52]	14.00	51.00	67.00	85.00	94.00
ELF [14]	12.00	43.00	60.00	81.00	93.00

ITML [9], as well as direct matching using ℓ_2 and ℓ_1 distance. Table 3 reports the recognition rates at rank 1, 10, 20, 50, and 100. As compared to the other methods with all the recognition rates below 30%, the proposed method achieves a rank-1 recognition 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 [25], as well as two baseline methods using

Figure 9: CMC curves on the VIPeR dataset for the proposed method and the other methods. Best viewed in color.

Table 2: Recognition 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 [18]	19.51	56.44	71.09	10.88	37.69	51.64
PRDC [74]	12.64	44.28	59.95	9.12	34.4	48.55

504 ℓ_1 distance and ℓ_2 distance. Table 4 reports the recognition rates at rank 1,
 505 5, 10, 20, and 50. Fig. 11 compares the CMC curve of our method and
 506 the other methods. The performance comparison from Table 4 and Fig. 11
 507 suggest that compared to the other methods, our method has over 10% im-
 508 provement in terms of matching accuracy at different ranks. Compared to

Table 3: Recognition 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

Figure 10: CMC curves on the CUHK Campus dataset for the proposed method and the other methods. Best viewed in color.

509 the baseline methods using ℓ_1 -norm distance, ℓ_2 -norm-distance, and KISSME
510 metric [25], in which the low-level appearance features are used for match-
511 ing, the performance of the proposed method and RD [2] shows significant

Table 4: Recognition 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 [25]	16	38	49	60	92
ℓ_1 distance	13	35	47	58	89
ℓ_2 distance	11	33	42	57	87

Figure 11: CMC curves on the PRID dataset for the proposed method and the other methods. Best viewed in color.

512 advantage due to the adoption of new feature representation instead of using
513 low-level appearance features directly for matching.

514 4.4.4. Effects of ℓ_2 Regularization

515 To evaluate the effectiveness of the proposed sparse representation with
516 ℓ_2 regularization. Fig. 12 shows the re-identification performance in terms of

Figure 12: Comparisons of the rank-1 recognition rates on different datasets using the proposed method ($\ell_1 + \ell_2$), sparse representation only (ℓ_1), and ℓ_2 regularization only. Best viewed in color.

rank-1 recognition rates with different regularization terms. On the VIPeR dataset, when sparse representation is used with ℓ_1 term ($\gamma_2 = 0$ in Eq. 7 and Eq. 8), the rank-1 recognition rate is 25.95%. When the ℓ_1 regularization term is dropped while keeping the ℓ_2 regularization term ($\gamma_1 = 0$ in Eq. 7 and Eq. 8), a recognition 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 recognition 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 recognition 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 recognition rates of 26.39% and 28.87%, respectively. Similar observations hold for the PRID dataset for which $\ell_1 + \ell_2$ produces a better performance (27%) compared with using each of the regularization terms alone.

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

Figure 13: Comparisons of the average intra-class similarity on different datasets after PCA and CCA projections. Best viewed in color.

537 features of the same person from different cameras. The similarity is com-
 538 puted by Eq. (9). The average intra-class similarities after PCA and CCA
 539 projection, for different datasets, are reported in Fig. 13. The similarities
 540 are computed from the testing data, while the PCA and CCA projections
 541 are learned from the training data. As observed, after CCA projection, the
 542 intra-class similarity is notably increased. In other words, CCA makes the
 543 features of the same subject in different views much closer in the transformed
 544 feature space, and this is very important in subsequent matching.

545 4.4.6. Computational Cost

546 Regarding computational cost, we implement our method in Matlab on
 547 a computer with 2.4GHz CPU and 8GB memory. On average, extracting
 548 feature from one image takes about 0.36 second. For dictionary learning, it
 549 takes about 10.23, 10.52, and 10.04 seconds on the VIPeR, CHUK Campus,
 550 and PRID dataset, respectively. The dictionary learning is based on the
 551 implementation from [45]. For matching one probe, less than 0.05 second is
 552 required on all the datasets. Further speedup can be achieved, for example,
 553 through parallel feature extraction.

554 4.4.7. Discussion

555 In our case, it is assumed that a probe is included in the gallery, which
 556 is also referred to as closed-set re-identification. This is also the assumption
 557 in most previous work. In case a probe is not included in the gallery, or
 558 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 recognition 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.

References

- [1] L. An, B. Bhanu, Face image super-resolution using 2D CCA, *Signal Processing* 103 (0) (2014) 184 – 194.
- [2] L. An, M. Kafai, S. Yang, B. Bhanu, Reference-based person re-identification, in: *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 Transactions on Circuits and Systems for Video Technology* PP (99) (2015) 1–1.

- 591 [4] L. An, S. Yang, B. Bhanu, Person re-identification by robust canonical
592 correlation analysis, *IEEE Signal Processing Letters* 22 (8) (2015) 1103–
593 1107.
- 594 [5] S. Bak, G. Charpiat, E. Corvee, F. Bremond, M. Thonnat, Learning
595 to match appearances by correlations in a covariance metric space, in:
596 *European Conference on Computer Vision (ECCV)*, 2012, pp. 806–820.
- 597 [6] A. Bedagkar-Gala, S. K. Shah, A survey of approaches and trends in
598 person re-identification, *Image and Vision Computing* 32 (4) (2014) 270
599 – 286.
- 600 [7] X. Chen, L. An, B. Bhanu, Reference set based appearance model for
601 tracking across non-overlapping cameras, in: *ACM/IEEE International
602 Conference on Distributed Smart Cameras (ICDSC)*, 2013.
- 603 [8] D. S. Cheng, M. Cristan, M. Stoppa, L. Bazzani, V. Murino, Custom
604 pictorial structures for re-identification, in: *British Machine Vision Con-
605 ference (BMVC)*, 2011, pp. 68.1–68.11.
- 606 [9] J. V. Davis, B. Kulis, P. Jain, S. Sra, I. S. Dhillon, Information-
607 theoretic metric learning, in: *International Conference on Machine
608 learning (ICML)*, 2007, pp. 209–216.
- 609 [10] M. Dikmen, E. Akbas, T. S. Huang, N. Ahuja, Pedestrian recogni-
610 tion with a learned metric, in: *Asian conference on Computer vision
611 (ACCV)*, 2011, pp. 501–512.
- 612 [11] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression,
613 *Annals of Statistics* 32 (2004) 407–499.
- 614 [12] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Per-
615 son re-identification by symmetry-driven accumulation of local features,
616 in: *IEEE Conference on Computer Vision and Pattern Recognition
617 (CVPR)*, 2010, pp. 2360–2367.
- 618 [13] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recog-
619 nition, reacquisition, and tracking, in: *IEEE International Workshop on
620 Performance Evaluation of Tracking and Surveillance (PETS)*, 2007.

- 621 [14] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an
 622 ensemble of localized features, in: European Conference on Computer
 623 Vision (ECCV), 2008, pp. 262–275.
- 624 [15] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? Metric learning
 625 approaches for face identification, in: IEEE International Conference on
 626 Computer Vision (ICCV), 2009, pp. 498–505.
- 627 [16] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE
 628 Transactions on Pattern Analysis and Machine Intelligence 28 (2) (2006)
 629 316–322.
- 630 [17] M. Hirzer, C. Beleznaï, P. M. Roth, H. Bischof, Person re-identification
 631 by descriptive and discriminative classification, in: Scandinavian Con-
 632 ference on Image analysis (SCIA), 2011, pp. 91–102.
- 633 [18] M. Hirzer, P. M. Roth, M. Köstinger, H. Bischof, Relaxed pairwise
 634 learned metric for person re-identification, in: European Conference on
 635 Computer Vision (ECCV), 2012, pp. 780–793.
- 636 [19] H. Hotelling, Relations between two sets of variates, Biometrika 28 (3/4)
 637 (1936) pp. 321–377.
- 638 [20] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou, S. Maybank, Principal axis-
 639 based correspondence between multiple cameras for people tracking,
 640 IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (4)
 641 (2006) 663–671.
- 642 [21] H. Huang, H. He, X. Fan, J. Zhang, Super-resolution of human face
 643 image using canonical correlation analysis, Pattern Recognition 43 (7)
 644 (2010) 2532 – 2543.
- 645 [22] O. Javed, K. Shafique, Z. Rasheed, M. Shah, Modeling inter-camera
 646 space-time and appearance relationships for tracking across non-
 647 overlapping views, Computer Vision and Image Understanding 109 (2)
 648 (2008) 146 – 162.
- 649 [23] X.-Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, B. Xu, Super-
 650 resolution person re-identification with semi-coupled low-rank discrimi-
 651 nant dictionary learning, in: IEEE Conference on Computer Vision and
 652 Pattern Recognition (CVPR), 2015, pp. 695–704.

- 653 [24] M. Khedher, M. El Yacoubi, B. Dorizzi, Multi-shot SURF-based per-
654 son re-identification via sparse representation, in: IEEE International
655 Conference on Advanced Video and Signal Based Surveillance (AVSS),
656 2013, pp. 159–164.
- 657 [25] M. Köstinger, M. Hirzer, P. Wohlhart, P. Roth, H. Bischof, Large scale
658 metric learning from equivalence constraints, in: IEEE Conference on
659 Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2288–
660 2295.
- 661 [26] C.-H. Kuo, S. Khamis, V. Shet, Person re-identification using semantic
662 color names and rankboost, in: IEEE Workshop on Applications of
663 Computer Vision (WACV), 2013, pp. 281–287.
- 664 [27] I. Kviatkovsky, A. Adam, E. Rivlin, Color invariants for person reiden-
665 tification, IEEE Transactions on Pattern Analysis and Machine Intelli-
666 gence 35 (7) (2013) 1622–1634.
- 667 [28] R. Layne, T. Hospedales, S. Gong, Person re-identification by attributes,
668 in: Proceedings of the British Machine Vision Conference, 2012, pp.
669 24.1–24.11.
- 670 [29] A. Li, L. Liu, K. Wang, S. Liu, S. Yan, Clothing attributes assisted
671 person reidentification, IEEE Transactions on Circuits and Systems for
672 Video Technology 25 (5) (2015) 869–878.
- 673 [30] W. Li, V. Mahadevan, N. Vasconcelos, Anomaly detection and local-
674 ization in crowded scenes, IEEE Transactions on Pattern Analysis and
675 Machine Intelligence 36 (1) (2014) 18–32.
- 676 [31] W. Li, X. Wang, Locally aligned feature transforms across views,
677 in: IEEE Conference on Computer Vision and Pattern Recognition
678 (CVPR), 2013, pp. 3594–3601.
- 679 [32] W. Li, R. Zhao, X. Wang, Human reidentification with transferred met-
680 ric learning, in: Asian conference on Computer vision (ACCV), 2012,
681 pp. 31–44.
- 682 [33] W. Li, R. Zhao, T. Xiao, X. Wang, DeepReID: Deep filter pairing neural
683 network for person re-identification, in: IEEE Conference on Computer
684 Vision and Pattern Recognition (CVPR), No. 152-159, 2014.

- [34] S. Liao, Y. Hu, X. Zhu, S. Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.
- [35] G. Lisanti, I. Masi, A. Bagdanov, A. Del Bimbo, Person re-identification by iterative re-weighted sparse ranking, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (8) (2015) 1629–1642.
- [36] C. Liu, Discriminant analysis and similarity measure, Pattern Recognition 47 (1) (2014) 359 – 367.
- [37] C. Liu, S. Gong, C. Loy, X. Lin, Person re-identification: What features are important?, in: European Conference on Computer Vision Workshops and Demonstrations, 2012, pp. 391–401.
- [38] C. Liu, S. Gong, C. C. Loy, On-the-fly feature importance mining for person re-identification, Pattern Recognition 47 (4) (2014) 1602 – 1615.
- [39] C. Liu, C. C. Loy, S. Gong, G. Wang, POP: Person re-identification post-rank optimisation, in: IEEE International Conference on Computer Vision (ICCV), 2013.
- [40] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, J. Bu, Semi-supervised coupled dictionary learning for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3550–3557.
- [41] C. C. Loy, C. Liu, S. Gong, Person re-identification by manifold ranking, in: IEEE International Conference on Image Processing (ICIP), 2013, pp. 3567–3571.
- [42] P. Luo, X. Wang, X. Tang, Pedestrian parsing via deep compositional network, in: IEEE International Conference on Computer Vision (ICCV), 2013, pp. 2648–2655.
- [43] A. Ma, J. Li, P. Yuen, P. Li, Cross-domain person reidentification using domain adaptation ranking svms, IEEE Transactions on Image Processing 24 (5) (2015) 1599–1613.
- [44] B. Ma, Y. Su, F. Jurie, BiCov: a novel image representation for person re-identification and face verification, in: British Machine Vision Conference (BMVC), 2012, pp. 57.1–57.11.

- 717 [45] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning
718 for sparse coding, in: International Conference on Machine Learning
719 (ICML), 2009, pp. 689–696.
- 720 [46] N. Martinel, C. Micheloni, Re-identify people in wide area camera net-
721 work, in: IEEE Conference on Computer Vision and Pattern Recogni-
722 tion Workshops (CVPRW), 2012, pp. 31–36.
- 723 [47] N. Martinel, C. Micheloni, C. Picciarelli, Learning pairwise feature dis-
724 similarities for person re-identification, in: ACM/IEEE International
725 Conference on Distributed Smart Cameras (ICDSC), 2013.
- 726 [48] A. Mignon, F. Jurie, PCCA: A new approach for distance learning from
727 sparse pairwise constraints, in: IEEE Conference on Computer Vision
728 and Pattern Recognition (CVPR), 2012, pp. 2666–2672.
- 729 [49] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and
730 rotation invariant texture classification with local binary patterns, IEEE
731 Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002)
732 971–987.
- 733 [50] B. A. Olshausen, D. J. Field, Sparse coding with an overcomplete basis
734 set: A strategy employed by V1?, Vision Research 37 (23) (1997) 3311
735 – 3325.
- 736 [51] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discrim-
737 inant analysis for pedestrian re-identification, in: IEEE Conference on
738 Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3318–
739 3325.
- 740 [52] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by
741 support vector ranking, in: British Machine Vision Conference (BMVC),
742 2010, pp. 21.1–21.11.
- 743 [53] A. Sharma, M. A. Haj, J. Choi, L. S. Davis, D. W. Jacobs, Robust
744 pose invariant face recognition using coupled latent space discriminant
745 analysis, Computer Vision and Image Understanding 116 (11) (2012)
746 1095 – 1110.

- 747 [54] Z. Shi, T. M. Hospedales, T. Xiang, Transferring a semantic representation for person re-identification and search, in: IEEE Conference on
748 Computer Vision and Pattern Recognition (CVPR), 2015.
749
- 750 [55] D. Tao, L. Jin, Y. Wang, Y. Yuan, X. Li, Person re-identification by regularized smoothing kiss metric learning, IEEE Transactions on Circuits
751 and Systems for Video Technology 23 (10) (2013) 1675–1685.
752
- 753 [56] R. Vezzani, D. Baltieri, R. Cucchiara, People re-identification in surveillance and forensics: a survey, ACM Computing Surveys 46 (2) (2013)
754 29:1–29:3.
755
- 756 [57] S. Wang, D. Zhang, Y. Liang, Q. Pan, Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis, in: IEEE Conference on Computer Vision and Pattern Recognition
757 (CVPR), 2012, pp. 2216–2223.
758
759
- 760 [58] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, Shape and appearance context modeling, in: IEEE International Conference on
761 Computer Vision (ICCV), 2007, pp. 1–8.
762
- 763 [59] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, Shape and appearance context modeling, in: IEEE 11th International Conference
764 on Computer Vision (ICCV), 2007, pp. 1–8.
765
- 766 [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
767 Transactions on Multimedia PP (99) (2015) 1–1.
768
- 769 [61] K. Q. Weinberger, L. K. Saul, Distance metric learning for large margin nearest neighbor classification, Journal of Machine Learning Research
770 10 (2009) 207–244.
771
- 772 [62] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and
773 Machine Intelligence 31 (2) (2009) 210–227.
774
- 775 [63] F. Xiong, M. Gou, O. Camps, M. Sznajder, Person re-identification using kernel-based metric learning methods, in: European Conference on
776 Computer Vision (ECCV), 2014, pp. 1–16.
777

- 778 [64] D. Xu, H. Zheng, Person re-identification by multi-resolution saliency-
779 weighted color histograms and local structural sparse coding, in: Seventh
780 International Conference on Image and Graphics (ICIG), 2013, pp. 477–
781 482.
- 782 [65] J. Yang, Z. Wang, Z. Lin, S. Cohen, T. Huang, Coupled dictionary train-
783 ing for image super-resolution, IEEE Transactions on Image Processing
784 21 (8) (2012) 3467–3478.
- 785 [66] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, S. Li, Salient color names for
786 person re-identification, in: European Conference on Computer Vision
787 (ECCV), 2014, pp. 536–551.
- 788 [67] L. Zhang, D. V. Kalashnikov, S. Mehrotra, R. Vaisenberg, Context-
789 based person identification framework for smart video surveillance, Ma-
790 chine Vision and Applications (2013) 1–15.
- 791 [68] R. Zhao, W. Ouyang, X. Wang, Person re-identification by salience
792 matching, in: IEEE International Conference on Computer Vision
793 (ICCV), 2013.
- 794 [69] R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for per-
795 son re-identification, in: IEEE Conference on Computer Vision and Pat-
796 tern Recognition (CVPR), 2013, pp. 3586–3593.
- 797 [70] R. Zhao, W. Ouyang, X. Wang, Learning mid-level filters for person
798 re-identification, in: IEEE Conference on Computer Vision and Pattern
799 Recognition (CVPR), No. 144-151, 2014.
- 800 [71] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person
801 re-identification: A benchmark, in: The IEEE International Conference
802 on Computer Vision (ICCV), 2015.
- 803 [72] W. Zheng, S. Gong, T. Xiang, Towards open-world person re-
804 identification by one-shot group-based verification, IEEE Transactions
805 on Pattern Analysis and Machine Intelligence PP (99) (2015) 1–1.
- 806 [73] W.-S. Zheng, S. Gong, T. Xiang, Transfer re-identification: From person
807 to set-based verification, in: IEEE Conference on Computer Vision and
808 Pattern Recognition (CVPR), 2012, pp. 2650–2657.

- 809 [74] W.-S. Zheng, S. Gong, T. Xiang, Reidentification by relative distance
810 comparison, *IEEE Transactions on Pattern Analysis and Machine Intel-*
811 *ligence* 35 (3) (2013) 653–668.
- 812 [75] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, S. Gong, Partial person re-
813 identification, in: *IEEE International Conference on Computer Vision*
814 *(ICCV)*, 2015.
- 815 [76] X. Zhou, B. Bhanu, Integrating face and gait for human recognition at
816 a distance in video, *IEEE Transactions on Systems, Man, and Cyber-*
817 *netics, Part B: Cybernetics* 37 (5) (2007) 1119–1137.
- 818 [77] H. Zou, T. Hastie, Regularization and variable selection via the elastic
819 net, *Journal of the Royal Statistical Society, Series B* 67 (2005) 301–320.