Person Re-identification by Unsupervised ℓ_1 Graph Learning

Elyor Kodirov, Tao Xiang, Zhenyong Fu and Shaogang Gong

School of EECS, Queen Mary University of London, UK
{e.kodirov, t.xiang, z.fu, s.gong}@qmul.ac.uk

Abstract. Most existing person re-identification (Re-ID) methods are based on supervised learning of a discriminative distance metric. They thus require a large amount of labelled training image pairs which severely limits their scalability. In this work, we propose a novel unsupervised Re-ID approach which requires no labelled training data yet is able to capture discriminative information for cross-view identity matching. Our model is based on a new graph regularised dictionary learning algorithm. By introducing a ℓ_1-norm graph Laplacian term, instead of the conventional squared ℓ_2-norm, our model is robust against outliers caused by dramatic changes in background, pose, and occlusion typical in a Re-ID scenario. Importantly we propose to learn jointly the graph and representation resulting in further alleviation of the effects of data outliers. Experiments on four benchmark datasets demonstrate that the proposed model significantly outperforms the state-of-the-art unsupervised learning based alternatives whilst being extremely efficient to compute.

Keywords: unsupervised person re-id, dictionary learning, robust graph regularisation, graph learning

1 Introduction

The problem of matching people across non-overlapping cameras, known as person re-identification (Re-ID), has drawn a great deal of attention recently [53, 20]. It remains an unsolved problem due to two reasons: (1) A person's appearance often changes dramatically across cameras views due to occlusion, lighting, illumination and pose changes; (2) Many people in public spaces wear similar clothes (e.g. dark coats, jeans) thus having similar visual appearance.

Most recent Re-ID methods are based on supervised learning. Given a set of labelled training data consisting of images of people paired across camera views according to identity, a distance metric is learned either using hand-crafted features [9, 60, 46, 37, 49, 38, 48, 19, 25, 14, 31, 56, 58, 55], or end-to-end using deep neural networks [2, 36]. However, they require images of hundreds or more people to be paired across each pair of camera views which is both tedious and sometimes not possible – some people do not reappear in other camera views. This severely limits the scalability of the existing methods making them unsuitable for practical large scale Re-ID tasks. To overcome this problem, a number of unsupervised Re-ID methods have been proposed [57, 54, 30, 41]. However, without

(a) (b)

Fig. 1: An illustration of graph learning for person re-id. (a) A graph constructed in the original low-level feature space; (b) A graph learned using the proposed model in this work. One graph node and its five connected neighbours are shown, with the neighbour capturing the same person highlighted in red.

labelled training data, they can only focus on learning salient and view invariant representations. Their performance is thus much weaker compared to the supervised methods. This is because they are unable to learn the cross-view discriminative information effectively, critical for matching the same person whilst separating the person from imposters of similar appearance. Due to their uncompetitiveness in published benchmarking metrics, these unsupervised learning models have received little attention when practicality and scalability are not considered in current benchmarking.

In this work, we propose to learn a low-dimensional feature representation from a set of unlabelled data that can be easily collected. To learn a feature representation that is both view-invariant and discriminative, we exploit dictionary learning models that are shared across camera views. It is easy to understand how a representation obtained by dictionary learning can be view-invariant and low-dimensional – dictionary learning is widely used as an unsupervised model for dimensionality reduction [28, 1, 43]; and by sharing the same dictionary across camera views, it intrinsically requires that the learned representation to be view-invariant. It is the discriminative part that is non-trivial: How can we enforce that the learned representation is good for matching people across camera views, without the discriminative information from a set of paired training data?

Our solution is to relax the definition of discriminativity. Consider each dictionary word as a new feature dimension, a learned dictionary defines a subspace, into which the original data points represented by high-dimensional low-level feature vectors are projected. Instead of enforcing that data points corresponding to the *same* person to be as close as possible whilst being further away from other people in the learned subspace as in supervised learning, we constrain the visually *similar* people to be close to each other. Without identity labels, this is obviously a weaker constraint but the best available. Specifically, discriminativ-

ity is achieved unsupervised via a visual similarity constraint, which is enforced by introducing a graph Laplacian regularisation term in the dictionary learning objective function [44].

However, two problems remain when the conventional graph Laplacian constraint is used in our problem context: (1) The conventional term has a squared ℓ_2-norm, which makes the term susceptible to data outliers. This is particularly unsuitable for the Re-ID problem as there are plenty of data outliers in Re-ID, caused by various reasons such as the person detection boxes being imperfect and severe (self-)occlusions. (2) The visual similarity is encoded in a graph whose topology and edge weights are all determined by distances computed using the original high-dimensional low-level features. However, these features are not ideal for people matching, hence learning a new representation in the first place. As illustrated in Fig. 1(a), a graph constructed using the low-level features connects many visually dissimilar neighbours to each node. This diminishes the power of the graph regularisation term as a visual similarity constraint.

To overcome these two problems, we introduce a robust graph regularisation term and propose to learn the new representation and the optimal graph jointly. Specifically, a ℓ_1-norm is introduced in our graph regularisation term to make it robust against outliers. With this ℓ_1-norm and joint graph and dictionary learning, our learning objective function is both non-smooth and non-convex. Solving this optimisation problem is thus non-trivial. An efficient iterative optimisation algorithm is formulated in this work to solve it. Once learned, our model can be used to compute a representation for each image much more efficiently than any existing unsupervised Re-ID method. The final matching is done by computing a simple cosine distance between a pair of the representation vectors.

1.1 Related Work

Most existing person Re-ID techniques are based on supervised learning: After hand-crafted features are extracted from each image, the optimal cross-view matching function is learned by either distance metric learning [14, 31], learning to rank [46, 7], or discriminant subspace learning [47, 19, 24, 25, 56, 58]. Recently representation and metric learning are combined end-to-end based on deep neural networks [2, 36] achieving state-of-the-art results when a large number of labelled training images are available. As mentioned early, all of them rely on hundreds of labelled data per camera pair. Considering that a modest-sized surveillance video network can easily have hundreds of cameras, these supervised learning Re-ID models are of very limited practical use. Our model is related to the discriminant subspace learning methods [47, 19, 24, 25, 56, 58]. However, none of them can be employed under the unsupervised setting. In addition, kernelisation is critical to make them work [55]. In contrast, no kernelisation is required for our model resulting in small memory footprint.

The existing models for unsupervised learning of either features or representations for Re-ID fall into three categories. (1) Many focus on designing hand-crafted appearance features [37, 10, 40, 39, 16, 42]. However, it is very challenging to design a set of view-invariant features which are suitable for all camera view

conditions. (2) Several methods exploit localised saliency statistics [57, 54]. Without being able to utilise cross-view identity-discriminative information, their performance is typically weak. Also, they are patch based methods and separate models are learned for every patch which makes them computational expensive. (3) There are also dictionary learning based methods which can intrinsically be used in an unsupervised setting [30, 41]. The key difference in this work is the use of robust graph Laplacian regularisation and joint graph and dictionary learning. We show experimentally that the proposed method is clearly superior to the existing unsupervised alternatives in both matching accuracy and running cost.

Beyond person Re-ID, dictionary learning [28, 1, 43] and graph regularisation [12, 18, 61] have been exploited in many different fields including unsupervised clustering [34], supervised face verification/recognition [21] and semi-supervised learning [5, 8, 33]. Graph learning has also been considered for subspace clustering [22, 45]. However, none of the existing models is directly applicable to the unsupervised cross-view person matching problem. Importantly none of them exploits both graph learning and robust graph regularisation. We show experimentally that both properties are critical for dictionary learning to be effective for solving the unsupervised Re-ID problem.

1.2 Contributions

Our contributions are two-fold: (1) We formulate a novel graph regularised dictionary learning model for unsupervised Re-ID with a new robust ℓ_1-norm graph regularisation term and joint graph and dictionary learning. The model only requires unlabelled training data, which makes it suitable for large-scale Re-ID problems. (2) We develop an efficient iterative optimisation algorithm for the non-smooth and non-convex objective function of our model. During test time, the model is linear and has a closed-form solution for inference; it is thus extremely efficient. Extensive experiments are conducted on four large benchmark datasets, and the results show that our method significantly outperforms existing unsupervised methods in terms of both matching accuracy and running cost.

2 Methodology

2.1 Problem Definition

Suppose we have a set of *unlabelled* training data collected from two camera views[1]. They are denoted as $\mathbf{X} = [\mathbf{X}^a,\ \mathbf{X}^b] \in \mathbb{R}^{n\times m}$, where $\mathbf{X}^a = [\mathbf{x}^a_1,\ ...\ , \mathbf{x}^a_{m_1}] \in \mathbb{R}^{n\times m_1}$ contains n-dimensional feature vectors of m_1 images in view A, and $\mathbf{X}^b = [\mathbf{x}^b_1,\ ...\ , \mathbf{x}^b_{m_2}] \in \mathbb{R}^{n\times m_2}$ of m_2 images in view B. We thus have $m = m_1+m_2$ data points in total. The objective of unsupervised person Re-ID is to learn a matching function f from $\mathbf{X}$, so that given $\mathbf{x}^a$ and $\mathbf{x}^b$ as two test person images from A and B respectively, $f(\mathbf{x}^a, \mathbf{x}^b)$ can match their identities.

[1] In practice our model is not restricted by the number of camera views. We use two here purely for notational simplicity.

2.2 Robust Graph Regularised Dictionary Learning

We solve the problem defined above by learning a dictionary $\mathbf{D} \in \mathbb{R}^{k \times n}$ shared by the two camera views using $\mathbf{X}$. Every atom of the learned dictionary (column of $\mathbf{D}$) can be considered as a latent appearance attribute that is invariant to camera view condition changes. Therefore, with this dictionary, each n-dimensional low-level feature vector, regardless which view it comes from, is represented by the coefficients of the k dictionary atoms. This is equivalent to projecting the original n-dimensional low-level feature vectors to a lower-dimensional ($k < n$) latent attribute space. The matching is done by computing a simple cosine distance between two coefficient vectors in the space. Formally, we aim to learn the optimal dictionary $\mathbf{D}$, such that the latent attribute representation of $\mathbf{X}$, denoted as $\mathbf{Y} = [\mathbf{Y}^a,\ \mathbf{Y}^b] \in \mathbb{R}^{k \times m}$, where $\mathbf{Y}^a = [\mathbf{y}^a_1,\ \dots\ , \mathbf{y}^a_{m_1}] \in \mathbb{R}^{k \times m_1}$ and $\mathbf{Y}^b = [\mathbf{y}^b_1,\ \dots\ , \mathbf{y}^b_{m_2}] \in \mathbb{R}^{k \times m_2}$, are optimised for matching the training data. We expect the same $\mathbf{D}$ can be generalised to match unseen test data across camera views.

Conventional dictionary learning methods estimate the dicitionary $\mathbf{D}$ and the representation $\mathbf{Y}$ simultaneously by solving the following optimisation problem:

$$(\mathbf{D}^*, \mathbf{Y}^*) = \min_{\mathbf{D},\mathbf{Y}} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_F^2 + \lambda_1 \Omega(\mathbf{Y}) \quad s.t. \quad \|\mathbf{d}_i\|_2^2 \leq 1, \tag{1}$$

where $\|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_F^2$ is the reconstruction error evaluating how well a linear combination of the learned atoms can approximate the input data, and $||.||_F$ denotes the matrix Frobenious norm. $\Omega(\mathbf{Y})$ is a regularisation term that is weighted by λ_1. Different models differ mainly in the choice of the regularisation term on $\mathbf{Y}$. The sparsity term, $\Omega(\mathbf{Y}) = \|\mathbf{Y}\|_1$ is widely used which favours a small number of atoms for reconstruction. The constraint $\|\mathbf{d}_i\|^2 \leq 1$ ($\mathbf{d}_i$ is a column of $\mathbf{D}$, $i = 1, ..., k$) enforces the learned dictionary atoms to be compact. It is clear from this formulation that a conventional dictionary learning model only cares about how to best reconstruct $\mathbf{X}$ using $\mathbf{D}$ and $\mathbf{Y}$, without taking into account whether the representation $\mathbf{Y}$ is discriminative. For learning a discriminative dictionary for cross-view Re-ID, one must exploit cross-view identity discriminative information.

A learned dictionary can be made discriminative by using a graph regularisation term which dictates that visually similar people will be close to each other in the learned latent attribute space [11]. Let $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ be an undirected graph connecting between the data points where $\mathbf{V}$ and $\mathbf{E}$ are a set of graph vertices representing the data points and an edge set, respectively. This graph can be encoded by an affinity matrix $\mathbf{W} \in \mathbb{R}^{m \times m}$ for m data points where $\mathbf{W}_{i,j} \neq 0$ if the two vertices are connected, *i.e.* the corresponding data points are in a local neighbourhood. Note: (1) In the context of person Re-ID, we focus on the cross-view discriminative dictionary learning, thus restricting the graph edges to connecting cross-view nodes only. (2) We use the graph regularisation term to replace the commonly used sparsity constraint $\|\mathbf{Y}\|_1$, for reasons to be explained later.

A standared graph regularisation term $\Omega(\mathbf{Y})$ is defined as:

$$\Omega(\mathbf{Y}) = \sum_{ij}^{m} \mathbf{W}_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2. \tag{2}$$

This regularisation essentially requires that the projected data points in the learned latent attribute space to be smooth with regards to the graph, that is, their distances need to conform to the visual similarity relationship embedded in the graph. However, we find that Eq. (2) has two critical limitations that make it unsuitable for the unsupervised Re-ID problem. First, the distance between two projected data points is calculated with a *squared ℓ_2-norm*. It is well-known that a square-based regularisation function can be easily dominated by outlying data samples. Unfortunately outlying samples are commonplace in Re-ID because of background in person detection bounding boxes, detector errors, and (self-)occlusions. Another limitation arises from how the graph is constructed. Most existing methods build the graph in the original high dimensional low-level feature space using $\mathbf{X}$. This is suboptimal – if the low-level feature space is good for measuring cross-camera visual similarity, we would have already solved the Re-ID problem. Learning a discriminative latent attribute space is precisely due to the fact that measuring visual similarity in the original space is unreliable and error-prone, as illustrated in Fig. 1. To tackle both limitations simultaneously, we introduce a robust graph regularisation formulation and a joint graph and dictionary learning method.

Robust graph regularisation. This new term is designed to alleviate the effect of outlying samples during model learning. To derive our robust graph regularisation, let us first rewrite Eq. (2) in a matrix form with trace notation:

$$\Omega(\mathbf{Y}) = \sum_{ij}^{m} \mathbf{W}_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 = tr(\mathbf{Y}\mathbf{L}_{\mathbf{W}}\mathbf{Y}^{\mathbf{T}}), \tag{3}$$

where $\mathbf{L}_{\mathbf{W}} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix, $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ is a degree matrix. Let $\mathbf{L}_{\mathbf{W}} = \mathbf{U}_{\mathbf{W}}\mathbf{S}_{\mathbf{W}}\mathbf{U}_{\mathbf{W}}^{\mathbf{T}}$ using the eigen decomposition technique, and after some matrix manipulation, we have

$$\begin{aligned} tr(\mathbf{Y}\mathbf{L}_{\mathbf{W}}\mathbf{Y}^{\mathbf{T}}) = tr(\mathbf{Y}\mathbf{U}_{\mathbf{W}}\mathbf{S}_{\mathbf{W}}\mathbf{U}_{\mathbf{W}}^{\mathbf{T}}\mathbf{Y}^{\mathbf{T}}) = \\ tr(\mathbf{Y}\mathbf{U}_{\mathbf{W}}\mathbf{S}_{\mathbf{W}}^{\frac{\mathbf{1}}{\mathbf{2}}}\mathbf{S}_{\mathbf{W}}^{\frac{\mathbf{1}}{\mathbf{2}}}\mathbf{U}_{\mathbf{W}}^{\mathbf{T}}\mathbf{Y}^{\mathbf{T}}) = \|\mathbf{Y}\mathbf{A}_{\mathbf{W}}\|_F^2, \end{aligned} \tag{4}$$

where $\mathbf{A}_{\mathbf{W}} = \mathbf{U}_{\mathbf{W}}\mathbf{S}_{\mathbf{W}}^{\frac{\mathbf{1}}{\mathbf{2}}}$. Eq. (4) above is quadratic. To promote sparsity and suppress effects of outlying samples, we adopt a ℓ_1-norm instead of the Frobenius norm. This gives the proposed graph weighted ℓ_1-norm regularisation term

$$\Omega_{\mathrm{R1}}(\mathbf{Y}) = \|\mathbf{Y}\mathbf{A}_{\mathbf{W}}\|_1. \tag{5}$$

Replacing $\Omega(\mathbf{Y})$ with $\Omega_{\mathrm{R1}}(\mathbf{Y})$ in Eq. (1), we have a robust graph regularised dictionary learning model:

$$\min_{\mathbf{D},\mathbf{Y}} \quad \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_F^2 + \lambda_1 \|\mathbf{Y}\mathbf{A}_{\mathbf{W}}\|_1 \quad s.t. \quad \|\mathbf{d}_i\|^2 \leq 1 \tag{6}$$

The key advantages of the proposed robust graph regularisation in this work over the conventional regularisation formulation, including the existing dictionary learning based Re-ID model DLLAP [30], are as follows:

1. *Non-linearity.* Robust graph regularisation introduces non-linearity into the objective, i.e. $\mathbf{Y}$ is non-linear with respect to the original data $\mathbf{X}$, whilst the conventional graph regularisation is linear.
2. *Sparsity.* It is well-known that ℓ_1-norm has a shrinkage property thus promotes sparsity [27, 29]. Intuitively, in the presence of noise and outliers, the magnitude of $\|\mathbf{Y}\mathbf{A}_W\|_F^2$ of the regularisation becomes very big for those outlying data points, and as a result the whole objective function could be dominated by the noise and outliers. In contrast, $\|\mathbf{Y}\mathbf{A}_W\|_1$ becomes sparse due to the use of ℓ_1-norm, consequently suppressing the impact of outliers and noises. Moreover, in the proposed robust regularisation model, explicit sparsity constraint such as $\|\mathbf{Y}\|_1$ is no longer needed[2].

Joint graph and dictionary learning. Instead of computing $\mathbf{W}$ using $\mathbf{X}$ and fixing it during model learning, we assume that $\mathbf{W}$ (hence the graph $\mathbf{G}$ as $\mathbf{W}$ depends on the topology of $\mathbf{G}$) is unknown and has to be learned together with $\mathbf{D}$ and $\mathbf{Y}$. Our objective function thus becomes:

$$\begin{aligned} \min_{\mathbf{D},\mathbf{W},\mathbf{Y}} & \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{Y}\|_F^2 + \lambda_1\|\mathbf{Y}\mathbf{A}_{\mathbf{W}}\|_1 + \lambda_2\|\mathbf{W}\|_F^2 \\ s.t. \quad & \|\mathbf{d}_i\|_2^2 \leq 1,\ \mathbf{W}_i^T\mathbf{1} = 1,\ \mathbf{W}_i \geq 0. \end{aligned} \tag{7}$$

where $\lambda_2\|\mathbf{W}\|_F^2$ is a regularisation term on $\mathbf{W}$ weighted by λ_2 to prevent trivial solutions. The constraints, $\mathbf{W}^\mathbf{T}\mathbf{1} = 1$ and $\mathbf{W} \geq 0$, ensure the validity of the learned graph. We show in our experiments (Sec. 3.2) that this novel joint learning of graph and dictionary has significant advantage over the existing dictionary learning based Re-ID model DLLAP [30]

2.3 Optimisation

The optimisation problem in (7) is non-convex and non-smooth. Solving it is thus more difficult than (1) due to the ℓ_1-norm used in $\Omega_{\mathrm{R1}}(\mathbf{Y})$ and the additional unknown variable $\mathbf{W}$. Next, we develop an efficient solver for (7) based on the Alternating Direction Method of Multipliers (ADMM) [6].

First, we transform (7) by letting $\mathbf{U} = \mathbf{Y}\mathbf{A}_\mathbf{W}$, then the Augmented Lagrangian function of (7) with the introduced constraint is:

$$\begin{aligned} \mathcal{L}_{(\mathbf{D},\mathbf{Y},\mathbf{U},\mathbf{W})} = & \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{Y}\|_F^2 + \lambda_1\|\mathbf{U}\|_1 + \langle \mathbf{F}, \mathbf{U}-\mathbf{Y}\mathbf{A}_\mathbf{W}\rangle \\ & + \frac{\gamma}{2}\|\mathbf{U}-\mathbf{Y}\mathbf{A}_\mathbf{W}\|_F^2 + \lambda_2\|\mathbf{W}\|_F^2 \\ s.t. \quad & \|\mathbf{d}_i\|^2 \leq 1,\ \mathbf{W}^\mathbf{T}\mathbf{1} = 1,\ \mathbf{W} \geq 0. \end{aligned} \tag{8}$$

[2] Empirically we found that adding an extra $\|\mathbf{Y}\|_1$ term makes little difference to the Re-ID performance, but results in more complex solver and higher computational cost.

where $\mathbf{F}$ is Lagrangian multiplier, and γ is a penalty parameter. Now, we can solve it alternatingly with the following five steps with respect to $\mathbf{D}$, $\mathbf{Y}$, $\mathbf{U}$, and $\mathbf{W}$, respectively.

1) Solving for D: To learn $\mathbf{D}$ for a given $\mathbf{Y}$, the objective function reduces to:

$$\min_{\mathbf{D}} \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{Y}\|_F^2 \quad s.t. \;\; \|\mathbf{d}_i\|_2^2 \leq 1 \tag{9}$$

To solve this, we use the Lagrange dual method as in [32]. The analytical solution of $\mathbf{D}$ can be computed as: $\mathbf{D}^* = \mathbf{X}\mathbf{Y}^\mathbf{T}(\mathbf{Y}\mathbf{Y}^\mathbf{T}+\mathbf{\Lambda}^*)^{-1}$, where $\mathbf{\Lambda}^*$ is a diagonal matrix constructed from all the optimal dual variables.

2) Solving for Y: For a given $\mathbf{D}$, solve the following objective to estimate $\mathbf{Y}$:

$$\min_{\mathbf{Y}} \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{Y}\|_F^2 + \frac{\gamma}{2}\|\mathbf{U}-(\mathbf{Y}\mathbf{A}_\mathbf{W}-\frac{\mathbf{F}}{\gamma})\|_F^2.$$

Since each term in this objective is quadratic, we can take its derivative and set it to zero which gives

$$(\mathbf{D}^\mathbf{T}\mathbf{D}\mathbf{Y}+\gamma\mathbf{Y}\mathbf{A}_\mathbf{W}\mathbf{A}_\mathbf{W}^\mathbf{T}) = \mathbf{D}^\mathbf{T}\mathbf{X}+\gamma\mathbf{U}\mathbf{A}_\mathbf{W}^\mathbf{T}+\mathbf{F}\mathbf{A}_\mathbf{W}^\mathbf{T}.$$

This is a standard Sylvester equation, which is solved using the Bartels-Stewart algorithm [4].

3) Solving for U: For a given $\mathbf{Y}$, solve the following objective to estimate $\mathbf{U}$:

$$\min_{\mathbf{U}} \lambda_1\|\mathbf{U}\|_1 + \frac{\gamma}{2}\|\mathbf{U}-(\mathbf{Y}\mathbf{A}_\mathbf{W}-\frac{\mathbf{F}}{\gamma})\|_F^2.$$

We can use the soft-thresholding operator to get $\mathbf{U}$:

$$\mathbf{U} = \text{sign}\left(\mathbf{Y}\mathbf{A}_\mathbf{W}-\frac{\mathbf{F}}{\gamma}\right)\max(\left|\mathbf{Y}\mathbf{A}_\mathbf{W}-\frac{\mathbf{F}}{\gamma}\right|-\frac{\lambda_1}{\gamma}). \tag{10}$$

4) Solving for W: Given $\mathbf{Y}$, the objective function with respect to $\mathbf{W}$ is:

$$\min_{\mathbf{W}} \lambda_1\sum_{ij}^{m}\mathbf{W}_{ij}\|\mathbf{y}_i-\mathbf{y}_j\|_1 + \lambda_2\|\mathbf{W}\|_F^2 \quad s.t. \;\; \mathbf{W}_i^T\mathbf{1}=1, \mathbf{W}_i \geq 0.$$

We set $\lambda_1 = 1$ for easiness, and denote $\mathbf{d}_{ij} = \frac{\|\mathbf{y}_i-\mathbf{y}_j\|_1}{2\lambda_2}$ and $\|\mathbf{W}\|_F^2 = \sum_{ij}\mathbf{W}_{ij}^2$, then

$$\min_{\mathbf{W}} \sum_{ij}^{m}\mathbf{W}_{ij}\mathbf{d}_{ij} + \sum_{ij}^{m}\mathbf{W}_{ij}^2 \quad s.t. \;\; \mathbf{W}_i^\mathbf{T}\mathbf{1}=1, \mathbf{W}_i \geq 0.$$

The above optimisation problem is composed of independent problems with respect to i, and therefore can be rewritten in a vector form:

$$\min_{\mathbf{W}_i} \|\mathbf{W}_i + \mathbf{d}_i\|_2^2 \quad s.t. \quad \mathbf{W}_i\mathbf{1} = 1, \mathbf{W}_i \geq 0.$$

There is a closed-form solution using Lagrange multipliers [45, 22] for this problem:

$$\mathbf{W}_i = \left(\frac{1 + \sum_{j=1}^{K} \tilde{\mathbf{d}}_j}{K} \mathbf{1} - \mathbf{d}_i \right)_+ \tag{11}$$

where the operator $(\mathbf{q})_+$ projects negative elements in $\mathbf{q}$ to 0. K is the parameter that controls the number of neighbours. $\tilde{\mathbf{d}}_i$ is $\mathbf{d}_i$ but with ascending order. After obtaining $\mathbf{W}$, we symmetrise it, and do eigen-decomposition to get $\mathbf{U}_\mathbf{W}$ and $\mathbf{S}_\mathbf{W}$. Then we set $\mathbf{A}_\mathbf{W} = \mathbf{U}_\mathbf{W}\mathbf{S}_\mathbf{W}^{\frac{1}{2}}$. Note that the regularisation parameter λ_2 can be determined by [45]:

$$\lambda_2 = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{K}{2} \mathbf{d}_{i,K+1} - \frac{1}{2} \sum_{j=1}^{K} \mathbf{d}_{i,j} \right). \tag{12}$$

5) Updating multipliers: $\mathbf{F}, \gamma$,

$$\mathbf{F} = \mathbf{F}^{old} + \gamma(\mathbf{U} - \mathbf{Y}\mathbf{A}_\mathbf{W}), \quad \gamma = \rho\gamma^{old}$$

In this work, we set ρ to 1.1 and initialise γ to 0.1. Typically the value for ρ is set between 1.0 and 1.8 [6].

We continue to alternate solving for $\mathbf{D}$, $\mathbf{Y}$, $\mathbf{U}$, $\mathbf{W}$ until a maximum number of iterations is reached or a predefined threshold (10^{-3}) is satisfied.

Convergence Analysis. The theoretical convergence proof of ADMM does not exist. However, in practice it is guaranteed that the objective function converges to at least a stable point [6]. This is validated by our experiments (see Sec. 3). In particular, it is observed that the proposed algorithm has a stable convergence behaviour, always converging after 10-25 iterations.

Remark on Computational Complexity and Scalability. Due to space limit we leave the computational complexity analysis and scalability with respect to the number of samples in the supplementary material.

2.4 Cross-view Matching

After learning the dictionary $\mathbf{D}$ using the unlabelled training data $\mathbf{X}$, given a pair of test samples $\mathbf{x}_i^a$ and $\mathbf{x}_i^b$, we first compute their collaborative representations $\mathbf{y}_i^a$ and $\mathbf{y}_i^b$ by solving the following problems:

$$\mathbf{y}_i^{a*} = \arg min_{\mathbf{y}_i^a} \|\mathbf{x}_i^a - \mathbf{D}\mathbf{y}_i^a\|_F^2 + \lambda\|\mathbf{y}_i^a\|_2^2 \tag{13}$$

$$\mathbf{y}_i^{b*} = \arg min_{\mathbf{y}_i^b} \|\mathbf{x}_i^b - \mathbf{D}\mathbf{y}_i^b\|_F^2 + \lambda\|\mathbf{y}_i^b\|_2^2 \tag{14}$$

These are standard ℓ_2-norm regularised least squares problems with closed-form solutions: $\mathbf{y}_i^{a*} = \mathbf{P}\mathbf{x}_i^a$ and $\mathbf{y}_i^{b*} = \mathbf{P}\mathbf{x}_i^b$, where $\mathbf{P} = (\mathbf{D^T D} + \lambda \mathbf{I})^{-1}\mathbf{D^T}$. Then, after obtaining $\mathbf{y}_i^{a*}$ and $\mathbf{y}_i^{b*}$ their cosine distance is used to measure the visual similarity for Re-ID. Hence, our model is very efficient in testing.

2.5 Extension to Supervised Re-ID

Although our model is designed for unsupervised Re-ID, it can be easily extended if labelled cross-view pairs become available. More specifically, the label information can be encoded in the graph $\mathbf{W}$. That is, instead of learning $\mathbf{W}$, it is now fixed so that if the corresponding cross-view pair (i, j) is labelled as containing the same person, we set $\mathbf{W}_{i,j}$ to 1, otherwise it is set as 0. This essentially gives thus the ideal graph and the relaxed visual similarity constraint becomes a more stringent identity constraint which requires that people of the same identity to be close in the learned attribute space and vice versa.

3 Experiments

3.1 Datasets and Settings

Datasets. Four widely used benchmark datasets are used for the experiments. *VIPeR* [15] contains 632 image pairs of people captured outdoor from two non-overlapping camera views. Following the standard setting which is single-shot i.e., one image per person per view, the dataset is randomly split into two sets of 316 image pairs, one for training and the other for testing. For the test set, all images from one view is used as the gallery set and the others probe set. The results for all evaluations were obtained by averaging over 10 splits. *PRID* [23] is different from the other available datasets in that the gallery and probe sets have different numbers of people. In our experiments, we use the single-shot version of the dataset as in [19, 26, 46]. Specifically, out of the 749 people captured in two camera views, only 200 people appear in both views. In each data split, 100 out of that 200 people are chosen randomly for training, while the remaining 100 of one view are used as the probe set, and the remaining 649 people's images of the other view are used as gallery, which thus includes the 100 people in the probe set. Experiments are carried out on the same 10 splits as in [19, 26] with the average results reported. *CUHK01* [35] consists of 971 people with two images per person per camera view i.e. multi-shot. We follow the standard setting [35]: 486 persons for training, while 485 persons for test. *CUHK03* [36] contains 13,164 images of 1,467 people. Two versions exist which differs in whether the images were obtained by manual cropping or automatically by applying the DPM person detector [17]. The detector-generated images are used as they reflect better the real-world application scenarios for testing the robustness of our model against outliers. There are in total six camera views but each person is observed in only two out of the six views, and has 4.8 images on average for each view. We used the same setting and random splits as in [36] with

a single-shot setting: for the probe set we randomly select 100 people with two images each, whilst images of the remaining people are used for training. Note that out of the four datasets, CHUK03 is much bigger than the other three in terms of both the number of identities and the number of images in the training set.

Settings. *Features*: The features introduced in [19] are adopted. Each image is scaled to 128×48 in all datasets, and then histogram-based image descriptors are computed consisting of three types: (1) Colour histogram using HS, RGB, and Lab colour spaces (2880-D colour vector), (2) HOG (1040-D) [13], and (3) LBP (1218-D) [3]. The final image feature vector, 5138-D, is obtained as the concatenation of these three types of features. *Evaluation metrics*: We obtain the Cumulative Matching Characteristics (CMC) curves. Due to space constraint, we only report matching accuracies at Rank 1 here and leave the full CMC curves in the supplementary material. *Parameter settings*: There are a number of parameters in our model. As an unsupervised learning method, there are no other means but setting them manually. For the dictionary size k, we do not tune it carefully and set it to 256 for the two small datasets VIPeR and PRID, and 512 for the larger CUHK01 and CUHK03 dataset. Its effects on the performance will be discussed later. In the objective function (Eq. (7)), there are two weights λ_1 and λ_2 for the two regularisation terms respectively. As explained in Sec. 2.3, λ_2 is set automatically using Eq. (12) in the ADMM algorithm, whilst for λ_1 we simply set it to 1 throughout, as we found that the algorithm is insensitive to its value. Similarly for the initial construction of graph $\mathbf{G}$, we use a KNN graph with cosine distance and $K = 5$ for all datasets.

3.2 Evaluation of Unsupervised Learning based Re-ID

Compared methods. Under this setting, we compared our approach with state-of-the-art unsupervised alternatives which fall into four categories: (1) The hand-crafted feature-based methods including SDALF [16] and CPS [10]. (2) The saliency learning-based eSDC [57] and GTS [54]; (3) The dictionary learning (DLLAP) [30] which uses the same 5138-D features for fair comparison. (4) The codebook learning method (BGG) [59].

Results. Table 1 compares the results of the proposed method against the six alternatives and a non-learning ℓ_1 distance based baseline. From Table 1, the following observations can be made: (1) Our robust graph regularised dictionary learning model outperforms all existing unsupervised methods on all four datasets, and often by a big margin. (2) The margin is in general bigger on the two larger datasets CUHK01 and CUHK03, which indicates that our model can benefit more from larger unlabelled training data. (3) Among the alternatives, the dictionary learning based method (DLLAP) [30] is the most competitive. However, the gap is still significant due to the introduced two novel components: robust graph regularisation and joint graph and dictionary learning. This result also suggests that learning a low-dimensional latent attribute representation is more suited for unsupervised Re-ID than the alternative models. In particular, the difference between Ours and ℓ_1 is large which means that matching people

Table 1: Unsupervised Re-ID results measured in Rank-1 matching accuracy (%) on VIPeR, PRID, CUHK01, CUHK03, where '-' denotes no reported result.

Datasets	*VIPeR*	*PRID*	*CUHK01*	*CUHK03*
ℓ_1	15.6	13.9	10.9	12.5
SDALF [16]	19.9	16.3	9.9	4.9
DLLAP [30]	29.6	21.1	28.4	22.3
eSDC [57]	26.7	-	26.6	7.7
CPS [10]	22.0	-	-	-
GTS [54]	25.2	-	-	-
BGG [59]	21.7	-	-	18.9
Ours	**33.5**	**25.0**	**41.0**	**30.4**

is made much easier in this learned discriminative subspace with less than one tenth of the original dimensions. The advantage of our method's computational efficiency over other methods will be discussed later.

3.3 Evaluation of Supervised Learning based Re-ID

Compared methods. Since the performance of different existing methods on different datasets often vary drastically[3], we choose the best methods for each dataset separately to better reflect the state-of-the-art. All methods are published in the last two years. Note that multi-feature fusion based methods are separated from single feature or deep models as typically any method can benefit from multi-feature fusion. As mentioned in Sec. 2.5, our model can also operate in the supervised mode; denoted as Ours_sup, this can be considered as the upper bound of our model's performance under the unsupervised setting when the graph is learned perfectly.

Results. We have the following key findings from Table 2: (1) The gap between Ours_un and Our_sup is moderate. This indicates that our graph learning method is very effective and the performance of the unsupervised model is not far off from its upper bound. (2) On the two smaller datasets, VIPeR and PRID, our model is very competitive under the supervised setting: on VIPeR it beats all single feature-based methods and on PRID, it outperforms all existing supervised methods, often significantly. Even our unsupervised model outperforms some very recent supervised models. Note that this is without any kernalisation which could further improve our model's performance. (3) On the two larger datasets CUHK01 and CUHK03 (with detected person images), the gap between our method and the state-of-the-art begins to appear [4]. Our model (both supervised and unsupervised) remains competitive on CUHK01, but on CUHK03, the gap

[3] For example, deep learning based methods often perform stronger on the large datasets than the small ones due to the need for large training data.

[4] The gap is much smaller if more powerful features are used - see supplementary material for details.

Table 2: Comparison state-of-the-art supervised methods

Datasets	*VIPeR*		*PRID*		*CUHK01*		*CUHK03*	
	Ref.	Rank 1	Ref.	Rank 1	Ref.	Rank 1	Ref.	Rank 1
Single-feature Methods	[37]	40.0	[19]	14.5	[58]	34.3	[37]	46.4
	[2]	34.8	[55]	19.7	[2]	47.5	[2]	44.9
	[38]	40.7	[30]	25.2	[36]	29.4	[38]	51.2
	[9]	36.1	[48]	16.0	[36]	27.8	[36]	19.9
	[50]	40.9	[51]	18.0	[37]	63.5	[50]	52.1
	[60]	30.2	[38]	12.3	[38]	**64.2**	[52]	**59.2**
Multi-feature Fusion	[46]	**45.9**	[46]	17.9	[46]	53.4	–	–
Ours_un	33.5		25.0		41.0		30.4	
Ours_sup	41.5		**30.1**		50.1		39.0	

Table 3: The contributions of individual model components

Methods	Ours_*DL*	Ours_ℓ_2	Ours_ℓ_2_graph	Ours_ℓ_1	Ours_full
VIPeR	19.6	29.4	30.1	32.0	33.5
CUHK01	17.4	36.9	37.5	38.7	41.0

is big, in particular to our unsupervised model. This is expected: with over 10,000 labelled training images from 1,367 people, an unsupervised model cannot compete with a supervised one, especially those based on deep learning. However, we would like to point out that in practice collecting hundreds of labelled training samples is very difficult and collecting thousands would be near impossible across even just a handful of camera views.

3.4 Further Analysis

The contributions of individual components. Our proposed method has two key components and to see the impact of each we compare our full model with various striped-down versions of the model under the unsupervised setting: (1) Ours_*DL* – without graph regularisation which is the same as conventional dictionary learning; (2) Ours_ℓ_2 – the graph is fixed and ℓ_2-norm is used for graph regularisation; (3) Ours_ℓ_2_graph – the graph is learned and ℓ_2-norm is used for graph regularisation; (4) Ours_ℓ_1 – the graph is fixed and ℓ_1-norm is used for graph regularisation; (5) Ours_full – our full proposed model in which the graph is learned and ℓ_1-norm is used for graph regularisation. Table 3 shows that both using robust ℓ_1-norm graph regularisation and joint graph and dictionary learning contribute positively toward the final performance. The result (comparing Ours_*DL* with the other models) also shows that adding a graph regularisation term to learn cross-view discriminative information in general is critical for dictionary-learning-based Re-ID.

Effect of dictionary size and convergence analysis. The only parameter we tuned for each dataset is the dictionary size. Figure 2(Left) shows that when

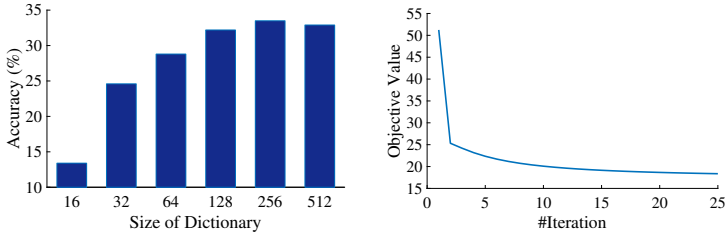

Fig. 2: (Left) Rank 1 accuracies with different dictionary sizes on VIPeR dataset; (Right) Objective function value with respect to the number of iterations on CUHK01.

Table 4: Average testing time of different methods on VIPeR

Stage	SDALF	eSDC	BGG	Ours
Feature Extraction (s)	2.92	0.76	0.62	0.03
Matching (s)	550.80	9.7	0.44	0.01

the size is over 100, its effect is small. Furthermore, Fig. 2(Right) shows the proposed method converges rapidly. Although there is no theoretically proof, convergence is observed in all our experiments within 25 iterations.

Running cost. Our experiments were conducted in MATLAB on a PC with two 3.40 GHz CPUs and 16G RAM. The training of the model on VIPeR takes 178.3 second but during test it is very efficient: once the 5138-D features are extracted, it takes only 0.01 second to match one probe image against 316 images from the gallery. Table 4 compares the running time of feature extraction and matching during test time against a number of alternative unsupervised methods. It is clear that our method is often a few magnitudes faster than its competitors.

4 Conclusion

We have proposed a novel unsupervised Re-ID model based on dictionary learning. The key contributions are the introduction of a robust ℓ_1-norm graph regularisation term in the dictionary learning formulation so that cross-view discriminative information can be learned. In addition, a joint graph and dictionary learning algorithm is developed which further improves the ability of the proposed model to deal with outlying samples abundant in person Re-ID data. Extensive experiments on four benchmark datasets show that the proposed method significantly outperforms existing unsupervised methods.

Acknowledgments

This project was partly funded by the EU FP7 Project SUNNY (grant no. 313243).

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing (2006)
2. Ahmed, E., Jones, M., Marks, T.K.: An improved deep learning architecture for person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
3. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2006)
4. Bartels, R.H., Stewart, G.W.: Solution of the matrix equation ax + xb = c [f4]. Commun. ACM 15(9), 820–826 (Sep 1972), http://doi.acm.org/10.1145/361573.361582
5. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. The Journal of Machine Learning Research 7, 2399–2434 (2006)
6. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning 3(1), 1–122 (2011)
7. Bryan, P., Wei-Shi, Z., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: Proc. BMVC (2010)
8. Cai, D., He, X., Han, J.: Semi-supervised discriminant analysis. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. pp. 1–7. IEEE (2007)
9. Chen, D., Yuan, Z., Hua, G., Zheng, N., Wang, J.: Similarity learning on an explicit polynomial kernel feature map for person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
10. Cheng, D.S., Cristani, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: Proc. BMVC (2011)
11. Chung, F.R.: Spectral graph theory, vol. 92. American Mathematical Soc. (1997)
12. Daitch, S.I., Kelner, J.A., Spielman, D.A.: Fitting a graph to vector data. In: Proceedings of the 26th Annual International Conference on Machine Learning. pp. 201–208. ACM (2009)
13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. CVPR (2005)
14. Dikmen, M., Akbas, E., Huang, T.S., Ahuja, N.: Pedestrian recognition with a learned metric. In: Proc. ACCV (2011)
15. Douglas, G., Shane, B., Hai, T.: Evaluating appearance models for recognition, reacquisition and tracking. In: PETS (2007)
16. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re- identification by symmetry-driven accumulation of local features. In: Proc. CVPR (2010)
17. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32(9), 1627–1645 (2010)
18. Gao, S., Tsang, I., Chia, L., Zhao, P.: Local features are not lonely laplacian sparse coding for image classification. In: Proc. CVPR (2010)
19. Giuseppe, L., Iacopo, M., Alberto, D.B.: Matching people across camera views using kernel canonical correlation analysis. In: Proc. ICDSC (2014)

20. Gong, S., Cristani, M., Yan, S., Loy, C.C.: Person re-identification, vol. 1. Springer (2014)
21. Guo, H., Jiang, Z., Davis, L.S.: Discriminative dictionary learning with pairwise constraints. In: Proc. ACCV (2014)
22. Guo, X.: Robust subspace segmentation by simultaneously learning data representations and their affinity matrix. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI) (2015)
23. Hirzer, M., Beleznai, C., Roth, M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: Proc. SCIA (2011)
24. Hirzer, M., Roth, M., Bischof, H.: Person re-identification by efficient impostor-based metric learning. In: Proc. AVSS (2012)
25. Hirzer, M., Roth, M., Koestinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: Proc. ECCV (2012)
26. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: Computer Vision–ECCV 2012, pp. 780–793. Springer (2012)
27. James, G., Witten, D., Hastie, T., Tibshirani, R.: An introduction to statistical learning, vol. 112. Springer
28. Kenneth, K., M.Joseph, Bhaskar, R., Kjersti, E., Te-Won, L., Terrence, S.: Dictionary learning algorithms for sparse representation. Neural Computing 15(2) (Feb 2003)
29. Kim, S.J., Koh, K., Boyd, S., Gorinevsky, D.: \ell_1 trend filtering. SIAM review 51(2), 339–360 (2009)
30. Kodirov, E., Xiang, T., Gong, S.: Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification. In: Xianghua Xie, M.W.J., Tam, G.K.L. (eds.) Proceedings of the British Machine Vision Conference (BMVC). pp. 44.1–44.12. BMVA Press (September 2015), https://dx.doi.org/10.5244/C.29.44
31. Layne, R., Hospedales, T., Gong, S.: Re-id: Hunting attributes in the wild. In: Proc. BMVC (2014)
32. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Advances in neural information processing systems. pp. 801–808 (2006)
33. Li, C.G., Lin, Z., Zhang, H., Guo, J.: Learning semi-supervised representation towards a unified optimization framework for semi-supervised learning. In: The IEEE International Conference on Computer Vision (ICCV) (December 2015)
34. Li, C.G., Vidal, R.: Structured sparse subspace clustering: A unified optimization framework. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
35. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: Computer Vision–ACCV 2012, pp. 31–44. Springer (2012)
36. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 152–159 (2014)
37. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
38. Liao, S., Li, S.Z.: Efficient psd constrained asymmetric metric learning for person re-identification. In: The IEEE International Conference on Computer Vision (ICCV) (December 2015)

39. Lisanti, G., Masi, I., Bagdanov, A.D., Bimbo, A.D.: Person re-identification by iterative re-weighted sparse ranking. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2013)
40. Liu, C., Gong, S., Loy, C.C.: On-the-fly feature importance mining for person re-identification. Pattern Recognition 47(4), 1602–1615 (2014)
41. Liu, X., Song, M., Tao, D., Zhou, X., Chen, C., Bu, J.: Semi-supervised coupled dictionary learning for person re-identification. In: Proc. CVPR (2014)
42. Ma, B., Su, Y., Jurie, F.: Bicov: a novel image representation for person re-identification and face verification. In: Proc. BMVC (2012)
43. Mairal, J., Bach, F., Ponce, J., Sapiro., G.: Online learning for matrix factorization and sparse coding. In: Journal of Machine Learning Research, vol. 11, pp. 19–60 (2010)
44. Nie, F., Wang, H., Huang, H., Ding, C.: Unsupervised and semi-supervised learning via l_1-norm graph. In: ICCV (2011)
45. Nie, F., Wang, X., Huang, H.: Clustering and projected clustering with adaptive neighbors. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 977–986. ACM (2014)
46. Paisitkriangkrai, S., Shen, C., van den Hengel, A.: Learning to rank in person re-identification with metric ensembles. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
47. Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification. In: Proc. CVPR (2013)
48. Roth, P.M., Hirzer, M., Köstinger, M., Beleznai, C., Bischof, H.: Mahalanobis distance learning for person re-identification. In: Person Re-Identification, pp. 247–267. Springer (2014)
49. Shen, Y., Lin, W., Yan, J., Xu, M., Wu, J., Wang, J.: Person re-identification with correspondence structure learning. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3200–3208 (2015)
50. Shi, H., Zhu, X., Liao, S., Lei, Z., Yang, Y., Li, S.Z.: Constrained deep metric learning for person re-identification. CoRR abs/1511.07545 (2015), `http://arxiv.org/abs/1511.07545`
51. Su, C., Yang, F., Zhang, S., Tian, Q., Davis, L.S., Gao, W.: Multi-task learning with low rank attribute embedding for person re-identification. In: The IEEE International Conference on Computer Vision (ICCV) (December 2015)
52. Ustinova, E., Ganin, Y., Lempitsky, V.S.: Multiregion bilinear convolutional neural networks for person re-identification. CoRR abs/1512.05300 (2015), `http://arxiv.org/abs/1512.05300`
53. Vezzani, R., Baltieri, D., Cucchiara, R.: People reidentification in surveillance and forensics: A survey. ACM Computing Surveys (CSUR) 46(2), 29 (2013)
54. Wang, H., Gong, S., Xiang, T.: Unsupervised learning of generative topic saliency for person re-identification. In: Proc. BMVC (2014)
55. Xiong, F., Gou, M., Camps, O., Sznaier, M.: Person re-identification using kernel-based metric learning methods. In: Proc. ECCV (2014)
56. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., S, S.L.: Salient color names for person re-identification. In: Proc. ECCV (2014)
57. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: Proc. CVPR (2013)
58. Zhao, R., Ouyang, W., Wang, X.: Learning mid-level filters for person re-identification. In: Proc. CVPR (2014)

59. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1116–1124 (2015)
60. Zheng, L., Wang, S., Tian, L., He, F., Liu, Z., Tian, Q.: Query-adaptive late fusion for image search and person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
61. Zheng, M., Bu, J., Chen, C., Wang, C., Zhang, L., Qiu, G., Cai, D.: Graph regularized sparse coding for image representation. In: IEEE Transactions on Image Processing (2011)