Temporal Model Adaptation for Person Re-identification

Niki Martinel[1,3(✉)], Abir Das[2], Christian Micheloni[1], and Amit K. Roy-Chowdhury[3]

[1] University of Udine, 33100 Udine, Italy
niki.martinel@uniud.it
[2] University of Massatchussets Lowell, Lowell, MA 01852, USA
[3] University of California Riverside, Riverside, CA 92507, USA

Abstract. Person re-identification is an open and challenging problem in computer vision. Majority of the efforts have been spent either to design the best feature representation or to learn the optimal matching metric. Most approaches have neglected the problem of adapting the selected features or the learned model over time. To address such a problem, we propose a temporal model adaptation scheme with human in the loop. We first introduce a similarity-dissimilarity learning method which can be trained in an incremental fashion by means of a stochastic alternating directions methods of multipliers optimization procedure. Then, to achieve temporal adaptation with limited human effort, we exploit a graph-based approach to present the user only the most informative probe-gallery matches that should be used to update the model. Results on three datasets have shown that our approach performs on par or even better than state-of-the-art approaches while reducing the manual pairwise labeling effort by about 80 %.

Keywords: Person re-identificaion · Metric learning · Active learning

1 Introduction

Person re-identification is the problem of matching a person acquired by disjoint cameras at different time instants. The problem has recently gained increasing attention (see [1] for a recent survey) due to its open challenges like changes in viewing angle, background clutter, and occlusions. To address these issues, existing approaches seek either the best feature representations (e.g., [2–4]) or propose to learn optimal matching metrics (e.g., [5–7]). While they have obtained reasonable performance on commonly used datasets (e.g., [8–10]), we believe that these approaches have not yet considered a fundamental related problem: how to learn from the data being continuously collected in an installed system and

Electronic supplementary material The online version of this chapter (doi:10.1007/978-3-319-46493-0_52) contains supplementary material, which is available to authorized users.

B. Leibe et al. (Eds.): ECCV 2016, Part IV, LNCS 9908, pp. 858–877, 2016.
DOI: 10.1007/978-3-319-46493-0_52

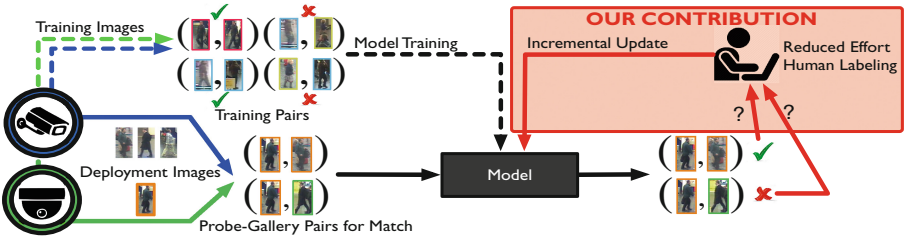

Fig. 1. Illustration of the re-identification pipeline highlighting our contribution. Dashed lines indicate the training stage, solid lines the deployment stage. Existing methods do not consider the information provided by a matched probe-gallery pair to update the model. We propose to use such information to improve the model performance by adapting it to the dynamic environmental variations.

adapt existing models to this new data. This is an important problem to address if re-identification methods have to work on long time-scales.

To illustrate such a problem, let us consider a simplified scenario in which at every time instant a conspicuous amount of visual data is being generated from two cameras. From each camera we obtain a large set of *probe* and *gallery* persons that have to be matched. Since this is a task that evolves over time, it is unlikely that the *a-priori* selected features or the learned model return the correct gallery match for every probe at any instant. In addition, after each of such matches is computed, the information provided by the considered images is discarded. This results in a loss of valuable information which could have been used to update the model, thus ideally yielding better performance over time.

The above problem could be overcome if the data could be exploited in a continuous learning process in which the model can be updated with every single *probe-gallery* match. Since we do not know whether a match is correct or not, the model might be updated with the wrong information. To tackle this issue, manual labeling of each match can be performed, but, doing so with a large corpus of data is clearly impossible. However, if the human labor is kept to a minimum, the model can ideally be adapted over time without compromising performance. Thus, *the main idea of the paper is a person re-identification solution based on an incremental adaptation of the learned model with human in the loop.*

Contributions: As shown in Fig. 1, this work brings in two main contributions: (i) an incremental learning algorithm that allows the model to be adapted over time, and (ii) a method to reduce the human labeling effort required to properly update the model. These objectives are achieved as follows.

(i) We propose a low-rank sparse similarity-dissimilarity metric learning method (Sect. 3.2) which
  (a) learns two low-rank projections onto discriminant manifolds providing optimal embeddings for a similarity and a dissimilarity measure;
  (b) introduces sparsity inducing regularizers that allow identification and exploitation of the most discriminative dimensions for matching; and

(c) is trained in an incremental fashion through a stochastic derivation of the Alternating Directions Methods of Multipliers (ADMM) [11].

(ii) We introduce an unsupervised graph-based approach which, for every probe, identifies only the most relevant gallery persons among a large set of available ones (Sect. 3.3). Such a set, obtained by exploiting dominant sets clustering [12], contains the most informative gallery persons which are first provided to the human labeler, then exploited to update the model.

To substantiate our contributions we have conducted the experiments on three benchmark datasets for person re-identification. Results demonstrate that (i) the proposed approach for identifying the most informative gallery persons yields better re-identification performance than using completely labeled data; (ii) the proposed low-rank sparse similarity-dissimilarity approach trained in an incremental fashion with such informative gallery persons, hence with significantly less manual labor, performs on par or even better than state-of-the-art methods trained on 100 % labeled data. In fact, with only 15 % labeled data we improve the previous best rank 1 results by more than 8 % on the PRID450S dataset. These experiments show how re-identification models can be continuously adapted over time with limited human effort and without sacrifice in performance.

2 Relation to Existing Work

The person re-identification problem has been studied from different perspectives, ranging from partially seen persons [13] to low resolution images [14] – also considered in camera networks [15], which can eventually be synthesized in the open-world re-identification idea [16]. In the following, we focus on metric and active learning methods relevant to our work.

Metric Learning approaches focus on learning discriminant metrics which aim to yield an optimal matching score/distance between a gallery and a probe image.

Since the early work of [17], many different solutions have been introduced [18]. In the re-identification field, metric learning approaches have been proposed by relaxing [19] or enforcing [20] positive semi-definite (PSD) conditions as well as by considering equivalence constraints [21–23]. While most of the existing methods capture the global structure of the dissimilarity space, local solutions [24–27] have been proposed too. Following the success of both approaches, methods combining them in ensembles [5,7,28] have been introduced.

Different solutions yielding similarity measures have also been investigated by proposing to learn listwise [29] and pairwise [30] similarities as well as mixture of polynomial kernel-based models [9]. Related to these similarity learning models are the deep architectures which have been exploited to tackle the task [31–33].

With respect to all such methods, the closest ones to our approach are [6,20]. Specifically, in [6], authors jointly exploit the metric in [21] and learn a low-rank projection onto a subspace with discriminative Euclidean distance. The solution

is obtained through generalized eigenvalue decomposition. In [20], a soft-margin PSD constrained metric with low-rank projections is learned via a proximal gradient method. Both works exploit a batch optimization approach.

Though sharing the idea of finding discriminative low-rank projections, there are significant differences with our method. Specifically, we introduce (i) an incremental learning procedure along with a stochastic ADMM solver which can handle noisy observations of the true data; (ii) a low-rank similarity-dissimilarity metric learning which brings significant performance gain with respect to each of its components; (iii) additional sparsity regularizers on the low-rank projections that allow self-discovery of the relevant components of the underlying manifold.

Active Learning: In an effort to bypass tedious labeling of training data there has been recent interest in "active learning" [34] to intelligently select unlabeled examples for the experts to label in an interactive manner.

This can be achieved by choosing one sample at a time by maximizing the value of information [35], reducing the expected error [36], or minimizing the resultant entropy of the system [37]. More recently, works selecting batches of unlabeled data by exploiting classifier feedback to maximize informativeness and sample diversity [38,39] were proposed. Specific application areas in computer vision include, but are not limited to, tracking [40], scene classification [35,41], semantic segmentation [42], video annotation [43] and activity recognition [44].

Active learning has been a relatively unexplored area in person re-identification. Including the human in the loop has been investigated in [8,45,46]. These methods focused on post-ranking solutions and exploit human labor to refine the initial results by relying on full [8] or partial [45] image selection. In [46], authors introduce an active learning strategy that exploits mid level attributes to train a set of attribute predictors aiding active selection of images.

Different from such approaches, in our proposed method human labor is not required to improve the post-rank visual search, but to reliably update the learned model over time. We do not rely on additional attribute predictors which require a proper training that calls for a large number of annotated attributes. Thus bypassing the need for attribute annotation, we reduce both the computational complexity as well as the additional manual effort. We introduce a graph-based solution that exploits the information provided by a single probe-gallery match as well as the information shared between all the persons in the entire gallery. With this, a small set of highly informative probe-gallery pairs is delivered to the human, whose effort is thus limited.

3 Temporal Model Adaptation for Re-identification

An overview of the proposed solution is illustrated in Fig. 2. Specifically, to achieve model adaptation over time, we first introduce a similarity-dissimilarity metric learning approach which can be trained in an incremental fashion (Sect. 3.2). Then, to limit the human labeling effort required to properly update the model, we propose an unsupervised graph-based approach that identifies only the most informative probe-gallery samples (Sect. 3.3).

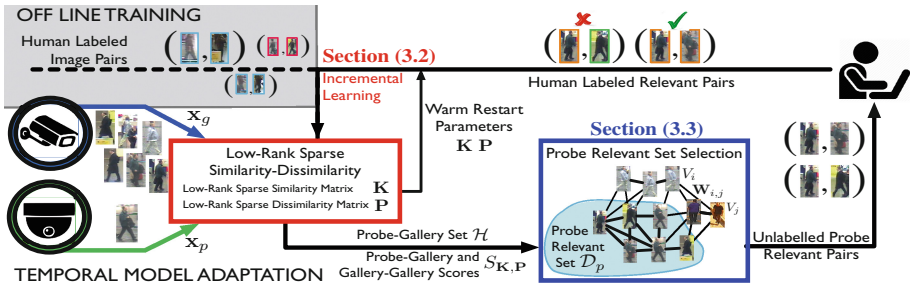

Fig. 2. Proposed temporal model adaptation scheme. An off-line procedure exploits labeled image pairs to train the initial similarity-dissimilarity model. As new unlabeled pairs are obtained, a score for each of those is obtained using the learned model. These are later used to identify a relevant set of gallery persons for each probe. Such a set, containing the most informative samples, is exploited to construct the relevant pairs which are first provided to the human annotator, then considered to update the model.

3.1 Preliminaries

Let $\mathcal{P} = \{\mathbf{I}_p\}_{p=1}^{|\mathcal{P}|}$ and $\mathcal{G} = \{\mathbf{I}_g\}_{g=1}^{|\mathcal{G}|}$ be the set of probe and gallery images acquired by two disjoint cameras. Let $\mathbf{x}_p \in \mathbb{R}^d$ and $\mathbf{x}_g \in \mathbb{R}^d$ be the feature representations of $\mathbf{I}_p$ and $\mathbf{I}_g$ of two persons p and g. Let $\mathcal{X} = \{(\mathbf{x}_p, \mathbf{x}_g; y_{p,g})^{(i)}\}_{i=1}^{n}$ denote the training set of $n = |\mathcal{P}| \times |\mathcal{G}|$ probe-gallery pairs where $y_{p,g} \in \{-1, +1\}$ indicates whether p and g are the same person $(+1)$ or not (-1). Finally, let an *iteration* be a parameter update computed by visiting a single sample and let an *epoch* denote a complete cycle on the training set.

3.2 Low-Rank Sparse Similarity-Dissimilarity Learning

Objective: The image feature representations $\mathbf{x}$ might be very high-dimensional and contain non-discriminative components. Hence, learning a metric in such a feature space might yield to non-optimal generalization performance. To overcome such a problem we propose to learn a low-rank metric which self-determines the discriminative dimensions of the underlying manifold.

Towards such an objective, inspired by the success of similarity learning on image retrieval tasks [47–49], we propose to learn a similarity function

$$\sigma_{\mathbf{K}}(\mathbf{x}_p, \mathbf{x}_g) = \mathbf{x}_p^T \mathbf{K}^T \mathbf{K} \mathbf{x}_g \tag{1}$$

parameterized by the low-rank projection matrix $\mathbf{K} \in \mathbb{R}^{r \times d}$, with $r \ll d$. This provides an embedding in which the dot product between the projected feature vectors is "large" if p and g are the same person, "small" otherwise. The similarity function is then coupled with the output of a metric learning solution that aims to find a matrix $\mathbf{P} \in \mathbb{R}^{r \times d}$ that projects the high-dimensional vectors to a low-dimensional manifold with a discriminative Euclidean dissimilarity

$$\delta_{\mathbf{P}}(\mathbf{x}_p, \mathbf{x}_g) = \|\mathbf{P}\mathbf{x}_p - \mathbf{P}\mathbf{x}_g\|_2^2 = (\mathbf{x}_p - \mathbf{x}_g)^T \mathbf{P}^T \mathbf{P} (\mathbf{x}_p - \mathbf{x}_g) \tag{2}$$

which is "small" if p and g are the same person, "larger" otherwise. This results in the score function

$$S_{\mathbf{K},\mathbf{P}}(p,g) = y_{p,g}(\underbrace{\sigma_{\mathbf{K}}(\mathbf{x}_p,\mathbf{x}_g)}_{\uparrow \text{for } p=g,\ \downarrow \text{for } p\neq g} - \underbrace{(1/2)\delta_{\mathbf{P}}(\mathbf{x}_p,\mathbf{x}_g)}_{\downarrow \text{for } p=g,\ \uparrow \text{for } p\neq g}) \tag{3}$$

which included in a margin hinge loss yields

$$\ell_{\mathbf{K},\mathbf{P}}(p,g) = \max\left(0, 1 - S_{\mathbf{K},\mathbf{P}}(p,g)\right). \tag{4}$$

Notice that zero loss is achieved if $S_{\mathbf{K},\mathbf{P}}(p,g) \geq 1$, i.e., when the difference between $\sigma_{\mathbf{K}}$ and $\frac{1}{2}\delta_{\mathbf{P}}$ is either greater than or equal to 1 for positive pairs or less than or equal to -1 for negative ones. In other cases a linear penalty is paid.

Obtaining the low-rank projections through Eq. (4) with fixed r implies that such a value should be carefully selected before the learning process begins. To overcome such a problem, we impose additional constraints on the low-rank projection matrices. In particular, the $\ell_{2,1}$ norm has shown to perform robust feature selection through the induced group sparsity [50–53]. Motivated by such findings, we can set $r = d$, then leverage on an $\ell_{2,1}$ norm regularizer to drive the rows of $\mathbf{P}$ and $\mathbf{K}$ to decay to zero. This corresponds to rejecting non discriminative dimensions of the underlying manifold.

Let $\Omega_{\mathbf{K},\mathbf{P}} = \alpha\|\mathbf{K}\|_{2,1} + \beta\|\mathbf{P}\|_{2,1}$ be the cost associated with the low-rank projection matrix regularizers where α and β are the corresponding trade-off parameters controlling the regularization strength. Then, considering that we want to optimize the empirical risk over $\mathcal{X}$, we can write our objective as

$$\arg\min_{\mathbf{K},\mathbf{P}} \mathcal{J}_{\mathbf{K},\mathbf{P}} + \Omega_{\mathbf{K},\mathbf{P}} \qquad \text{where} \qquad \mathcal{J}_{\mathbf{K},\mathbf{P}} = \frac{1}{n}\sum_{i=1}^{n} \ell_{\mathbf{K},\mathbf{P}}\left(p^{(i)}, g^{(i)}\right) \tag{5}$$

and $p^{(i)}$ and $g^{(i)}$ denote the identities of persons p and g in the i-th pair of $\mathcal{X}$.

Incremental Learning: The objective function in Eq. (5) is a sum of two functions which are both convex but non-smooth. A solution to such kind of a problem that allows us to perform incremental updates can be obtained using the ADMM optimization algorithm [11].

ADMM solves optimization problems defined by means of the corresponding augmented Lagrangian. By introducing two additional constraints $\mathbf{K} - \mathbf{U} = \mathbf{0}$ and $\mathbf{P} - \mathbf{V} = \mathbf{0}$ we can define the augmented Lagrangian for Eq. (5) as

$$\begin{aligned} L_{\mathbf{K},\mathbf{P},\mathbf{U},\mathbf{V},\mathbf{\Lambda},\mathbf{\Psi}} = \mathcal{J}_{\mathbf{K},\mathbf{P}} + \Omega_{\mathbf{U},\mathbf{V}} + \langle \mathbf{\Lambda}, \mathbf{K} - \mathbf{U}\rangle + \langle \mathbf{\Psi}, \mathbf{P} - \mathbf{V}\rangle \\ + \frac{\rho}{2}\left(\|\mathbf{K} - \mathbf{U}\|_F^2 + \|\mathbf{P} - \mathbf{V}\|_F^2\right) \end{aligned} \tag{6}$$

where $\mathbf{\Lambda} \in \mathbb{R}^{r\times d}$ and $\mathbf{\Psi} \in \mathbb{R}^{r\times d}$ are two Lagrangian multipliers, $\langle\cdot,\cdot\rangle$ denote the inner product, $\|\cdot\|_F$ is the Frobenius norm and, $\rho > 0$ is a penalty parameter.

To solve the optimization problem, at each epoch s, ADMM alternatively minimizes L with respect to a single parameter, $\mathbf{K}$, $\mathbf{P}$, $\mathbf{U}$, $\mathbf{V}$, $\mathbf{\Lambda}$ or $\mathbf{\Psi}$, keeping others fixed. The result of each minimization gives the updated parameter.

Standard deterministic ADMM implicitly assumes true data values are available, hence overlooking the existence of noise [54]. Noticing that only $\mathbf{K}$ and $\mathbf{P}$ depend on the data samples, we define the corresponding update rules using the scalable stochastic ADMM approach [55,56] which can handle such an issue.

Update K and P: Let $\frac{\partial}{\partial \mathbf{K}} \mathcal{J}_{\mathbf{K},\mathbf{P}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \mathbf{K}} \ell_{\mathbf{K},\mathbf{P}}(p^{(i)}, g^{(i)})$ and $\frac{\partial}{\partial \mathbf{P}} \mathcal{J}_{\mathbf{K},\mathbf{P}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \mathbf{P}} \ell_{\mathbf{K},\mathbf{P}}(p^{(i)}, g^{(i)})$ denote the subgradients components of Eq. (4) computed for all samples with respect $\mathbf{K}$ and $\mathbf{P}$, respectively. Then, at each iteration t, i.e., for the t-th random sample, we compute

$$\tilde{\mathbf{K}}^{(t+1)} = \tilde{\mathbf{K}}^{(t)} - \eta \Bigg(\frac{\partial}{\partial \tilde{\mathbf{K}}^{(t)}} \ell_{\tilde{\mathbf{K}}^{(t)}, \tilde{\mathbf{P}}^{(t)}}(p^{(t)}, g^{(t)}) - \frac{\partial}{\partial \mathbf{K}^{(s)}} \ell_{\mathbf{K}^{(s)}, \mathbf{P}^{(s)}}(p^{(t)}, g^{(t)}) + \frac{\partial}{\partial \mathbf{K}^{(s)}} \mathcal{J}_{\mathbf{K}^{(s)}, \mathbf{P}^{(s)}} + \rho \left(\tilde{\mathbf{K}}^{(t)} - \mathbf{U}^{(s)} + \mathbf{\Lambda}^{(s)}/\rho \right) \Bigg) \tag{7}$$

$$\tilde{\mathbf{P}}^{(t+1)} = \tilde{\mathbf{P}}^{(t)} - \eta \Bigg(\frac{\partial}{\partial \tilde{\mathbf{P}}^{(t)}} \ell_{\tilde{\mathbf{K}}^{(t+1)}, \tilde{\mathbf{P}}^{(t)}}(p^{(t)}, g^{(t)}) - \frac{\partial}{\partial \mathbf{P}^{(s)}} \ell_{\mathbf{K}^{(s)}, \mathbf{P}^{(s)}}(p^{(t)}, g^{(t)}) + \frac{\partial}{\partial \mathbf{P}^{(s)}} \mathcal{J}_{\mathbf{K}^{(s)}, \mathbf{P}^{(s)}} + \rho \left(\tilde{\mathbf{P}}^{(t)} - \mathbf{V}^{(s)} + \mathbf{\Psi}^{(s)}/\rho \right) \Bigg) \tag{8}$$

where η is the step size and $\tilde{\mathbf{K}}^{(t)}$ and $\tilde{\mathbf{P}}^{(t)}$ denote the parameters for a specific iteration t, while $\mathbf{K}^{(s)}$ and $\mathbf{P}^{(s)}$ represent the parameters obtained for epoch s. Once T iterations are completed, the two low-rank matrices are updated as

$$\mathbf{K}^{(s+1)} = \frac{1}{T} \sum_{t=1}^{T} \tilde{\mathbf{K}}^{(t)} \qquad \mathbf{P}^{(s+1)} = \frac{1}{T} \sum_{t=1}^{T} \tilde{\mathbf{P}}^{(t)} \tag{9}$$

Update U and V: To derive the updates for the two regularizers, we first compute the partial derivatives of Eq. (6) with respect to $\mathbf{U}$ and $\mathbf{V}$ while keeping other parameters fixed. Then, solving for a stationary point yields

$$\mathbf{U}^{(s+1)} = \left(\mathbf{K}_{i,:}^{(s+1)} + \mathbf{\Lambda}_{i,:}^{(s)}/\rho \right) \max\left(0, 1 - \alpha / \left(\rho \left\| \mathbf{K}_{i,:}^{(s+1)} + \mathbf{\Lambda}_{i,:}^{(s)}/\rho \right\|_2 \right)\right) \tag{10}$$

$$\mathbf{V}^{(s+1)} = \left(\mathbf{P}_{i,:}^{(s+1)} + \mathbf{\Psi}_{i,:}^{(s)}/\rho \right) \max\left(0, 1 - \beta / \left(\rho \left\| \mathbf{P}_{i,:}^{(s+1)} + \mathbf{\Psi}_{i,:}^{(s)}/\rho \right\|_2 \right)\right) \tag{11}$$

whose closed form solutions have been obtained using the group soft-thresholding technique [51] and $i = 1, \cdots, r$ denotes the i-th row of a parameter matrix.

Update $\mathbf{\Lambda}$ and $\mathbf{\Psi}$: Results from Eq. (9) and Eqs. (10–11) can be finally used to update the duals for the Lagrangian multipliers as

$$\mathbf{\Lambda}^{(s+1)} = \mathbf{\Lambda}^{(s)} + \rho(\mathbf{K}^{(s+1)} - \mathbf{U}^{(s+1)}) \tag{12}$$

$$\mathbf{\Psi}^{(s+1)} = \mathbf{\Psi}^{(s)} + \rho(\mathbf{P}^{(s+1)} - \mathbf{V}^{(s+1)}) \tag{13}$$

To conclude, after S epochs have been performed, the optimal estimates for the two low-rank projection matrices are given by $\mathbf{K}^{(S)}$ and $\mathbf{P}^{(S)}$.

3.3 Model Adaptation with Reduced Human Effort

In the previous section we have presented a similarity-dissimilarity learning model which can be trained in an incremental fashion. To achieve model adaptation over time, we propose to perform incremental steps to minimize Eq. (6) with new image pairs that are progressively acquired as time passes. This requires human labeling of such pairs. To limit such a manual effort and improve model generalization, we aim to select only a small set of informative gallery persons to update the model. These are persons for which the positive/negative association with the probe is very uncertain. Given a probe, such gallery persons form its *probe relevant set*.

Probe Relevant Set Selection: Let $\mathcal{H} = \{\mathbf{x}_p, \mathbf{x}_g \mid g = 1, \ldots, |\mathcal{G}|\}$ denote the probe-gallery set for probe p. We represent such a set as an undirected graph with no loops. More precisely, let $G = (V, E, \mathbf{W})$ denote a graph where $V = \{p, g | g = 1, \ldots, |\mathcal{G}|\}$ is the set of vertices, $E \subseteq V \times V$ is the set of edges and $\mathbf{W} \in \mathbb{R}_+^{|V|\times|V|}$ denotes the adjacency symmetric matrix of positive edge weights such that, for any two vertices i and j, $\mathbf{W}_{i,j} = f(S_{\mathbf{K},\mathbf{P}}(i,j))$ if $i \neq j$, $\mathbf{W}_{i,j} = 0$, otherwise. $f(\cdot)$ is the Platt function [57] used to ensure a positive edge weight.

To obtain the probe relevant set, we aim to cluster G in such a way that (i) a cluster contains the probe and gallery persons which are similar to each other, and (ii) all persons outside a cluster should be dissimilar to the ones inside. To achieve such an objective, we exploit the dominant sets clustering technique [12].

Dominant set clustering partitions a graph into dominant sets on the basis of the coherency between vertices as measured by the edge weights. A dominant set is a subset of the graph nodes having high internal and low external coherency.

To obtain such partitions, the dominant sets approach is based on the participation vector $\mathbf{h}$. It expresses the probability of participation of the corresponding person in the cluster. More precisely, the objective is

$$\hat{\mathbf{h}} = \arg\max_{\mathbf{h}} \mathbf{h}^T \mathbf{W} \mathbf{h} \qquad \text{s.t.} \quad \mathbf{h} \in \mathcal{S} \tag{14}$$

where $\mathcal{S}$ is the standard simplex of $\mathbb{R}^{|V|}$.

Let the participation vector be initialized to a uniform distribution, i.e., $h_i = 1/|V|$, for $i = 1, \ldots, |V|$[1]. Then, as shown in [12], a solution to the optimization problem can be obtained by an iterative procedure that, at each iteration k, updates the participation vector as

$$h_i^{(k+1)} = h_i^{(k)} \frac{(\mathbf{W}\mathbf{h}^{(k)})_i}{(\mathbf{h}^{(k)})^T \mathbf{W} \mathbf{h}^{(k)}} \qquad \text{for } i = 1, \ldots, |V| \tag{15}$$

The iterative updates are applied until the objective function difference between two consecutive iterations is higher than a predefined threshold ϵ. When such a condition is not satisfied a local optima is obtained and the non-zero

[1] Effect of this initialization is checked by adding random noise to each element of $\mathbf{h}$. Results show that in 96 % of the cases the output cluster is the same.

Algorithm 1. Temporal Model Adaptation for Person Re-Identification

Off-Line Training

Input: $\mathcal{X}$, $\eta > 0$, $\rho > 0$, $T > 0$, $S > 0$
Output: Discriminative low rank projection matrices $\mathbf{K}$ and $\mathbf{P}$
Initialize: $\mathbf{K}^{(1)}$ and $\mathbf{P}^{(1)}$ to random, $\mathbf{\Lambda}^{(1)}$ and $\mathbf{\Psi}^{(1)}$ to $\mathbf{0}$
Set: $\mathbf{U}^{(1)} = \mathbf{K}^{(1)}$,$\mathbf{V}^{(1)} = \mathbf{P}^{(1)}$
Iterate *for* $s = 1, \ldots, S$
- **1.** Consider all the n training samples to pre-compute the average hinge loss subgradients with respect to $\mathbf{K}^{(s)}$ and $\mathbf{P}^{(s)}$
- **2.** Set $\tilde{\mathbf{K}}^{(t)} = \mathbf{K}^{(s)}$, $\tilde{\mathbf{P}}^{(t)} = \mathbf{P}^{(s)}$, then run T iterations and update $\tilde{\mathbf{K}}^{(t)}$ and $\tilde{\mathbf{P}}^{(t)}$ as in Eqs. (7) and (8)
- **3.** Average over the T updates as in Eq. (9) to obtain $\mathbf{K}^{(s+1)}$ and $\mathbf{P}^{(s+1)}$
- **4.** Update the constraints $\mathbf{U}^{(s)}$ and $\mathbf{V}^{(s)}$ using Eqs. (10) and (11)
- **5.** Compute the dual updates for the Lagrangian multipliers as in Eqs. (12) and (13)

6. Obtain the optimal estimates $\mathbf{K} = \mathbf{K}^{(S)}$ and $\mathbf{P} = \mathbf{P}^{(S)}$

Temporal Model Adaptation

Input: $\mathbf{K}$, $\mathbf{P}$, $\mathcal{H}$, $\eta > 0$, $\rho > 0$, $\hat{T} > 0$, $\hat{S} > 0$, $\epsilon > 0$
Output: Updated discriminative low rank projection matrices $\mathbf{K}$ and $\mathbf{P}$
1. Compute the scores for each possible probe-gallery pair via $S_{\mathbf{K},\mathbf{P}}$ to obtain $\mathbf{W}$
2.Solve the problem in Eqs. (14) – using Eq. (15)– to obtain the probe relevant set $\mathcal{D}_p$
3. Form the set of probe relevant pairs
4. Update $\mathbf{K}$ and $\mathbf{P}$ by performing *off-line training* steps **1**–**6** with the probe relevant pairs, $S = \hat{S}$ and $T = \hat{T}$

entries in the participation vector $\hat{\mathbf{h}}$ specify the relevant nodes included in the dominant set. Notice that the dominant sets clustering can be easily extended to cluster a graph in multiple dominant sets. This is obtained by removing the person identities included in the current dominant set from $\mathcal{H}$, creating the new graph structure and then repeating the process. In our approach such a procedure is applied until the dominant set containing the probe person p is found. This is the probe relevant set for person p and is denoted as $\mathcal{D}_p = \{i \mid i \neq p \wedge h_i > 0\}$.

Incremental Model Update: Armed with the probe relevant set, we can now achieve temporal model adaptation by performing the incremental learning steps described in Sect. 3.2. Towards this objective, we first ask the human annotator to label only the probe relevant pairs in $\{(\mathbf{I}_p, \mathbf{I}_g) \mid g \in \mathcal{D}_p\}$. Then, using the current parameters $\mathbf{K}$ and $\mathbf{P}$ as a "warm-restart", we exploit the newly labeled samples to run $\hat{S}$ epochs, each providing $\hat{T}$ incremental iterations. When such a process is completed the updated model parameters $\mathbf{K}$ and $\mathbf{P}$ are obtained.

3.4 Discussion

Through the preceding sections we have introduced two main contributions that allow us to obtain model adaptation over time. Specifically, the goal has been achieved (i) by proposing a stochastic similarity-dissimilarity metric learning procedure that can be incrementally updated and (ii) by introducing a graph-based approach that allows to identify the most informative pairs that should be labeled by the human. All the steps are summarized in Algorithm 1.

MLAPG [20] and XQDA [6], which learn a discriminant subspace as well as a distance function in the learned subspace, are close to the proposed approach.

(a) (b) (c)

Fig. 3. 15 image pairs from the (a) VIPeR, (b) PRID450S and (c) Market1501 datasets. Columns correspond to different persons, rows to different cameras.

However, both of them do not update the model over time. In addition, our solution differs in the stochastic ADMM optimization, the combination of both a similarity and a dissimilarity measure, as well as the sparsity regularization.

4 Experimental Results

Datasets: We evaluated our approach on three publicly available benchmark datasets[2], namely VIPeR [58], PRID450S [59], and Market1501 [60] (see Fig. 3 for few sample images). Following the literature, we run 10 trials on the VIPeR and PRID450S dataset, while we use the available partitions for Market 1501. We report on the average performance using the Cumulative Matching Characteristic (CMC). We refer to our method as Temporal Model Adaptation (TMA).

VIPeR [58] is considered one of the most challenging datasets. It contains 1,264 images of 632 persons viewed by two cameras. Most image pairs have viewpoint changes larger than 90°. Following the general protocol, we split the dataset into a training and a test set each including 316 persons.

PRID450S [59] is a more recent dataset containing 450 persons viewed by two disjoint cameras with viewpoint changes, background interference and partial occlusion. As performed in literature [61,62], we partitioned the dataset into a training and a test set each containing 225 individuals.

Market1501 [60] is the largest currently available person re-identification dataset. It contains 32,668 images of 1,501 persons taken from 6 disjoint cameras. Multiple images of a same person have been obtained by means of a state-of-the-art detector, thus providing a realistic setup. To run the experiments, we used the available code[3] to get the same BoW feature representation as well as the same train/test partitions containing 750 and 751 person identities each.

Implementation: To model person appearance we adopted the Local Maximal Occurrence (LOMO) representation [6]. We selected $\alpha = 0.001$, $\beta = 0.001$, $\eta = 1$, and $\rho = 1$ by performing 5-fold cross validation on $\{1, 0.5, 0.1, 0.05, 0.01, 0.001\}$. The temporal model adaptation followed the common batch framework used in

[2] See supplementary for additional results on the 3DPeS and CUHK03 datasets.
[3] http://www.liangzheng.com.cn.

Table 1. Comparison with state-of-the-art methods on the VIPeR dataset. Best results for each rank are in boldface font.

Rank →	1	10	20	50	Labeled [%]	Reference
TMA_4+LADF	**48.19**	**87.65**	93.54	98.41	20.32+100	Proposed + [24]
TMA_0	43.83	83.86	91.45	97.47	100	Proposed
LMF+LADF	43.29	85.13	**94.12**	–	100	CVPR 2014 [63]+[24]
TMA_4	41.46	82.65	92.46	**99.65**	20.32	Proposed
MLAPG	40.73	82.34	92.37	–	100	ICCV 2015 [20]
XQDA	40.00	80.51	91.08	–	100	CVPR 2015 [6]
TMA_3	37.97	75.00	87.66	96.52	12.56	Proposed
SCNCDFinal	37.80	81.20	90.40	97.0	100	ECCV 2014 [61]
PKFM	36.8	83.7	91.7	97.8	100	CVPR 2015 [9]
TMA_2	36.08	71.84	81.96	94.62	6.78	Proposed
TMA_1	35.13	69.94	81.01	93.35	4.91	Proposed
QALF	30.17	62.44	73.81	–	100	CVPR 2015 [60]
ISR	27.43	61.06	72.92	86.69	100	TPAMI 2015 [3]
WFS	25.81	69.56	83.67	95.12	100	TPAMI 2015 [64]
KISSME	19.60	62.20	77.00	91.80	100	CVPR 2012 [21]

active learning [34]. It partitioned each training set into 4 disjoint *batches*. Due to the adopted randomization procedure, each batch contains approximately $z = (|\mathcal{P}|/4) \times (|\mathcal{G}|/4)$ pairs[4]. We have used the first batch to train the initial model with $T = 2z$, $S = 200$, and no further stopping criteria. The remaining ones have been used for the batch-incremental updates with $\hat{T} = 2z$ and $\hat{S} = 150$ (in the following, the subscript of TMA indicates the number of model updates that are achieved for every probe in each batch). Finally, to select the relevant gallery images in each batch we have set $\epsilon = 0.1$ (see Table 5).

4.1 State-of-the-art Comparisons

In the following we compare the results of our approach with existing methods. In addition to the incremental performance, we also provide our results when no model adaptation is exploited and all the training data is included in one single batch (TMA_0).

VIPeR: Results in Table 1 show that our approach has better performance than recent solutions even in the case only about 5 % of the data is used. This result indicates that, partially due to the feature representation (see results of KISSME in Table 4), our approach produces a robust solution to viewpoint variations. Incremental updates bring TMA_4 to be the second best. In such a case, only

[4] The percentage of labeled pairs is computed with respect to n.

Table 2. Comparison with state-of-the-art methods on the PRID 450S dataset. Best results for each rank are in boldface font.

Rank →	1	5	10	20	50	Labeled [%]	Reference
TMA_0	**54.22**	73.78	83.11	90.22	97.33	100	Proposed
TMA_4	52.89	**76.00**	**85.78**	**93.33**	**97.78**	14.25	Proposed
TMA_3	50.22	75.56	85.33	92.89	97.64	10.18	Proposed
TMA_2	48.89	75.33	84.01	91.1	97.33	8.64	Proposed
TMA_1	45.33	72.00	83.11	89.78	96.02	6.42	Proposed
CSL	44.4	71.6	82.2	89.8	96.0	100	ICCV2015 [62]
SCNCDFinal	41.6	48.9	79.4	87.8	95.4	100	ECCV2014 [61]
SCNCD	41.5	66.6	75.9	84.4	92.4	100	ECCV2014 [61]
KISSME	33	–	71	79	90	100	CVPR2012 [21]

LMF+LADF performs better. However, such an approach is a combination of two methods, which, as shown in [5], generally improves the performance. Indeed, a rank 1 recognition rate of 48.19 % is achieved by summing TMA_0 and LADF scores. If the same batches as TMA_{1-4} are considered to train LADF, the fused rank 1 performances are of 35.6 %, 37.9 %, 40.8 % and 43.4 %, respectively – which represent an average improvement of 11 % over standalone LADF.

Finally, results obtained with TMA_0 show that the best rank 1 is achieved, but performance on higher ranks is slightly worse than the one obtained using incremental updates (TMA_4). Hence, using all the available data requires additional manual labor and might also drive to decreasing performance. This strengthens our contribution showing that, by identifying the most informative samples to train with, better results can be achieved with reduced human effort.

PRID450S: In Table 2 we report on the performance comparisons between existing method and our approach on the PRID450S dataset. Results show that our solution outperforms the methods used for comparisons regardless of the amount of data used for training. In particular, using only 14.25 % of the data an 8 % improvement with respect to the best existing approach is obtained at rank 1. By training only with the initially available data (i.e., TMA_1), our solution outperforms SCNCDFinal [61], which, on the VIPeR dataset, had better performance (until the 3^{rd} batch update). This may suggest that our approach is robust to background clutter and occlusions which PRID450S suffer from.

Market1501: Comparisons of our approach with existing methods on the Market 1501 dataset are shown in Table 3. The obtained performance are consistent with the ones achieved on the VIPeR and PRID450S datasets. Our approach has significantly better performance than methods used for comparisons even by using 5.23 % of labeled data. Incremental updates bring in relevant improvements and with TMA_4 we achieve the best rank 1 recognition rate, i.e., 44.74 %. Using the LOMO feature representation instead of the BoW one provided by [60], about a 3 % rank 1 performance gain is obtained. Results on such dataset

Table 3. Rank 1 and mAP performance comparison with existing methods on the Market 1501 dataset. Best result is in boldface font.

Method	BoW [60]	LMNN BoW [60]	ITML BoW [60]	KISSME BoW [60]	TMA_1 BoW	TMA_2 BoW	TMA_3 BoW	TMA_4 BoW	TMA_4 LOMO
Rank 1	42.64	38.91	27.08	43.03	28.77	34.68	39.81	44.74	**47.92**
mAP	19.47	17.34	8.13	19.98	8.69	14.56	17.89	20.92	**22.31**
Labeled [%]	100	100	100	100	5.23	8.71	12.09	14.36	13.58

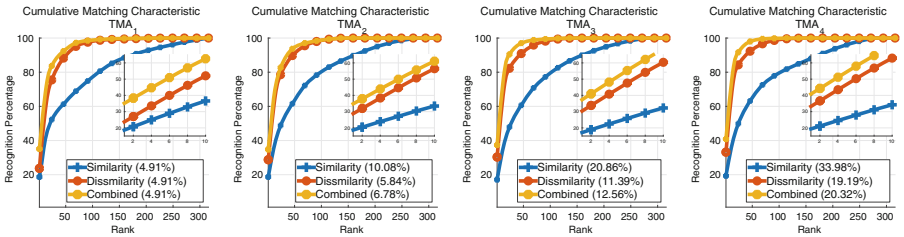

Fig. 4. Comparison of the similarity-dissimilarity learning components. (a)–(d) show the results on the VIPeR dataset computed using incremental batch updates. For each curve, the percentage of manually labeled samples is indicated in parenthesis. The inside picture show the results for rank range 1–10.

demonstrate that our approach can scale to a real scenario and achieve competitive performance with significantly less manual labor. The reason for the improved performance with much less training data is because our method identifies the most discriminating examples to train with, and does not waste labeling effort on those that will add little or no value to the re-identification accuracy.

4.2 Influence of the Temporal Model Adaptation Components

To better understand the achieved performance, we have run additional experiments by separately considering the similarity-dissimilarity metric learning approach and the probe relevant set selection method.

Similarity-Dissimilarity Metric: In the following, we first analyze the contribution of the similarity and the dissimilarity components. Then, we compare our performance with existing methods using the same LOMO representation.

Contribution of the components: In Fig. 4, we report on the results obtained using either the learned similarity, the learned dissimilarity or both. Results show that most of the performance contribution is provided by the dissimilarity. The similarity has significantly lower performance and calls for more labeled pairs. This is due to the fact that the majority of the edges of the corresponding graph have weak weights, thus causing the maximization procedure to select more samples before the stop condition. Enforcing agreement on a specific pair by jointly optimizing the similarity and the dissimilarity measure results in the best

Table 4. Comparison with metric learning approaches on the VIPeR dataset. Results obtained using truncated projections (100 dimensions) are given for three representative ranks. Last row shows the percentage of manually labeled samples. Best results for each rank are in bold. Most of the results are from [20].

Rank ↓	MLAPG	XQDA	KISSME	LMNN	LADF	ITML	LDML	PRDC	TMA_1	TMA_2	TMA_3	TMA_4
1	39.21	38.23	33.54	28.42	27.63	19.02	13.99	12.15	32.28	34.81	36.07	**39.88**
10	**81.42**	81.14	79.30	72.31	75.47	52.31	38.64	35.82	69.62	73.10	76.27	81.33
20	**92.50**	92.18	90.47	85.32	88.29	67.34	48.73	48.26	81.33	58.79	90.19	91.46
Labeled [%]	100	100	100	100	100	100	100	100	4.91	6.91	11.48	15.77

performances. With respect to the dissimilarity approach, this yields negligible increase of manual labor and improved results (7 % at rank 1).

Comparison with existing methods: In Table 4, we report on the comparison of our similarity-dissimilarity approach with general state-of-the-art metric learning approaches, namely ITML [65], LMNN [66], LDML [67], and re-identification tied ones namely, PRDC [30], KISSME [21], LADF [24], XQDA [6], and MLAPG [20]. To provide a fair comparison, we used the same settings in [20]. Precisely, the 100 principal components found by PCA have been exploited to train LMNN, ITML, KISSME, and LADF. Since other methods, i.e., XQDA, PRDC, LDML, MLAPG and TMA, are able to discover the discriminative features, we used all the principal components. For a fair comparison, projection learned by XQDA, MLAPG and TMA were truncated to 100 dimensions.

Results in Table 4 show that our approach, trained with only 4.91 % of the available data, has the 4^{th} best rank 1 result. As shown in Fig. 4, such a successful result is due to the competition between the similarity and the dissimilarity approaches. Performing incremental updates yields significant improvements and, after the 4^{th} update is completed, the best rank 1 recognition rate is achieved. At higher ranks, TMA performs on par with other methods but with substantially less labeled pairs (i.e., 15.77 % of all possible annotations).

Discussion: Results have demonstrated that, while the dissimilarity metric has more impact on the performance, by enforcing competition with the similarity measure better results can be obtained. Additional evaluations showed that by removing the $\ell_{2,1}$ norms the degradation is of 3 %. Comparisons with existing approaches have shown that, under the same conditions, our approach achieves good results using only 1/6 of the data. Incremental updates produce considerable improvements with a significantly reduced human effort. This substantiates the benefits of the proposed similarity-dissimilarity learning approach and demonstrate the feasibility of temporal model adaptation for the task.

Probe Relevant Set Selection: In the following, we provide an analysis of the graph-based solution to identify the most informative gallery persons. We report on the effects of the ϵ parameter, then we compare with three approaches.

Influence of ϵ: To verify the influence of the ϵ parameter, we have computed the results in Table 5. These show that, large values of ϵ produce coarse under-segmented sets, hence identify a large number of relevant pairs to label.

Table 5. Analysis of the ϵ parameter used to obtain the probe relevant set. Each entry in the table shows the rank 1 performance as well as the percentage of labeled data (in brackets). Best results for each rank are in bold.

$\epsilon \rightarrow$	0.5	0.3	0.1	0.05	0.01
TMA_1	35.13 (4.91)	35.13 (4.91)	35.13 (4.91)	35.13 (4.91)	35.13 (4.91)
TMA_2	**36.08** (7.87)	35.76 (7.26)	**36.08** (6.78)	34.49 (6.62)	34.49 (6.20)
TMA_3	37.03 (11.36)	36.23 (10.59)	**37.97** (12.56)	36.71 (8.97)	34.81 (7.95)
TMA_4	38.61 (14.74)	38.92 (13.50)	**41.46** (20.32)	39.87 (11.49)	37.97 (9.81)

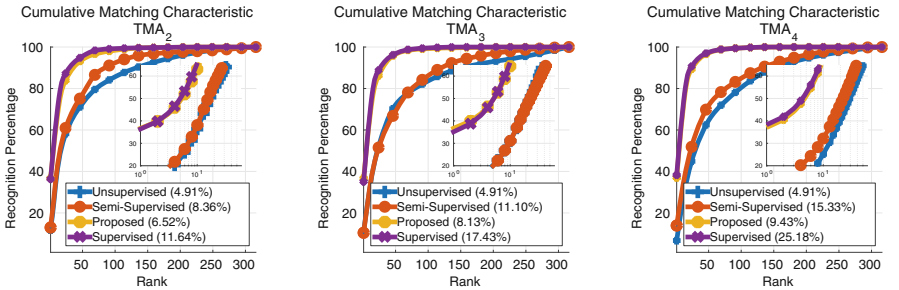

Fig. 5. Re-Identification performance on the VIPeR dataset computed using four different probe relevant set selection criteria. (a)–(c) show the performances achieved using the 2^{nd}-4^{th} batch incremental updates. The percentage of manually labeled samples is given within parenthesis. The inside picture show the results on a log-scale reduced rank range, i.e., 1–50.

Small values of ϵ, e.g., 0.01, produce over segmented-graphs, hence small dominant sets. Indeed, after the 4^{th} update, less than 10 % of all the available pairs has been used for training. This results in achieving similar performance improvements, but with a different manual effort. The reason behind this is that, in the former case, the probe relevant sets contain additional persons which are not "similar" to the probe and any other gallery person. This causes the model to be updated with uninformative pairs which weaken its discriminative power. In the latter, too few informative pairs are found and the model overfits such samples.

Selection Criteria Comparison: In Fig. 5, we compare our probe relevant set selection approach with three different criteria. Before exploiting such criteria, we applied Platt scaling [57] to the obtained scores to get the probability of each probe-gallery pair being positive.

(i) *Unsupervised*: Each pair having probability less than 0.5 has been assigned the negative label, remaining ones have been assigned the positive label.

(ii) *Semi-Supervised*: Top and bottom 20 ranked pairs have been labeled as positive or the negative, respectively. Remaining pairs have been human labeled.
(iii) *Supervised*: Every pair has been human labeled.

Results show that using the unsupervised or the semi-supervised criteria, the performance obtained with incremental updates tends to decrease. This behavior is due to the fact that, right after the first update, the produced scores induce very small or very large probabilities. This yields zero manual labor, but, as a consequence, the model is updated with a large portion of mislabeled samples. Using our solution, performance reaches the ones obtained using a fully-supervised approach. In particular, with the 4^{th} batch update our approach yields the highest rank 1 recognition rate (41.46 % vs 39.87 %) with 5 % less manual labor. Additional experiments considering the human mislabeling error $C \in \{5, \ldots, 95\}\%$ show that the model update is effective when $C \leq 15\,\%$.

Discussion: In this section, we have shown that our approach is moderately sensible to the selection of ϵ, which to some extent, controls the human effort. In addition, it performs better than a fully supervised approach in which all the samples are manually labeled. This demonstrates that the proposed approach identifies the most informative pairs that should be used to update the model.

4.3 Computational Complexity

In Table 6, we compare the computational performance of deterministic ADMM and our stochastic solution. While achieving similar rank 1 performance, deterministic ADMM brings in more complexity, hence the training time is considerably higher. In particular, while d might be arbitrarily large, n and K are usually small (those depend on the number of samples which are manually labeled), thus our solution is more desirable in a continuous learning scenario.

Finally, notice that, while the initial training is more expensive than existing approaches, e.g., KISSME [21], the proposed incremental learning solution is more effective in the long term since it does not require re-training like others.

Table 6. Comparison between deterministic ADMM and our stochastic solution. VIPeR result computed by running MATLAB code on an Intel Xeon 2.6 GHz. Complexity is computed for the parameters updates which differs from the two solutions

Method ↓	TMA_1 - Rank 1	Per-Epoch complexity	Training time [s]
Deterministic ADMM	34.84	$\mathcal{O}\left(2(n2d^2 + d^3)\right)$	12051.19
Stochastic ADMM	35.15	$\mathcal{O}\left(2(n3d^2 + K3d^2)\right)$	2948.38

5 Conclusion

In this paper we have proposed a person re-identification approach based on a temporal adaptation of the learned model with human in the loop. First, to

allow temporal adaptation, we have proposed a similarity-dissimilarity metric learning approach which can be trained in an incremental fashion by means of a stochastic version of the ADMM optimization method. Then, to update the model with the proper information, we have included the human in the loop and proposed a graph-based approach to select the most informative pairs that should be manually labeled. Informative pairs selection has been obtained through the dominant sets graph partition technique. Results conducted on three datasets have shown that similar or better performances than existing methods can be achieved with significantly less manual labor.

Acknowledgment. The work was partially supported by US NSF grant IIS-1316934.

References

1. Vezzani, R., Baltieri, D., Cucchiara, R.: People reidentification in surveillance and forensics. ACM Comput. Surv. **46**(2), 1–37 (2013)
2. Wu, Z., Li, Y., Radke, R.J.: Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features. IEEE Trans. Pattern Anal. Mach. Intell. **37**(5), 1095–1108 (2015)
3. Lisanti, G., Masi, I., Bagdanov, A.D., Bimbo, A.D.: Person re-identification by iterative re-weighted sparse ranking. IEEE Trans. Pattern Anal. Mach. Intell. **37**(8), 1629–1642 (2015)
4. Martinel, N., Micheloni, C.: Sparse matching of random patches for person reidentification. In: International Conference on Distributed Smart Cameras, pp. 1–6 (2014)
5. Xiong, F., Gou, M., Camps, O., Sznaier, M.: Using kernel-based metric learning methods. In: European Conference Computer Vision, pp. 1–16 (2014)
6. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: International Conference on Computer Vision and Pattern Recognition (2015)
7. Paisitkriangkrai, S., Shen, C., Hengel, A.V.D.: Learning to rank in person reidentification with metric ensembles. In: International Conference on Computer Vision and Pattern Recognition (2015)
8. Liu, C., Loy, C.C., Gong, S., Wang, G.: POP: person re-identification post-rank optimisation. In: International Conference on Computer Vision (2013)
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: International Conference on Computer Vision and Pattern Recognition (2015)
10. Garcia, J., Martinel, N., Micheloni, C., Gardel, A.: Person re-identification ranking optimisation by discriminant context information analysis. In: International Conference on Computer Vision (2015)
11. Boyd, S., Parikh, N., E Chu, B.P., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2010)
12. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 167–172 (2007)
13. Zheng, W.S., Li, X., Xiang, T., Liao, S., Lai, J., Gong, S.: Partial person reidentification. In: International Conference on Computer Vision, pp. 4678–4686 (2015)

14. Li, X., Zheng, W.S., Wang, X., Xiang, T., Gong, S.: Multi-scale learning for low-resolution person re-identification. In: International Conference on Computer Vision, pp. 3765–3773 (2015)
15. Martinel, N., Foresti, G.L., Micheloni, C.: Person reidentification in a distributed camera network framework. IEEE Trans. Cybern. 1–12 (in press, 2016)
16. Zheng, W.S., Gong, S., Xiang, T.: Towards open-world person re-identification by one-shot group-based verification. IEEE Trans. Pattern Anal. Mach. Intell. **8828**(2), 1–1 (2015)
17. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning, with application to clustering with side-information. Adv. Neural Inf. Process. Syst. **15**, 505–512 (2002)
18. Bellet, A., Habrard, A., Sebban, M.: A Survey on Metric Learning for Feature Vectors and Structured Data. ArXiv e-prints, June 2013
19. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part VI. LNCS, vol. 7577, pp. 780–793. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33783-3_56
20. Liao, S., Li, S.Z.: Efficient PSD constrained asymmetric metric learning for person re-identification. In: International Conference on Computer Vision, pp. 3685–3693 (2015)
21. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: International Conference on Computer Vision and Pattern Recognition, pp. 2288–2295 (2012)
22. Tao, D., Jin, L., Wang, Y., Yuan, Y., Li, X.: Person re-identification by regularized smoothing KISS metric learning. IEEE Trans. Circ. Syst. Video Technol. **23**(10), 1675–1685 (2013)
23. Tao, D., Jin, L., Wang, Y., Li, X.: Person reidentification by minimum classification error-based KISS metric learning. IEEE Trans. Cyber. **45**(2), 1–11 (2014)
24. Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.: Learning locally-adaptive decision functions for person verification. In: International Conference on Computer Vision and Pattern Recognition, pp. 3610–3617. IEEE, June 2013
25. Pedagadi, S., Orwell, J., Velastin, S.: Local fisher discriminant analysis for pedestrian re-identification. In: International Conference on Computer Vision and Pattern Recognition, pp. 3318–3325 (2013)
26. Martinel, N., Micheloni, C.: Classification of local eigen-dissimilarities for person re-identification. ieee sig. process. lett. **22**(4), 455–459 (2015)
27. García, J., Martinel, N., Gardel, A., Bravo, I., Foresti, G.L., Micheloni, C.: Modeling feature distances by orientation driven classifiers for person re-identification. J. Vis. Commun. Image Representation **38**, 115–129 (2016)
28. Martinel, N., Micheloni, C., Foresti, G.L.: Kernelized saliency-based person re-identification through multiple metric learning. IEEE Trans. Image Process. **24**(12), 5645–5658 (2015)
29. Chen, J., Zhang, Z., Wang, Y.: Relevance metric learning for person re-identification by exploiting listwise similarities. IEEE Trans. Image Process. **7149**(c), 1–1 (2015)
30. Zheng, W.S., Gong, S., Xiang, T.: Re-identification by relative distance comparison. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 653–668 (2013)
31. Li, W., Zhao, R., Xiao, T., Wang, X.: DeepReID: deep filter pairing neural network for person re-identification. In: Conference on Computer Vision and Pattern Recognition, pp. 152–159, June 2014

32. Ahmed, E., Jones, M., Marks, T.K.: An improved deep learning architecture for person re-identification. In: IEEE International Conference on Computer Vision and Pattern Recognition (2015)
33. Zhang, R., Lin, L., Zhang, R., Zuo, W., Zhang, L.: Bit-Scalable deep hashing with regularized similarity learning for image retrieval and person re-identification. IEEE Trans. Image Process. **24**(12), 4766–4779 (2015)
34. Settles, B.: Active learning. Synth. Lect. Artif. Intell. Mach. Learn. **6**(1), 1–114 (2012)
35. Joshi, A.J., Porikli, F., Papanikolopoulos, N.P.: Scalable active learning for multi-class image classification. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2259–2273 (2012)
36. Aodha, O.M., Campbell, N.D.F., Kautz, J., Brostow, G.J.: Hierarchical subquery evaluation for active learning on a graph. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)
37. Biswas, A., Parikh, D.: Simultaneous active learning of classifiers & attributes via relative feedback. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
38. Chakraborty, S., Balasubramanian, V.N., Panchanathan, S.: Optimal batch selection for active learning in multi-label classification. In: ACM International Conference on Multimedia, pp. 1413–1416 (2011)
39. Elhamifar, E., Sapiro, G., Yang, A., Sasrty, S.S.: A convex optimization framework for active learning. In: IEEE International Conference on Computer Vision, pp. 209–216 (2013)
40. Vondrick, C., Ramanan, D.: Video annotation and tracking with active learning. In: Advances in Neural Information Processing Systems (2011)
41. Vijayanarasimhan, S., Grauman, K.: Large-scale live active learning: training object detectors with crawled data and crowds. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
42. Vezhnevets, A., Buhmann, J.M., Ferrari, V.: Active learning for semantic segmentation with expected change. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3162–3169 (2012)
43. Karasev, V., Ravichandran, A., Soatto, S.: Active frame, location, and detector selection for automated and manual video annotation. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)
44. Hasan, M., Roy-Chowdhury, A.K.: Context aware active learning of activity recognition models. In: IEEE International Conference on Computer Vision (2015)
45. Wang, Z., Hu, R., Liang, C., Leng, Q., Sun, K.: Region-based interactive ranking optimization for person re-identification. In: Ooi, W.T., Snoek, C.G.M., Tan, H.K., Ho, C.-K., Huet, B., Ngo, C.-W. (eds.) PCM 2014. LNCS, vol. 8879, pp. 1–10. Springer, Heidelberg (2014). doi:10.1007/978-3-319-13168-9_1
46. Das, A., Panda, R., Roy-Chowdhury, A.: Active image pair selection for continuous person re-identification. In: International Conference on Image Processing (2015)
47. Chechik, G., Sharma, V., Shalit, U., Bengio, S.: Large scale online learning of image similarity through ranking. J. Mach. Learn. Res. **11**, 1109–1135 (2010)
48. Guo, Z.C., Ying, Y.: Guaranteed classification via regularized similarity learning. Neural Comput. **26**(3), 497–522 (2013)
49. Xia, H., Hoi, S.C.H., Jin, R., Zhao, P.: Online multiple kernel similarity learning for visual search. IEEE Trans. Pattern Anal. Mach. Intell. **36**(3), 536–549 (2014)
50. Nie, F., Huang, H., Cai, X., Ding, C.H.: Efficient and robust feature selection via joint L2, 1-norms minimization. In: Advances in Neural Information Processing Systems, pp. 1813–1821 (2010)

51. Bach, F., Jenatton, R., Mairal, J., Obozinski, G.: Convex optimization with sparsity-inducing norms. In: Sra, S., Nowozin, S., Wright, S.J. (eds.) Optimization for Machine Learning, pp. 1–35. The MIT Press (2011)
52. Cao, X., Zhang, H., Guo, X., Liu, S., Chen, X.: Image retrieval and ranking via consistently reconstructing multi-attribute queries. In: European Conference on Computer Vision, pp. 569–583 (2014)
53. Zhang, C., Fu, H., Liu, S., Liu, G., Cao, X.: Low-Rank tensor constrained multiview subspace clustering. In: International Conference on Computer Vision, pp. 1582–1590 (2015)
54. Ouyang, H., He, N., Tran, L., Gray, A.: Stochastic alternating direction method of multipliers. In: International Conference on Machine Learning, pp. 80–88 (2013)
55. Zhao, S.Y., Li, W.J., Zhou, Z.H.: Scalable stochastic alternating direction method of multipliers. arXiv preprint arXiv:1502.03529 (2015)
56. Johnson, R., Zhang, T.: Accelerating stochastic gradient descent using predictive variance reduction. Adv. Neural Inf. Process. Syst. **1**(3), 1–9 (2013)
57. Platt, J.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Smola, A.J., Bartlett, P., Schölkopf, B., Schuurmans, D. (eds.) Advances in large margin classifiers, pp. 61–74. The MIT Press (1999)
58. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recongnition, reacquisition and tracking. In: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), Rio De Janeiro, Brazil, October 2007
59. Roth, P.M., Hirzer, M., Koestinger, M., Beleznai, C., Bischof, H.: Mahalanobis distance learning for person re-identification. In: Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.) Person Re-Identification, pp. 247–267 (2014)
60. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: International Conference on Computer Vision (2015)
61. Yang, Y., Jimei, Y., Junjie, Y., Liao, S.: Salient color names for person re-identification. In: European Conference on Computer Vision (2014)
62. Shen, Y., Lin, W., Yan, J., Xu, M., Wu, J., Wang, J.: Person re-identification with correspondence structure learning. In: International Conference on Computer Vision, pp. 3200–3208 (2015)
63. Zhao, R., Ouyang, W., Wang, X.: Learning mid-level filters for person re-identification. In: International Conference on Computer Vision and Pattern Recognition, pp. 144–151. IEEE, June 2014
64. Martinel, N., Das, A., Micheloni, C., Roy-Chowdhury, A.K.: Re-identification in the function space of feature warps. IEEE Trans. Pattern Anal. Mach. Intell. **37**(8), 1656–1669 (2015)
65. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: International Conference on Machine Learning, pp. 209–216. ACM Press, New York (2007)
66. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res. **10**, 207–244 (2009)
67. Guillaumin, M., Verbeek, J., Schmid, C.: Is that you? metric learning approaches for face identification. In: International Conference on Computer Vision, pp. 498–505. IEEE, September 2009