JOINT INSTANCE AND FEATURE IMPORTANCE RE-WEIGHTING FOR PERSON REIDENTIFICATION

Qin Zhou[1,2]*, Shibao Zheng*[1,2]*, Hua Yang*[1,2]*, Yu Wang*[1,2] *and Hang Su*[3]

[1]Department of Electronic Engineering, Shanghai Jiao Tong University
[2]Institution of Image Communication and Network Engineering, Shanghai Jiao Tong University
[3]Department of Computer Science and Technology, Tsinghua University
{zhou.qin.190, sbzh, hyang, txtxs}@sjtu.edu.cn, suhangss@mail.tsinghua.edu.cn

ABSTRACT

Person reidentification refers to the task of recognizing the same person under different non-overlapping camera views. Presently, person reidentification based on metric learning is proved to be effective among various techniques, which exploits the labeled data to learn a subspace that maximizes the inter-person divergence while minimizes the intra-person divergence. However, these methods fail to take the different impacts of various instances and local features into account. To address this issue, we propose to learn a projection matrix such that the importance of different instances and local features are re-weighted jointly. We also come up with a simplified formulation of the proposed algorithm, thus it can be solved by the efficient UDFS optimization algorithm. Extensive experiments on the VIPeR and iLIDS datasets demonstrate the effectiveness and efficiency of our algorithm.

***Index Terms*—** Person reidentification, instance importance re-weighting, feature importance re-weighting, optimization, metric learning

1. INTRODUCTION

Person reidentification is an important problem with many applications. Modern long-term tracking systems often need to verify whether two tracklets under different camera views belong to the same person, which is especially important in smart video surveillance systems. Besides, with more and more surveillance cameras in our city collecting large amount of surveillance videos every day, it is laborious and tedious to require human labors to recognize people across cameras, making the developing of an automatic person reidentification system vitally imperative. Although with great application prospect, person reidentification is confronted with great challenges in real world scenarios. The illumination and camera settings often bear great variations across cameras, in which case the appearance of different people can be much more alike than appearance of the same person across different views.

The existing person reidentification algorithms mainly can be categorized into two types: one tries to tackle this problem by seeking descriptive and robust representations of the human appearance. For instance, Farenzena et al. [1] model three complementary aspects of the human appearance: the overall chromatic content, the spatial arrangement of colors into stable regions, and the presence of recurrent local motifs with high entropy. They take into account the symmetry and asymmetry structure of the human appearance, which proves to be effective in modeling human appearance. Other feature designing algorithms include BiCov [2], Covariance Descriptors [3], attribute based features [4] et al. However, discrimination of this kind of methods is limited since human body is nonrigid and real scenarios are too complex to manually model. The other type of methods are based on sophisticated learners, the aim of which is either to learn a discriminative feature model [5] [6], or to learn a desired subspace/similarity-measure [7] [8]. Learning a discriminative feature model mainly involves feature selection which tries to weigh local features differently according to their performance on the training set. While learning a desired subspace/similarity-measure mainly refers to metric learning, which tries to exploit the instance label information to pull features of same person closer and push features of different people apart in the learned subspace. For detailed survey on person reidentification, please refer to [9] [10] [11].

Relation to prior work: In this paper, we try to incorporate the advantage of feature selection into metric learning. The goal of LMNN [12] metric learning algorithm is that the k-nearest neighbors always belong to the same class while examples from different classes are separated by a large margin. This cannot be directly applied for person reidentification, the aim of which is to ensure that all images of the same person having smaller distances than image pairs of different people. Thus we modify the LMNN [12] metric learning algorithm by re-weighting instances and re-selecting features with $L_{2,1}$ regularization, making it adapt to the specific person reidentification task.

We build our algorithm based on the following observations: (1) The input of metric learning algorithms are feature pairs of images under different camera views, making the computational cost grow as the square of the training image number, which is impractical for large dataset. Considering that only a small amount of image pairs of different people tend to be mistakenly recognized as more similar than true matched image pairs, we propose to put more emphasis on the more indistinguishable instance pairs; (2) Traditional metric learning algorithms often consider each local feature equally during the learning procedure. Intuitively, we assume that a certain camera captures specific view of the pedestrians, which is restricted by many factors such as the erecting height, angle and view of field of the camera. Therefore, images under the same view tend to have similar body part structure (which body part locates at which position of the image) in the image plane (see Figure 1 for an example). Consequently, different view settings may lead to feature misalignment if we directly concatenate local features extracted from sequential local patches. Thus, we introduce the $L_{2,1}$ regularization to automatically exploit the correspondence pattern of body part structure between camera pairs. **Overall, the aim of our proposed algorithm is to**

This work was supported by the NSFC(No.61221001, No.61171172 and No.61571261).

learn a projection matrix which puts more emphasis on separating difficult negative image pairs (through instance importance re-weighting) and tries to alleviate the influence of feature misalignment (through feature importance re-weighting).

2. ALGORITHM DESCRIPTION

In this section, we elaborate on how we perform joint instance and feature importance re-weighting. The instance importance re-weighting part is built on LMNN [12], thus we give a brief review on the LMNN metric learning algorithm and point out the difference between our instance importance re-weighting and LMNN. Then we elaborate on why we introduce the $L_{2,1}$ norm for feature importance re-weighting. Finally, we come up with a simplified formulation of the proposed algorithm and adopt the efficient UDFS algorithm in [13] to solve the optimization problem.

2.1. LMNN to Instance Importance Re-weighting

To better understand the LMNN algorithm, we introduce some important terms frequently used in this algorithm.

Target neighbors: Target neighbors of a specific instance refer to the k nearest neighbors with the same class label in the Euclidean space. It is a fixed prior and do not change during the learning process. We use the notation $j \rightarrow i$ to indicate that input x_j is a target neighbor of input x_i. Note that this relation is not symmetric: $j \rightarrow i$ does not imply $i \rightarrow j$.

Impostors: In mathematical terms, impostors are defined by a simple inequality. For an input x_i with label y_i and target neighbor x_j, an impostor is any input x_l with $y_i \neq y_l$ such that

$$||L'(x_i - x_l)||^2 \leq ||L'(x_i - x_j)||^2 + 1 \tag{1}$$

where $L \in R^{d \times k}$ is the transformation matrix to be learned, d is the input feature dimension and $k \ll d$ is the output feature dimension. 1 fixes the scale of L. The aim of the LMNN algorithm is that kNN classification errors in the original input space are corrected by learning an appropriate linear transformation.

To achieve the goal mentioned above, the loss function to be minimized is formulated as follows:

$$\begin{aligned}
&\varepsilon_a(L) = \sum_{j \rightarrow i} ||L'(x_i - x_j)||^2 \\
&\varepsilon_b(L) = \sum_{i,j \rightarrow i,l} (1 - y_{il})[1 + ||L'(x_i - x_j)||^2 - ||L'(x_i - x_l)||^2]_+ \\
&\varepsilon(L) = (1 - \mu)\varepsilon_a(L) + \mu\varepsilon_b(L)
\end{aligned} \tag{2}$$

where $\mu \in [0, 1]$ is the balance coefficient, y_{il} is an indicator variable, and $y_{il} = 1$ if and only if $y_i = y_l$. In Eq.(2), the $\varepsilon_a(L)$ term tries to pull target neighbors closer, while the $\varepsilon_b(L)$ term tries to push impostors faraway. The hinge loss function of the $\varepsilon_b(L)$ term in Eq.(2) actually plays the role of **instance selection**, which means only the feature pairs violating the constraints are selected to update the transformation matrix during the learning procedure. However, all impostors are equivalently considered regardless of the degree of violation. On the contrary, we believe that impostors should have different weights according to the degree of violation, such that more difficult impostors (more likely to invade the boundary of target neighbors) can be more carefully considered. As for the $\varepsilon_a(L)$ term, to adapt it to the person reidentification task, we set all the instances with the same label to be target neighbors instead of only considering the kNN nearest neighbors as target neighbors. We

Fig. 1: Illustration of the reason for feature misalignment. Images of first row are from camera A, and second row are from camera B. We can see that directly concatenating the features extracted from sequential patches in the image plane will lead to feature misalignment (e.g. features extracted from the red patches in the first row mainly correspond to the upper part of the right arm, while features extracted from the purple patches in the same positions mainly correspond to the background or backpack).

therefore formulate the loss function as follows:

$$\begin{aligned}
&\varepsilon_a(L) = \sum_{i,j,y_j=y_i} ||L'(x_i - x_j)||^2 \\
&\varepsilon_b(L) = \sum_{i,j,y_j=y_i,l} (1 - y_{il})W_{ijl}[1 + ||L'(x_i - x_j)||^2 - ||L'(x_i - x_l)||^2]_+ \\
&\varepsilon(L) = (1 - \mu)\varepsilon_a(L) + \mu\varepsilon_b(L)
\end{aligned} \tag{3}$$

where W_{ijl} indicates the degree of violation, we can simply define W_{ijl} as follows:

$$W_{ijl} = \max(||L'(x_i - x_j)||^2 - ||L'(x_i - x_l)||^2, 0) \tag{4}$$

2.2. Feature Importance Re-weighting

Feature importance re-weighting tries to address the feature misalignment problem by putting more emphasis on those local features bearing true body part correspondences and suppressing the influence of features extracted from local patches which are mostly misaligned among image pairs in the training set. Some existing algorithms try to solve the feature misalignment problem by designing body-part related feature, which is heavily dependent on the body part detector. The failure of part detection can lead to unexpected result. Since body part correspondence in the image plane is related to camera view (see Figure 1 for visualized explanation), and the feature representation used in many existing person reidentification algorithms are sequential concatenation of local features extracted from local patches [8] [7], which is likely to cause feature misalignment due to body-part in-correspondence in the image plane, **it is important to automatically find out features extracted from which positions are beneficial to the reidentification task**.

Inspired by the work of [13], we adopt the $L_{2,1}$ regularization term to perform feature importance re-weighting. Denote l^i as the i_{th} row of L, then $L_{2,1}$ norm can be formulated as:

$$|L||_{2,1} = \sum_{i=1}^{d} ||l^i||_2 \tag{5}$$

Combined with the instance importance re-weighting term introduced in Sect. 2.1, the loss function for joint instance and feature

importance re-weighting algorithm is formulated as follows:

$$\underset{L'L=I}{\varepsilon(L)} = \mu_1 \sum_{i,j,y_j=y_i} ||L'(x_i - x_j)||^2 + \mu_3||L||_{2,1} + \mu_2 \sum_{i,j,y_j=y_i,l} (1 - y_{il})W_{ijl}[1 + ||L'(x_i - x_j)||^2 - ||L'(x_i - x_l)||^2]_+ \quad (6)$$

where the orthogonal constraint is imposed to avoid arbitrary scaling and avoid the trivial solution of all zeros, μ_1, μ_2, μ_3 are the balance coefficients.

After learning, **many rows of the optimal L will shrink to zero**. Consequently, given two feature vectors $x_i, y_j \in R^d$, the difference of transformed features in the learned subspace can be formulated as: $x_i^{'} - y_j^{'} = L'(x_i - y_j)$, where the resulting difference only uses a small set of selected local features. Therefore, the influence of features extracted from local patches which are mostly misaligned among image pairs in this dataset is suppressed, while influence of features extracted from local patches which are mostly correct corresponding parts is enhanced. Besides, as stated in [13], we can rank each local feature d_i according to l^i in descending order.

2.3. Optimization

Through detailed derivation, Eq.(6) can be reformulated as follows:

$$\underset{L'L=I_k}{\varepsilon(L)} = tr(L'(\mu_1 \sum_{i,j,y_j=y_i} C_{ij})L) + \mu_3||L||_{2,1} + tr(L'(\mu_2 \sum_{i,j,y_j=y_i} \sum_{l} (1 - y_{il})W_{ijl}[\frac{1}{k}I_k + C_{ij} - C_{il}]_+)L) \quad (7)$$

where $tr(.)$ refers to the trace operation, $C_{ij} = (x_i - x_j) * (x_i - x_j)'$ and I_k is the k dimensional identity matrix. During derivation trace cyclic permutation is utilized. Set $\mu_1 \sum_{i,j,y_j=y_i} C_{ij} + \mu_2 \sum_{i,j,y_j=y_i,l} (1 - y_{il})W_{ijl}[\frac{1}{k}I_k + C_{ij} - C_{il}]_+$ to M, we finally arrive at the simplified formulation:

$$\underset{L'L=I}{\varepsilon(L)} = tr(L'ML) + \mu_3||L||_{2,1} \quad (8)$$

Eq.(8) can be efficiently solved by the UDFS algorithm proposed in [13].

3. EXPERIMENTS AND DISCUSSIONS

We evaluate the proposed joint instance and feature re-weighting algorithm on two publicly available challenging person reidentification datasets: the VIPeR [6] dataset and the iLIDS dataset [14]. In our experiments, we adopt a Single-Shot experiment setting as in [15]. All the datasets are randomly divided into two subsets so that the test set contains all the images of p individuals. This partition is performed 10 times and average performance is recorded. Under each partition, one image for each individual in the test set is randomly selected as the reference image set and the rest of the images are used as query images. This process is performed 10 times as well, and it can be seen as the recall at each rank. The performance of our algorithm is evaluated by the Cumulative Matching Characteristic (CMC) curve, which represents the expected probability of finding the correct match in the top r matches. We evaluate the necessity of instance and feature importance re-weighting,

Fig. 2: Some visualized results on the VIPeR dataset. Images in the first column are probe images and the right images are the first 10 ranked images according to the Euclidean distances in the learned subspace. Images in the red bounding boxes are true matches.

also we compare our algorithm with some state-of-the-art methods, which validates the efficacy of our algorithm. We elaborate on the experimental details as follows. Figure 2 shows some visualized results of the proposed algorithm.

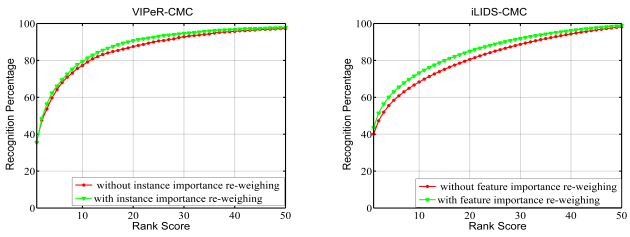
VIPeR dataset: VIPeR is the largest and most challenging person re-identification dataset consisting of 632 people with two images from two cameras for each person. It bears great variations in pose and illumination, most of the examples contain a viewpoint change of more than 90 degrees.

The aim of our algorithm is to learn the optimal transformation matrix L, thus all the existing feature designing method can be directly used to extract features. For fair comparison with the state-of-the-art methods, we use the same feature when comparing with specific method (e.g. when compared with [16], we use the LOMO feature provided by the authors, and we use the $HSV + LBP + LAB$ feature when compared with [8]). On this dataset, p is set to 316. We fix μ_1 to 1 and adjust μ_2, μ_3 by cross validation. The detailed parameter setting is as follows: $\mu_2 = 0.001, \mu_3 = 1$.

iLIDS dataset: The iLIDS dataset is another publicly available dataset captured at an airport arrival hall. It contains 479 images of 119 pedestrians, with each image subjected to great illumination changes and occlusions. For simplification, we extract the simple $HSV + LBP + LAB$ feature as in [8] on this dataset throughout the experiments and compare the results with some state-of-the-arts. We adjust the parameters in the same way mentioned above, and the detailed parameter setting is as follows: $\mu_2 = 0.005, \mu_3 = 0.5$. On this dataset, p is set to 60.

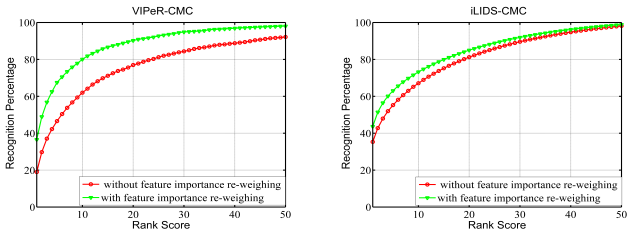
3.1. Evaluations and Analysis

Evaluation of Instance Importance Re-weighting: The instance importance re-weighting procedure of our algorithm is performed



(a) VIPeR (b) iLIDS

Fig. 3: Evaluation results of instance importance re-weighting

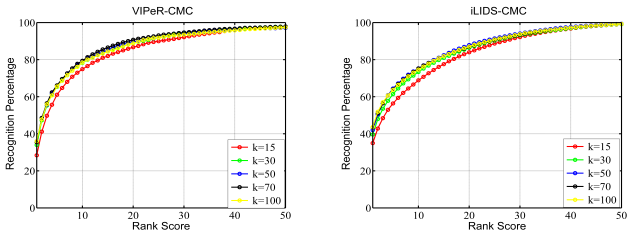


(a) VIPeR (b) iLIDS

Fig. 4: Evaluation results of feature importance re-weighting

by adding weights to the second item (whose coefficient is μ_2) in the loss function (refer to W_{ijl} in Eq.(7)). We first initialize L with the k largest principle components of the training data (k is the output feature dimension). The transformed features $L'x$ are used to calculate the weights by Eq.(4), then the weights are fixed to compute M in Eq.(8). Setting all the weights to 1 and compare the results with our algorithm, we find it necessary to weigh the differently labeled image pairs differently according to the constraint violation degree. The comparison results are shown in Figure 3. As shown, the instance importance re-weighting slightly boost the performance on both the two datasets, but this procedure avoids the calculation of the correctly classified image pairs, making our algorithm efficient.

Evaluation of Feature Importance Re-weighting: Feature importance re-weighting is mainly achieved by the $L_{2,1}$ regularization term, therefore, we compare between the performance of our algorithm with and without the $L_{2,1}$ norm. The reidentification results are demonstrated in Figure 4. As illustrated in the figures, the feature importance re-weighting procedure can significantly boost the performance on the two datasets. This also validates the assumption that images in the same dataset share some common body part part correspondence across different views. By feature importance re-weighting with the $l_{2,1}$ norm, we can enhance the influence of local features extracted from true corresponding parts in the image plane , while suppress the influence of misaligned local features caused by view change.



(a) VIPeR (b) iLIDS

Fig. 5: Performance with different output dimension

Influence of the Dimension: An extra bonus of the proposed algorithm is that we can perform feature dimension reduction by setting k to much smaller value. Figure 5 demonstrates the influence of different output feature dimension k, through which we can see that generally the reidentification performance improves with bigger k. As shown, our algorithm is robust to the feature dimension and we can achieve quite good result with very low feature dimension (the performance of $k = 15$ does not drop much compared with $k = 100$).

Evaluation of the efficiency: We also report the computation time of the proposed algorithm on a $E5 - 2.1GHZ$ computer. The average time cost for one trial on the VIPeR dataset is $0.83s$ (which includes all the training and reidentification procedure, but feature extraction is not included), and the average time cost on the iLIDS dataset is only $64.6ms$.

Method\Rank	r=1	r=5	r=10	r=20
KISSME [8]	19.8	47.8	62.2	76.2
Our	23	49.3	63.8	79.5
LOMO+Metric [16]	40	68.1	80.5	91
Our	36.6	67.5	80.1	90.2

Table 1: Comparison results on the VIPeR dataset

Method\Rank	r=1	r=5	r=10	r=20
KISSME [8]	29.4	54.9	68.8	82.1
PRDC [7]	37.8	63.7	75.1	88.3
kLFDA [15]	38	65.1	77.4	89.2
PCCA [17]	24.5	53.2	68.8	84.9
SVMML [18]	22.3	51.1	66.7	83
Our	43.5	63.9	75.2	86.8

Table 2: Comparison results on the iLIDS dataset

Comparison with State-of-the-arts: We compare our algorithm with LOMO+Metric Learning [16], KISSME [8] on the VIPeR dataset. For fair comparison, we apply the same features provided by the authors. On the iLIDS dataset, we compare with PCCA [17], SVMML [18], KISSME [8], kLFDA [15] and PRDC [7]. For simplification, we apply the simple feature introduced in KISSME [8] throughout the experiments on the iLIDS dataset. Detailed comparison results are shown in Table 1 and Table 2. As shown in Table 1, our algorithm achieves slightly better result than KISSME [8] and has competitive performance compared with LOMO+Metric [16]on the VIPeR dataset. And on the iLIDS dataset (Table 2), our algorithm performs better than all the listed state-of-the-art algorithms in the rank 1 recognition rate, which is more important in practical application and has competitive or better recognition performance in higher ranks.

4. CONCLUSION

This paper proposes a joint instance and feature importance re-weighting algorithm, which proves to be effective and efficient for person reidentification. Instance importance re-weighting pays more attention to those indistinguishable negative image pairs while neglects or pays less attention to the simple ones, which also makes our algorithm more efficient. The feature importance re-weighting tries to handle the problem of feature misalignment caused by pose or view changes. It is mainly achieved by the $L_{2,1}$ regularization term. Experimental results demonstrate that the $L_{2,1}$ norm leads to a notable performance improvement to our algorithm, which implies that feature importance re-weighting is indeed effective in handling the feature misalignment problem. Comparison results with some state-of-the-art algorithms on both the two datasets show that our algorithm has competitive or even better results than the existing algorithms.

5. REFERENCES

[1] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. CVPR*, June 2010, pp. 2360 –2367.

[2] Bingpeng Ma, Yu Su, and Frédéric Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *British Machine Vision Conference, BMVC*, 2012, pp. 1–11.

[3] Slawomir Bak and François Brémond, "Re-identification by covariance descriptors," in *Person Re-Identification*, pp. 71–91. 2014.

[4] Ryan Layne, Timothy M. Hospedales, and Shaogang Gong, "Attributes-based re-identification," in *Person Re-Identification*, pp. 93–117. 2014.

[5] Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. SCIA*, 2011, pp. 91–102.

[6] Douglas Gray and Hai Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. ECCV*, 2008, pp. 262–275.

[7] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proc. CVPR*, 2011, pp. 649–656.

[8] Martin Köstinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof, "Large scale metric learning from equivalence constraints," in *Proc. CVPR*, 2012, pp. 2288–2295.

[9] Shaogang Gong, Marco Cristani, Shuicheng Yan, and Chen Change Loy, ," in *Person Re-Identification*. 2014.

[10] Apurva Bedagkar-Gala and Shishir K. Shah, "A survey of approaches and trends in person re-identification," *Image Vision Comput.*, vol. 32, no. 4, pp. 270–286, 2014.

[11] Roberto Vezzani, Davide Baltieri, and Rita Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 29, 2013.

[12] Kilian Q. Weinberger and Lawrence K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[13] Yi Yang, Heng Tao Shen, Zhigang Ma, Zi Huang, and Xiaofang Zhou, "$l_{2,1}$-norm regularized discriminative feature selection for unsupervised learning," in *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22*, 2011, pp. 1589–1594.

[14] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Associating groups of people," in *Proc. BMVC*, 2009, pp. 1–11.

[15] Fei Xiong, Mengran Gou, Octavia I. Camps, and Mario Sznaier, "Person re-identification using kernel-based metric learning methods," in *Proc. ECCV*, 2014, pp. 1–16.

[16] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. CVPR*, 2015.

[17] Alexis Mignon and Frédéric Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. CVPR*, 2012, pp. 2666–2672.

[18] Zhen Li, Shiyu Chang, Feng Liang, Thomas S. Huang, Liangliang Cao, and John R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. CVPR*, June 2013.