Robust Partial Person Re-Identification Based on Similarity-Guided Sparse Representation

Min Ren[1,2,3], Lingxiao He[1,2,3], Haiqing Li[1,2], Yunfan Liu[1,2], Zhenan Sun[1,2], and Tieniu Tan[1,2]

[1]Center for Research on Intelligent Perception and Computing, CRIPAC
[2]Institute of Automation, Chinese Academy of Sciences, CASIA
[3]University of Chinese Academy of Sciences, UCAS, Beijing P.R.China, 100190
renmin2017@ia.ac.cn, lingxiao.he@nlpr.ia.ac.cn, hqli@nlpr.ia.ac.cn, yunfan.liu@cripac.ia.ac.cn, znsun@nlpr.ia.ac.cn, tieniu.tan@ia.ac.cn

Abstract. In this paper, we study the problem of partial person re-identification (re-id). This problem is more difficult than general person re-identification because the body in probe image is not full. We propose a novel method, similarity-guided sparse representation (SG-SR), as a robust solution to improve the discrimination of the sparse coding. There are three main components in our method. In order to include multi-scale information, a dictionary consisting of features extracted from multi-scale patches is established in the first stage. A low rank constraint is then enforced on the dictionary based on the observation that its subspaces of each class should have low dimensions. After that, a classification model is built based on a novel similarity-guided sparse representation which can choose vectors that are more similar to the probe feature vector. The results show that our method outperforms existing partial person re-identification methods significantly and achieves state-of-the-art accuracy.

Keywords: partial person re-identification, low rank constraint, similarity-guided sparse representation

1 Introduction

Person re-identification, which aims to re-identify a specific person in scenes captured by camera after he/she disappears from another disjoint camera view, is particularly an important topic in camera networks study. In recent years, impressive progress has been made in person re-identification techniques [1].

But there are still some challenges for person re-identification techniques in real-world applications, such as view angle, illumination, posture, and occlusion. And situation in which only partial body available is common. This situation can be caused by occlusions of other people, obstacles that are stationary or moving, and the boundary of the camera view. Given a partial probe image, the goal of re-identification is to find the same individual in a collection of whole body appearance in other views (gallery images). This problem has been addressed by Wei-Shi Zheng et al.[2].

Partial person re-identification is more difficult than the general person re-identification problem. First of all, since the amount of information contained by a probe image is less than that in general problem, distinctive features are more likely to be lost, thus increase the chance of mismatching. Secondly, it is hard to normalize and compare the probe images and gallery images due to the scaling issues, as we do not know the proportion of the partial body to the full appearance. The last challenge is that we are unable to know which part of the person is covered by objects and which part of the whole body the partial probe should match with.

To tackle these problems, we propose a new re-identification framework which contains three components. The first component is a multi-scale feature extractor, which extracts feature vectors without normalization of images. The second part of our method is enforcing a low rank constraint on the components of dictionary of each person, so that the features invariant with respect to disturbances could be distilled and the noise could be suppressed. The third component of our framework is a classifier based on similarity-guided sparse representation which is constrained by coefficient of similarity.

The main contributions of this work are: (1) a new partial person re-identification model is proposed in which low rank method is applied, (2) a novel similarity-guided sparse representation is proposed and it is found to be effective, (3) the model proposed achieves state-of-the-art performance.

2 Related Works

Person re-identification has become increasingly popular in computer vision community. To solve this problem, various approaches have been proposed throughout years, including methods based on transferred metric learning [3], post-rank [4], and spatial-temporal [5]. Recently, deep learning [6] [7] and video-based modelling [8] were also introduced for person re-identification. However, these algorithms are all based on the assumption that whole body is available in the probe and gallery images. Part-based models are proposed to handle the problem of incomplete body appearance in probe images, offering robust solutions to person re-identification under partial occlusions. In [9], the authors provided a part-based deep hashing model to solve the problem. Sparse representation based method and low-rank attribute embedding model for person re-identification were discussed in [11] and [10], respectively.

The problem of partial person re-identification remains unsolved as much more difficulties will be encountered than in general re-identification problems. As far as we know, partial person re-identification was first addressed in [2], and a matching framework named AMC-SWM was proposed in the work. This framework consists of two parts: a local patch-level matching model and a global part-based matching model. For super-resolution re-identification, an approach which applied low rank regularization and dictionary learning was proposed in [21]. The experiments in [21] shown that low rank regularization could improve the matching rate. Beyond person re-identification, challenge of occlusions has been studied broadly in other computer vision problems, especially in face recognition. A maximum correntropy criterion based sparse algorithm was proposed in [12]. Low-rank matrix recovery [13] and low-rank dictionary learning [14] were also used for robust face recognition. Multi-modal low-rank dictionary learning method for face recognition was proposed in [15]. An alignment free approach for partial face recognition was proposed in [16]. Robust partial face recognition using instance-to-class distance model was proposed in [17]. Sparse representation and collaborative representation were discussed in [18] and a new framework for face recognition was proposed in the same work. Low rank approach and sparse representation were applied to image classification [22]. The methods based on low rank and sparse representation inspired us to employ low rank constraints in partial person re-identification problem and propose a novel sparse representation as a classifier.

3 Methodology

In our work, we consider the partial person re-identification as matching the probe images containing only part of the body with the gallery of whole body images. We manually cropped out the regions of visible body parts from the probe images and then apply the entire algorithm. Our framework consists of three components (Fig. 1): a multi-scale feature extractor, constraint on feature dictionary and similarity-guided sparse representation.

(1) The multi-scale feature extractor computes features from patches sampled from image pyramids. After that, a feature dictionary is established based on the feature vectors extracted from all gallery images.

(2) The second component is to enforce low rank constraint on the dictionary. To be concrete, every part of the dictionary that belongs to the same person could be represented as the summation of a low rank matrix and a sparse matrix. The matrix with low dimensions contains features invariant with respect to disturbances, and the sparse matrix could be considered as noise. Hence, we are able to reduce the influence of the noise by constituting a new dictionary with the low dimension matrix of each class.

(3) A classifier based on similarity-guided sparse representation is then built upon the purified feature dictionary. We measure the similarities between the feature vector of probe image and vectors in the dictionary by a similarity-guided term. And we incorporate this similarity-guided term in the objective function in the sparse representation proposed.

3.1 Multi-Scale Feature Extraction

People could appear at various distances that are unknown to the monitoring cameras. Lacking prior knowledge on the location of the target would result in scaling issue as we do not know how to normalize the probe and gallery images. Besides, in partial person re-identification problem, we are also unaware of the proportion of the partial body to the whole body in another camera view.

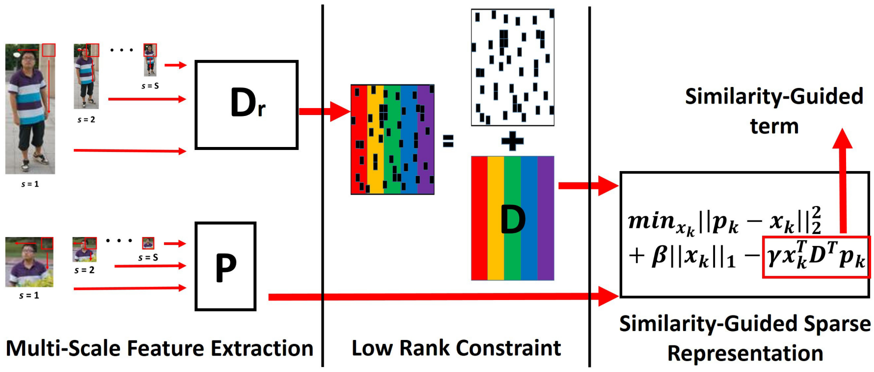

Fig. 1. Framework of our method. Multi-scale feature is extracted in the first portion, low rank constraint is enforced on the feature dictionary in the second portion, and then similarity-guided sparse representation is used to classify the probe images in the third portion

Therefore, we can not normalize the probe images and gallery images accordingly. To solver this problem, following operations are taken to every gallery image as follows.

(1) If the size of the input image is $W \times H$ and the image pyramid is built with scale S, there will be S images in each image pyramid and the size of the images are $\frac{sW}{S} \times \frac{sH}{S}$, where $s = \{1, 2, ..., S\}$.

(2) Segment the images in image pyramid into patches with the same size $L \times L$, where $L = \frac{min(W,H)}{S}$. The sample stride is selected to be L so that there is no overlap between patches.

(3) Resize all the patches to be 20×20 and then extract feature vectors from all the patches.

The feature vector of each patch consists of three components. The first component is the 20-bin histogram of 8 color channels (RGB, YCbCr, HS). The second part is the HOG descriptors (279 dimensions). LBP histogram (256 dimensions) is also included in the feature vector. Concatenating these descriptors gives us a feature vector of 695 dimensions for each patch.

Suppose there are M persons in the gallery, each person has C images and there are $K_{m,c}$ patches in one image. Grouping the feature vectors of the m-th person gives us $D_m = \{D_{m,1}, D_{m,2}, ..., D_{m,C}\}$, where $D_{m,c}$ is the collection of feature vectors of the c-th image, every column of $D_{m,c}$ correspond to a patch of the image pyramid and $D_{m,c} \in R^{695 \times K_{m,c}}$. The feature dictionary of the gallery images could be represented as $D_r = \{D_1, D_2, ..., D_M\}$, and $D_m \in R^{695 \times \sum\limits_{c} K_{m,c}}$.

Same operations are applied to the probe image as well. The collection of feature vectors extracted from the probe image is: $P = \{p_1, p_2, ..., p_K\}$, where p_k is the feature of a patch.

3.2 Low-Rank Constraint on Feature Dictionary

As mentioned in section 3.1, D_m contains information of the m-th person and could be decomposed into two separate parts: (1) the features of the m-th person that are invariant with respect to view angle, posture, lighting conditions and other variations, (2) undesired resultant noise. We assume that the first component A_m is low-rank while the second component E_m is sparse, and A_m is more discriminative than D_m for robust matching. Therefore, we could write the decomposition of D_m as follow:

$$D_m = A_m + E_m \tag{1}$$

With restrictions on rank and sparsity, this problem could be represented as:

$$\min \ rank(A_m) + \lambda ||E_m||_0 \quad s.t. \quad D_m = A_m + E_m \tag{2}$$

This is a NP-hard problem, while remaining convex, it could be relaxed as:

$$\min \ ||A_m||_* + \lambda ||E_m||_1 \quad s.t. \quad D_m = A_m + E_m \tag{3}$$

Where $||A_m||_*$ is the trace-norm of A_m. We implement the augmented Lagrange multiplier method mentioned in [14] to solve the decomposition problem and the new feature dictionary could then be written as $D = \{A_1, A_2, ..., A_M\}$.

3.3 Similarity-Guided Sparse Representation

We could then transform the matching problem into a sparse representation and classification problem:

$$\min_{x_k} \ ||x_k||_1 \quad s.t. \quad p_k = Dx_k \tag{4}$$

where p_k is the feature of a patch of the probe image, and x_k is the sparse coding of p_k. In order to eliminate the constraint, we relax the optimization problem as follows:

$$\min_{x_k} \ ||p_k - Dx_k||_2^2 + \beta ||x_k||_1 \tag{5}$$

When rebuilding the probe feature vector, we would like to choose vectors which are more similar to the probe feature vector to improve the discrimination of the sparse coding.

The angular similarity between the probe feature vector and vectors in the dictionary could be evaluated using inner product of D^T and p_k. Hence, elements of $D^T p_k$ measure the angular similarity between p_k and corresponding column of D in vector space, in other words, the similarity between patch of probe image and corresponding patch in gallery. And $x_k^T D^T p_k$ is inner product of x_k and $D^T p_k$, it measures the oriented similarity between x_k and $D^T p_k$ in vector space. Hence, the larger $x_k^T D^T p_k$ is, the vectors of D chosen to rebuild p_k is more similar to the probe feature vector. Accordingly, the above equation could be extended to:

$$\min_{x_k} ||p_k - Dx_k||_2^2 + \beta||x_k||_1 - \gamma {x_k}^T D^T p_k \tag{6}$$

where γ is the coefficient of similarity. Unfolding this equation gives us:

$$\min_{x_k} \frac{1}{2} x_k^T D^T D x_k - (1 + \frac{1}{2}\gamma) {x_k}^T D^T p_k + \beta||x_k||_1 \tag{7}$$

According to the discussion in [19], this optimization problem could be solved by feature-sign search method. Following this algorithm we could obtain the sparse coding matrix: $X = \{x_1, x_2, ..., x_K\}$. The residual error could then be computed as r_m:

$$r_m = ||P - D_m X_m||_F^2 \tag{8}$$

where X_m is the submatrix of X that contains only the columns corresponding to the m-th person. The classifying problem could finally be represented as:

$$m = \arg\min_m \frac{r_m}{K} \tag{9}$$

in which K is the number of the columns of X_m.

4 Experiments

4.1 Datasets and Benchmark Methods

The Partial REID Dataset [2] is the only publicly available partial person re-id dataset. It contains 600 images of 60 people, with 5 full-body images and 5 partial images per person. These images are collected at university campus with different viewpoints as well as background settings, and the partial images contain different types of severe occlusions (Fig. 2).

The QMUL underGround Re-IDentification (GRID) dataset [20] used frequently in re-identification study contains 250 persons and each person is shown by a image pair. Each pair has two images of the same individual seen from different camera views. All images are captured from 8 disjoint camera views installed in a busy underground station. Fig. 2 shows a snapshot of camera views in the station and sample images in the GRID dataset. This dataset is challenging due to variations of postures, colours, and illumination conditions; as well as poor image quality caused by low spatial resolution. We randomly crop half of the images of each person to simulate severe occlusions (Fig. 2).

For benchmarking, five existing methods were considered, including AMC-SWM proposed [2], collaborative representation based classification method (CRC-RLS) [18], an alignment-free approach named MTSR proposed [16], discriminative low-rank dictionary learning (DLR) method [14], and method using instance-to-class distance (I2C-Distance) proposed in [17]. We chose some of the partial face recognition methods because they were also designed to solve partial target recognition.

Fig. 2. Examples of partial person images (first row) and the corresponding full images (second row). Columns $1-5$ are from Partial REID Dataset, and columns $6-10$ are from GRID dataset.

4.2 Experimental Settings and Evaluations

Both single-shot and multi-shot experiments were conducted. For efficiency of experiments, we randomly selected 100 people from GRID dataset for experiments and all 60 people in Partial REID Dataset were used. In single-shot experiments, only one ($N = 1$) image of each person is included in the gallery set. Multi-shot re-identification indicates that more than one ($N = 2$, $N = 3$) images of each person are used in the gallery. The multi-shot experiments are conducted only on Partial REID Dataset.

There are three parameters in our model, the multi-scale number S, coefficient of low rank (λ in Eq. 3), and coefficient of similarity (γ in Eq. 6). The multi-scale number S was set as 5 in all the experiments. In single-shot experiments, coefficient of low rank was set as 0.4, coefficient of similarity was set as 1.2. In multi-shot experiments, coefficient of low rank was set as 0.7, coefficient of similarity was set as 1.7. Discussions of these parameters can be seen in Section 4.4.

We employed cumulative match characteristic (CMC) to measure the matching performance in closed-set setting. To evaluate the performance of our method in open-set setting, we randomly removed 30% (the same percentage with [2]) people and their corresponding images from the gallery and provide the area-under-curve (AUC) values of ROC curves for evaluation.

4.3 Results

Single-shot experiments In single-shot re-identification experiments, we compared SLR-SRM to other existing methods including AMC-SWM, CRC-RLS, MTSR, DLR, and I2C-Distance. The results shown in Table 1 demonstrate clearly that the performance of our method is better than other five methods on both of the two datasets (achieved 55% and 32% matching rate at rank-1). AMC-SWM is the method with minimal gap comparing to our approach (achieved 45% and 20% matching rate at rank-1).

Table 1. Single-shot experiments

	Partial REID N=1				GRID N=1			
method	rank-1	rank-5	rank-10	AUC	rank-1	rank-5	rank-10	AUC
SG-SR	**55.0%**	**78.3%**	**86.7%**	**0.925**	**32.0%**	**45.0%**	**54.0%**	**0.860**
AMC-SWM	45.0%	60.0%	70.0%	0.806	20.0%	41.0%	52.0%	0.782
DLR	23.3%	58.3%	66.7%	0.783	14.0%	34.0%	49.0%	0.779
I2C-Distance	15.0%	46.7%	63.3%	0.645	15.0%	31.0%	47.0%	0.734
MTSR	15.0%	30.0%	40.0%	0.668	12.0%	37.0%	50.0%	0.765
CRC-RLS	1.67%	15.0%	28.3%	0.548	2.00%	13.0%	23.0%	0.617

Multi-shot experiments In multi-shot re-identification experiments ($N = 2$, $N = 3$), as shown in Table 2, the matching rate at rank-1 increases to 61.7% ($N = 2$), 65.0% ($N = 3$). Again, our approach outperforms all other methods.

Table 2. Multi-shot experiments

	Partial REID N=2				Partial REID N=3			
method	rank-1	rank-5	rank-10	AUC	rank-1	rank-5	rank-10	AUC
SG-SR	**61.7%**	**81.7%**	**86.7%**	**0.934**	**65.0%**	**81.7%**	**88.3%**	**0.956**
AMC-SWM	53.3%	70.0%	76.7%	0.884	55.0%	71.7%	78.3%	0.915
DLR	31.7%	66.7%	75.0%	0.848	31.7%	65.0%	73.3%	0.901
I2C-Distance	26.7%	50.0%	71.7%	0.706	25.0%	51.7%	71.7%	0.733
MTSR	25.0%	46.7%	61.7%	0.808	28.3%	65.0%	75.0%	0.902
CRC-RLS	5.00%	15.0%	38.3%	0.616	20.0%	46.7%	58.3%	0.737

4.4 Discussions

The matching rate at rank-1 on Partial REID Dataset under different multi-scale number S was shown in Fig. 3 (left). The results shown that the matching rate increased as S increased. But the time of recognition was also increased as S increased. So we need to choose an appropriate value of S.

The matching rate under different value of coefficient of low rank was shown in Fig. 3 (middle). The matching rate without low rank constraint was expressed by the imaginary line. The results shown that appropriate value of coefficient of low rank could increase the matching rate by 3.3%.

The matching rate under different value of coefficient of similarity was shown in Fig. 3 (right). The results shown that the matching rate could be increased 8.3% compared with $\gamma = 0$. It demonstrate that similarity-guided sparse representation proposed could improve the discrimination of the sparse coding.

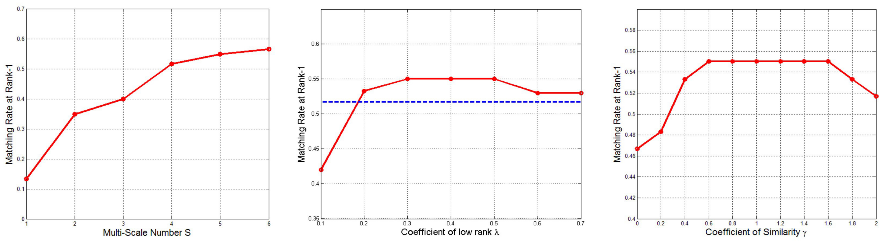

Fig. 3. Evaluation of different multi-scale number S (left), coefficient of low rank λ (middle) and coefficient of similarity γ (right) using matching rate at rank-1 on Partial REID (N=1).

5 Conclusions

In this work, we have proposed a novel framework for partial person re-identification problem. This framework consisted of multi-scale feature extraction, low rank constraint and similarity-guided sparse representation. Our experiments have shown that our approach is more effective than existing methods to solve partial person re-identification. The reasons can be summarized as follow: (1) patches sampled from the image pyramid contain information of different scales, (2) the low rank constraint makes the feature vectors of dictionary more discriminative, (3) the similarity-guided sparse representation makes the vectors of dictionary which are more similar to the probe are more likely to be chosen, thus improves the matching accuracy.

Acknowledgements

This work is supported by the Beijing Municipal Science and Technology Commission (No.Z161100000216144) and the National Natural Science Foundation of China (Grant No.61427811, 61573360).

References

1. Zheng, Y.Yang, A.G.Hauptmann: Person Re-identification: Past, Present and Future. CoRR abs/1610.02984 (2016)
2. W.S.Zheng, X.Li, T.Xiang, S.Liao, J.Lai, S.Gong: Partial Person Re-Identification. in IEEE International Conference on Computer Vision, pp. 4678-4686 (2015).
3. W. Li, R. Zhao, X. Wang: Human Reidentification with Transferred Metric Learning. in Asian Conference on Computer Vision, pp. 31-44 (2012)
4. C. Liu, C. C. Loy, S. Gong, and G. Wang: POP: Person Re-identification Post-rank Optimisation. in IEEE International Conference on Computer Vision, pp. 441-448 (2013)
5. G. Lian, J. Lai, W. S. Zheng: Spatial-temporal consistent labeling of tracked pedestrians across non-overlapping camera views. Pattern Recognition, vol. 44, pp. 1121-1136 (2011)

6. L. Wu, C. Shen, and A. V. D. Hengel: PersonNet: Person Re-identification with Deep Convolutional Neural Networks. CoRR abs/1601.07255 (2016)
7. Li. Dangwei, Chen. Xiaotang, Zhang. Zhang, Huang. Kaiqi: Learning Deep Context-aware Feature over Body and Latent Parts for Person Re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017)
8. J. You, A. Wu, X. Li, W. S. Zheng: Top-Push Video-Based Person Re-identification. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1345-1353 (2016)
9. F. Zhu, X. Kong, L. Zheng, H. Fu, Q. Tian: Part-based Deep Hashing for Large-scale Person Re-identification. IEEE Transactions on Image Processing, vol. PP, pp. 1-1 (2017)
10. L. An, X. Chen, S. Yang, B. Bhanu: Sparse representation matching for person re-identification. Information Sciences, vol. 355-356, pp. 74-89 (2016)
11. C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, W. Gao: Multi-Task Learning with Low Rank Attribute Embedding for Person Re-Identification. In IEEE International Conference on Computer Vision, pp. 3739-3747 (2016)
12. R. He, W. S. Zheng, B. G. Hu: Maximum Correntropy Criterion for Robust Face Recognition. IEEE Transactions on Pattern Analysis, Machine Intelligence, vol. 33, pp. 1561-76 (2011)
13. C. F. Chen, C. P. Wei, Y. C. F. Wang: Low-rank matrix recovery with structural incoherence for robust face recognition. In Computer Vision and Pattern Recognition, pp. 2618-2625 (2012)
14. H. Nguyen, W. Yang, B. Sheng, and C. Sun: Discriminative low-rank dictionary learning for face recognition. Neurocomputing, vol. 173, pp. 541-551 (2016)
15. H. Foroughi, M. Shakeri, N. Ray, H. Zhang: Face Recognition using Multi-Modal Low-Rank Dictionary Learning. Computing Research Repository (2017)
16. S. Liao, A. K. Jain, S. Z. Li: Partial face recognition: alignment-free approach. IEEE Transactions on Pattern Analysis, Machine Intelligence, vol. 35, p. 1193 (2013)
17. J. Hu, J. Lu, Y. P. Tan: Robust partial face recognition using instance-to-class distance. In Visual Communications and Image Processing, pp. 1-6 (2014)
18. L. Zhang, M. Yang, X. Feng: Sparse representation or collaborative representation: Which helps face recognition?. In International Conference on Computer Vision, pp. 471-478 (2011)
19. H. Lee, A. Battle, R. Raina, A. Y. Ng: Efficient sparse coding algorithms. In International Conference on Neural Information Processing Systems, pp. 801-808 (2006)
20. C. Liu, S. Gong, C. C. Loy, and X. Lin: Person Re-Identification: What Features are Important. International Workshop on Re-Identification, European Conference on Computer Vision, Firenze, Italy (2012)
21. Jing, Xiao-Yuan, Xiaoke Zhu, Fei Wu, Ruimin Hu, Xinge You, Yunhong Wang, Hui Feng and Jing-Yu Yang: Super-Resolution Person Re-Identification With Semi-Coupled Low-Rank Discriminant Dictionary Learning. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 695-704 (2015)
22. Zhu, Xiaoke, Xiao-Yuan Jing, Fei Wu, Di Wu, Li Cheng, Sen Li and Ruimin Hu: Multi-Kernel Low-Rank Dictionary Pair Learning for Multiple Features Based Image Classification. AAAI (2017).