Person Re-Identification Postprint Based on Post-fusion of Discriminative Features

Authors: Liu Qi, Hou Li

Date: 2018-05-24T00:00:00+00:00

Abstract

Cross-camera persons exhibit significant appearance variations due to differences in illumination, viewpoint, and pose, posing severe challenges to person re-identification research. Based on multi-feature fusion and distance metric learning techniques, we propose a late fusion algorithm and apply it to person re-identification. First, we extract Local Maximal Occurrence (LOMO) features and Salient Color Names based Color Descriptor (SCNCD) features from cross-camera person sample images respectively to represent the appearance of persons across cameras. Then, based on the extracted LOMO and SCNCD features, we train Cross-view Quadratic Discriminant Analysis (XQDA) distance metric learning models respectively to obtain feature-specific optimized distances for each pair of cross-camera persons. Finally, we apply a min-max normalized distance fusion algorithm to obtain the final distance between persons across cameras for matching. Experiments are conducted on two challenging public datasets, VIPeR and PRID450S, and the results demonstrate that the proposed person re-identification algorithm effectively improves the accuracy of person re-identification.

Full Text

Discriminative Feature Based Late Fusion for Person Re-Identification

Liu Qi[1,2], Hou Li[1,2]
(1. School of Information Engineering, Huangshan University, Huangshan 245041, China; 2. School of Communication & Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: Pedestrians may vary greatly in appearance due to differences in illumination, viewpoint, and pose across cameras, which poses serious challenges for person re-identification. This paper proposes a discriminative feature based

late fusion algorithm using multiple-feature fusion and distance metric learning techniques for person re-identification. First, we extract Local Maximal Occurrence (LOMO) features and Salient Color Names-based Color Descriptor (SCNCD) features from cross-camera pedestrian sample images to represent pedestrian appearance. Then, based on the extracted LOMO and SCNCD features, we train separate Cross-view Quadratic Discriminant Analysis (XQDA) distance metric learning models to obtain optimized distances for each feature for each pair of pedestrians across cameras. Finally, we apply a min-max normalization based distance fusion algorithm to obtain the final distance between cross-camera pedestrians for matching. Experimental results on two challenging public datasets, VIPeR and PRID450S, demonstrate that the proposed algorithm effectively improves the accuracy of person re-identification.

Key words: person re-identification; multiple-feature fusion; distance metric learning; distance fusion; min-max normalization

0 Introduction

Person re-identification aims to enable computers to recognize whether pedestrians captured by different cameras share the same identity, matching pedestrians across cameras based on their appearance. This technology provides critical support for large-scale video applications such as automated pedestrian tracking in camera networks and urban automated video surveillance. However, pedestrians across different cameras exhibit significant appearance variations due to differences in illumination, viewpoint, and pose, as illustrated in Figure 1 [Figure 1: see original paper] (the upper and lower rows show pedestrian images from two different cameras in the VIPeR benchmark dataset, with each column representing the same pedestrian). These variations pose severe challenges for person re-identification research.

Current research on person re-identification primarily focuses on two aspects [1–3]: first, extracting discriminative visual features to represent cross-camera pedestrian appearance; second, finding discriminative distance metric learning algorithms to optimize feature distances between cross-camera pedestrians. The vast majority of person re-identification methods employ multi-feature fusion to improve accuracy and robustness [4–7]. Farenzena et al. [4] proposed the Symmetry-Driven Accumulation of Local Features (SDALF) algorithm to describe cross-camera pedestrian appearance. SDALF encodes three complementary visual features: overall color content represented by HSV color histograms, stable regions in color spatial arrangement represented by maximally stable color regions, and local patterns with high entropy represented by periodically structured blocks. It handles viewpoint changes through symmetric and asymmetric properties. Kviatkovsky et al. [5] proposed using color invariants (ColorInv) for person re-identification. ColorInv describes pedestrian appearance using color histograms, covariance descriptors, and part-based shape context descriptors in Log color space. The part-based shape context descriptor, as an invariant shape feature descriptor, uses different parts of the human body to describe the

intrinsic structural distribution of discriminative color distributions. Yang et al. [6] proposed the Salient Color Names-based Color Descriptor (SCNCD) for person re-identification, which fuses SCNCD and color histograms computed in four different color spaces (original RGB, normalized RGB, l1l2l3, and HSV) to jointly describe cross-camera pedestrian appearance. Liao et al. [7] proposed the Local Maximal Occurrence (LOMO) feature representation for person re-identification. LOMO consists of HSV color histograms and Scale Invariant Local Ternary Pattern (SILTP) texture descriptors. It analyzes the probability of color and texture features appearing at the same horizontal strip position in human body parts and maximizes this probability to obtain robust feature representations, effectively handling viewpoint changes in cross-camera pedestrians.

Based on these feature representation methods, using standard distance metrics such as Euclidean distance for cross-camera pedestrian matching severely affects re-identification accuracy. To mitigate appearance feature discrepancies in cross-camera pedestrian matching and improve accuracy, distance metric learning methods [7–10] have been widely applied to person re-identification in recent years. Inspired by likelihood ratio test statistics, Koestinger et al. [8] proposed KISSME (Keep It Simple and Straightforward MEtric) for person re-identification. This method is simple and efficient, requiring only the computation of two small covariance matrices from dissimilar and similar pairs, making it easily scalable to large datasets. Pedagadi et al. [9] combined unsupervised Principal Component Analysis (PCA) dimensionality reduction with supervised Local Fisher Discriminant Analysis (LFDA), proposing a low-dimensional manifold distance metric learning framework for person re-identification. LFDA preserves local neighborhood structure while maximizing between-class separation, enabling multi-modal distribution of sample data, and estimates the LFDA transformation through generalized eigenvalue decomposition. However, when applied to relatively small datasets, this framework may produce undesirable compression of the most discriminative features, affecting re-identification accuracy. To address this issue, Xiong et al. [10] further employed Kernel LFDA (KLFDA) for person re-identification by fully utilizing kernel tricks and the advantages of LFDA. KLFDA is a closed-form nonlinear learning method that uses kernel tricks to handle high-dimensional feature vectors while maximizing the Fisher optimization criterion. This method preserves the most discriminative features during dimensionality reduction and improves re-identification accuracy through flexible kernel selection. However, the computational speed is very slow when using nonlinear kernels.

Liao et al. [7] proposed a discriminative metric learning method called Cross-view Quadratic Discriminant Analysis (XQDA) for person re-identification. XQDA aims to learn a discriminative low-dimensional subspace using cross-camera training data while simultaneously learning an optimized distance function in this subspace to measure feature similarity between cross-camera pedestrians.

To further address the significant appearance variations of cross-camera pedestrians and improve re-identification accuracy, this paper proposes a discriminative feature based late fusion algorithm on the basis of the aforementioned multi-feature fusion and distance metric learning techniques. Specifically, each discriminative feature first undergoes separate metric learning, after which their respective feature distances are fused for person re-identification.

The proposed person re-identification algorithm is illustrated in Figure 2 [Figure 2: see original paper]. First, we extract two features—Local Maximal Occurrence (LOMO) and Salient Color Names-based Color Descriptor (SCNCD)—from pedestrian images captured by different cameras. Then, we train separate Cross-view Quadratic Discriminant Analysis (XQDA) distance metric learning models using these two features to obtain optimized distance metrics for each feature between cross-camera pedestrians. Finally, to balance the contribution of each optimized distance metric, we apply min-max normalization to standardize each feature' s optimized distance metric to the range [0, 1] before fusing them to obtain the final distance metric between cross-camera pedestrians for matching and re-identification.

1 Feature Extraction

To effectively handle significant appearance variations in cross-camera pedestrians, this paper selects visual features robust to illumination and viewpoint changes—LOMO and SCNCD—to jointly represent cross-camera pedestrian appearance.

The LOMO feature extraction process is illustrated in Figure 3 [Figure 3: see original paper] [7]. A 10×10 sliding sub-window is used to represent local regions of a pedestrian image. In each sub-window, we extract an 8×8×8-bin joint HSV color histogram and SILTP texture histograms at two scales. We then maximize the local occurrence probability of each pattern (each histogram bin) across all sub-windows at the same horizontal position, effectively handling viewpoint changes in pedestrian appearance. Additionally, we employ a three-scale pyramid representation with 2×2 local pooling to downsample the original pedestrian image. The above feature extraction steps are repeated, and finally, all computed local maximal feature vectors at each horizontal position are concatenated to form the final LOMO feature for a pedestrian image.

Considering that color features are more discriminative for cross-camera illumination changes, we further extract SCNCD features from the same pedestrian image to enhance the handling of illumination variations. The SCNCD feature extraction process is illustrated in Figure 4 [Figure 4: see original paper] [6]. There are M color name indices to describe color information, where the numbers on the color names represent probability distributions of being close to a certain color (here referring to red) in the color name set. Only the few closest color names (salient color names) have non-zero probability values. The SCNCD feature of a pedestrian image is characterized by the probability distribution of

these salient color names.

2 XQDA Distance Metric Learning

To mitigate significant appearance variations in cross-camera pedestrians, this paper selects the fast and effective XQDA distance metric learning method to obtain optimized distance metrics.

XQDA learns a low-dimensional feature subspace using cross-camera pedestrian sample data while simultaneously learning an optimized distance function in this subspace to measure similarity between cross-camera pedestrians [7]. The distance function between a pair of cross-camera pedestrians in the low-dimensional feature subspace is defined as:

$$d_{\mathbf{W}}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^{\top} \mathbf{W} (\mathbf{x} - \mathbf{y})$$

where $\mathbf{W}$ represents the learned distance metric matrix.

Based on LOMO and SCNCD features, we can learn separate distance metric matrices $\mathbf{W}_m$ for each feature using cross-camera pedestrian training samples (assuming each camera has N training samples per pedestrian). Let $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{Nm}$ and $\mathcal{Y} = \{\mathbf{y}_i\}_{i=1}^{Nm}$ denote the training sample sets from two camera views. Using the distance function defined in Equation (1), we can obtain optimized distance metrics $d_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y})$ for each feature.

3 Min-Max Normalization Based Distance Fusion

Considering that the optimized distance metrics obtained from LOMO and SCNCD features have different numerical ranges, to balance the contribution of each feature' s optimized distance metric, we first apply min-max normalization to standardize each feature' s optimized distance metric to the range [0, 1] before performing distance summation as the final distance metric between cross-camera pedestrians.

Assume $d_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y})$ represents the optimized distance metric based on LOMO and SCNCD features. The normalized distance metric $d^*_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y})$ after min-max normalization is shown in Equation (3):

$$d^*_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y}) = \frac{d_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y}) - \min_{\mathbf{y}' \in \mathcal{Y}} d_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y}')}{\max_{\mathbf{y}' \in \mathcal{Y}} d_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y}') - \min_{\mathbf{y}' \in \mathcal{Y}} d_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y}')}$$

Note that min-max normalization is performed along the dimension of query samples in the test set. For example, when there is one test sample and 100 query samples, the min-max normalization algorithm is executed using 100 distance values. Here, $d_{\mathbf{W}_m}(\mathbf{x}, \mathcal{Y})$ represents a distance vector consisting of distances between one test sample and all query samples; $\min d_{\mathbf{W}_m}(\mathbf{x}, \mathcal{Y})$ and

$\max d_{\mathbf{W}_m}(\mathbf{x}, \mathcal{Y})$ denote the minimum and maximum elements in this vector, respectively. Through this linear transformation, values are mapped to the range [0, 1].

The final distance metric $d(\mathbf{x}, \mathbf{y})$ between cross-camera pedestrians is computed by summing the normalized distance metrics for each feature, as shown in Equation (4):

$$d(\mathbf{x}, \mathbf{y}) = \sum_{m \in \{\text{LOMO,SCNCD}\}} d^*_{\mathbf{W}_m}(\mathbf{x}, \mathbf{y})$$

4 Experimental Results

Experiments were conducted on a Lenovo computer with an Intel i7 processor running Windows 7 64-bit operating system, using MATLAB 2017a software. We evaluated the proposed person re-identification algorithm on two challenging public datasets: VIPeR [11] and PRID450S [12], estimating the Cumulative Matching Characteristics (CMC). Table 1 provides a brief introduction to these two datasets. We randomly selected half of the pedestrians as the training set and the other half as the test set. The training set was used to learn different feature kernel matrices $\mathbf{M}_m$, while the test set was used to measure similarity between cross-camera pedestrian samples. To ensure stable and reliable results, each experiment was repeated 10 times and the average recognition rate was computed when calculating the CMC curve.

Table 1 Brief introduction to VIPeR and PRID450S datasets

Dataset	Training Images	Camera Views	Images per Person per View
VIPeR	-	2	1
PRID450S	-	2	1

4.1 Feature Sensitivity Analysis

Using the same XQDA distance metric learning method, we analyzed the contribution sensitivity of LOMO and SCNCD features. We conducted experiments on VIPeR and PRID450S datasets using LOMO alone, SCNCD alone, and our proposed method (Figure 2). Figure 5 [Figure 5: see original paper] shows the CMC performance comparison results.

As shown in Figure 5, our proposed method achieves the best performance, with rank-1 recognition rates of 48.4% on VIPeR and 73.7% on PRID450S. LOMO features are more discriminative than SCNCD features, achieving rank-1 recognition rates of 40.8% on VIPeR and 65.0% on PRID450S, while SCNCD features achieve 27.3% on VIPeR and 38.4% on PRID450S. Clearly, the proposed discriminative feature based late fusion method can effectively improve person re-identification accuracy.

4.2 Comparison with State-of-the-Art Algorithms

We compared our method with recently proposed state-of-the-art person re-identification algorithms on both VIPeR and PRID450S datasets in terms of CMC performance. Tables 2 and 3 present the comparative results.

Table 2 Recognition rates of recently published person re-identification algorithms on VIPeR dataset (only cumulative matching scores at ranks 1, 5, 10, 20 are listed) /%

Rank	Our approach	MR [13]	LSSCDL [14]	DLSR [15]	MED_VR [16]	DRML [17]	LOMO [7]	ECM [18]	SCNCD [19]	CMWCE [20]	CSL [21]	KPLS [22]
1												
5												
10												
20												

Table 3 Recognition rates of recently published person re-identification algorithms on PRID450S dataset (only cumulative matching scores at ranks 1, 5, 10, 20 are listed) /%

Rank	Our approach	LSSCDL [14]	DRML [17]	MLMR [13]	KPLS [22]	MED_VR [16]	DLSR [15]	CSL [21]	ECM [18]	SCNCD [19]	CMWCE [20]
1											
5											
10											
20											

As shown in Tables 2 and 3, our proposed method (Figure 2) significantly outperforms other compared state-of-the-art person re-identification algorithms on both VIPeR and PRID450S datasets, demonstrating that discriminative feature late fusion contributes to improved re-identification performance.

5 Conclusion

Based on LOMO and SCNCD features and XQDA distance metric learning, this paper proposes a discriminative feature based late fusion method for person re-identification. Experimental results on the challenging VIPeR and PRID450S benchmark datasets demonstrate the effectiveness and superiority of the proposed algorithm.

References

[1] Wang Xiaogang. Intelligent multi-camera video surveillance: a review [J]. Pattern Recognition Letters, 2013, 34 (1): 3-19.

[2] Yu Tianshu, Wang Ruisheng. Enhancing scene parsing by transferring structures via efficient low-rank graph matching [C]// Proc of ACM SIGSPATIAL International Conference on Geographic Information Systems. 2014: 247-267.

[3] Zhang Xin, Ding Meng, Fan Guoliang. Video-based human walking estimation using joint gait and pose manifolds [J]. IEEE Trans on Circuits and Systems for Video Technology, 2017, 27 (7): 1540-1554.

[4] Farenzena M, Bazzani L, Perina A, et al. Person re-identification by symmetry-driven accumulation of local features [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010: 2360-2367.

[5] Kviatkovsky I, Adam A, Rivlin E. Color invariants for person re-identification [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35 (7): 1622-1634.

[6] Yang Yang, Yang Jimei, Yan Junjie, et al. Salient color names for person re-identification [C]// Proc of European Conference on Computer Vision. 2014: 536-551.

[7] Liao Shengcai, Hu Yang, Zhu Xiangyu, et al. Person re-identification by local maximal occurrence representation and metric learning [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 2197-2206.

[8] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2012: 2288-2295.

[9] Pedagadi S, Orwell J, Velastin S, et al. Local fisher discriminant analysis for pedestrian re-identification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3318-3325.

[10] Xiong Fei, Gou Mengran, Camps O, et al. Person re-identification using kernel-based metric learning methods [C]// Proc of European Conference on Computer Vision. 2014: 1-16.

[11] Gray D, Brennan S, Tao Hai. Evaluating appearance models for recognition, reacquisition, and tracking [C]// Proc of IEEE International Workshop on Performance Evaluation for Tracking and Surveillance. 2007: 1-7.

[12] Roth P M, Hirzer M, Köstinger M, et al. Mahalanobis distance learning for person re-identification [M]// Person Re-Identification. London: Springer, 2014: 247-267.

[13] Chen Yingcong, Zheng Weishi, Lai Jianhuang. Mirror representation for modeling view-specific transform in person re-identification [C]// Proc of International Joint Conference on Artificial Intelligence. 2015: 3402-3408.

[14] Zhang Ying, Li Baohua, Lu Huchuan, et al. Sample-specific SVM learning for person re-identification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1278-1287.

[15] Shi Zhiyuan, Hospedales T M, Xiang Tao. Transferring a semantic representation for person re-identification and search [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4184-4193.

[16] Yang Yang, Lei Zhen, Zhang Shifeng, et al. Metric embedded discriminative vocabulary learning for high-level person representation [C]// Proc of AAAI Conference on Artificial Intelligence. 2016: 3648-3654.

[17] Yao Wenbin, Weng Zhenyu, Zhu Yuesheng. Diversity regularized metric learning for person re-identification [C]// Proc of IEEE International Conference on Image Processing. 2016: 4264-4268.

[18] Liu Xiaokai; Wang Hongyu; Wu Yi, et al. An ensemble color model for human re-identification [C]// Proc of IEEE Winter Conference on Applications of Computer Vision. 2015: 868-875.

[19] Yang Yang, Yang Jimei, Yan Junjie, et al. Salient color names for person re-identification [C]// Proc of European Conference on Computer Vision. 2014: 536-551.

[20] Yang Yang, Liao Shengcai, Lei Zhen, et al. Color models and weighted covariance estimation for person re-identification [C]// Proc of International Conference on Pattern Recognition. 2014: 1874-1879.

[21] Shen Yang, Lin Weiyao, Yan Junchi, et al. Person re-identification with correspondence structure learning [C]// Proc of IEEE International Conference on Computer Vision. 2015: 3200-3208.

[22] Prates R, Oliveira M, Schwartz W R. Kernel partial least squares for person re-identification [C]// Proc of IEEE International Conference on Advanced Video and Signal Based Surveillance. 2016: 249-255.

Note: Figure translations are in progress. See original paper for figures.