Person Re-identification via Learning Visual Similarity on Corresponding Patch Pairs

Hao Sheng[1], Yan Huang[1,*], Yanwei Zheng[1], Jiahui Chen[1], and Zhang Xiong[1]

[1]State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, 100191 Beijing, P.R.China
{shenghao,yanhuang,zhengyw,chenjiahui1991,xiongz}@buaa.edu.cn

Abstract. Since humans concentrate more on differences between relatively small but salient body regions in matching person across disjoint camera views, we propose these differences to be the most significant character in person re-identification (Re-ID). Unlike existing methods focusing on learning discriminative features to adapt viewpoint variation using global visual similarity, we propose a learning visual similarity algorithm via corresponding patch pairs (CPPs) for person Re-ID. The novel CPPs method is introduced to represent the corresponding body patches of the same person in different images with good robustness to body pose, viewpoint and illumination variations. The similarity between two people is measured by an improved bi-directional weight mechanism with a TF-IDF like patches weight. At last, a complementary similarity measure and a mutually-exclusive regulation are presented to enhance the performance of Re-ID. With quantitative evaluation on public datasets, the best rank-1 matching rate on the VIPeR dataset is improved by 4.14%.

Keywords: person re-identification,corresponding patch pairs, patch weight, visual similarity, cross-camera

1 Introduction

Person re-identification (Re-ID), from a series of candidate images captured in distributed locations or different times, is essential to several computer vision problems such as image retrieval, person identification, re-acquisition and cross-camera person tracking [1, 2],[2],[3],[4],[5]. To associate the images of the same individual and discard irrelevant ones, person Re-ID mostly relies on pairwise data to evaluate the similarity. Designing robust person descriptor is one of the key technologies in person Re-ID, including the entire image descriptor [2],[6],[7],[8] and the one without background [4],[9],[10],[11],[12]. However, despite years of research, state-of-the-art algorithms are no match for human's abilities in person Re-ID for the existence of the variations of viewpoint, illumination, pose, and other factors.

To eliminate the effects caused by the variations, many discriminative features are proposed for designing the person descriptors on the entire image. Gray

Fig. 1. The CPPs sampled from the same person in different camera views.

et al. [2] introduce the ensemble of localized features via boosting algorithm. To absorb the illumination variation in different camera views, a descriptor that combines biologically inspired feature and covariance descriptor is given by Ma et al. [6],[7]. Prosser et al. [13] formulate person Re-ID as a ranking problem, and learn the feature weight based on RankSVM; Zheng et al. [8] propose a relative distance comparison model to maximize the pairwise likelihood in person Re-ID. Both of them use the entire image features, a mixture of color (RGB, YCbCr, HSV) and texture (Gabor [14] and Schmid [15]) histogram features, designed in [2].

To minimize the effect of background clutters, many effective body segmentation methods [10],[11],[12] have been put forward for person Re-ID. Considering that background region is far away from the body center, Farenzena et al. [4] downgrade the weights of patches which are more distant from the symmetry axis of body. Bak et al. [12] present the body appearance based on haar-like features and dominant color descriptors. A gaussian descriptor includes the mixing characteristics of color, texture and spatial structure of body is demonstrated in [16]. However, confined by the strongly inhomogeneous background and low-resolution image, existing body segmentation methods cannot accurately capture the body contour. Good body segmentation method can efficiently improve the person Re-ID matching rate, but it is not the emphasis in our research. In this work, we try to solve person Re-ID problem based on body region. Instead of developing a better body segmentation method, a complementary similarity measure is proposed to enhance the performance of person Re-ID.

In this paper, we discuss a novel corresponding patch pairs (CPPs) method to learn visual similarity for person Re-ID. On the present state of studies, patch-based person Re-ID methods have obtained good performance in several works. Cai et al [17] collect signatures of patches along body contour. These patches are matched with corresponding patches that aid by dominant color representation and geometric constraints. Since salient regions can effectively distinguish the appearance of different persons and are general enough to identify person across different camera views. Zhao et al [18] divide images into small patches and extract color and SIFT feature in each patch for salience matching. In order to automatically discover effective patches to learn mid-level filter for person Re-ID, patches are qualitatively measured and classified with their discriminative

power in [19]. Different from previous works, CPPs is proposed to learn visual similarity which can effectively represent all the patch pairs variations in visual for the same person in different views (shown in Fig. 1).

To sample CPPs which are robust to illumination, background and pose variations, three processes are proposed including: visual pattern feature clustering, Hungarian adjacency constrained search and CPPs candidate quantification after image preprocessing. An improved bi-directional weight mechanism that respectively exploits a TF-IDF like method and the learned distance metric on CPPs to calculate the patch weight and patch pair similarity, is used to measure the similarity of individuals. Furthermore, a complementary similarity measure that can reduce the effect of background clutters caused by inaccurate body segmentation, is used to enhance our approach. At last, a mutually-exclusive regulation is proposed to achieve better rank-1 performance. We demonstrate the validity of our approach on the VIPeR [20] and CAVIAR4REID [5] datasets. The results show that our approach achieve very competitive results by quantitative evaluation.

2 Image Preprocessing

Considering the original person images easily change with the variations of background and illumination. In the first stage, the pedestrian parsing method is used to discard the background and the illumination equalization method is used to reduce the influence caused by the illumination.

As shown in Fig. 2, after normalizing the V channel in HSV color space for illumination equalization, the Deep Decompositional Network (DDN) [10] is used to remove background and parse person body into five parts, including hair, head, body, arms and legs. Then, the foreground mask and boundary between upper and lower body can be obtained by the parsing result. Due to the lack of differentiation, the head and foot parts are ignored. DDN is able to accurately estimate complex pose variation with good robustness to occlusions and background clutters, details can be found in [10]. In addition, erosion is used to erase the burrs in foreground region to get smooth body contour. At last, the upper and lower bodies are segmented respectively into grid of local patches. We set the size of patch is 8x8 and the grid step is 4. A 672-dimensional dColorSIFT feature which is robust to the variations of viewpoint and illumination is extracted from each patch. Many researches such as [18],[19] and [21] show great effect by using the dColorSIFT feature as patch descriptor.

3 Sampling Corresponding Patch Pairs

3.1 Visual Pattern Feature Clustering

Patches with different visual information are mixed together. Therefore, clustering is performed to group patches into subsets with coherent visual information.

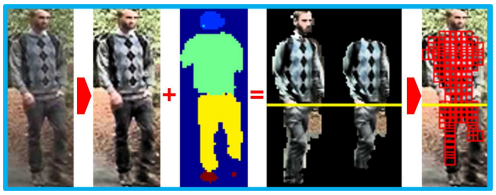


Fig. 2. Image preprocessing: The first column shows the original image. The second column shows the illumination equalization result. The third column shows the parsing result by DDN. The body contour and boundary given by the column 2 and 3 are shown in the fourth column. The last column shows the grid of patch partition.

CPPs are sampled from the same cluster. Patches clustering, between probe image and gallery image, is the first step of sampling CPPs.

In this task, the feature dimension of patch is high, and the distributions of data clusters are often in different densities and sizes. Therefore, graph degree linkage (GDL) algorithm [22] is employed for patch clustering since it can well cope with these problems. However, due to the difficulty to determine the appropriate cluster granularity, some patch clusters have mixed visual patterns [19]. It cannot accurately sample CPPs from these clusters. Thus two kinds of patch clustering trees (PCTs) are constructed for upper and lower bodies respectively. Given a probe image A and a a gallery image B, upper body patch set $P_u^{A,B} = P_u^A \cup P_u^B$ and lower body patch set $P_l^{A,B} = P_l^A \cup P_l^B$. Two PCTs are built with order O and depth D, e.g. each parent node in upper PCT has O_u children and there are D_u layers of nodes. Considering a certain degree of differences among upper and lower body, we set $O_u = 2, D_u = 3$ and $O_l = 3, D_l = 2$ in our experiment. The root nodes of upper and lower PCT contain all the patches in $P_u^{A,B}$ and $P_l^{A,B}$, respectively. Specially, the second-layer node of upper PCT, in which the number of patches are less than τ_{min}. We set τ_{min} equals a quarter of the patches in $P_u^{A,B}$. The clustering results is denoted as c_i where i is the index of clusters. As shown in Figure 3, c_i is corresponding to the leaf nodes of PCT.

3.2 Hungarian Adjacency Constrained Search

Based on the clustering results, Hungarian adjacency constrained search is proposed to select CPPs candidates in relaxed adjacent vertical space. Specially, our experiments only use the upper body patches because of the good discriminative power they have. To go against multi-to-one patch matching, Hungarian algorithm is introduced to improve adjacency constrained search strategy [18]. The CPPs candidates between the probe image A and the corresponding gallery image $A^{'}$ should belong to the same cluster c_i, which is defined as $S^{A,A^{'}}$:

$$S^{A,A^{'}} = \{s_x^{A,A^{'}} | x = 1...N_{row}\}, s_x^{A,A^{'}} = \boldsymbol{H}(\boldsymbol{M}(P_x^{A,c_i}, \varepsilon(P_x^{A,c_i}))), \quad (1)$$

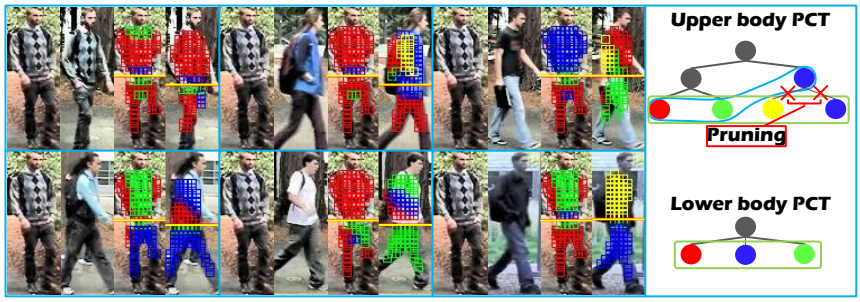


Fig. 3. Different colors of leaf nodes represent different clusters. Specially, upper body patches are clustered to 3 or 4 clusters under our pruning condition.

where x is the row number and $P_x^{A,c_i} \in c_i$ denotes a patch set at the x-th row of image A, $\varepsilon(P_x^{A,c_i})$ is the corresponding patch set of P_x^{A,c_i} in image $A^{'}$, which is formulated as:

$$\varepsilon(P_x^{A,c_i}) = \{p_{k,y}^{A^{'},c_i} | p_{k,y}^{A^{'},c_i} \in c_i, y = 1...N_{col}, k = max(0, x-\delta)...min(N_{row}, x+\delta)\}, \tag{2}$$

$p_{k,y}^{A^{'},c_i}$ represents the patch at the k-th row and the y-th column in image $A^{'}$, δ is the height of vertical search space ($\delta = 2$ in our experiment, the setup is the same as in [18]). In Equ. 1, $\boldsymbol{H}(\cdot)$ represents the Hungarian assignment algorithm and $\boldsymbol{M}(\cdot)$ is cost matrix among patches between P_x^{A,c_i} and $\varepsilon(P_x^{A,c_i})$ sets. The Hungarian assignment algorithm is used for finding CPPs candidate with minimum total cost. Supposing there are M patches in P_x^{A,c_i}, and N patches in $\varepsilon(P_x^{A,c_i})$. $\boldsymbol{M}$ is formulated as:

$$\boldsymbol{M} = [d_{m,n}(F_m^A, F_n^{A^{'}})]_{M \times N}, m = 1...M, n = 1...N, \tag{3}$$

where F is the feature of patch, d is Radial Basis function.

3.3 Quantitative Description on CPPs Candidates

Some incorrect CPPs which lead to erroneous estimations of visual similarity are sampled in CPPs condidate. Therefore, visual similarity score (VSS) is presented to quantify the visual similarity of each pair in CPPs candidates. The pairs, whose VSSs are higher than the mean value of all CPPs candidates' VSSs are selected as CPPs for visual similarity learning. Intuitively, a large number of similar patches in corresponding images (e.g. A and $A^{'}$) with small difference in visual pattern are most likely to be selected as CPPs. In another aspect, patches in non-corresponding images are similar with each other should be penalized.

Thus, we define the VSS as follows:

$$VSS(p_j^{A,c_i}, p_j^{A',c_i}) = \frac{d(F_j^{A,c_i}, F_j^{A',c_i}) \cdot (\frac{N_{c_i}^A}{N^A} + \frac{N_{c_i}^{A'}}{N^{A'}})}{\sum_{id=1}^{N_{per}} (\frac{N_{c_{i'}}^A}{N^A} + \frac{N_{c_{i'}}^{G_{id}}}{N^{G_{id}}}) \cdot \zeta\{N_{c_{i'}}^{G_{id}} \neq 0\} - (\frac{N_{c_i}^A}{N^A} + \frac{N_{c_i}^{A'}}{N^{A'}})}, \quad (4)$$

where j is the pair index of CPPs candidates, d is Radial Basis function, F_j^{A,c_i} and F_j^{A',c_i} are features of patches in p_j^{A,c_i} and p_j^{A',c_i}. N_α^β is the number of patches, where α and β represent the cluster and person image respectively. N^β denotes the number of patches in β (contains all clusters). N_{per} is the number of person in gallery set. If $N_{c_{i'}}^{G_{id}} \neq 0$ indicator function $\zeta\{\cdot\}$ equals 1 and 0 otherwise. Specially, when the patch p_j^{A,c_i} is clustered with patches in image G_{id}, c_i becomes $c_{i'}$ which represents a common cluster between A and G_{id}.

4 Matching Based on CPPs

4.1 Patch Weight Calculation

Patches with distinct features are discriminative in finding the same person in different views [18]. The frequency of patch clustering with different persons can effectively reflect the discriminative power of the patch. Patches with high weight have higher discriminative power than others. Based on that reason, a TF-IDF like function is proposed to calculate patch weight on upper and lower body, which is defined as:

$$\omega(p_{x,y}^{A,c_{i'}}) = exp(\frac{T(p_{x,y}^{A,c_{i'}})/I(p_{x,y}^{A,c_{i'}})}{2\sigma^2}), \quad (5)$$

where $p_{x,y}^{A,c_{i'}} \in c_{i'}$ is the patch at the x-th row and the y-th column in image A, and σ is bandwidth. $T(p_{x,y}^{A,c_{i'}})$, which is measured by the frequency of visual similar patches clustering in image A itself, is defined as:

$$T(p_{x,y}^{A,c_{i'}}) = \sum_{id=1}^{N_{per}} \frac{N_{c_{i'}}^A}{N^A} \cdot \zeta\{N_{c_{i'}}^{G_{id}} = 0\}. \quad (6)$$

If $N_{c_{i'}}^{G_{id}} = 0$ indicator function $\zeta\{\cdot\}$ equals 1 and 0 otherwise. $I(p_{x,y}^{A,c_{i'}})$, which measures whether the patch is common or rare according to the patch clustering frequency between persons in Equ. 5, is defined as:

$$I(p_{x,y}^{A,c_{i'}}) = ln(N_{per}/(1 + N_{per} - N_{per'})), \quad (7)$$

$N_{per'}$ is the number of persons who do not have any patches to cluster with $p_{x,y}^{A,c_{i'}}$.

4.2 Integrated Matching

In our experiment, Locally-Adaptive Decision Function (LADF) [23] is used to learn the similarity of CPPs. With the bi-directional weighting mechanism [18], a person matching model based on LADF training result and patch weight is defined as:

$$d_{CPPs}(A,B) = \sum_{k=1}^{N} \frac{\omega(p_k^A) \cdot s(p_k^A, p_k^B) \cdot \omega(p_k^B)}{1 + \left|\omega(p_k^A) - \omega(p_k^B)\right|}, \tag{8}$$

where $s(p_k^A, p_k^B)$ represents the distance between patch p_k^A and p_k^B which is measured with LADF training result, with k being the index of patch pair. For sampling patch pairs of unlabeled persons, visual pattern features clustering step is ignored for keeping enough patch pairs in similarity measure. Hungarian adjacency constrained search method is used directly in finding the patch pairs between A and B (testing set) without considering clusters. Supposing N is the number of patch pairs between A and B. For fair comparison, we use the first N_{pair} distances. N_{pair} is defined as:

$$N_{pair} = min(N_i), i = 1...N_{per}, \tag{9}$$

where N_i is the number of patch pairs between A and the i-th person in gallery set.

5 Enhancing Re-ID Performance

5.1 Complementary Similarity Measure

Integrating several types of features with complementary nature are presented in most Re-ID methods [4],[2],[24],[25],[26]. Therefore, a complementary similarity measure is designed to make up for the shortcomings (In general, body segmentation method cannot accurately find the body contour.) of the entire image (EI) and body (HB) matching. The distance $d_{eCPPs}(*)$ between two images can be computed as:

$$d_{eCPPs}(A,B) = d_{CPPs}(A,B) + d_{EI}(F_{EI}^A, F_{EI}^B) + d_{HB}(F_{HB}^A, F_{HB}^B), \tag{10}$$

where $d(*)$ is distance measure between person A and B. d_{CPPs} is in Equ. 10. d_{EI} and F_{EI} correspond to the distance measure and feature in [23] and [27], respectively. d_{HB} and F_{HB} correspond to the distance measure and feature in [16].

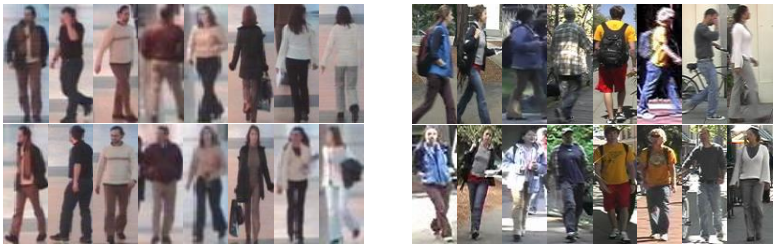
5.2 Mutually-Exclusive Regulation

A mutually-exclusive regulation (MER) is proposed to find the best-matching person in Single versus Single (S vs. S), only one exemplar per individual is available both in probe and gallery sets. Two steps of MER are presented as follows. First, if d_{CPPs}, d_{EI} and d_{HB} between person A (in probe set) and B

(in gallery set) are all minimum respectively, it means B is the most similar and the best matching with A. So B is removed from matching candidates and keep matching with A only. The result is verified by the three independence distance. Second, if one (denoted as G_B) in gallery set is the best matching with multiple individuals (denoted as P) in probe set measured by d_{eCPPs}, then the best-matching person in P have a relative lower similarity probability compared with others in gallery set. After that, a rank swaps between G_B and the second most similar candidates of the rest in P achieve better rank-1 performance.

6 Results and Analysis

Dataset. We demonstrate the evaluation results on two public datasets: VIPeR [20] and CAVIAR4REID [5]. The CAVIAR4REID dataset contains 1220 images of 72 persons with different viewpoints, poses and resolutions, each image captured from cameras inside a shopping mall, as shown in Fig. 4(a). The CAVIAR4REID dataset is measured in Single versus Multiple(S vs. M), multiple exemplars per individual are available in the probe set and only the first exemplar per individual in the gallery set. For a fair comparison, we use the same training and testing protocol mentioned in [23] that randomly divide the 72 persons into two sets, 36 persons for training and the rest for testing on CAVIAR4REID dataset, . The VIPeR is another challenging benchmark dataset for person Re-ID. This dataset contains 1264 images of 632 persons, each person has two images in different views with complex background and illumination variations, as shown in Fig. 4(b). The same experiment setup mentioned in [4] is used on VIPeR dataset. Ten trials of evaluation are conducted on two datasets and the average result is reported. In addition, the result is shown using the cumulative matching characteristic (CMC) curve [20], where the horizontal coordinate exploits top ranking k and the vertical coordinate donates the correct matching rates.

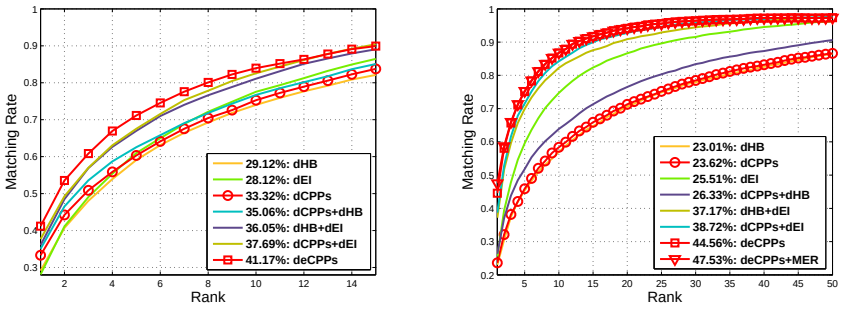


(a) CAVIAR4REID dataset (b) VIPeR dataset

Fig. 4. A set of 16 randomly taken images pairs from the CAVIAR4REID and VIPeR dataset. The images in the same column belong to the same person.

6.1 Framework Verification

Tab. 1 illustrates several intermediate results and final results on two datasets. Because there are more than two images of each person, MER is ignored on CAVIAR4REID dataset. From the Re-ID results, the separate results of d_{EI}, d_{HB} and d_{CPPs} can generate the low matching rates. From the analysis of d_{CPPs} on two datasets, the good performance of CPPs is shown on CAVIAR4REID dataset, which is easy to achieve relatively perfect body segmentation since the background changes small with fewer clutter. Nonetheless, it is not relatively efficient on VIPeR dataset because the background changes considerably. The results of $d_{EI} + d_{HB}$, $d_{HB} + d_{CPPs}$ and $d_{EI} + d_{CPPs}$ manifest that the fusion of different features can make up for the shortcomings of any one. Compared with above results of the features fusion, $d_{EI} + d_{CPPs}$ achieves the best result since the result of d_{EI} is designed with the entire image features and d_{CPPs} pays more attention to the features of human body. Since both of d_{HB} and d_{CPPs} are based on the features of human body, the improvement of $d_{HB} + d_{CPPs}$ is not significant. Compared with results of $d_{HB} + d_{EI}$, the matching rate of d_{eCPPs} increases from 36.05% to 41.17% on CAVIAR4REID dataset. And it also has similar results on VIPeR dataset. In addition, the validity of mutually-exclusive regulation is shown by the increase of the matching rate from 44.56% to 47.53% on VIPeR dataset. The CMCs are shown in Fig. 5.



(a) CAVIAR4REID (b) VIPeR

Fig. 5. Comparing with intermediate results. CMC on the CAVIAR4REID dataset (a) and the VIPeR dataset (b).

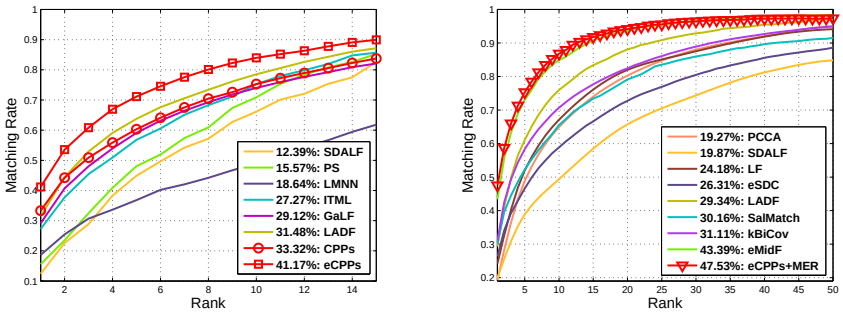
6.2 Quantitative Evaluation on Public Datasets

Fig. 6(a) shows that our approach achieves the best rank-1 accuracy 41.17% and outperforms other methods on CAVIAR4REID dataset in S vs. M scenario, including SDALF [4], PS [5], LMNN [28], ITML [29], GaLF [16] and LADF [23]. Comparison with PCCA [30], SDALF [4], LF [31], eSDC [18], LADF [23],

Method	CAVIAR4REID(%)(G=36)				VIPeR(%)(P=316,G=316)			
	r=1	r=3	r=5	r=7	r=1	r=5	r=10	r=15
d_{HB}	29.12	48.04	59.10	66.45	23.00	45.39	57.00	64.73
d_{EI}	28.12	49.06	60.97	68.87	25.50	59.57	74.73	82.34
d_{CPPs}	**33.31**	**50.88**	**60.32**	**67.54**	23.62	45.90	58.37	65.88
$d_{CPPs} + d_{HB}$	35.05	53.59	62.56	69.05	26.32	51.80	63.87	71.31
$d_{HB} + d_{EI}$	36.05	56.90	66.99	74.02	37.16	69.92	82.27	87.57
$d_{CPPs} + d_{EI}$	37.69	56.99	67.53	75.31	38.72	71.52	84.24	89.95
d_{eCPPs}	**41.17**	**60.84**	**71.14**	**77.57**	**44.56**	**75.11**	**86.75**	**91.72**
d_{eCPPs}+MER					**47.53**	**75.32**	**86.84**	**91.77**

Table 1. Performance comparisons on CAVIAR4REID and VIPeR dataset

SalMatch [21], kBiCov [7] and eMidF [19] on VIPeR dataset is shown in Fig. 6(b). The result shows our approach improves the best state-of-the-art rank-1 matching rate on the VIPeR dataset by 4.14%.



(a) (b)

Fig. 6. Comparing with state-of-the-art methods. CMC on the CAVIAR4REID dataset (a) and the VIPeR dataset (b).

7 Conclusion

We have presented a learning visual similarity algorithm via CPPs for person Re-ID. Contrary to global visual similarity measure methods, a novel CPPs description method represent the corresponding body patch pairs of the same person with good robustness to body pose, viewpoint and illumination variations. After calculating the score of salience and matched patches with a TF-IDF like patches weight and a learning CPPs method, the similarity between two people is evaluated by an improved bi-directional weight mechanism. To further enhance Re-ID performance, a hierarchical descriptor makes our approach insensitive to background changes and a mutually-exclusive regulation improves the

rank-1 performance. We demonstrate the validity of our approach on two public datasets, achieving very competitive results in terms of quantitative evaluation.

Acknowledgement. This study was partially supported by the National Natural Science Foundation of China (No. 61472019), the National High Technology Research and Development Program of China (No. 2013AA01A603) and the National Aerospace Science Foundation of China (No. 2013ZC51). Supported by the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under grant #SKLSDE-2015ZX-21.

References

1. Weiming Hu, Min Hu, Xue Zhou, Tieniu Tan, Jianguang Lou, and Stephen J. Maybank, "Principal axis-based correspondence between multiple cameras for people tracking," *PAMI*, vol. 28, pp. 663–671, 2006.
2. Douglas Gray and Hai Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *ECCV*. Springer, 2008, vol. I, pp. 262–275.
3. William Robson Schwartz and Larry S. Davis, "Learning discriminative appearance-based models using partial least squares," in *SIBGRAPI*. IEEE, 2009, vol. XXII, pp. 322–329.
4. Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *CVPR*. IEEE, 2010, pp. 2360–2367.
5. Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino, "Custom pictorial structures for re-identification," in *BMVC*. BMVA, 2011, pp. 1–11.
6. Bingpeng Ma, Yu Su, and Frederic Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *BMVC*. BMVA, 2012, pp. 1–11.
7. Bingpeng Ma, Yu Su, and Frederic Jurie, "Covariance descriptor based on bio-inspired features for person re-identification and face verification," *Image Vision Computing*, vol. 32, pp. 379–390, April 2014.
8. Weishi Zheng, Shaogang Gong, and Tao Xiang, "Reidentification by relative distance comparison," *PAMI*, vol. 35, pp. 653–668, February 2013.
9. Bak S, Suresh S, Brmond F, and Thonnat M, "Fusion of motion segmentation with online adaptive neural classifier for robust tracking," in *VISAPP*. INSTICC, 2009, vol. II, pp. 410–416.
10. Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Pedestrian parsing via deep decompositional network," in *ICCV*. IEEE, 2013, pp. 2648–2655.
11. Nebojsa Jojic, Alessandro Perina, Marco Cristani, Vittorio Murino, and Brendan J. Frey, "Stel component analysis: Modeling spatial correlations in image class structure," in *CVPR*. IEEE, 2009, pp. 2044–2051.
12. Bak S, Corvee E, Brmond F, and Thonnat M, "Person re-identification using haar-based and dcd-based signature," in *AVSS*. IEEE, 2010, pp. 1–8.
13. Bryan Prosser, Weishi Zheng, Shaogang Gong, and Tao Xiang, "Person re-identification by support vector ranking," in *BMVC*. BMVA, 2010, vol. I, pp. 1–11.

14. Itzhak Fogel and Dov Sagi, "Gabor filters as texture discriminator," *Biological cybernetics*, vol. 61, pp. 103–113, 1989.
15. Cordelia Schmid, "Constructing models for content-based image retrieval," in *CVPR*, 2001, pp. 39–45.
16. Ma Bingpeng, Qian Li, , and Hong Chang, "Gaussian descriptor based on local features for person re-identification," in *ACCV*. Springer, 2014.
17. Yinghao Cai, Kaiqi Huang, and Tieniu Tan, "Human appearance matching across multiple non-overlapping cameras," in *ICPR*. IEEE, 2008, pp. 1–4.
18. Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Unsupervised salience learning for person re-identification," in *CVPR*. IEEE, 2013, pp. 3586–3593.
19. Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Learning mid-level filters for person re-identification," in *CVPR*. IEEE, 2014, pp. 144–151.
20. Douglas Gray, Shane Brennan, and Hai Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *IEEE International workshop on performance evaluation of tracking and surveillance*. Citeseer, 2007.
21. Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Person re-identification by salience matching," in *ICCV*. IEEE, 2013, pp. 2528–2535.
22. Wei Zhang, Xiaogang Wang, Deli Zhao, and Xiaoou Tang, "Graph degree linkage: Agglomerative clustering on a directed graph," in *ECCV*. Springer, 2012, vol. I, pp. 428–441.
23. Zhen Li, Shiyu Chang, Feng Liang, Thomas S. Huang, Liangliang Cao, and John R. Smith, "Learning locally-adaptive decision functions for person verification," in *CVPR*. IEEE, 2013, pp. 3610–3617.
24. Loris Bazzani, Marco Cristani, Alessandro Perina, and Vittorio Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 898–903, 2012.
25. Xiaogang Wang, Gianfranco Doretto, Thomas Sebastian, Jens Rittscher, and Peter Tu, "Shape and appearance context modeling," in *ICCV*. IEEE, 2007, pp. 1–8.
26. Chen Change Loy, Chunxiao Liu, and Shaogang Gong, "Person re-identification by manifold ranking," in *ICIP*. Citeseer, 2013, vol. 1, p. 5.
27. Xi Zhou, Na Cui, Zhen Li, Feng Liang, and Thomas S. Huang, "Hierarchical gaussianization for image classification," in *ICCV*. IEEE, 2009, pp. 1971–1977.
28. Kilian Q. Weinberger, John Blitzer, and Lawrence K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *NIPS*, 2005, pp. 1473–1480.
29. Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon, "Information-theoretic metric learning," in *ICML*. ACM, 2007, pp. 209–216.
30. Alexis Mignon and Frdric Jurie, "Pcca: A new approach for distance learning from sparse pairwise constraints," in *CVPR*. IEEE, 2012, pp. 2666–2672.
31. Sateesh Pedagadi, James Orwell, Sergio A. Velastin, and Boghos A. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *CVPR*. IEEE, 2013, pp. 3318–3325.