Saliency-Based Person Re-identification by Probability Histogram

Zongyan Zhang[1], Cairong Zhao[1], Duoqian Miao[1(✉)], Xuekuan Wang[1], Zhihui Lai[2], and Jian Yang[3]

[1] Department of Computer Science and Technology, Tongji University, Shanghai, China
dqmiao@tongji.edu.cn
[2] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
[3] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

Abstract. Person re-identification has attracted increasing interests due to its broad application in automatic search and video surveillance. It is easy for humans to recognize person identities but difficult for computers. Thus, knowing how humans to recognize person identities is helpful to improve the performance of the computers person re-identification. In this paper, we propose an effective feature representation based on salient regions and a pool of multiple metric learning for person re-identification. The proposed feature representation extracts local details (salience regions) and global distribution of pedestrian images. To reduce the effects of illumination changes, we apply probability histogram in four kinds of color spaces where similar color can be characterized by the similar histogram distribution to different color spaces. Moreover, a pool of multiple metric learning is applied to all features captured from different spaces and models. The proposed method has been evaluated on two public datasets. Experimental results show that the proposed method outperforms others.

1 Introduction

Person re-identification is a task to find a person from a series of non-overlapping camera views in which a person has the same identity. It is a challenging task because the person seems quite different under the change of viewpoints, poses, appearance and illumination. Besides, non-overlapping camera views make background information useless, thus the used pedestrian images contain lots of noises.

To find the correct person, an effective feature representation which should be robust to environment, poses and viewpoints changes is necessary. Color and texture are the most pervasive and based features used in person re-identification (e.g. [1–5]). These descriptors have been successfully applied to solve the problem

Z. Zhang and C. Zhao—Authors contributed equally.

C.-S. Chen et al. (Eds.): ACCV 2016 Workshops, Part III, LNCS 10118, pp. 315–329, 2017.
DOI: 10.1007/978-3-319-54526-4_24

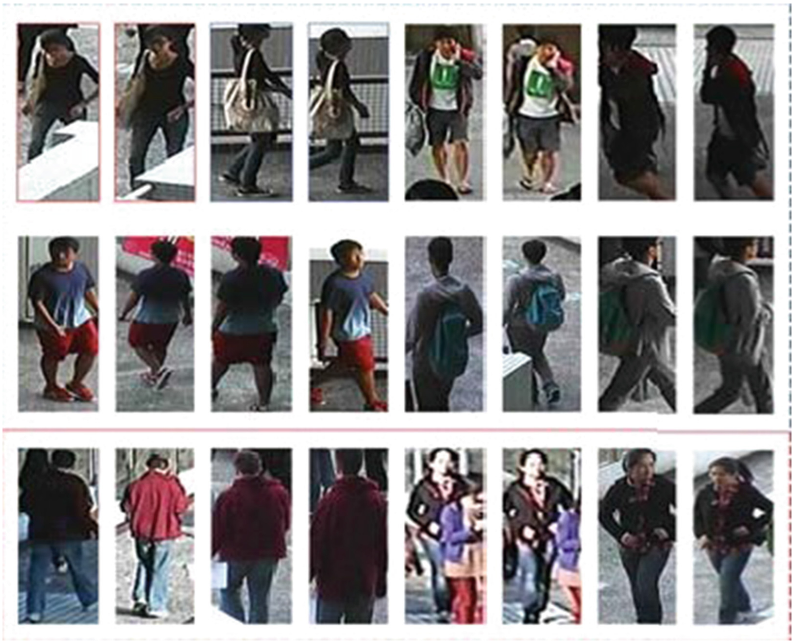

Fig. 1. Samples of pedestrian images in different camera views of person re-identification.

of person re-identification. In addition, misalignments caused by variations of viewpoints and poses are commonly exist in person re-identification as Fig. 1. The remarkable spatial difference should be considered in feature representation.

Several effective approaches greatly improve the robustness of feature representation and advance the person re-identification, such as LOMO [5], SDALF [6], salience match [7], SCNCD [8]. These feature describe pedestrian image from local details or the whole image which neglect the complementarity between local details with global geometric correspondence. Moreover, because of the variations of pedestrian images, an effective feature representation is still a challenging problem in person re-identification.

The other aspect of person re-identification is learning a distance or similarity function to divide persons into similar pairs and dissimilar pairs. Many metric learning algorithms have been proposed to address this problem, for example KISSME [1], PCCA [9], LFDA [10], MFA [11]. However, existing approaches have not considered that different feature extracted by various model may not be effectively handled by the single metric.

In this paper, we propose a novel person re-identification model which fuses a hybrid feature representation with a multiple metric learning. The hybrid feature representation describes a pedestrian image from the local detail and global geometric distribution, which fusion probability histogram feature extracted from each overlapping sliding windows as global description and some small salient regions as local detail. The probability histogram based on different color spaces is robust to variations of illumination. Salient regions have great invariance against viewpoints and poses changes. To learning a discriminant metric, we propose a novel model where different features from different space are handled by different metric respectively.

The rest of the paper is organized as follow. In Sect. 2, a brief review of related works in the person re-identification field is given. The proposed person re-identification approach is described in Sect. 3. The experiment and comparison between our method with exist ones are showed in Sect. 4. Finally, conclusion are drawn in Sect. 5.

2 Related Work

In the past few years, many researchers have proposed different approaches to tackle the challenge of person re-identification. These approaches mainly follow two different aspects: feature representation and metric learning.

Feature Representation. To stably describe pedestrians appearance in disjoint camera views, various approaches have been proposed. Farenzena et al. [6] proposed the Symmetry-Driven Accumulation of Local Features (SDALF) method which exploits the symmetry and dissymmetry in pedestrian image to select the major part of the body figure. The application of the symmetry and dissymmetry is considered to handle viewpoint variations. Zheng et al. [12] proposed to segment pedestrian image into six horizontal stripes equally and compare pedestrian image in each stripe. This feature is robust to the viewpoint changes in horizontal direction. Yang et al. [8] proposed a salient color names based color descriptor (SCNCD) which designs a novel method to describe colors. SCNCD utilizes the probability distribution to represent the distance of color with color name. A higher probability means the color name is nearer to the color. SCNCD is robust to the variations of illumination because similar colors have similar probability distributions. Liao et al. [5] proposed an efficient Local Maximal Occurrence (LOMO) approach which considers the spatial details within a horizontal stripe. To handle viewpoint changes, LOMO maximizes the local occurrence of the same histogram bin among sub-windows which at the same horizontal location. These feature representations describe pedestrian images from global distribution, consider little about local details and none of them combines the local details with global geometric correspondence to represent pedestrian images.

Moreover, Zhao et al. [7] proposed the salience match where salience is used to handle the change of poses and viewpoints. They considered the saliency information in pedestrian images and the salience patches match each other by adjacency constrained search. These methods use salience region to describe pedestrian images, but the time complexity is very high.

Metric Learning. On the other hand, discriminative learning models have been extensively studies for person re-identification. Prosser et al. [13] proposed a RankSVM method which reformulates the person re-identification problem as ranking problem. RankSVM learns a subspace where potential true match is given higher ranking. Pedagadi et al. [10] apply local fisher discriminant analysis (LFDA) to the person re-identification. Koestinger et al. [1] utilize KISSME to learn a distance metric from equivalence constraints. KISSME learns a Mahalanobis distance to minimize the intra-class distance and maximize the inter-class distance to compute the match in different feature space. Liao et al. [5] extended KISSME approach and proposed a cross-view quadratic discriminant analysis (XQDA). XQDA further learns a discriminant subspace and a metric together.

3 Saliency-Based Person Re-identification by Probability Histogram

A person from a series of non-overlapping camera views usually has quite different appearance. For example, some parts of a pedestrian on different camera view can be misaligned due to the diverse poses and viewpoints. To handle such problems, some small salient regions, that can represent the certain pedestrian as specific as possible, would be detected as local features from pedestrian images, and these features describe the major color information of body or legs as well as the characteristic of clothing. Moreover inspired by Spatial Pyramid Matching [14], we utilize a multi-scale sliding windows to describe global geometric correspondence of a person image. As show in Fig. 2, the proposed person re-identification approach consists of five phases: (1) feature extraction, (2) background suppression, (3) multiple metric learning, (4) salient region detection, and (5) distance fusion.

3.1 Feature Extraction

Before feature extraction, each pedestrian image has been segmented into a grid of multi-scale local patches. In this paper, we use a patch size of $10 * 10/15 * 15$, with overlapping step of 5 pixels. Within each patch, we extract color probability histogram in four color space and a texture feature: LBP.

Since there is no single color space and descriptor that could be against the variations of illumination, we project the original RGB images into each color space $S \in \{\mathrm{HSV}, \mathrm{normalizedrgb}, \mathrm{YUV}, \mathrm{rgs}\}$ and compute color features in each color space respectively. Since color value in each channel is smooth and successive, we divide the channel into 8 sections and the color space into $8 * 8 * 8 = 512$ subspaces equally. Throughout the paper, each channel in all color space is normalized to the range [0, 1]. $S^i_{c_j}$ denotes the i-th section of color channel c_j (j = 1, 2, 3). $m^i_{c_j}$ is the center of section that can be computes by:

$$m^i_{c_j} = \frac{\sum v}{n} (\forall v \in s^i_{c_j}, n = 32) \tag{1}$$

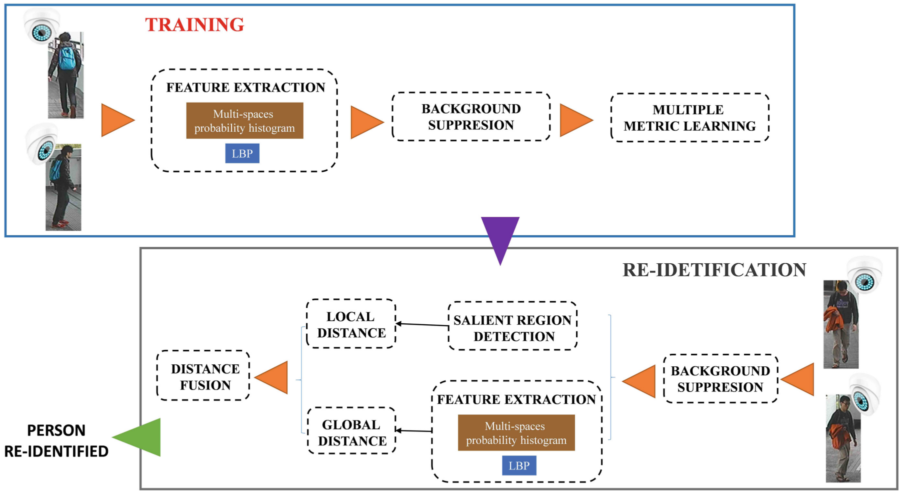

Fig. 2. The system architecture based on five main stages: (1) feature extraction, (2) background suppression, (3) multiple metric learning, (4) salient region detection, and (5) distance fusion.

where n is the number of values in $S^i_{c_j}$. The collection of all space center can be denoted as $z = [m^1_{c_1}m^1_{c_2}m^1_{c_3}, m^1_{c_1}m^1_{c_2}m^2_{c_3}, m^1_{c_1}m^1_{c_2}m^3_{c_3}, ..., m^8_{c_1}m^8_{c_2}m^7_{c_3}, m^8_{c_1}m^8_{c_2}m^8_{c_3}]$.

The distance between color c and k-nearest neighbor color space centers is computed by:

$$dist(c, z_p) = \sum_{i=1}^{3} \|c_i - z_{p_i}\|^2 \tag{2}$$

where i means the i-th channel of color c and z_p.

The probability distribution based the above distance is used to describe color c, defined as:

$$d_p(c) = \begin{cases} \frac{exp(-dist(c,z_p))}{\sum_{l=1}^{k} exp(-dist(c,z_l))} & , z_p \in KNN(C) \\ 0 & , z_p \notin KNN(C) \end{cases} \tag{3}$$

where K means the number of nearest neighbors (we set k = 8 in this paper), p = [1,512] refers to index of z. z_p and $z_l(p, l = 1, 2, .., K)$ belong to K nearest color space center of color c. To calculate the probability distribution of color c, we first use KNN algorithm to find K nearest space centers in Euclidean space. Then the distance from color c to each KNN color space center is utilized to embody the similarity between color c and each KNN color space center. After normalization, the probability distribution of color c over 512 space is defined as d_p. Moreover, it is easy to prove that the sum of the probability distribution of color c over all color space is 1. Finally, the color probability histogram feature of patch Xb in p-th part of z is the sum of all pixels in this patch, which can be defined as:

$$Xb^p = \sum d_p(c_i) \tag{4}$$

where c_i is the color of a pixel in patch Xb. So the color feature dimension of patch Xb is $4 * 512 = 2048$.

In addition, Local Binary Pattern (LBP) [15] is extracted from each patch. A feature which fuses color and texture feature is effective to represent the pedestrian image.

3.2 Salient Region Evaluate

In Sect. 3.1, we proposed a global pedestrian image representation model based on a continuity and effective color representation method. Thus, local characteristics also embedded in the image has been lost, but these features has rich discriminative information.

In real life, human could recognize person identities based on some small salient regions. These small salient regions represent pedestrian main characteristic which is distinctive and reliable in person re-identification. Besides, pedestrian body usually contain more distinctive information than legs. Thus, we choose these specific regions as our salient regions:

(a) Body salient region which represents major color information of body;
(b) Legs salient region which represents major color information of legs;
(c) Characteristic salient region which contains characteristics of body, as show in Fig. 3;

The example of salient region detected by our method is shown in Fig. 4.

Fig. 3. Examples of salience region which contain characteristic of body. These regions has the invariable ability to against the changes of poses and viewpoints.

To detect these small salient regions, we propose a detect method called similarrank to compute the saliency of patch. Similarrank is improved from the pagerank [16] which is first applied to person re-identification to our knowledge. Serval suppositions should be made in similarrank:

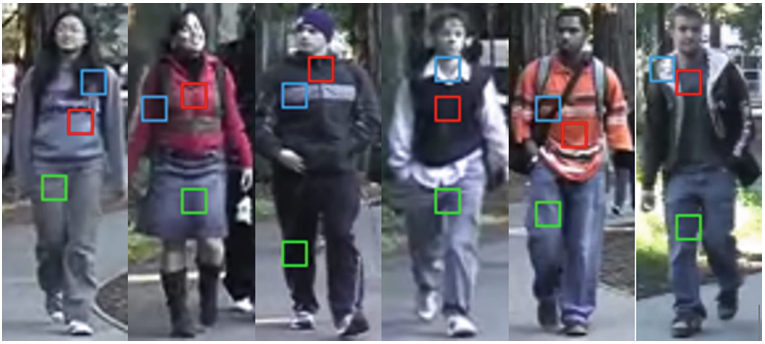

Fig. 4. Examples of salient regions detected by our method. The body salient regions are shown in the red boxes. The legs salient regions are in the green boxes. The Characteristic salient regions are shown in the blue boxes. (Color figure online)

For body/legs salient region:

(a) the more patches similar to patch A, the more salient/important patch A is;
(b) the more salient/important patches similar to patch A, the more salient/important patch A is;

For characteristic salient region:

(a) the more patches dissimilar to patch A, the more salient/important patch A is;
(b) the more salient/important patches dissimilar to patch A, the more salient/important patch A is;

The patch b_i is similar to patch b_j means that similar score for patch b_i to b_j is much large than zero. The similar score for patch b_i to b_j is computed as:

$$S(b_i, b_j) = \begin{cases} exp(-dist(f_i, f_j)) & , exp(-dist(f_i, f_j)) \geq \theta \\ 0 & , exp(-dist(f_i, f_j)) < \theta \end{cases} \tag{5}$$

where θ is a similarity threshold (we choose $\theta = 0.9$ in this paper), f_i and f_j denotes the feature of patch b_i and b_j respectively, $dist(f_i, f_j)$ is the Euclidean distance between f_i and f_j. Furthermore, we divide each pedestrian images into M horizontal stripes. The similar score vector of patch b_i in stripe m can be denoted as $ds_{im} = \{S(b_i, b_1), ..., S(b_i, b_j), ..., S(b_i, b_n)\}$ where $j \neq i$ and n is the number of patches in stripe m, patch b_j represents all other patches in the horizontal stripe positioned at similar height. In addition, the $dist(f_i, f_j)$ in ds_{im} has been normalized by L2-norm, thus $S(b_i, b_j)$ and $S(b_j, b_i)$ are not equal. The set of similar score vectors of all patches in a stripe m can be written as $ds_m = \{ds_{1m}, ..., ds_{im}, ..., ds_{nm}\}$, where ds_{im} has been normalized by L1-norm. The set of similar score vectors of all stripes can be denoted as $ds = \{ds_1, ..., ds_m, ..., ds_M\}$, where M is the number of stripes. A set of similarRank score can be denoted as $SR = \{SR_1, ..., SR_m, ..., SR_M\}$, where $SR_m = \{SR_{11}, ..., SR_{im}, ..., SR_{nM}\}$ is a set of similarRank score in stripe m. Finally, the similarRank score of each patch b_i can be compute as Algorithm 1. The details are shown in Fig. 5:

Algorithm 1 SimilarRank algorithm

Input: The set of all stripes' similar vector: *ds;* The max Iterations Times: *max;*

Output: A set of similar Rank score *SR;*

Initialize: $SR_{im}=1$, $\forall$ $SR_{im} \in sR$;

Iterations Times *Iter*=1;

Repeat

For each SR_{im} in SR

$$SR_{im} = \frac{1-q}{|SR|} + q \sum_{\forall S(b_j, b_i) \in ds} \frac{S(b_j, b_i)}{n_j}, \quad n_j = \left|\{s \mid s \in ds_{jm} \,\&\, s \neq 0\}\right| ;$$

End

Iter++;

Until Iter=max or $|\Delta SR| < \gamma$

Fig. 5. Algorithm 1 to compute the similarRank score of each patch b_i.

Body/Legs Salient Region. For stripe m, the patch which has the maximum of SR_m is denoted as $bmax_m$. The final score is computed as

$$SI(bmax_m) = \sum_{\forall m' \neq m} S(bmax_m, bmax_{m'}) \tag{6}$$

Top half of stripes are selected as body region, and the patch b_m which displays the maximum of $SI(b_m)$ in body region is chosen as the body salient region. Bottom half of stripes are selected as legs region, and the patch b_m which shows the maximum of $SI(b_m)$ in legs region is chosen as the legs salient region.

Characteristic Salient Region. For stripe m, the patch which has the minimum of SR_m is denoted as $bmin_m$. The final score is computed as

$$ST(bmin_m) = \sum_{\forall m' \neq m} S(bmin_m, bmin_{m'}) \tag{7}$$

The patch b_m which is the minimum of $ST(b_m)$ in body region is chosen as the characteristic salient region.

To handle misalignment caused by large viewpoints and poses variations, we compute Euclidean distance between query salient patch b and all patches in gallery images and choose the minimum as the final salient distance $dist_{salient}$.

3.3 Salient Region Evaluate

In Sects. 3.1 and 3.2, we proposed a model to represent the global distribution and local details of pedestrian image. But some patches in background hide interferential information, which have a great negative impact on pedestrian representation. Because person re-identification is a task to find a person from a series of non-overlapping camera views which inevitably make the background of the pedestrian image inconstant. Such negative impact of background will reduce the accuracy of person re-identification. To handle this problem, we set different weight for each patch according to their column. The weight w_c of c-th patch bc in horizontal is defined as:

$$w_c = exp(-\frac{(\mu_c - \mu)^2}{2l^2}) \tag{8}$$

where l is the half of patch size, μ is the half of image width and μ_c is the center column of b_c in pixel. Therefore, the representation near the edge of the image has been suppressed by weight w.

3.4 Metric Learning and Distance Fusion

In person re-identification, the feature of a pedestrian usually input into metric learning as a vector. But as there is no single color space could be against to the variations of illumination, there is no single transformation function suit for a vector which contain different kinds of feature. The combined feature space may also be too complex to be robustly handle by single metric. Therefore, we propose metric each feature space separately and train a pool of multiple metrics.

In [1], Kostinger et al. proposed a metric method called KISSME based on the log likelihood ratio test of two Gaussian distributions. In [5] Liao et al. extend the KISSME approaches to a cross-view metric learning called XQDA. XQDA first learn a subspace $W \in R^{d*n}$ with cross-view data where d means the original dimensional and n means the dimensional subspace. And then it learns a kernel matrix M in the n dimensional subspace. The distance between probe X_p with gallery X_g is computed as

$$distM(X_p, X_g) = (X_p - X_g)^T WMW^T (X_p - X_g) \tag{9}$$

where X_p, X_g mean the feature representation of pedestrian image I_q, I_g.

We assume the result of metric learning in feature space k is defined as W_k, M_k, for k = 1, ..., n (n is the number of space). The result of a probe person p and a gallery person g being the same in feature space k can be defined as the credibility P_k which transform from distance

$$P_k(p = g|M_k) = \sigma((X_p - X_g)^T W_k M_k W_k^T (X_p - X_g)) \tag{10}$$

where $\sigma(z) = exp(-Z)$. The final global credibility can be denoted as:

$$P_k(p = g|M_1, ..., M_k) = \sum_{k=1}^{4} \beta_k P_k(p = g|M_k) \tag{11}$$

where β_k is parameter of feature space weight.

In real life, saliency various in pedestrian images. Only salient regions in salient pedestrian image has the higher credibility. Differences among salient regions can be considered as the saliency of pedestrian image which computed as:

$$\lambda = (1 - \frac{f_b f_l}{\sqrt{|f_b|}\sqrt{|f_l|}})(1 - \frac{f_c f_l}{\sqrt{|f_c|}\sqrt{|f_l|}})(1 - \frac{f_c f_b}{\sqrt{|f_c|}\sqrt{|f_b|}}) \quad (12)$$

where f_b, f_l, f_c represent the feature of body salient region, legs salient region, characteristic salient region respectively. The more different among salient regions, the more salient this pedestrian image is.

In order to output a final decision, all result must be pooled. The pooled result can be obtained by computing the integrated credibility considering local and global result. The salient value λ of pedestrian image can be considered as the weight of local result. Thus, the final decision is computed as

$$P_k(p = g|M_1, ..., M_k) = (1 - 0.7\lambda) \sum_{k=1}^{4} \beta_k P_k(p = g|M_k) + 0.7\lambda dist_{salient} \quad (13)$$

Formula (13) is used to compute the final ranking for re-identification.

4 Experiments

In this section, we evaluate our method on two publicly available datasets, i.e. the VIPeR dataset [17], and the CUHK01 dataset [18]. These datasets have been selected because they both are very challenge datasets for person re-identification and contain many real scenarios, i.e. viewpoints, poses and illumination changes, different backgrounds, image resolutions, occlusions, etc. In our experiments on both datasets, we randomly partition the dataset into two equal part, one for training and another for testing, without overlap on every single person. For each dataset, evaluation procedure is repeated 10 times. We set $\beta = [0.5, 0.3, 0.3, 0.1]$ in each color space i.e. HSV, normal rgb, YUV, rgs.

All the results are shown on the basis of recognition rate by the Cumulative Matching Characteristic (CMC) curve. Images from one camera are used as probe and images from other camera are used as gallery. Each probe image is matched with every gallery image. The rank of correct match is expressed in the CMC curve. In our experiments, our approach is compared with the state-of-the-art methods, i.e. LOMO+XQDA, MFA, kLFDA, KISSME, LFDA.

4.1 Experiment on VIPeR

The VIPeR dataset is captured by two cameras in outdoor environment. It contains 632 low spatial resolution image pairs which relate to 632 persons and are captured by two different cameras. Images from camera A are mostly captured from 0° to 90° while others from camera B mostly from 90° to 180°. All images in VIPeR dataset are 128 * 48 pixels. Some examples from VIPeR dataset show in Fig. 6.

Fig. 6. Examples from VIPeR dataset. The two rows show the different appearances of the same person in different camera views.

We compare our method with the state-of-the-art methods including LOMO+XQDA, MFA, kLFDA, KISSME, LFDA. The experiment results show in Table 1 and Fig. 7. Each methods are following the same evaluation protocol as ours. Results demonstrate that our method achieves better result than the state-of-art methods on higher ranks. As shown, our method achieves the highest rank 1 score by reaching a recognition rate of 44.3% which more 4.1% than the second. From rank 1 to 20, our method performs better than all other methods.

Table 1. Comparison with the state-of-the-art methods on VIPeR dataset.

Method	Rank = 1	Rank = 5	Rank = 10	Rank = 15	Rank = 20
Ours	**44.3**	**74.6**	**85.8**	**91.8**	**94.9**
LOMO+XQDA	40.2	68.3	80.8	87.1	91.1
MFA	20.5	49.0	63.4	73.1	78.2
LFDA	18.3	44.6	57.3	66.7	73.0
KISSME	22.5	49.6	64.1	71.8	83.5
kLFDA	22.2	47.2	60.3	69.0	76.0

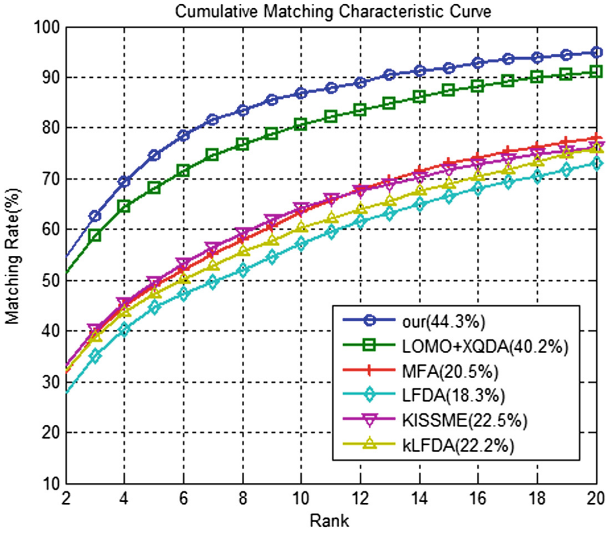

Fig. 7. The CMC curve of ours method and state-of-the-art method on VIPeR dataset

4.2 Experiment on CUHK-01

The CUHK-01 dataset has images captured by non-overlapping camera views in a campus environment. Images in this dataset are of higher resolution. The dataset contain 971 persons images captured from two different cameras. Each person has two images in each camera view. Camera A is from a frontal view and camera B is from a side view. All images are normalized to 160 * 60 pixel for evaluations. It also contain significant variations on viewpoints, poses, illumination, and their images are with occlusions and background clutters. Some example from CUHK-01 dataset show in Fig. 8.

We split the dataset into a training set which contain 485 person and a testing set which contain 486 person. We compare our method with five state-of-the-art approaches include LOMO+XQDA, MFA, kLFDA, KISSME, LFDA. The experimental results are shown in Fig. 9 and Table 2. From the experimental results, we can see our method performs better than other ones by reaching a correct recognition rate of 65.3% at rank 1. As shown, in the most rank, our method performs always better recognition rate than all other existing methods. From the analysis of all the results, we can conclude that, in general, our method is outperform the state-of-the-art approaches and more robust to the viewpoint and illumination variations.

The intrinsic reasons for significant improvement is as follows. Firstly, we analysis person images from different scales. It describes pedestrian image from local details and global distribution. Secondly, we extract probability histogram from different color space to represent pedestrian images. It is much more stable and effective for the obvious change of illumination. In addition, we model each feature space separately. A pool of multiple metric learning is built to metric an optimal distance.

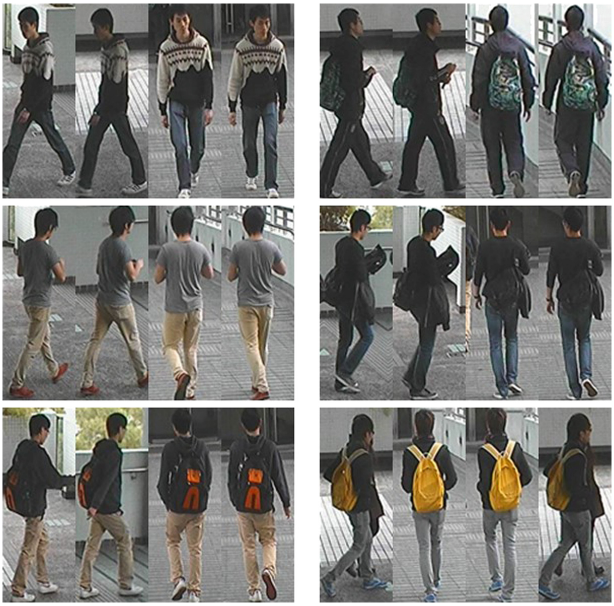

Fig. 8. Examples from CUHK-01 dataset. Each region show the different appearances of the same person in different camera views.

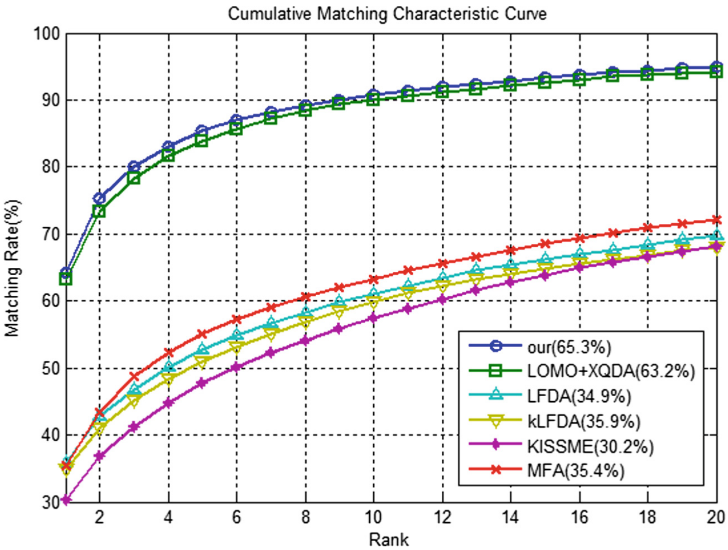

Fig. 9. The CMC curve of ours method and state-of-the-art method on CUHK-01 dataset

Table 2. Comparison with the state-of-the-art methods on CUHK-01 dataset.

Method	Rank = 1	Rank = 5	Rank = 10	Rank = 15	Rank = 20
Ours	**65.3**	**85.6**	**91.15**	**94.0**	**94.6**
LOMO+XQDA	63.2	83.9	90.0	92.6	94.2
MFA	35.4	55.1	63.3	68.5	72.1
LFDA	34.9	50.9	59.9	64.8	68.0
KISSME	30.2	47.7	57.5	63.9	68.2
kLFDA	35.9	52.7	61.1	66.2	69.8

5 Conclusion

In this paper, we have proposed a novel feature representation which consider the local detail and global geometric correspondence of pedestrian image. Some small salient regions detected to describe the local detail of pedestrian image. Probability histograms is computed in different color space to address the illumination problem. Besides, we proposed a pool of multiple metric learns. All the distance metric results are translated into reliability credibility and pooled as final decision. The evaluated on two public datasets VIPeR, CUHK-01, showing outperforms the state-of-the-art on both datasets. In the future work, we are considering how to learn a color transformation between different camera views and how salient region can be used to weight patches.

References

1. Koestinger, M., Hirzer, M., Wohlhart, P., et al.: Large scale metric learning from equivalence constraints. In: 2012 IEEE Conference on Computer Vision, Pattern Recognition (CVPR), pp. 2288–2295. IEEE (2012)
2. Prates, R., Dutra, C.R.S., Schwartz, W.R.: Predominant color name indexing structure for person re-identification. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 779–783. IEEE (2016)
3. Lisanti, G., Masi, I., Bagdanov, A.D., et al.: Person re-identification by iterative re-weighted sparse ranking. IEEE Trans. Pattern Anal. Mach. Intell. **37**(8), 1629–1642 (2015)
4. Chen, D., Yuan, Z., Hua, G., et al.: Similarity learning on an explicit polynomial kernel feature map for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1565–1573 (2015)
5. Liao, S., Hu, Y., Zhu, X., et al.: Person re-identification by local maximal occurrence representation, metric learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)
6. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: CVPR (2010)
7. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: Proceedings of the CVPR (2013)

8. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., Li, S.Z.: Salient color names for person re-identification. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 536–551. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10590-1_35
9. Mignon, A., Jurie, F.: PCCA: a new approach for distance learning from sparse pairwise constraints. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2666–2672. IEEE (2012)
10. Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3318–3325. IEEE (2013)
11. Yan, S., Xu, D., Zhang, B., Zhang, H.J., Yang, Q., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
12. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 649–656. IEEE (2011)
13. Prosser, B., Zheng, W.S., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: BMVC (2010)
14. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178. IEEE (2006)
15. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
16. Page, L., Brin, S., Motwani, R., et al.: The PageRank citation ranking: bringing order to the web (1999)
17. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: PETS (2007)
18. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7724, pp. 31–44. Springer, Heidelberg (2013). doi:10.1007/978-3-642-37331-2_3