Gaussian Descriptor based on Local Features for Person Re-identification

Bingpeng Ma[1], Qian Li[2], Hong Chang[3]

[1] School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing, 100049, China.
[2] School of Computer Science & Technology, Huazhong University of Science & Technology, Wuhan, 430074, China.
[3] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China.
(E-Mail: bpma@ucas.ac.cn, qian.li@vipl.icat.ac.cn, changhong@ict.ac.cn).

Abstract. This paper proposes a novel image representation for person re-identification. Since one person is assumed to wear the same clothes in different images, the color information of person images is very important to distinguish one person from the others. Motivated by this, in this paper, we propose a simple but effective representation named Gaussian descriptor based on Local Features (GaLF). Compared with traditional color features, such as histogram, GaLF can not only represent the color information of person images, but also take the texture and spatial structure as the supplement. Specifically, there are three stages in extracting GaLF. First, pedestrian parsing and lightness constancy methods are applied to eliminate the influence of illumination and background. Then, a very simple 7-d feature is extracted on each pixel in the person image. Finally, the local features in each body part region are represented by the mean vector and covariance matrix of a Gaussian model. After getting the representation of GaLF, the similarity between two person images are measured by the distance of two set of Gaussian models based on the product of Lie group. To show the effectiveness of the proposed representation, this paper conducts experiments on two person re-identification tasks (VIPeR and i-LIDS), on which it improves the current state-of-the-art performance.

1 Introduction

The task of person re-identification can be defined as finding the correspondences between a *probe set* of images representing a single person and those from a *gallery set*. In recent years, person re-identification has attracted a lot of attentions because of its importance in many real-world applications, such as video surveillance. Depending on the number of available images per individual (the size of the probe set), the scenarios in person re-identification can be categorized as: (a) single-shot [1] [2], if only one frame per individual is available in both probe and gallery sets; and (b) multiple-shot [1] [2], if multiple frames per individual are available in both probe and gallery sets. In this paper, we just care about the single-shot scenario.

One key issue of person re-identification is how to represent the human images in the bounding boxes. Coming from different cameras in a distributed network or from the same camera at different time, the human images may have great variations in illumination, pose, viewpoint, background, partial occlusions and resolution. These variations increase the difficulty of person re-identification. To this end, researchers have proposed a lot of representation methods which are somewhat robust to the above variations. Generally speaking, these representations are based on (i) color [1], usually encoded within histograms of RGB or HSV values [1], (ii) shape, e.g. using HOG based signature [3] [4], (iii) texture, often represented by Gabor filters [5] [6] [7], differential filters [7], Harr-like representations [8] and Co-occurrence Matrices [4], (iv) interest points, e.g. SURF [9] and SIFT [10] [11] and (v) image regions [3] [1].

Since these elementary features (color, shape, texture, etc.) capture different aspects of the information contained in the images, they are often combined to give a richer signature. For example, [5] combines 8 color features with 21 texture filters (Gabor and differential filters). [1] and [2] combine the descriptors of Maximally Stable Colour Regions (MSCR) with weighted Color Histograms (wHSV), achieving the state-of-the-art results on several widely-used person re-identification datasets. Similarly, [12], [13] and [14] combine their features with wHSV and MSCR, respectively.

Among these representations, the color features of the human images, as simple but efficient visual signatures, are the most commonly used representation. In the task of person re-identification, it is often assumed that person wears the same clothes in different images. Under this assumption, compared with other representations, the color feature of the person images, such as the histograms of the different color channels, is the primary key to distinguish one person from the others. However, on one side, the color features change easily with illumination variations and camera parameters. On the other side, the color histograms lose the information of the texture and spatial structures of the human images. These drawbacks limit the applications of the color features in the real person re-identification systems.

To overcome the above drawbacks of the color features, this paper presents a novel image representation named Gaussian descriptor based on Local Features, GaLF for short. Specially, GaLF includes three stages. In the first stage, pedestrian parsing and lightness constancy method are applied to the human images to eliminate the influence of the background and illumination, respectively. In the second stage, a very simple 7-dimensional local feature is extracted on each pixel in the human images. The 7-d features include the information of color, texture and spatial structure at the same time. In the third stage, the distribution of local features in each body part region are modeled by a Gaussian model. It is easy to know that the mean vector can keep the color information and spatial structure while the covariance matrix keep the texture as the supplement.

Therefore, by integration of the mean vector and the covariance matrix, GaLF can keep the information of color, texture and spatial structure at the same time.

After getting the representations of GaLF, the similarity measurement of two human images can be obtained by computing the similarity between two sets of part-wise Gaussian models. In this paper, based on product of Lie group, both the difference of the mean vectors and the LOG-Euclidean distance of the covariance matrices are computed in the Riemannian space. Finally, the similarity of two human images are decided by the sum of the difference of the mean vectors and the LOG-Euclidean distance of the covariance matrices of two Gaussian sets. Similar to the combination of the different features to give a richer signature, GaLF can be also combined with some other representations. GaLF and its combination are experimentally validated on two challenging public datasets for person re-identification: VIPeR and i-LIDS. On both databases, the results of the proposed representations outperform the current state-of-the-art.

There are three main contributions in this paper. First, we propose a novel pixel-wise feature representation which can express color, texture and spatial structure at the same time. Second, we propose local Gaussian descriptors, GaLF, based on pedestrian parsing and lightness constancy results, and the similarity between two sets of local Gaussian descriptors is measured properly based on the product of Lie group. Third, the proposed representation method is successfully applied to the task of person re-identification, achieving even higher performance than state-of-the-art methods.

The remaining of this paper is organized as follows: Section 2 describes the proposed representation in details. Experimental validations are given in Section 3. Finally, Section 4 concludes the paper and some future works are also discussed.

2 Gaussian descriptor based on Local Features

This section introduces the proposed novel image representation: GaLF. GaLF is a three stage representation. In the following parts, we introduce each stage of GaLF in details, followed by how to improve the performance of GaLF by combining it with other representations.

2.1 data processing

Considering the traditional color features are easily varied with the variations of illuminations and camera parameters, in the first stage of GaLF, we use the pedestrian parsing method to discard the background and the lightness constancy method to eliminate the influence caused by the illumination, respectively.

Specially, for each image, we use a method named Deep Decompositional Network (DDN) [15] for parsing pedestrian images into six semantic regions, such as hair, head, body, arms, legs and background. DDN is able to effectively characterize the boundaries of body parts and accurately estimate complex

Fig. 1. The images after data processing. The images in the same column belong to the same person. The images in the left are the origin image in the database and the images in the right are the images after data processing.

pose variations with good robustness to occlusions and background clutters. The further details about DDN can be found in [15].

After getting the human body regions, we use Gray World (GW), a very simple lightness constancy adaptation approach, to eliminate the influence of the illumination. GW makes us perceive the objects as medium gray which reflect the average luminance of a scene. In terms of histogram properties, it seeks to equalize the mean of the different channels and this corresponds to a level distribution which has its center mass around the middle value. Generally speaking, GW can eliminate some global chromatic dominant.

We show the images after person parsing and color constancy processing in Fig. 1. The image pair in the same column belongs to the same person. The images in the left sub-figure are the original images in the database. We can find that there are the great variations of background and illumination between the original images of the same person. It is easy to understand that these variations increase the difficulty of the task of person re-identification. The images in the right sub-figure are the images after data processing. For the images after data processing, background is eliminated roughly and the similarity of the same person's images is improved greatly. On the whole, the images in Fig. 1 affirm that data processing can improve the similarity of the images belong to the same person.

2.2 local features

In the second stage of GaLF, we extract the local features on each pixel in the human images. There are some traditional local features in computer vision, such as SIFT. They have gained the great success in many tasks. But in GaLF, considering the computation complex and the importance of the cloth color information in person re-identification, we design a very simple 7-dimensional local features to keep more color information.

Specially, in this paper, we extract a 7-dimensional local feature $f(x, y)$ for the pixel at position (x, y). The 7-d local descriptor can be computed as:

$$\begin{aligned} f(x,y) =& [y, L(x,y), A(x,y), B(x,y), \\ & d_{L_y}(x,y), d_{A_y}(x,y), d_{B_y}(x,y)] \end{aligned} \tag{1}$$

In $f(x, y)$, y is the coordinate in the vertical direction, which can be used to keep the spatial structure of the body. $L(x, y)$, $A(x, y)$ and $B(x, y)$ are the intensity of L, A and B color channels at position (x, y), respectively. They can be taken as the color information. $d_{L_y}(x, y)$, $d_{A_y}(x, y)$ and $d_{B_y}(x, y)$ are the gradient in the vertical direction for the different channels and can be taken as the texture information. So, $f(x, y)$ includes the simple information about position, color and textures.

In Eq. 1, we only use the information in the vertical direction, not the information in the horizontal direction. The reason can be attributed to the misalignment. Since the size of the human image is the same in the database and the body is full of the image in the vertical direction, we can argue that human body has been aligned coarsely in the vertical direction. But, in the horizontal direction, the misalignment happens because the human bodies have the pose variations or the images are captured at the different views. We believe that the position information in the horizontal direction can not be used directly in Eq. 1. To validate our idea, we design the experiments in Section 3. The results confirm our argues.

2.3 Gaussian models

In the third stage of GaLF, we use a Gaussian model to model the 7-d local features in the the same semantic body region, which is gained by DDN. Then, the representation of GaLF is the concatenation of the mean vector and the covariance matrix of the Gaussian model.

Specially, for the local features in the n-th body region of image i, their mean vector and covariance matrix can be computed and denoted by μ_{in} and Σ_{in}, respectively. In GaLF, we use the parameter $\mathbf{g}_{in} = (\mu_{in}, \Sigma_{in})$ of the Gaussian model to represent the n-th region of image i. Finally, for human image i, its GaLF representation $\mathbf{D}_i$ is represented by a set of Gaussian models: $\{\mathbf{g}_{in} = (\mu_{in}, \Sigma_{in}), n = 1, \cdots, N\}$, where N is the total regions of the human images.

After getting the representation of GaLF, the distance between region I_{in} and I_{jn} can be obtained by computing the distance between their representations $\mathbf{g}_{in}$ and $\mathbf{g}_{jn}$. Recently, by regarding the space of Gaussians into as a product of Lie groups, [16] measure the intrinsic distance between Gaussians in the underlying Riemannian manifold. They have gained the great success in the image retrieval task. This paper follows their ways and the similarity of Gaussian model $\mathbf{g}_{in}$ and

$\mathbf{g}_{jn}$ can be computed by:

$$\begin{aligned} d_\theta(\mathbf{g}_{in}, \mathbf{g}_{jn}) &= (1-\theta)a(\mathbf{g}_{in}, \mathbf{g}_{jn}) + \theta b(\mathbf{g}_{in}, \mathbf{g}_{jn}) \\ a(\mathbf{g}_{in}, \mathbf{g}_{jn}) &= \left((\mu_{in} - \mu_{jn})^T(\Sigma_{in}^{-1} + \Sigma_{jn}^{-1})(\mu_{in} - \mu_{jn})\right)^{1/2} \\ b(\mathbf{g}_{in}, \mathbf{g}_{jn}) &= \|\log(\Sigma_{in}) - \log(\Sigma_{jn})\|_F \end{aligned} \tag{2}$$

In Eq. 2, θ is a constant in the range $[0, 1]$ to balance the weight of $a(\mathbf{g}_{in}, \mathbf{g}_{jn})$ and $b(\mathbf{g}_{in}, \mathbf{g}_{jn})$. In this paper, θ is set to the constant 0.4. $a(\mathbf{g}_{in}, \mathbf{g}_{jn})$ measures the difference of mean vectors μ_{in} and μ_{jn}. According to the method of the Lie group, the distance between μ_{in} and μ_{jn} should be appropriately weighted by the associative covariance matrices Σ_{in} and Σ_{jn}. $b(\mathbf{g}_{in}, \mathbf{g}_{jn})$ measures the Log-Euclidean distance between covariance matrix Σ_{in} and Σ_{jn}, which is the geodesic distance in the Riemannian space. $\|\cdot\|_F$ denotes the matrix Frobenius norm. Satisfying all the metric axioms, $d_\theta(\mathbf{g}_{in}, \mathbf{g}_{jn})$ is the metric between $\mathbf{g}_{in}$ and $\mathbf{g}_{jn}$.

Finally, the distance between image I_i and I_j is obtained by sum the similarity between their regions:

$$d(\mathbf{D}_i, \mathbf{D}_j) = \sum_{n=1}^{n} d_\theta(\mathbf{g}_{in}, \mathbf{g}_{jn}) \tag{3}$$

Compared with one Gaussian Model, Gaussian Mixed Model (GMM) can represent the region more well and has been widely used. However, in GaLF, for each region, we only use one Gaussian model. Three reasons make us believe that it is possible to just use one Gaussian model to represent the body region. First, in this paper, we just put our attentions on the single-shot scenario. Under this setting, there is just one image in the galley set for each person. Second, since the body region is gained by the pedestrian method, the pixels in the same region are very similar to each other. Finally, compared with GMM, the computational complex of one Gaussian model is decreased greatly.

2.4 enriched GaLF

As mentioned in Section 1, better person re-identification performance was usually obtained by combining different type of image descriptors. In this paper, we follow the same methodology and combine our GaLF representation with other two representations: (a) MSCR, as defined in [1] and (b) Local Descriptors encoded by Fisher Vectors (LDFV) [13]. For symbolic simplicity, we name this combination as eGaLF (enriched eGaLF). In eGaLF, the difference between two image signatures $\mathbf{eD}_1 = (GaLF_1, MSCR_1, LDFV_1)$ and $\mathbf{eD}_2 = (GaLF_2, MSCR_2, LDFV_2)$ is computed as:

$$\begin{aligned} d_{eGaLF}(\mathbf{eD}_1, \mathbf{eD}_2) = &\frac{1}{3} d(GaLF_1, GaLF_2) \\ &+ \frac{1}{3} d_{MSCR}(MSCR_1, MSCR_2) \\ &+ \frac{1}{3} d_{LDFV}(LDFV_1, LDFV_2) \end{aligned} \tag{4}$$

Fig. 2. Some images in the VIPeR database. The images in the same column are belonging to the same person.

Obviously, improvements could be obtained by optimizing the weighs based on additional information like class labels, but as we are looking for an unsupervised method, we let them fixed once for all. Regarding the definition of d_{MSCR} and d_{LDFV}, we use the one given in [1] and [13], respectively.

3 Experiment

In this section, the proposed representations are experimentally validated on the VIPeR [17] database and the i-LIDS database [11].

3.1 Pedestrian re-identification on VIPeR Database

The VIPeR database has been widely used and is considered to be one of the benchmarks for person re-identification. It contains $1,264$ images of 632 pedestrians. There are exactly two views per pedestrian, taken from two non overlapping viewpoints. All the images have been normalized to 128×48 pixels. The VIPeR database contains a high degree of viewpoint and illumination variations: most of the examples contain a viewpoint change of 90 degrees, as it can be seen in Fig. 2.

We use Cumulative Matching Characteristic (CMC) curve [18], the standard performance measurements for person re-identification, to show the performance of the proposed representation. CMC measures the expectation of the correct match being at rank r. For the fair comparison, we follow the same experimental protocol [1] and report the average performance over 10 different random sets of 316 pedestrians.

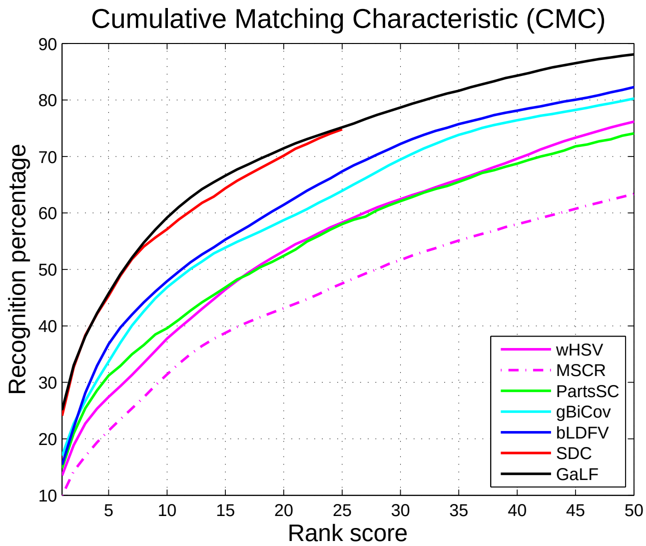

Fig. 3. VIPeR database: CMC curves of the single representations.

Firstly, we compare GaLF with the following single representations: wHSV [1], MSCR [1], PartsSC [19], gBiCov [20], bLDFV [13] and SDC [14]. wHSV is the histogram of HSV channels and can be seen as the baseline in this task. MSCR are extracted from the body region and described by their area, centroid, second moment matrix and average color of the region. bLDFV is the global representation integrated by the local features based on the Fisher Vectors method. SDC is a method based on unsupervised salience learning [14], which can be seen as the state-of-the-art on VIPeR database for the single representation.

Fig. 3 shows the CMC plots of GaLF as well as other single representations. We just show the CMC curve of SDC from rank 1 to rank 25 as it shown in [14]. From the figure, we can see that first, the accuracies of GaLF are better than those of wHSV about 15% at each rank. These obvious advantage shows that by using the textures and spatial structure as the supplement, compared with wHSV, GaLF can gain the better results compared with wHSV. Then, for the different single representations, the performance of GaLF are the best of all. The rank 1 matching rate for GaLF is 25.08% while that of bLDFV is 15.40%. The rank 10 matching rate for GaLF is 59.18% while that of bLDFV is 47.93%. Though the rank 1 matching rates of SDC is very close to that of GaLF, the rank 10 matching rate for SDC is around 56%, which is less about 3% compared with GaLF. These scenes show the good performance of the proposed representation.

Secondly, we compare eGaLF with the representations of SDALF [1], Comb [19], eBiCov [20], eLDFV [13], and eSDC [14]. As shown in Tab. 1, those representations are all the combinations of the different features. From the table, we can know that for other representations, they often select wHSV(Hist) as

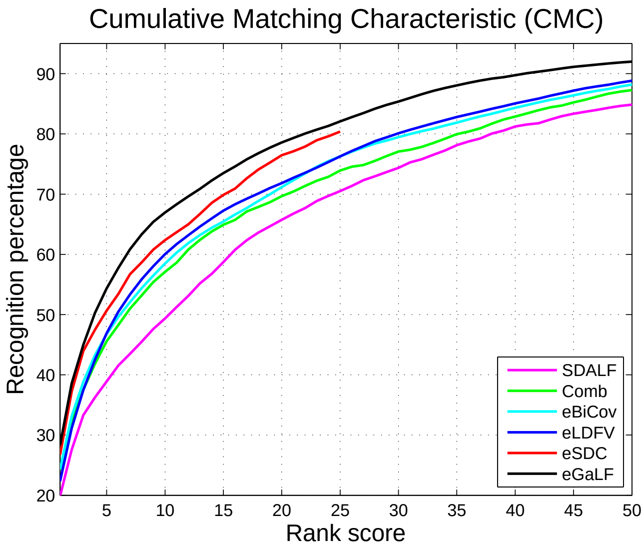

Fig. 4. VIPeR database: CMC curves of the combination of the different features.

Table 1. The combination of the different features.

Representations	features
SDALF [1]	wHSV, MSCR, RHSP
Comb [19]	Hist, PartsSC, Cov
eBiCov [12]	wHSV, MSCR, gBiCov
eLDFV [13]	wHSV, MSCR, bLDFV
eSDC [14]	wHSV, MSCR, SDC
eGaLF	GaLF, MSCR, bLDFV

their color component. But for eGaLF, we use the proposed GaLF as the color component, not the color histogram.

Fig. 4 shows the CMC plots of eGaLF as well as other representations. From Fig. 4, we can know that the results of eGaLF are much better than those of other methods. The rank 1 matching rate for eGaLF is 28.34% while that of eLDFV and eSDC is 22.34% and 26.74%, respectively. The rank 10 matching rate for eGaLF is 66.94% while that of eLDFV and eSDC is 60.04% and 62.37%, respectively. These results show the good performance of the proposed representation. Specially, compared with eLDFV, eGaLF just use GaLF to replace the histogram feature wHSV in eLDFV. So, the advantage of eGaLF shows that GaLF is much better than the histogram features again.

As shown in Eq. 1, in GaLF, we only use the information in the vertical direction, not the horizontal direction. We argue that the position information of the body in the horizontal direction can not be used directly because the misalignment often happens in the task of person re-identification. To validate

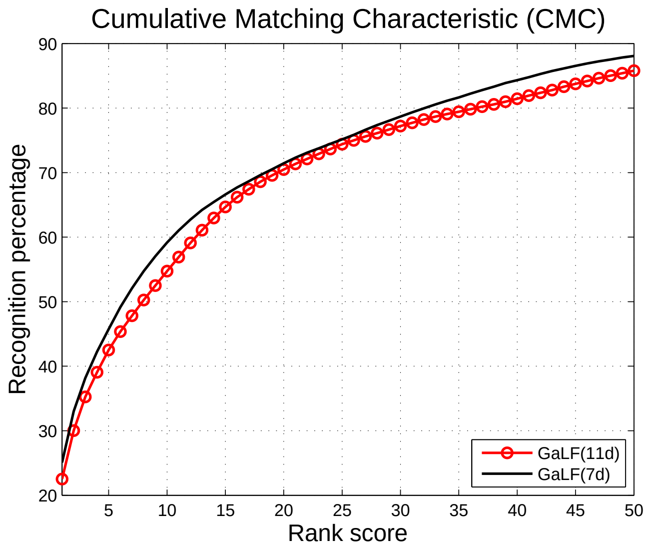

Fig. 5. VIPeR database: CMC curve using the different local features.

this idea, similar to Eq. 1, we also design a 11-d local feature which includes the information in the horizontal direction:

$$f'(x,y) = [f(x,y), x, d_{L_x}(x,y), d_{A_x}(x,y), d_{B_x}(x,y)] \tag{5}$$

In $f'(x,y)$, x is the coordinate in the horizontal direction. $d_{L_x}(x,y)$, $d_{A_x}(x,y)$ and $d_{B_x}(x,y)$ are the gradient in the horizontal direction for the different channels.

We repeat the experiment by using $f'(x,y)$ to replace $f(x,y)$ in GaLF. In Fig. 5, we show the CMC curve of the 7-d and 11-d local features while at the same time. From the figure, we can know that the performance of GaLF is decreased when using the information in the horizontal direction. This scene is accorded with our judgement. So, based on this result, we discard the information in the horizontal direction in GaLF.

In Eq. 2, parameter θ decides the weight of the similarity of mean vectors and covariance matrixes in the distance of two Gaussian models. To show the influence of θ to the performance, we also repeat the experiments by using the different θ. The results are shown in Fig. 6. In the figure, $\theta = 0$ means only using the similarity of $a(\cdot)$ while $\theta = 1$ using $b(\cdot)$. From the figure, we can know that from 0 to 0.4, the performance is improved gradually. When θ is set to 0.4, GaLF can get the best performance. Then, the performance is decreased gradually when θ varies from 0.4 to 1. Based on the results of GaLF using the different θ, in this paper, we set θ to 0.4 for simplicity. We also can know that the performance of $a(\cdot)$ is better than that of $b(\cdot)$ though the advantage is not so obviously. Considering the color information is kept in $a(\cdot)$, the good performance of $a(\cdot)$ also show the color information is the key issue in GaLF.

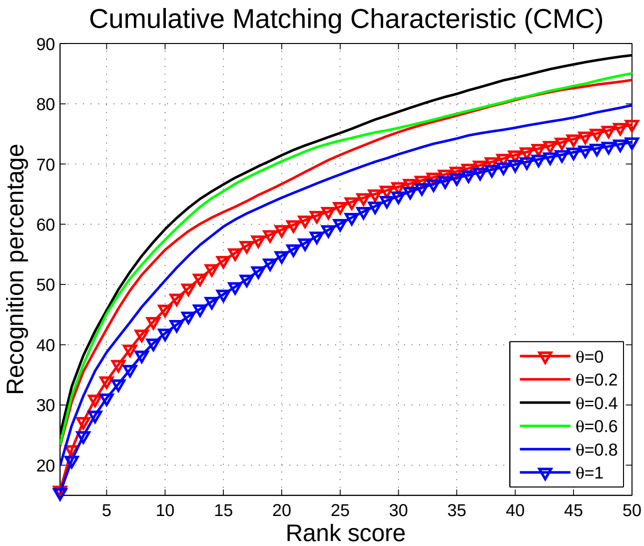

Fig. 6. VIPeR database: CMC curve of GaLF using the different θ.

3.2 Person re-identification on i-LIDS database

Besides VIPeR database, we also test the proposed GaLF and eGaLF on the i-LIDS database. The i-LIDS database has been captured by multiple non-overlapping cameras at a busy airport arrival hall. There are 119 pedestrians with total 476 images. All the images are normalized to the size of 128×64 pixels. Many of these images undergo quite large illumination changes and occlusions (see Fig. 7).

We follow the same experimental settings of [1] [2] and test the proposed descriptors in the single-shot scenario. Considering there are 4 images on average for each pedestrian, we randomly select one image for each pedestrian to build the gallery set, while the rest (357 images) form the probe set. We repeat this procedure 10 times and compute the average CMC. Fig. 8 shows the CMC curves given by GaLF, eGaLF, SDALF [1], Custom Pictorial Structures (PS) [2], gBiCov [20] and SCR [21]. On the i-LIDS database, the best single-shot published performance is obtained by a covariance-based technique (SCR).

From Fig. 8, we can know that GaLF and eGaLF outperform other representations on this database. The rank 1 matching rate for GaLF and eGaLF is 34.50% and 44.34%, respectively, while that of SCR is about 30%. The rank 10 matching rate for GaLF and eGaLF is 69.18% and 72.94%, respectively, while that of SCR is around 63%. Compared with the results on the VIPeR database, the advantages of the proposed GaLF and eGaLF are more obvious. These results show the good performance of the proposed representations again.

Fig. 7. Some images in the i-LIDS database. The images in the same column are belonging to the same person.

4 Conclusion

This paper propose a novel image representation named GaLF for the task of person re-identification. In contrast to the pervious color features, such as color histogram, using the Gaussian model to represent the local features of the body region, GaLF can keep the color features of the input human image while make the texture and spatial structure as the supplement. One important advantage of the proposed representation is its simplicity. Experiments on two pedestrian databases (VIPeR and i-LIDS) show that the proposed GaLF can achieve the state-of-the-art performances in unsupervised setting.

There are several aspects to be further studied in the future. First, in this paper, we just care about the single-shot scenario and use one Gaussian model to represent the human region. For the multi-shot scenario, since there are many samples for one person, we can use GMMs to model the regions of the same person. But this idea need to be validated by experiments. Then, the proposed representations is an unsupervised method. How to use the metric learning method to improve the performance of the proposed representations should be researched further.

Acknowledgment

This paper is partially supported by National Natural Science Foundation of China under Contract nos. 61173065, 61272319 and 61332016 and the President Fund of UCAS.

References

1. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. Proc. IEEE Conference on Computer Vision and Pattern Recognition (2010) 2360–2367

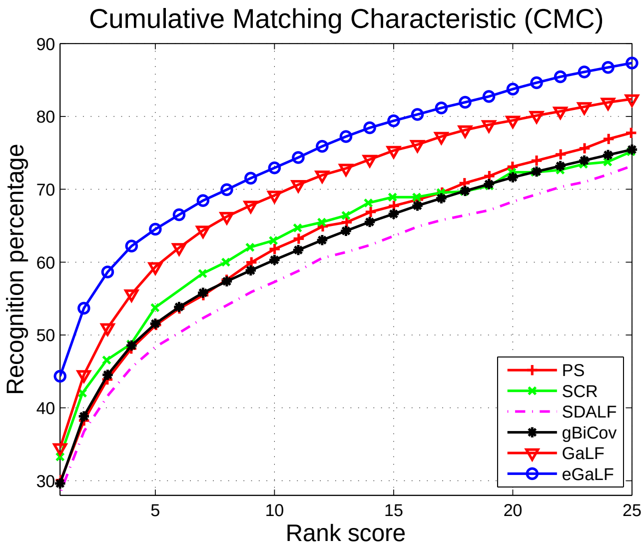

Fig. 8. i-LIDS database: CMC curves of the different methods in the single shot scenario.

2. Cheng, D., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. Proc. British Machine Vision Conference (2011)
3. Oreifej, O., Mehran, R., Shah, M.: Human identity recognition in aerial images. Proc. IEEE Conference on Computer Vision and Pattern Recognition (2010) 709–716
4. Schwartz, W., Davis, L.: Learning discriminative appearance based models using partial least squares. Brazilian Symposium on Computer Graphics and Image Processing (2009)
5. Prosser, B., Zheng, W., Gong, S., Xiang, T.: Person re-identification by support vector ranking. Proc. British Machine Vision Conference (2010)
6. Zhang, Y., Li, S.: Gabor-LBP based region covariance descriptor for person reidentification. International Conference on Image and Graphics (2011) 368–371
7. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. Proc. European Conference on Computer Vision (2008) 262–275
8. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using haar-based and DCD-based signature. Proc. International Workshop on Activity Monitoring by Multi-camera Surveillance Systems (2010)
9. Gheissari, N., Sebastian, T., Tu, P., Rittscher, J., Hartley, R.: Person reidentification using spatiotemporal appearance. Proc. IEEE Conference on Computer Vision and Pattern Recognition **2** (2006) 1528–1535
10. Kai, J., Bodensteiner, C., Arens, M.: Person re-identification in multi-camera networks. Proc. IEEE Conference on Computer Vision and Pattern Recognitio Workshops (2011) 55–61
11. Zheng, W., Gong, S., Xiang, T.: Associating groups of people. Proc. British Machine Vision Conference (2009)

12. Ma, B., Su, Y., Jurie, F.: Bicov: a novel image representation for person re-identification and face verification. Proc. British Machine Vision Conference (2012) 57.1–57.11
13. Ma, B., Su, Y., Jurie, F.: Local descriptors encoded by fisher vectors for person re-identification. International Workshop conjunction with European Conference on Computer Vision (2012) 413–422
14. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. (2013) 3586–3593
15. Luo, P., Wang, X., Tang, X.: Pedestrian parsing via deep decompositional neural network. Proc. IEEE International Conference on Computer Vision (2013)
16. Li, P., Wang, Q., Zhang, L.: A novel earth mover's distance methodology for image matching with Gaussian mixture models. Proc. IEEE International Conference on Computer Vision (2013)
17. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (2007)
18. Moon, H., Phillips, P.: Computational and performance aspects of PCA-based face-recognition algorithms. Perception **30** (2001) 303–21
19. Kviatkovsky, I., Adam, A., Rivlin, E.: Color invariants for person reidentification. IEEE Trans. on Pattern Analysis and Machine Intelligence **35** (2013) 1622–1634
20. Ma, B., Su, Y., Jurie, F.: Covariance descriptor based on bio-inspired features for person re-identification and face verification. Image and Vision Computing **32** (2014) 379–390
21. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using spatial covariance regions of human body parts. Proc. International Conference on Advanced Video and Signal-Based Surveillance (2010)