



Feature Fusion and Ellipse Segmentation for Person Re-identification

Meibin Qi[✉], Junxian Zeng^(✉), Jianguo Jiang, and Cuiqun Chen

School of Computer and Information, Hefei University of Technology,
Hefei 230009, Anhui, China
qimeibin@163.com, junxian@mail.hfut.edu.cn, jgjjiang@hfut.edu.cn,
chencuiqun@163.com

Abstract. Person re-identification refers a task of associating the same person in different camera views. Due to the variance of camera angles, pedestrian posture and lighting conditions, the appearance of the same pedestrian in different surveillance videos might change greatly, which becomes a major challenge for person re-identification. To solve the above problems, this paper proposes a feature fusion and ellipse segmentation algorithm for person re-identification. First of all, in order to reduce the impact of changes in light illumination, an image enhancement algorithm is used to process pedestrian images. Then the ellipse segmentation algorithm is applied to reduce the influence of background clutter in the image. After that, we extract features which contain more abundant information and merge them together. Finally, bilinear similarity metric is combined with Mahalanobis distance as a distance metric function, and the final metric matrix is obtained by using optimization algorithm. Experiments are performed on three public benchmark datasets including VIPeR, PRID450s, CUHK01, and the results clearly show the significant and consistent improvements over the state-of-the-art methods.

Keywords: Person re-identification · Feature fusion
Ellipse segmentation

1 Introduction

Person re-identification matches persons across non-overlapping camera views at different time. It is applied to criminal investigation, pedestrian search, and multi-camera pedestrian tracking, etc. And person re-identification plays a crucial role in the field of video surveillance. Actually, the pedestrian images come from different cameras, and the appearance of pedestrians will change greatly when the lighting, background and visual angle vary. In order to solve the above

Supported by organization National Natural Science Foundation of China Grant 61632007 and Key Research and Development Project of Anhui Province, China 1704d0802183.

problems, many of the previous works mainly focus on two aspects: extracting features [11, 15, 16, 19] from images and measuring the similarity [3, 11, 20] between images. The former is to extract the robust feature to solve the change of the pedestrian appearance. The latter is to make the similarity of different pedestrians smaller and the similarity of the same pedestrian greater.

At present, most features use color and texture information. The SCSP (Spatially Constrained Similarity function on Polynomial feature map) [3] uses color and texture features, and combines global and local features. However, its features are relatively simple, and the information contained is not comprehensive enough. To solve these problems, based on SCSP, this paper adopts more informative LOMO (Local Maximal Occurrence) [11] feature and GOG (Gaussian of Gaussians Descriptor) [15] feature. The LOMO feature contains color and texture information and the GOG feature contains position coordinates and gradient that the LOMO feature does not have. So LOMO and GOG can achieve complementarity, we use the two fused features to replace the global features of SCSP which has better performance than the SCSP. In addition, in order to reduce background noise, this paper proposes ellipse segmentation which has the advantages of effectiveness and simplicity.

Our contributions can be summarized as follows:

- (1) We propose an effective feature representation that uses the fusion of LOMO and GOG features as the global feature and then combine the global and local features to form the final feature.
- (2) We present a new and simple segmentation method called ellipse segmentation, which can effectively reduce the impact of background interference.
- (3) We operate in-depth experiments to analyze various aspects of our approach, and the final results outperform the state-of-the-art over three benchmarks.

The rest of this paper is organized as followed. Section 2 reviews related works. Section 3 describes the details of the proposed method, include: how to extract the feature; the details of the image partition. The experiments and results are in Sect. 4. We finally make a conclusion and discuss possible future works in Sect. 5.

2 Related Works

Currently, person re-identification is mainly divided into two major research directions: deep learning methods [4, 9, 13, 22, 26] and traditional methods [2, 7, 11, 16, 19].

Deep Learning Methods. The deep learning model is a data-driven model that learns high-level features by constructing models to discover complex structures and implicit information in large amount of data. In other words, the key point of deep learning technology is how to efficiently learn the high-level semantic expression of features from a large amount of data. In [26], the recognition rate of Rank1 on the small dataset VIPeR is only 45.9% using the deep learning

method, while using the traditional method [3] is 53.54%. Therefore, traditional methods perform better on small datasets. In addition, the deep learning method cannot be used in the environment where the computing capacity of device is insufficient.

Traditional Methods. The traditional methods of person re-identification consist of two main steps: feature representation and metric learning.

The purpose of feature representation is to extract robust feature, thus solving pedestrian appearance changes. Liao et al. [11] propose an efficient feature representation called Local Maximal Occurrence (LOMO), which consists of color and SILTP (Scale Invariant Local Ternary Pattern) [12] histograms to represent person appearance. It uses sliding windows to describe local details. In each sub-window, SILTP histogram and HSV histogram are respectively counted, and the maximum probability value is taken as the final histogram value in the same horizontal sliding window. [16] divides the image into 6 non-overlapping horizontal regions and extracts four color features for each region and fuse them with LBP (Local Binary Pattern) texture features. [19] that is improved on the basis of [16] uses non-uniform image segmentation and extracts the feature which is the combination of four color features and SILTP texture features. The GOG (Gaussian of Gaussians Descriptor) feature is proposed in [15], which is based on a hierarchical Gaussian distribution of pixel features. In [11], it focuses on the global features. [15, 16, 19] divide the image horizontally, which obtain local features. Different from their work, we use the LOMO and GOG features and fuse them together. Furthermore, we use a combination of global and local features, which guarantees the integrity of the information as well as includes more detailed information.

Metric learning is also called similarity learning, which makes the similarity of different types of pictures as small as possible and the same type as large as possible. Most of the metric learning algorithms have the problems of time consuming and high complexity. To solve this problem, the KISSME (Keep It Simple and Straightforward Metric) method [20] learns metric matrix in Mahalanobis distance by considers the problem from the perspective of statistics. Based on KISSME, XQDA (Cross-view Quadratic Discriminant Analysis) [11] learns a discriminate low dimensional subspace by cross-view quadratic discriminate analysis and gets a QDA metric learned on the derived subspace at the same time. In [3], SCSP uses the combination of Mahalanobis distance and a bilinear similarity metric. The Mahalanobis distance compares the similarity of the same location. Bilinear similarity metric can be used to compare the similarity of different locations. This combined metric function is very robust. Therefore, we adopt this metric in this paper.

3 Our Approach

In this section, we describe our method in detail. This paper improves on the basis of SCSP [3], uses the ellipse segmentation and extracts the LOMO and GOG features from the segmented images, then fuses them to replace the global

feature in SCSP, then combines the local features proposed in SCSP to form the final feature. In terms of metric learning, when the number of training samples is too small in practical applications, it is easy to overfit. In order to solve this problem, this paper uses the metric function combining the bilinear similarity metric and the Mahalanobis distance, and finally adopts the ADMM (Alternating Direction Method of Multipliers) [3] optimization algorithm to obtain the optimal metric matrix.

3.1 Ellipse Segmentation

Due to the particularity of pedestrian images, most person re-identification datasets are manually cropped using rectangular frames. This leads to the fact that most pedestrian images contain redundant background information. Because pedestrians are generally in the center of the rectangular box, and the four right-angled areas of the rectangular box are basically background information. In order to tackle this problem, this paper proposes a new segmentation method called ellipse segmentation. It can preserve the effective information of pedestrians and reduce the impact of background interference. The specific segmentation method is shown in Fig. 1.

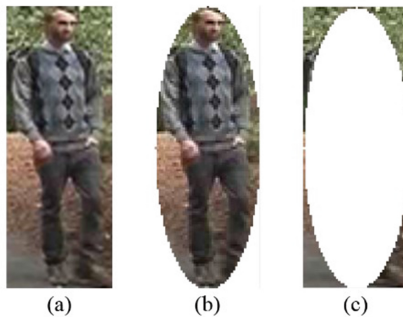


Fig. 1. Ellipse segmentation of image. (a) Original image: contains all the information for the entire image. (b) Ellipse area: retains valid pedestrian information after ellipse splitting and contains a small amount of background information. (c) Background area: contains background information and a small amount of pedestrian information.

3.2 Feature Extraction and Fusion

This paper combines global and local features. At the beginning, we use the image enhancement algorithm [19] to preprocess person images which can reduce the impact of the change of illumination. Then, we fuse the LOMO and GOG features as the global feature. Considering the LOMO feature contains the HSV color and SILTP texture information and the GOG feature contains four colors,

position coordinates, gradient and other information, we combine LOMO with GOG to achieve their complementary power, which makes the features more expressive and robust. According to SCSP, we can know that its local features have good complementarity and recognition effects. Therefore, we use its local features in our method.

Extracting LOMO Feature. First, we perform the ellipse segmentation operation on the image and then extract the LOMO feature, and denote it as LOMO(b), as shown in Fig. 2. Like the literature [11], we use a subwindow whose size is 10×10 and overlapping step is 5 pixels to locate local patches in 128×48 images. Within each subwindow, we extract two scales of SILTP histograms, and an $8 \times 8 \times 8$ -bin joint HSV histogram. To further consider the multi-scale information, we build a three-scale pyramid representation, which downsamples the original 128×48 image by two 2×2 local average pooling operations, and repeats the above feature extraction procedure. By concatenating all the computed local maximal occurrences, our LOMO(b) has $(8 \times 8 \times 8 \text{ color bins} + 3^4 \times 2 \text{ SILTP bins}) * (24 + 11 + 5 \text{ horizontal groups}) = 26960$ dimensions.

The elliptical region we selected has less background noise and more pedestrian information. However, it cannot completely accurately segment pedestrians. So, it may lost some useful information. The background noise in the elliptical region will correspondingly increase when the pedestrian's posture and camera angle change. Therefore, we also extract the LOMO feature from the original image to supplement the information, denote it as LOMO(a), as well as the improved mean LOMO (LOMO_mean) [6] to reduce the background noise in the elliptical region, which is denoted as LOMO(c). The mean value can increase the anti-interference of noise and improve the robustness, and reduce the randomness that brought by the maximum. Thus, we combine three LOMO features as LOMO(a+b+c).

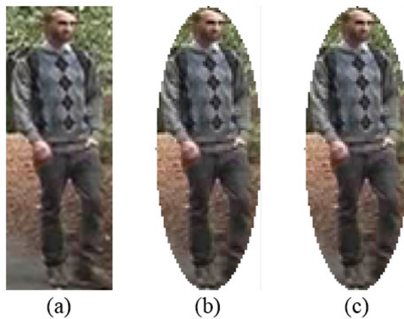


Fig. 2. LOMO feature composition: LOMO(a+b+c). (a) We extract the LOMO(a) feature from the whole picture. (b) We extract the LOMO(b) feature from the elliptical area. (c) We extract the LOMO(c) feature from the elliptical area.

Extracting GOG Feature. According to [15], the dimensionality of GOG descriptor is $27622 = 3 \times ((452 + 3 \times 45)/2 + 1) \times G + 1 \times ((362 + 3 \times 36)/2 + 1) \times G$, G represents the number of overlapping horizontal strips we divide the image. We extract the GOG feature from the whole image as GOG(a) to ensure the integrity of the information. Simultaneously, we also extract the GOG feature from the ellipse region as GOG(b). we combine two features as GOG(a+b).

In summary, for the global feature, the dimensions of LOMO(a), LOMO(b) and LOMO(c) are all reduced from 26960 to 300 dimensions by PCA [8]. GOG(a) and GOG(b) features are reduced from 27622 to 300 dimensions, and then the above five features are concatenated to form the global feature. For local features, we also reduce them to 300 dimensions and then concatenate them in series to form local features. In the end, we concatenate global and local features to form the final feature.

4 Experiments

4.1 Datasets and Settings

Datasets. Three widely used datasets are selected for experiments, including VIPeR [7], PRID450s [21] and CUHK01 [10]. Each dataset is separated into the training set and test set. The test set is further divided into probe set and gallery set, and the two sets contains the different images of the same person. Finally, we take the average results of the 10 experiments.

VIPeR. The VIPeR dataset is one of most challenging datasets for person re-identification task that has been widely used for benchmark evaluation. It contains 632 persons. For each person, there are two 48×128 images taken from camera A and B under different viewpoints, poses and illumination conditions. We randomly select 316 persons for training, and the rest persons for testing.

PRID450s. The PRID450s dataset captures a total of 450 pedestrian image pairs from two disjoint surveillance cameras. Pedestrian detection rectangular box is manually marked and the original image resolution is 168×80 pixels. Each pedestrian contains two images with strong lighting changes. This paper normalizes the image size to 48×128 . The dataset is randomly divided into two equal parts, one for training and the other for testing.

CUHK01. The CUHK01 dataset is captured by two cameras A and B in a campus environment. Each camera captures two pedestrian images. Namely, each pedestrian has four pedestrian images with a total of 971 pedestrians and 3884 pedestrian images. Camera A captures the front and back of the pedestrian, and camera B captures the side of the pedestrian. In the experiment, the persons are split to 485 for training and 486 for test.

Evaluation Metrics. We match each probe image with every image in gallery set, and rank the gallery images according to the similarity score. The results are evaluated by Cumulated Matching Characteristics (CMC) curves. In order

to compare with the published results more easily, we report the cumulated matching result at selected rank- i ($i \in 1, 5, 10, 20$) in following tables.

4.2 Comparison to State-of-the-Art Approaches

Results on VIPeR. The training set and the testing set contain 316 persons respectively. The algorithm of this paper is compared with the existing algorithms on the VIPeR dataset. From the results in Table 1, we can conclude that our algorithm, based on SCSP, has significantly improved the matching rates in comparison with other algorithms. The recognition rate is 9% higher than SCSP on Rank1. At the same time, Rank5, Rank10 and Rank20 have been improved. The Table 1 shows that our method has stronger expression ability and better recognition effect.

Table 1. Matching rates (%) of different methods on VIPeR.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
LOMO+XQDA [11]	40.00	68.13	80.51	91.08
S-SVM [32]	42.66	-	84.27	91.93
literature[19]	42.7	74.5	85.4	92.8
SSDAL [24]	43.50	71.80	81.50	89.00
ME [18]	45.89	77.40	88.87	95.84
Quadruplet+MargOHNM [5]	49.05	73.10	81.96	-
LRP [6]	49.05	74.08	84.43	93.10
Multi-Level Similarity [30]	50.10	73.10	84.35	-
NFST [31]	51.17	82.09	90.51	95.92
SCSP [3]	53.54	82.59	91.49	96.65
Fusion+SSM [1]	53.73	-	91.49	96.08
TLSTP [14]	59.17	73.49	78.62	-
Ours	62.56	87.53	93.89	97.97

Results on PRID450s. From the experimental data in Table 2, we can see that the algorithm in the PRID450s dataset has the highest recognition rate over the state-of-the-art methods. The best Rank1 identification rate of comparison methods is 68.47% [15], while we has achieved 73.29%, with an improvement by nearly 5%.

Results on CUHK01. In previous experiments, the VIPeR and PRID450s datasets are based on single image pair, that is single-shot. But such matching results are easily affected by the quality of single image, and the dataset is generally small. In order to fully embody the performance of the algorithm, a larger dataset CUHK01 is used under multi-shot. Table 3 shows the recognition

Table 2. Matching rates (%) of different methods on PRID450s.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
KISSME [20]	33.0	59.8	71.0	79.0
SCNCD [28]	41.6	68.9	79.4	87.8
DRML [29]	56.4	-	82.2	90.2
LSSCDL [32]	60.5	-	88.6	93.6
LOMO+XQDA [11]	62.60	85.60	92.00	96.60
FFN [25]	66.6	86.8	92.8	96.9
GOG [15]	68.47	88.80	94.50	97.80
Fusion+SSM [1]	72.98	-	96.76	99.11
Ours	73.29	91.78	95.11	97.73

rates of the proposed algorithm and the existing algorithm on the CUHK01 dataset.

Table 3. Matching rates (%) of different methods on CUHK01.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
PCCA [17]	17.8	42.4	55.9	69.1
KISSME [20]	17.9	38.1	48.0	58.8
kLFDA [27]	29.1	55.2	66.4	77.3
Semantic [23]	31.5	52.5	65.8	77.6
FFN [25]	55.5	78.4	83.7	92.6
Quadruplet+MargOHNM [5]	62.55	83.44	89.71	-
LOMO+XQDA [11]	63.2	-	90.8	94.9
GOG [15]	67.3	86.9	91.8	95.9
NFST [31]	69.09	86.87	91.77	95.39
LRP [6]	70.45	87.92	92.67	96.34
Ours	76.19	92.24	95.58	98.09

It can be seen that the algorithm still has significant improvements in the recognition rates in comparison with the existing algorithms on large datasets. Compared with LRP (Local Region Partition) [6], the algorithm of this paper improves about 6% on Rank1. Moreover, our method is 9% higher than the GOG. The improvement of our method is particularly significant on CUHK01.

4.3 Contribution of Major Components

In order to verify the effectiveness of the proposed method, we analyze the algorithm of this paper in detail on the VIPeR dataset. The two sets of Probe and Gallery both have 316 persons.

The Fusion of GOG and LOMO Features. To compare the combination of LOMO and GOG features and the global feature (SCSP-G) [3], Table 4 lists the comparison results of the algorithm in this paper when no segmentation is performed.

Table 4. Matching rates (%) of using LOMO+GOG features and SCSP-G.

Feature	Recognition rate of VIPeR (%)			
	Rank-1	Rank-5	Rank-10	Rank-20
LOMO(a)	40.16	71.33	83.67	92.50
GOG(a)	46.90	78.89	88.61	94.94
SCSP-G [3]	48.10	79.30	89.78	95.76
LOMO(a)+GOG(a)	53.92	83.70	92.06	96.65

It can be seen from the experimental data in Table 4. The LOMO(a)+GOG(a) features has a significant improvement over the SCSP-G [3]. In particular, there is an increase of 5.82% in Rank1, which is 13% and 7% higher than the LOMO(a) and GOG(a) features on Rank1 respectively. It shows that LOMO and GOG can complement each other. Therefore, the features selected in this paper contain more complete pedestrian information and have stronger robustness.

Ellipse Segmentation. To verify the validity of the ellipse segmentation, Table 5 lists the recognition results that the algorithm only uses the ellipse-segmented feature and the original feature without local features.

Table 5. Matching rates (%) of original image and ellipse segmentation.

Feature	Recognition rate of VIPeR (%)			
	Rank-1	Rank-5	Rank-10	Rank-20
GOG(a)	46.90	78.89	88.61	94.94
GOG(b)	47.85	77.31	88.01	94.11
GOG(a+b)	49.46	79.94	89.78	95.00
LOMO(a)	40.16	71.33	83.67	92.50
LOMO(b)	41.60	73.29	85.38	92.18
LOMO(a+b+c)	45.60	76.52	88.01	95.03

From the experimental data in Table 5, it can be seen that the GOG(b) can achieve 47.85% Rank1 matching rate, which outperforms the rate of the GOG(a), and the LOMO(b) increases by nearly 1.5% on Rank1 over the LOMO(a). It shows that the ellipse segmentation feature is more effective than the original. However, Rank5 and Rank10 have decreased. It is because that the segmentation causes the loss of part of the information. After the addition of original feature as supplemental information, the combined feature GOG(a+b) has significantly improved on the Rank1, Rank5, etc. Especially, our method has achieved 49.46%, with an improvement of 2.5%. Moreover, LOMO (a+b+c) increases by 5.5% on Rank1. Therefore, we can conclude that the combined feature with the ellipse segmentation performs better and has a better recognition effect.

5 Conclusions

In this paper, the proposed method fuses LOMO(a+b+c) and GOG(a+b) features as the global feature, and combines them with local features, thus forming more robust feature for the changes of illumination and visual angle. Meanwhile, the algorithm of ellipse segmentation reduces background noise. Furthermore, it can increase the proportion of effective area for pedestrians and enhance the robustness of final joint features. Experimental results show that the proposed algorithm significantly improves the recognition rate of pedestrian re-identification. The recognition rate on Rank10 in the VIPeR, PRID450s, and CUHK01 datasets all reach over 90%, which has practical application of great value.

References

1. Bai, S., Bai, X., Tian, Q.: Scalable person re-identification on supervised smoothed manifold, pp. 3356–3365 (2017)
2. Braz, J., Mestetskiy, L.: Proceedings of the International Conference on Computer Vision Theory and Application: Foreword (2011)
3. Chen, D., Yuan, Z., Chen, B., Zheng, N.: Similarity learning with spatial constraints for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1268–1277 (2016)
4. Chen, W., Chen, X., Zhang, J., Huang, K.: Beyond triplet loss: a deep quadruplet network for person re-identification, pp. 1320–1329 (2017)
5. Chen, W., Chen, X., Zhang, J., Huang, K.: Beyond triplet loss: a deep quadruplet network for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1320–1329 (2017)
6. Chu, H., Qi, M., Liu, H., Jiang, J.: Local region partition for person re-identification. *Multimed. Tools Appl.* **7**, 1–17 (2017)
7. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking (2007)
8. Jégou, H., Chum, O.: Negative evidences and co-occurrences in image retrieval: the benefit of PCA and whitening. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, pp. 774–787. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33709-3_55

9. Li, D., Chen, X., Zhang, Z., Huang, K.: Learning deep context-aware features over body and latent parts for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
10. Li, W., Wang, X.: Locally aligned feature transforms across views. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3594–3601 (2013)
11. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)
12. Liao, S., Zhao, G., Kellokumpu, V., Pietikäinen, M., Li, S.Z.: Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In: Computer Vision and Pattern Recognition, pp. 1301–1306 (2010)
13. Lin, J., Ren, L., Lu, J., Feng, J., Zhou, J.: Consistent-aware deep learning for person re-identification in a camera network. In: Computer Vision and Pattern Recognition, pp. 3396–3405 (2017)
14. Lv, J., Chen, W., Li, Q., Yang, C.: Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns. CoRR abs/1803.07293 (2018). <http://arxiv.org/abs/1803.07293>
15. Matsukawa, T., Okabe, T., Suzuki, E., Sato, Y.: Hierarchical Gaussian descriptor for person re-identification. In: Computer Vision and Pattern Recognition, pp. 1363–1372 (2016)
16. Mei-Bin, Q.I., Tan, S.S., Wang, Y.X., Liu, H., Jiang, J.G.: Multi-feature subspace and kernel learning for person re-identification. *Acta Automatica Sinica* **42**(2), 299–308 (2016)
17. Mignon, A.: PCCA: a new approach for distance learning from sparse pairwise constraints. In: Computer Vision and Pattern Recognition, pp. 2666–2672 (2012)
18. Paisitkriangkrai, S., Shen, C., Hengel, A.V.D.: Learning to rank in person re-identification with metric ensembles, vol. 1, pp. 1846–1855 (2015)
19. Qi, M., Hu, L., Jiang, J., Gao, C.: Person re-identification based on multi-features fusion and independent metric learning. *J. Image Graph.* (2016)
20. Roth, P.M., Wohlhart, P., Hirzer, M., Kostinger, M., Bischof, H.: Large scale metric learning from equivalence constraints. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2288–2295 (2012)
21. Roth, P.M., Hirzer, M., Kostinger, M., Beleznai, C., Bischof, H.: Mahalanobis distance learning for person re-identification, pp. 247–267 (2014)
22. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
23. Shi, Z., Hospedales, T.M., Xiang, T.: Transferring a semantic representation for person re-identification and search. In: Computer Vision and Pattern Recognition, pp. 4184–4193 (2015)
24. Su, C., Zhang, S., Xing, J., Gao, W., Tian, Q.: Deep attributes driven multi-camera person re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 475–491. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_30
25. Wu, S., Chen, Y.C., Li, X., Wu, A.C., You, J.J., Zheng, W.S.: An enhanced deep feature representation for person re-identification. In: Applications of Computer Vision, pp. 1–8 (2016)
26. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification, pp. 1249–1258 (2016)

27. Xiong, F., Gou, M., Camps, O., Sznai, M.: Person re-identification using kernel-based metric learning methods. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8695, pp. 1–16. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10584-0_1
28. 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, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_35
29. Yao, W., Weng, Z., Zhu, Y.: Diversity regularized metric learning for person re-identification. In: IEEE International Conference on Image Processing, pp. 4264–4268 (2016)
30. Guo, Y., Ngai-Man, C.: Efficient and deep person re-identification using multi-level similarity. In: Computer Vision and Pattern Recognition, pp. 2335–2344 (2018)
31. Zhang, L., Xiang, T., Gong, S.: Learning a discriminative null space for person re-identification. In: Computer Vision and Pattern Recognition, pp. 1239–1248 (2016)
32. Zhang, Y., Li, B., Lu, H., Irie, A., Xiang, R.: Sample-specific SVM learning for person re-identification. In: Computer Vision and Pattern Recognition, pp. 1278–1287 (2016)