

# Multiple Metric Learning in Kernel Space for Person Re-identification

Tongwen Xu, Yonghong Song<sup>(✉)</sup>, and Yuanlin Zhang

Institute of Artificial Intelligence and Robotics,  
Xi'an Jiaotong University, Xi'an 710049, China  
xutongwen2014@stu.xjtu.edu.cn, {songyh, ylzhangxian}@mail.xjtu.edu.cn

**Abstract.** In this paper, we present multiple metric learning in kernel space to preferably get more discriminative metrics. Usually, the kernel-based approaches exploit kernel trick to map vectors that usually thousands of dimensions into high dimensional space to enhance their linear capacity. But it could loss the discriminative information in the process of projecting. To address this problem, we propose to map these feature vectors into kernel space respectively and the metrics are learned in their corresponding space. The Relief algorithm is modified here to get the weights and we can get the final result by weighting multiple metrics based on features. Experiments on the public datasets demonstrate the performance of our proposed method outperforms some state of the art methods.

**Keywords:** Person re-identification · Multiple metric learning  
Kernel space

## 1 Introduction

Person re-identification is the task of finding out the same identity in the field of non-overlapping cameras [1, 2, 15]. Due to illumination variations, human pose changes and different background of multiple cameras, its hard to match people from one camera to another. Although lots of approaches have been proposed from various standpoints, it is still an open problem.

There are two common ways to solve problems above for person re-identification, feature representation [3–5, 27, 28] and metric learning [4, 6, 7, 16, 17]. The feature representation mainly focused on how to describe the discriminative information of the object. A color invariant model was proposed in [3] which designed a variable of the change is smaller than the original change to robust illumination variations. Lisanti [4] exploited five color and texture low level features to complementarily represent individuals. In [5], a novel salient color name was presented to describe colors which projected the pixels to probability distribution over sixteen color names. Besides, it is important to learn an efficient metric to measure the distance between two vectors [6, 29, 30]. The metric learned from the algorithm [6] which can minimize the distance

between features of pairs of true matches and maximizes the distance the same between pairs of wrong matches greatly improves the performance of the metric learning in person re-identification. The KISS (Keep It Simple and Straightforward) algorithm [6] was conducted the deeply research and was found that the commonness of two vectors could also describe the similarity of them [7].

Generally, the metric is learned from the feature vectors that exacted from images. However, due to the non-linear capability of these features, the metric learned is not discriminative. To solve this problem, the kernel-based approaches [4, 8, 9, 18] exploited kernel trick to map these features into kernel space to enhance the linear capability and then in kernel space, the metric learning is carried on. In [8], the feature vector was projected into kernel space and the metric was learned by KISS algorithm. To get an efficient kernel space, in [9], multiple kernels were used to map features into more than one subspace. These approaches prove the effectiveness of kernel trick mapping and achieve good performances. However, if the dimension of features mapped is high, the kernel representation gotten by kernel trick is not suitable and it may loss the discriminative information in the process of calculating the kernel matrix. Besides, different features may not be modeled by the same transformation function [10]. The jointly feature space may also be too complex to be robustly handled by a single metric. Therefore we propose to map these features respectively and the metric is learned in their corresponding kernel space. Finally, multiple metric measures can be weighted to get the final result.

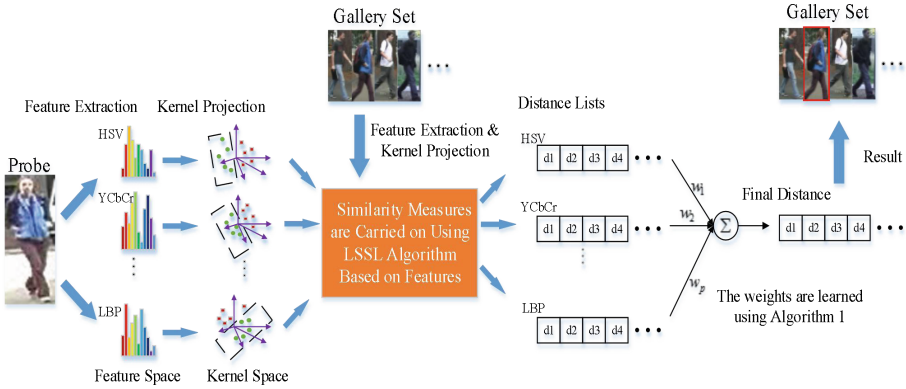
The overview in Fig. 1 depicts the architecture of our proposed work. Firstly, the color, shape and texture features are extracted from images. And then they are mapped into kernel space respectively. In each subspace, the similarity measure is done using LSSL (Large Scale Similarity Learning) algorithm [7] to get distance based on features. Finally, the distances are weighted which the weights were already learned using modified Relief algorithm to get the result.

The main contributions in this paper are threefold: (1) multiple metric learning in kernel space is proposed to solve the problem of losing discriminative information in the process of producing kernel matrix. And the Relief algorithm [11] is modified here to get the weights of metrics. (2) Grey-world normalization is introduced to enhance discrimination of the color descriptors. (3) The experiments are deeply carried on to analyze aspects of our approach and the final results outperform the state-of-the-arts on two benchmarks.

## 2 Our Approach

### 2.1 Person Representation

Color histograms [19] are very important features and are also common used in person re-identification. But, they are sensitive to changing illuminations and the color responses of cameras. The Grey-world normalization [12] is a common and effective technique to address this problem which assumes that the average color of a picture caught by the camera is grey. It divides each channel of every



**Fig. 1.** The overview of the proposed approach.

pixel by the average of this channel in RGB color space, i.e.  $R_1 = R/mean(R_1)$ . If the illumination and color responses change, the RGB values of pixels may change a lot, but  $R_1G_1B_1$  could change smaller than RGB. So the Grey-world normalization is robust to the changes. Here, we introduce this technique in RGB color space and then transformed into nRGB color space to enhance its robustness. The experiments following show its effectiveness.

Besides, the other three color spaces: HSV, RGS, YCbCr are also used to complementarily describe the color information. For each channel, 16-bin histogram is extracted from one of six non-overlapping strips in each image [6] i.e.  $16 \times 6 \times 3 = 288$  dimensions for three channels color space. For shape and texture features, HOG and LBP are used [4], but we only remove 8 pixels up and down of images in order to decrease background pixel influence.

After extracted features from images, these features are projected into kernel spaces respectively using Chi-Square exponential kernel to get their kernel representation [4].

### 2.2 Weighted Multiple Metrics in Kernel Space

In this section, multiple metric learning in kernel space based on features is proposed. The different distances gotten by LSSL algorithm for features are weighted to be the final distance:

$$d_f = \sum_{n=1}^p w_n d_n \tag{1}$$

where  $p$  is the number of features and  $d_n$  is the corresponding distance with features.  $w_n$  denotes the weight and  $\sum_{n=1}^p w_n = 1$ .

To make sure the weights of distances, the importance of features should be de-pendent on. The more important the feature is, the larger weight should

be given. In order to get  $w_n$ , the classic algorithm Relief [11] can be used to calculate the importance of each feature. But in the re-id problem, the Relief algorithm needs to be modified because it usually can deal with two classification problem. To solve this problem we can put the images belonging to the same person into one category and the other images into another. Besides, the sampling number in the Relief is not needed here, because the number of images of the same individual is few, e.g. two. We put all the training samples into the process of calculating the weights and the weights are cumulative sum. Finally, the values of the importance of each feature are normalized to get the weights. The specific algorithm is showed in following.

---

**Algorithm 1.** Modified Relief Algorithm

---

**Input:** Training Samples  $S$ ; Training Samples Pair Number  $num$ ; Features Number  $p$ ;

**Output:** Importance Value Vector  $w = [w_1, w_2, \dots, w_p]$ ;

```

1: for  $i = 1, 2, \dots, p$  do
2:    $w_i = 0$ ;
3:   for  $j = 1, 2, \dots, num$  do
4:     Choose the  $j_{th}$  training sample from  $S$ , labeled as  $A$ ;
5:     Choose the sample belonging to the same category closest to  $A$ , labeled
   as  $B$ ;
6:     Choose the sample belonging to the different category closest to  $A$ , labeled
   as  $C$ ;
7:     Choose the sample farthest to  $A$ , labeled as  $D$ ;
8:     Calculate  $md = (d_{AC} - d_{AB}) / (d_{AD} - d_{AB})$ ;
9:     Calculate  $w_i = w_i + md$ ;
10:  end for
11: end for

```

---

### 2.3 Metric Learning Algorithm

In our approach, LSSL [7] is used to learn the metrics in kernel spaces. Given two samples  $x$  and  $y$ , the similarity distance is defined by LSSL as:

$$d = (x + y)^T M(x + y) - \lambda(x - y)^T W(x - y) \quad (2)$$

$\lambda$  is a constant.  $M$  and  $W$  are two metrics learned as following.

$$M = \sum_{mD}^{-1} - \sum_{mS}^{-1} \quad (3)$$

$$W = \sum_{eS}^{-1} - \sum_{eD}^{-1} \quad (4)$$

where

$$\sum_{mS} = 1/N \sum_{i=1}^N (x_i + y_i)(x_i + y_i)^T \quad (5)$$

$$\sum_{eS} = 1/N \sum_{i=1}^N (x_i - y_i)(x_i - y_i)^T \quad (6)$$

$$\sum_{mD} = \sum_{eD} = 1/(2N) \sum_{i=1}^N [(x_i + y_i)(x_i + y_i)^T + (x_i - y_i)(x_i - y_i)^T] \quad (7)$$

In this paper, Algorithm 1 is used to get the weights of features. But in the experiments, it is noticed that there were different maximum distances with different features. So the multiple distances are not simply weighted to get the final distance. The maximum distances based on features should be normalized to the same and then were weighted.

### 3 Experiments and Discussions

Our proposed approach is tested on two public datasets: VIPeR [13] and CUHK01 [14]. Images have been normalized to pixels, is set to 1.5 on two datasets. The result of the average of cumulative match characteristic (CMC) curve for 10 trials of each experiment is reported in this work.

#### 3.1 Datasets

VIPeR dataset<sup>1</sup> is now considered the most challenging dataset which is also widely used in the field of person re-identification. It contains 1264 images from 632 pedestrians in two disjoint cameras with different illumination conditions and viewpoints. 632 pedestrians are randomly split into two groups, 316 pedestrians for training, the others for testing.

CUHK01 dataset<sup>2</sup> contains 971 individuals from two non-overlapping cameras. Each person has two images with different angles per camera. The dataset is also randomly split into two parts, 485 persons for training, the others for testing.

#### 3.2 Comparison with State-of-the-Art Methods

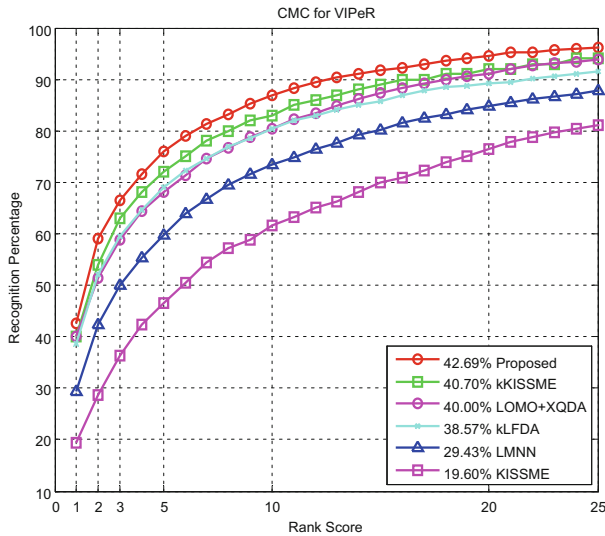
In this section, the performance of our method compared with other some state-of-the-art algorithms on two datasets is shown. In Fig. 2, Tables 1 and 2 as we can see, the performance of our proposed algorithm on VIPeR and CUHK01 outperforms other methods. Accurately, the rank-1 rate is increased from 40.7% to 42.69% after we use the multiple metric learning in kernel space. In Table 2, our method greatly improves the accuracy on CUHK01 dataset. The reasonable explanation is that the form of projecting features expresses more information than the traditional kernel mapping. Besides, the Grey-world normalization and the weights learned from nor-normalized modified Relief also prove their effectiveness.

<sup>1</sup> <http://vision.soe.ucsc.edu/?q=node!178>.

<sup>2</sup> <http://www.ee.cuhk.edu.hk/lxgwang/CUHKjdentification.html>.

### 3.3 Evaluation and Analysis

**Effect of Grey-world normalization.** To illustrate the effectiveness of Grey-world normalization fairly, we also use the KISS metric learning which was used in [6]. The four color feature vectors are concatenated to be a single vector. As shown in Fig. 3, the rank-1 recognition rate of color features extracted from VIPeR without normalization is 29.56%. After Grey-world normalization on images, the rate ups to 34.37%. In Fig. 4, the rank-1 rate of color features increases from 36.54% to 39.13%. The experiments show that the results with normalization are better than that without. This indicates that the normalization is robust to illuminations changing and color responses of cameras and it can effectively improve the accuracy of color features.



**Fig. 2.** CMC curves of some state-of-the-art methods on VIPeR dataset. The rank-1 identification rates are marked before the name of each curve.

**Effect of multiple metric learning.** Different from that multiple vectors are concatenated to be a single vector, the features were projected respectively and the corresponding metric is learned for each feature. For example, in Fig. 3, the four color feature vectors are concatenated to be a 1152 dimension vector and its rank-1 matching rate is 34.37%. Now, we learn four metrics for four color feature and their distances are weighted to be the final distance. The rate ups to 38.07%, improving 3.7% comparing with 34.37%. The HOG and LBP features without multiple metric achieve 5.16%, while the rate improves 2.28% to 7.44% using two metrics learning. Similarly, in Fig. 4 the experiments on CUHK01 dataset, the result of multiple metrics in color, shape and texture features are better than that without. This is because the features are mapped respectively instead

**Table 1.** The identification rates (%) on VIPeR at rank 1, 10, 20, 50, 100 are listed.

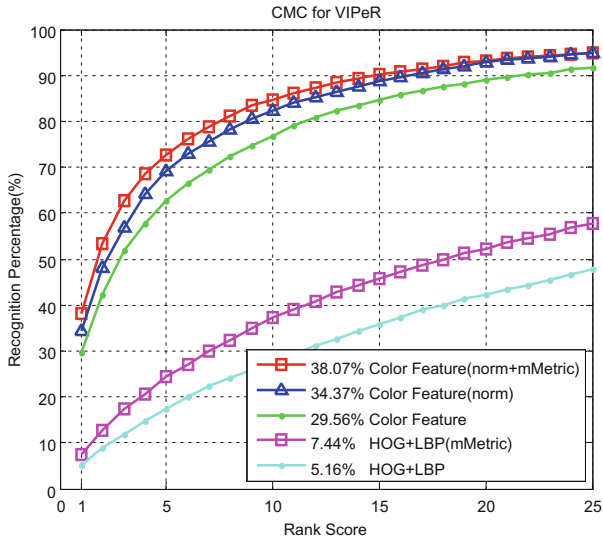
Method	Rank-1	Rank-10	Rank-20	Rank-50	Rank-100
Ours	42.69	86.99	94.65	98.80	99.65
KEPLER [10]	42.41	82.37	90.70	97.06	98.89
kKISSME [8]	40.07	83.95	92.08	98.02	-
XQDA [16]	40.00	80.51	91.08	98.54	99.72
kLFDA [18]	38.58	69.15	80.44	89.15	96.33
KCCA [4]	37.25	84.59	92.78	98.48	99.72
BSFR [20]	34.20	79.70	90.20	98.10	-
SalMatch [21]	30.16	65.54	79.15	91.49	98.10
RCCA [22]	30.75	87	96	99	
MtMCML [23]	28.83	75.82	88.51	-	-
KISSME [6]	19.60	62.20	74.92	91.80	98.00
PRDC [24]	15.66	53.86	70.09	87.79	92.84

**Table 2.** The identification rates (%) on CUHK01 at rank 1, 10, 20, 50, 100 are listed.

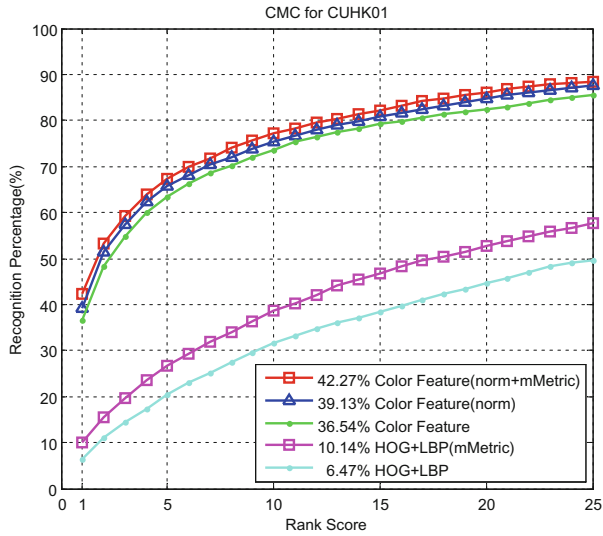
Method	Rank-1	Rank-10	Rank-20	Rank-50	Rank-100
Ours	49.75	83.77	91.15	96.16	98.37
KEPLER [10]	38.85	76.12	84.72	84.72	96.58
kKISSME [8]	36.10	72.61	81.90	-	-
SalMatch [22]	28.45	55.67	67.95	92.85	96.58
PatMatch [22]	20.53	41.09	51.56	72.46	87.91
TML [25]	20.53	56.61	69.62	85.74	93.75
SDALF [26]	9.90	30.33	41.03	55.99	67.39

of projecting a high dimensional vector. The multiple metrics can express more discriminative information to avoid loss information in the process of mapping.

**Effect of Relief algorithm.** In this small section, the four common algorithms used to make sure the weights are introduced to compare their performances. Adopt Weight refers to adopt weight learning and we get the rank-1 39.78%. It is because adopt weight algorithm only consider the rank-1 to get the importance of the feature and it do not consider the distinguished ability of features. Then the Most Probability is carried on. The result is much worse than our method. In Fig. 5, the performance of normalized modified relief is better than the modified relief. This is because we consider there are different maximum distances with different features and we normalize them to solve it. So the weights by normalized modified relief can get better complementarities.

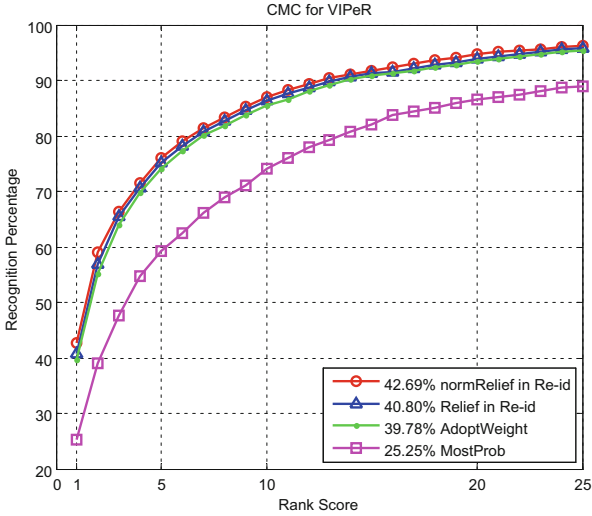


**Fig. 3.** The Effectiveness of Grey-world normalization and multiple metric learning in color, HOG and LBP features on VIPeR.



**Fig. 4.** The effectiveness of Grey-world normalization and multiple metric learning in color, HOG and LBP features on CUHK01.





**Fig. 5.** The effectiveness of Relief algorithm.

## 4 Conclusion

In this paper, we have proposed multiple metric learning in kernel space to address the problem of losing discriminative information in the process of calculating the kernel matrix. The experiments demonstrate the results of projecting features respectively can get better performance than that of mapping one high dimensional vector and the experiments also prove that our approach is effective and achieve better performance than some state-of-the-art methods. In the future, feature representation especially the color invariant feature will be deeply investigated.

**Acknowledgement.** This work was supported by the National Natural Science Foundation of China (91520301).

## References

1. Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. *Image Vis. Comput.* **32**(4), 270–286 (2014)
2. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: past, present and future. *J. Latex Class Files* **14**(8) (2015)
3. Chen, Y., Zhao, C., Wang, X., Gao, C.: Robust color invariant model for person re-identification. In: You, Z., Zhou, J., Wang, Y., Sun, Z., Shan, S., Zheng, W., Feng, J., Zhao, Q. (eds.) *CCBR 2016. LNCS*, vol. 9967, pp. 695–702. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46654-5\\_76](https://doi.org/10.1007/978-3-319-46654-5_76)
4. Lisanti, G., Masi, I., Del Bimbo, A.: Matching people across camera views using kernel canonical correlation analysis. In: *ICDSC*, pp. 1–6 (2014)

5. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., Li, S.Z.: Salient color name for person re-identification. In: ECCV, pp. 536–551 (2014)
6. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR, pp. 2288–2295 (2012)
7. Yang, Y., Liao, S., Lei, Z., Li, S.Z.: Large scale similarity learning using similar pairs for person verification. In: AAAI, pp. 3655–3661 (2016)
8. Qi, M., Tan, S., Wang, Y., Liu, H., Jiang, J.: Multi-feature subspace and kernel learning for person re-identification. *Acta Automatica Sinica* **42**(2), 299–308 (2015)
9. Syed, M.A., Jiao, J.: Multi-kernel metric learning for person re-identification. In: IEEE International Conference on Image Processing, pp. 784–788 (2016)
10. Martinel, N., Micheloni, C., Foresti, G.L.: Kernelized saliency-based person re-identification through multiple metric learning. *IEEE Trans. Image Process.* **24**(12), 5645–5658 (2015)
11. Kira, K., Rendell, L.A.: The feature selection problem: traditional methods and a new algorithm. In: AAAI, pp. 129–134 (1992)
12. Satta, R.: Appearance descriptors for person re-identification: a comprehensive review. Eprint Arxiv (2013)
13. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: Proceedings of PETS Workshops (2007)
14. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: Proceedings of ACCV, pp. 31–44 (2012)
15. Chen, X., Huang, K., Tan, T.: Object tracking across non-overlapping views by learning inter-camera transfer models. *Pattern Recogn.* **47**(3), 1126–1137 (2014)
16. Liao, S., Hu, Y., Zhu, X., et al.: Person re-identification by local maximal occurrence re-presentation and metric Learning. *Comput. Vis. Pattern Recognit.* **8**(4), 2197–2206 (2015)
17. Tao, D., Guo, Y., Song, M., et al.: Person re-identification by dual-regularized KISS metric learning. *IEEE Trans. Image Process.* **25**(6), 1–1 (2016)
18. Xiong, F., Gou, M., Camps, O., et al.: Person re-identification using kernel-based metric learning methods. *ECCV* **2014**, 1–16 (2014)
19. Zeng, M., Wu, Z., Tian, C., et al.: Efficient person re-identification by hybrid spatiogram and covariance descriptor. In: CVPR 2015, pp. 48–56 (2015)
20. Liu, H., Ma, L., Wang, C.: Body-structure based feature representation for person re-identification. In: IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, pp. 1389–1393 (2015)
21. Zhao, R., Ouyang, W., Wang, X.: Person re-identification by salience matching. In: IEEE International Conference on Computer Vision, IEEE, pp. 2528–2535 (2013)
22. An, L., Kafai, M., Yang, S., et al.: Reference-based person re-identification. In: IEEE International Conference on Advanced Video and Signal Based Surveillance, IEEE, pp. 244–249 (2013)
23. Ma, L., Yang, X., Tao, D.: Person re-identification over camera networks using multi-task distance metric learning. *IEEE Trans. Image Process.* **23**(8), 3656–70 (2014). A Publication of the IEEE Signal Processing Society
24. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(3), 653 (2013)
25. 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). [https://doi.org/10.1007/978-3-642-37331-2\\_3](https://doi.org/10.1007/978-3-642-37331-2_3)

26. Bazzani, L., Cristani, M., Murino, V.: Symmetry-driven accumulation of local features for human characterization and re-identification. *Comput. Vis. Image Underst.* **117**(2), 130–144 (2013)
27. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008*. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-88682-2\\_21](https://doi.org/10.1007/978-3-540-88682-2_21)
28. Ma, B., Su, Y., Jurie, F.: Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image Vis. Comput.* **32**(6), 379–390 (2014)
29. Prosser, B., Zheng, W.-S., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: *BMVC* (2010)
30. Li, W., Wang, X.: Locally aligned feature transforms across views. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2013)