Person Re-identification by Unsupervised Color Spatial Pyramid Matching

Yan Huang[1,*], Hao Sheng[1], Yang Liu[1], Yanwei Zheng[1], and Zhang Xiong[1]

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

Abstract. In this paper, we propose a novel unsupervised color spatial pyramid matching (UCSPM) approach for person re-identification. It is well motivated by our study on spatial pyramid to build effective structural object representation for person re-identification. Through the combination of illumination invariance color feature, UCSPM can well cope with the variations of viewpoint, illumination and pose. First, local superpixel regions are divided to accurately represent the color feature. Second, human body are divided into increasing fine vertical sub-regions to construct the spatial pyramid matching scheme. Third, the color feature and its spatial distribution information are used in a pyramid match kernel for calculating the similarity between person and person. The effectiveness of our approach is validated on the VIPeR dataset and CUHK campus dataset. Comparing with other approaches, our UCSPM improves the best unsupervised rank-1 matching rate on the VIPeR dataset by 3.08% with only one kind of feature—color.

Keywords: person re-identification, color spatial pyramid, structural object representation, unsupervised, cross-camera

1 Introduction

The problem of person re-identification (Re-ID) requires the ability to re-identify an individual across multiple disjoint camera views, is becoming one of the most challenging tasks in computer vision [1],[2],[3],[4],[5]. It is also important in the field of video surveillance by searching a person from large amounts of video sequences as accurately as possible. In recent years, the methods of Re-ID are dominated by supervised learning that aim to learn an optimal metric or distance function [6],[7],[3]. These works usually require identity label of person that must be annotated manually for each pair of camera views, as training data. By employing supervised models, discriminative features are extracted to cope with the appearance variations of the same person under cross-view cameras. However, since video surveillance system can capture hundreds of pedestrians for a while, some of them may have a similar appearance. In addition, the same person observed in different camera views often under significant variations in viewpoint, illumination, pose, background, etc. Therefore, the training data should be as sufficient and as diverse as possible to enhance the generalization ability of training model that can implicitly discover the visual features of intra-class variations. To this end, it needs lots of manual annotation and sample selection. That is difficult to implement in a large scale video surveillance system.

Another widely used method in person Re-ID is unsupervised learning that aim to explore intuitive feature of human appearance and match directly by distance function (Euclidean distance, Mahalanobis distance, Gaussian distance, etc). The unsupervised methods are much better adaptability to different camera pair setting, although it may sacrifice matching accuracy. To seek more stable feature that can cope with the intra-class variations, various visual techniques are presented in previous works [1],[8],[9],[5]. Among these techniques, color features extraction of person image, a simple but efficient and important technique is used in person Re-ID, are commonly employed to construct human representation. Beyond that, the spatial layout information of features is also important to confine the feature distribution. Building effective structural object representation can effectively presents the spatial layout information, especially when the same person under disjoint camera views in the presence of large viewpoint or pose variations. Therefore, [10] introduces a kernel based recognition method that works by computing rough geometric correspondence on a global scale using an efficient approximation technique adapted from the spatial pyramid matching scheme of Grauman and Darrell [11]. In this method, the global non-invariant representation based on aggregating statistics of local features over fixed subregions significantly improves the recognize performance over methods based on detailed geometric correspondence.

Inspired from the pyramid matching scheme, this paper proposes a new approach based on unsupervised color spatial pyramid matching (UCSPM). Although we pursue spatial pyramid matching, this work is different from previous attempts. The similarity of local features is computed at increasingly fine vertical stripes of sub-region as shown in Fig.1. Spatial layout information aided by spatial pyramid is used to confine feature distribution and match between images. Beyond that, in order to control variations and misalignment between images, Hungarian algorithm and mean distance vector are used in the pyramid match kernel [10]. The color feature used in this work with the property of illumination invariance [12]. That is particularly applicable to camera views under outdoor areas. Furthermore, color features are usually extracted from the entire image, human body (without background) or local patches [13],[9],[5]. In general, these areas including a variety of colors can hardly represent the color property of human. Therefore, in this paper, superpixel-based segmentation technique is used to subdivide person images into local superpixel regions. In comparison with local patch method, superpixel method shows a significant improvement in our experiments.

The contributions of this paper can be summarized in two-folds. First, an unsupervised color spatial pyramid scheme is proposed to build effective structural object representation for person Re-ID. Second, superpixel-based color feature is presented with the combination of spatial pyramid framework that achieves competitive performance. Although we only use the color information in our pyramid framework, the comparative evaluations on two public datasets (VIPeR [14] and CUHK campus [15]) demonstrate that our method outperforms not only existing unsupervised learning methods, but also quite a bit of supervised learning methods.

2 Related Work

Many supervised learning methods have been put forward for person Re-ID [2],[6],[7],[3],[16],[17]. Prosser et al. [2] use ensemble RankSVMs to learn pairwise similarity that formulate person Re-ID as a ranking problem. A Mahalanobis distance learning method that is optimal for k-nearest neighbour classification using a maximum margin formulation is proposed by Dikmen et al. [6]. In [7], a relaxed pairwise metric learning is presented which takes advantages of the structure of the data with reduced computational cost, and achieves the state-of-the-art with simple feature descriptors. Another idea is trying to solve metric learning in a probabilistic manner used in person Re-ID. Zheng et al. [3] focuses on maximizing the probability that a true match pair has a smaller distance than a false matched pair. In addition to these metric learning, transfer learning based methods are also popular for person Re-ID. Zheng et al [16] reformulate person Re-ID problem as verification task and show that discriminant information can be learnt from unlabelled data. A transfer RankSVM to adapt a model trained on the source domain to target domain is proposed in [17].

For supervised learning method, their performance is limited by the fact that it is based on the subtraction of misaligned feature vectors, which can cause significant information loss. In the case of existing methods, the research on unsupervised learning methods are another important branch of person Re-ID. Many effective researches of features are proposed in unsupervised methods. Farenzena et al. [1] exploit the symmetry property in person image and propose the symmetry driven accumulation of local featrues. A combination of biologically inspired features and covariance descriptors that handle both background and illumination variations is proposed in [8]. Considering certain features play more important role than others, Liu et al. [4] employ a feature mining framework to optimise the weights of global features. Since color, shape and texture features can capture different aspects of information contained in image, Ma et al. [9] extract 7-d features of each pixel based on color, position and gradient and represent them by Gaussian models. Inspired by human eye that recognize person identities based on salient regions, Zhao et al. [5] proposed a patch based feature to learn salient regions in human appearance. In this paper, an unsupervised color spatial pyramid scheme is given to build effective structural object representation, it achieves competitive performance with only one kind of feature—color.

3 Color Spatial Pyramid Matching

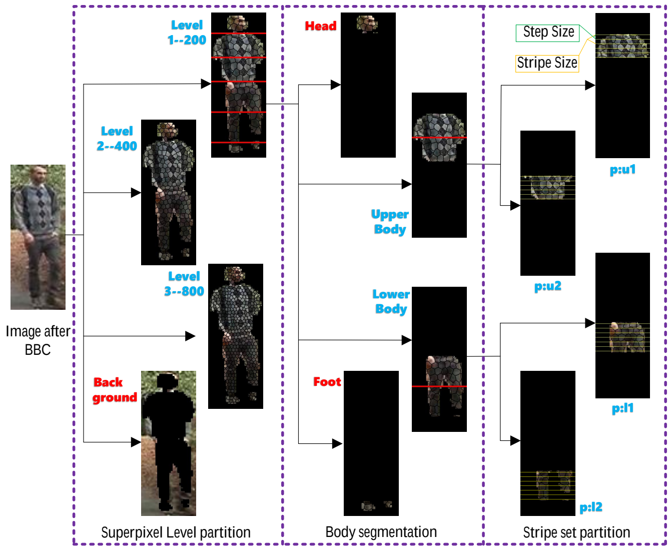


Fig. 1. Image in the leftmost column is processed by Bayesian Color Constancy method. After that, person image is divided into superpixel regions with increasing granularity from 200 to 800 (entire image, from coarse to fine). Aided by the Deep Decompositional Network, the background is discarded and the body is divided into four part. Considering the discriminative power, only upperbody and lowerbody are retained and divided into two vertical part (u1, u2, l1, l2) respectively. Stripes are divided in each part of vertical body regions.

3.1 Data Pre-processing

In the first stage of UCSPM, color constancy and pedestrian parsing methods are used to depress the influence caused by the variation of illumination and discard background clusters, respectively.

Specifically, Bayesian Color Constancy (BCC) [18] method is used to depress the influence of illumination for each image. BCC can perceive surface color consistently, despite variations in ambient illumination. Compared to Grey World Color Constancy method, used in the pre-processing step in person Re-ID [9], the performance of BCC outperforms it that proven by [18]. After depressing the influence of illumination, Deep Decompositional Network (DDN) [19] is used to parse pedestrian into five regions, and discard background. DDN is able to accurately estimate contour of body and parse it into semantic regions such as hair, head, body, arms and legs, with robustness to occlusions and background clutters.

3.2 Spatial Pyramid Matching

Given two sets of feature vectors, spatial pyramid matching is proposed to find an approximate correspondence between them. It works by placing a sequence of increasingly coarser grids over the feature space and taking a weighted sum of the number of matches that occur at each level of resolution [10]. In previous attempts, two points match if they fall into the same cell of the grid at any fixed resolution and high weight is given in finer resolutions than coarser ones. For these reasons, spatial pyramid can effective confine the features distribution and keep the spatial layout information for matching. In our work, we construct a sequence of level $1, ..., L$, each level corresponding to different vertical stripe size and superpixel granularity as shown in Fig.1.

Let $F_l^{A,U_p} = \left\{ f_{l_{x,y}}^{A,U_p} \right\}$ denotes the feature set in one stripe, and $f_{l_{x,y}}^{A,U_p}$ represents a d-dimensional feature of superpixel (detial is shown in section 3.3) at l-th stripe in p-th part ($p = u1, u2, l1$ or $l2$, shown in Fig.1) of person U from camera A, x and y represent the centroid of superpixel that should satisfy $(x, y) \in l$. We simple use the City Block distance to measure the similarity between $f_{l_{x,y}}^{A,U_p}$ and its corresponding superpixel in person V from camera B:

$$D(f_{l_{x,y}}^{A,U_p}, f_{l'(k)_{x',y'}}^{B,V_p}) = \sum_{i=1}^{d} \left| f_{l_{x,y}}^{A,U_p}(i) - f_{l'(k)_{x',y'}}^{B,V_p}(i) \right|, \tag{1}$$

where i is i-th element in feature vector, $l^{'}(k)$ represents k-th adjacency constrained search stripe corresponding to l:

$$k = max(0, l - \delta), ..., min(N_{sripe}^{p}, l + \delta), \tag{2}$$

N_{stripe}^{p} is the number of stripes in part p. If person in image do not exist vertical pose variation, $\delta = 0$. However, we set $\delta = 1$ in our experiment to tolerate the vertical spatial variation. In the following, we use a four-tuple e.g. (A, U, p, l) represents a stripe l in p-th part of person U from camera A.

Due to the pose variation between different images, matching superpixel directly can easily lead to that more than one regions are most similar to a single one between different images. Therefore, we specify matching in units of stripes instead of superpixel regions. In order to find the best matching stripes in adjacency search areas by Equ.2, the City Block distance (Equ.1) is calculated between any two superpixels from one stripe in (A, U, p, l) to another stripe in $(B, V, p, l^{'}(k))$.

$$\boldsymbol{M}^p = [D(F_l^{A,U_p}(m), F_{l'(k)}^{B,V_p}(n))]_{M \times N}, m = 1, ..., M, n = 1, ..., N, \tag{3}$$

where m(n) represent m-th(n-th) superpixel in stripe $l(l^{'}(k))$. The elements in $\boldsymbol{M}^p$ are the similarity between any two superpixels from l and $l^{'}(k)$ measured by Equ.1. Hungarian algorithm is used in $\boldsymbol{M}^p$ to find the best superpixel assignment $D_H^p(i), i = 1, ..., min(m, n)$ (i represents the i-th matching pair in $\boldsymbol{M}^p$, D is City Block distance) and cope with the case of multi-to-one matching between two stripes. Therefore, the similarity between two stripes can be quantized as a mean distance vector:

$$\boldsymbol{D}^p_{l,l'(k)} = \overline{\sum_{i=1}^{min(m,n)} (D^p_H(i))}, \tag{4}$$

At last, the similarity between one stripe in image U and its adjacency stripe in image V is:

$$\boldsymbol{D}^p_{l,l'} = min(\boldsymbol{D}^p_{l,l'(k)}), \tag{5}$$

where k is given by Equ.2.

The similarity between two person images U and V can be obtained in different body parts by the stripe similarity. Considering the discriminative power of different body parts, more higher weight is given to upperbody parts u1 and u2 than lowerbody parts l1 and l2:

$$\boldsymbol{D}(U,V) = \sum_{i=1}^{n1} \boldsymbol{D}^{u1}_{l(i),l'} + \sum_{j=1}^{n2} \boldsymbol{D}^{u2}_{l(j),l'} + 0.5 \cdot (\sum_{x=1}^{n3} \boldsymbol{D}^{l1}_{l(x),l'} + \sum_{y=1}^{n4} \boldsymbol{D}^{l2}_{l(y),l'}), \tag{6}$$

where $n1$, $n2$, $n3$ and $n4$ are the number of stripes in different body parts that have corresponding stripe in adjacency constrained search area by Equ.5. If one image U matching with multiple images V, there are uncertain number of $n1$, $n2$, $n3$ and $n4$ between one image U and different image V. For fair comparison, the minimum $n1$, $n2$, $n3$ and $n4$ between U and any one of V are selected for calculating the similarity between U and multiple V among different body part. Further, if any one of the $n1$, $n2$, $n3$ and $n4$ larger than the minimum value in its own body part when U is matching with one V, the first minimum $n1$, $n2$, $n3$ or $n4$ similarity distances will be selected for matching. In Equ.6, we simply consider that the discriminative power of upperbody is almost twice as lowerbody, so the weight is set to 0.5. The weight used in our experiment was somewhat haphazard, so it is likely that better weight may still be found by using a validation set.

For all the level $1, ..., L$, we want to penalize matches found in stripe with larger size and superpixel with coarser granularity, sine they involve increasingly dissimilar features. Putting all the level together, we get the following definition of a pyramid match kernel:

$$k^L(U,V) = \sum_{\iota=1}^{L} \frac{1}{2^{L-\iota+1}} \boldsymbol{D}^\iota(U,V), \tag{7}$$

where $\boldsymbol{D}^\iota(U,V)$ is defined in Equ.6.

3.3 Color Feature Extraction

The color feature is widely used in person Re-ID. However, due to the variation of illumination, it is hard to remain reliable and even vary significantly. As far as the present state of study, a wide range of color features have been proposed in [12],[20],[21]. Considering the illumination invariance property, two color features: hue histogram and opponent histogram from Van de Weijer and Schmid [12] are briefly reviewed in this paper. These two features are extracted for each superpixel and their performance will be given in section 4.

According to Van de Weijer and Schmid [12], the computation of hue with small value of saturation will bring uncertainties. Hence, it should be counted less in histogram. In the computation of hue histogram, hue is weighted by its saturation in [12]. Hue and saturation are computed as follows:

$$hue = \arctan(\frac{O_1}{O_2}) = \arctan(\frac{\sqrt{3}(R-G)}{R+G-2B}), \tag{8}$$

$$saturation = \sqrt{O_1^2 + O_2^2} = \sqrt{\frac{2}{3}(R^2 + G^2 + B^2 - RG - RB - GB)}, \tag{9}$$

where O_1 and O_2 are both from opponent color space:

$$O_1 = \frac{1}{\sqrt{2}}(R-G), O_2 = \frac{1}{\sqrt{6}}(R+G-2B), \tag{10}$$

In opponent color space, the opponent angle ang_x^O is supposed to be specular invariant in [12]. The ang_x^O is defined as:

$$ang_x^O = \arctan(\frac{O_{1x}}{O_{2x}}), \tag{11}$$

where O_{1x} and O_{2x} are the first order derivative of O_1 and O_2 respectively. Van de Weijer and Schmid [12] define ang_x^O as the weight for the opponent angle as an error analysis to the opponent angle:

$$\partial ang_x^O = \frac{1}{\sqrt{O_{1x}^2 + O_{2x}^2}}, \tag{12}$$

Finally, both of hue and opponent histograms are quantized to 36 bins.

A 72-dimensional color feature is extracted from each superpixel. The color features obtained in different body parts $(u1, u2, l1, l2)$ of all person are grouped together respectively to calculate the color dictionaries of local appearances by k-mean clustering. Four color dictionaries corresponding to $u1, u2, l1$ and $l2$ can be obtained, each dictionary includes the centers of cluster based on the number of clusters specified in advance. The feature used in Equ.1 is calculated by squared Euclidean distance between the 72-dimensional color histogram feature and its corresponding color dictionary in different body parts.

4 Experiments

We evaluate our unsupervised color spatial pyramid matching approach on two public datasets: the VIPeR dataset [14] and the CUHK campus dataset [15]. The VIPeR dataset is the most widely used person Re-ID dataset for evaluation.

The CUHK campus dataset contains more images than VIPeR. Both datasets are very challenging since they under significant variations in viewpoint, illumination, pose, and background. The quantitative results are presented in standard Cumulated Matching Characteristics (CMC) curves [14]. The rank-k in CMC indicates the percentage of the correct matches found in the top k ranks from probe images to gallery images.

For fair comparison, the same experiment setup mentioned in [13] is used which divides the dataset into two parts, 50% for training and 50% for testing, without overlap on person identities. Images captured from camera A are as probe and camera B as gallery. Each probe image is matched with every gallery image. We conduct 10 trials of evaluation in experiment. Tab.1 gives settings for different pyramid levels. It is worth noting that the superpixel obtained from the entire image, but we only use the ones on body region. In our experiment, the superpixel regions are divided using the method in [22]. In addition, to validate the usefulness of superpixel based color spatial pyramid, we construct a set of contrast experiments based on patch. Fig.2 shows a comparison between superpixel and patch. The feature extracted from each patch and the comparison results are given on both datasets. Specifically, these patches are divided in each stripe. The patch size and step size in horizontal direction are the same as stripe size and its step size as shown in Tab.1. Notice that the experiment steps are the same as in the contrast experiment except for the regions of color feature extraction. For different levels, we denote the patch and superpixel based approaches by *patchLevel1-3* and *spLevel1-3* respectively. Using the spatial pyramid, the patch and superpixel based approaches are denoted by *patchPyramid* and *spPyramid(UCSPM)*, respectively.

Pyramid Level (ι)	Superpixel Granularity (regions)	Stripe Size (pixels)	Stripe Step (pixels)	Number of Clusters
1	200	16	8	100
2	400	24	12	200
3	800	32	16	400

Table 1. Different pyramid level corresponding to different settings. The ι is used in Equ.7. Superpixel granularity, stripe size and stripe step can be found in Fig.1. The number of clusters is defined in section 3.3.

VIPeR Dataset. The VIPeR dataset contains 632 person image pairs captured in different camera views. Most pairs show significant variations of viewpoint, illumination and pose as shown in Fig.3(a). In our experiment, all images are normalized to 512×192. Fig 4(a) shows the comparison between different levels and their combinations by spatial pyramid approach. Since superpixel can better locate color regions than patch, much more better performance is achieved in different levels. Using the spatial pyramid match kernel, the performance of *spPyramid* is better than *patchPyramid*, which indicates superpixel is

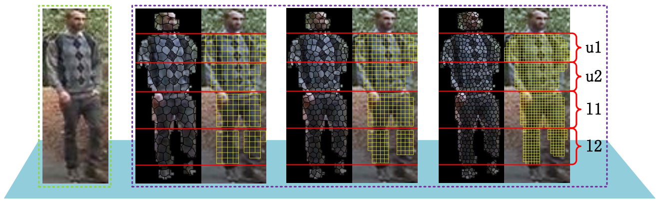


Fig. 2. Dividing person into superpixel regions and patches corresponding to different pyramid levels

(a) VIPeR dataset (b) CUHK campus dataset

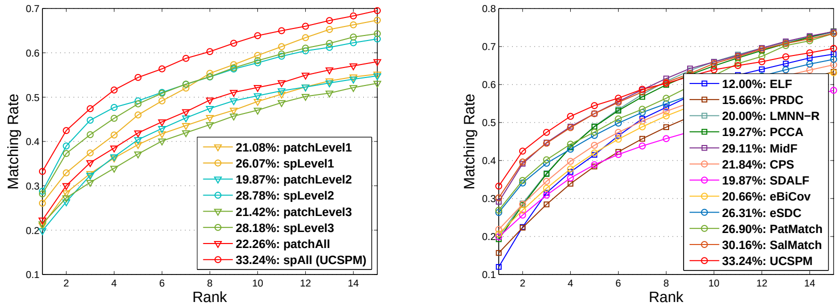
Fig. 3. Sets of 16 image pairs from the VIPeR and CUHK campus datasets

much effective than patch in color feature extraction in our color spatial pyramid matching approach.

We also compare *UCSPM* with several unsupervised approaches including CPS [23], SDALF [1], eBiCov [8], eSDC [5], PatMatch [24] and SalMatch [24], and five supervised learning approaches including ELF [13], PRDC [3], LMNN-R [6], PCCA [25] and MidF [26]. Our *UCSPM* achieves 33.24% at rank-1 and outperforms all these approaches. In addition, *UCSPM* outperforms the state-of-the-art unsupervised approach SalMatch [24] by 3.08%. What's more, our approach only use color and spatial information without the benefit of any salience information on human appearance proposed in [24].

CUHK campus Dataset. The CUHK campus dataset is also a very challenging dataset. It contains 971 persons, each person has two images in each camera view. Specially, camera A captures the frontal or back view of person and camera B is the side view. Different from VIPeR dataset, most persons in CUHK dataset have their salience regions i.e. color bags, clothes or shoes etc (as shown in Fig.3(b)). All the images are normalized to 512×192 in our experiment.

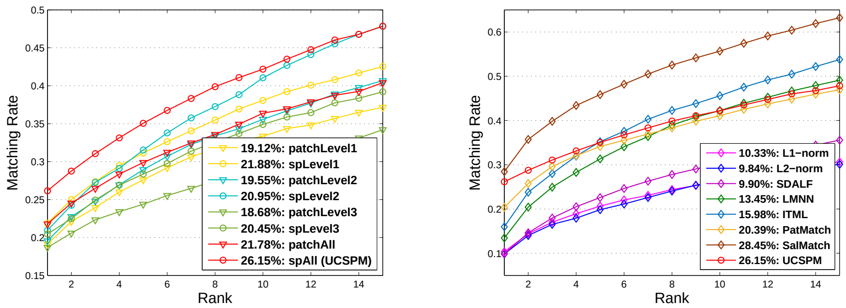
Besides comparison between patch and superpixel-based pyramid approaches Fig.5(a). We also compare our approach with available results including L1-norm [24], L2-norm [24], LMNN [15], ITML [15], SDALF [1], GenericMetric [15], PatMatch [24] and SalMatch [24] as shown in Fig.5(b). Since SalMatch [24] focuses on unsupervised salience matching and the CUHK campus dataset are



(a) Superpixel vs. patch (b) Ours vs. state-of-the-art

Fig. 4. CMC on the VIPeR dataset. Rank-1 matching rate is marked before the name of each approach.

more suitable to show the effectiveness of salience matching. The accuracy of our approach is slightly lower than SalMatch [24] by 2.3%. However, comparison with other approaches, our method shows effectiveness on CUHK campus dataset.



(a) Superpixel vs. patch (b) Ours vs. state-of-the-art

Fig. 5. CMC on the CUHK campus dataset. Rank-1 matching rate is marked before the name of each approach.

5 Conclusions

In this paper, we propose a color spatial pyramid approach for person Re-ID. We explore spatial pyramid to build effective structural object color representation and cope with large viewpoint or pose variations. We compute the similarity of local features at increasingly fine vertical stripes and use pyramid match kernel for matching. To depress the variation of illumination, Bayesian Color Constancy method and illumination invariance color feature are used for stable color feature extraction. To accurately represent color features, we use superpixel-based segmentation technique to subdivide person into local regions. Experimental results show our color spatial pyramid approach improves the performance of unsupervised person re-identification.

Acknowledgement. This study was partially supported by the National Natural Science Foundation of China (No.61272350), the National High Technology Research and Development Program of China (No.2013AA01A603) and the National Aerospace Science Foundation of China (No.2013ZC51). Supported by the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under grant #SKLSDE-2015ZX-21.

References

1. Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Conference on Computer Vision and Pattern*, 2010, pp. 2360–2367.
2. Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Person re-identification by support vector ranking," in *British Machine Vision Conference*, 2010, pp. 1–11.
3. Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Person re-identification by probabilistic relative distance comparison," in *Conference on Computer Vision and Pattern Recognition*, 2011, pp. 649–656.
4. Chunxiao Liu, Shaogang Gong, and Chen Change Loy, "On-the-fly feature importance mining for person re-identification," *Pattern Recognition*, vol. 47, no. 4, pp. 1602–1615, 2014.
5. Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Unsupervised salience learning for person re-identification," in *Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3586–3593.
6. Mert Dikmen, Emre Akbas, Thomas S. Huang, and Narendra Ahuja, "Pedestrian recognition with a learned metric," in *Asian Conference on Computer Vision*, 2010, pp. 501–512.
7. Martin Hirzer, Peter M. Roth, Martin Köstinger, and Horst Bischof, "Relaxed pairwise learned metric for person re-identification," in *European Conference on Computer Vision*, 2012, pp. 780–793.
8. Bingpeng Ma, Yu Su, and Frédéric Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *British Machine Vision Conference*, 2012, pp. 1–11.
9. Bingpeng Ma, Qian Li, and Hong Chang, "Gaussian descriptor based on local features for person re-identification," in *ACCV*, 2014, pp. 505–518.
10. Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2169–2178.
11. Kristen Grauman and Trevor Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2005, vol. 2, pp. 1458–1465.

12. Joost van de Weijer and Cordelia Schmid, “Coloring local feature extraction,” in *European Conference on Computer Vision*, 2006, pp. 334–348.
13. Douglas Gray and Hai Tao, “Viewpoint invariant pedestrian recognition with an ensemble of localized features,” in *European Conference on Computer Vision*, 2008, pp. 262–275.
14. Douglas Gray, Shane Brennan, and Hai Tao, “Evaluating appearance models for recognition, reacquisition, and tracking,” in *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007, vol. 3.
15. Wei Li, Rui Zhao, and Xiaogang Wang, “Human reidentification with transferred metric learning,” in *Asian Conference on Computer Vision*, 2012, pp. 31–44.
16. Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, “Transfer re-identification: From person to set-based verification,” in *Computer Vision and Pattern Recognition*, 2012, pp. 2650–2657.
17. Andy Jinhua Ma, Pong C. Yuen, and Jiawei Li, “Domain transfer support vector ranking for person re-identification without target camera label information,” in *International Conference on Computer Vision*, 2013, pp. 3567–3574.
18. Peter V. Gehler, Carsten Rother, Andrew Blake, Thomas P. Minka, and Toby Sharp, “Bayesian color constancy revisited,” in *Conference on Computer Vision and Pattern*, 2008.
19. Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang, “Deepreid: Deep filter pairing neural network for person re-identification,” in *Conference on Computer Vision and Pattern Recognition*, 2014, pp. 152–159.
20. Koen E. A. van de Sande, Theo Gevers, and Cees G. M. Snoek, “Evaluating color descriptors for object and scene recognition,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010.
21. Gertjan J. Burghouts and Jan-Mark Geusebroek, “Performance evaluation of local colour invariants,” *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 48–62, 2009.
22. Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurélien Lucchi, Pascal Fua, and Sabine Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.
23. Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino, “Custom pictorial structures for re-identification,” in *British Machine Vision Conference*, 2011, pp. 1–11.
24. Rui Zhao, Wanli Ouyang, and Xiaogang Wang, “Person re-identification by salience matching,” in *International Conference on Computer Vision*, 2013, pp. 2528–2535.
25. Alexis Mignon and Frédéric Jurie, “PCCA: A new approach for distance learning from sparse pairwise constraints,” in *Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2666–2672.
26. Rui Zhao, Wanli Ouyang, and Xiaogang Wang, “Learning mid-level filters for person re-identification,” in *Conference on Computer Vision and Pattern Recognition*, 2014, pp. 144–151.