Resolution Invariant Person Re-identification

Kai Li

Department of Electrical & Computer Engineering,
Northeastern University
kaili@ece.neu.edu

Abstract. In real world visual surveillance system, a person of interest can be captured in very different resolutions in different camera views. The resolution variances of human images cast additional difficulties on matching images of the same persons, besides the arbitrary of human poses and light conditions. Conventional person re-identification methods conduct image scaling and simple size normalization to couple the resolution discrepancy, which however may cause the loss of appearance information that are essential for matching the person. In this report, a new algorithm is introduced which matches human images of different resolutions by learning dictionaries directly from the original human images. Instead of learning only one pair of dictionaries from entire probe and gallery image sets, pairs of dictionaries are learned from homogeneous image patches to boost the feature representation power of the dictionaries. Latent high resolution human images corresponding to low resolution ones are incorporated into the learning framework to help better dictionary learning. Experiments show the proposed algorithm outperforms several existing algorithms.

Keywords: Person Re-identification, Dictionary Learning

1 Introduction

Person re-identification (Re-ID) is crucial for visual surveillance and has drawn increasing attentions in recent years [24]. One of the key factors complicating person Re-ID is that a same person may vary significantly in resolution in different camera views [9]. Fig. 1 shows an example of this situation. Traditional methods tackle this challenge simply by normalizing input images to a uniform frame, for the ease of feature description. However, many appearance details, which could be crucial for person Re-ID, are lost after the normalization.

In this report, a new algorithm is introduced which matches human images of different resolutions by learning dictionaries directly from the original human images. Instead of learning only one pair of dictionaries from entire probe and gallery image sets [18, 21], image patch cluster based dictionary learning is performed on pairs of homogeneous image patch clusters. The patch cluster specific dictionaries can therefore represent human image features better. Besides, latent high resolution human images corresponding to low resolution ones are incorporated into the learning framework to help better dictionary learning.

Fig. 1. The resolution of a persons images in two different camera views are significantly different. This figure is from [13].

2 Related Works

The proposed algorithm targets to solve person re-identification problem, and uses dictionary learning technique, so that it is related to this two categories of algorithms. This part is based on [12], with some adaptations.

2.1 Person Re-identification

Person re-identification has become a very hot topic in the computer vision community in recent years. Some traditional methods focus on learning effective metrics to measure the similarity between two images captured from different camera views [6, 25].Other research works focus on learning expressive features, which usually obtain better performance that the metric learning methods. They suggest that learning effective representations is the key in person re-identification. Some advanced features include salience features [22], mid-level features [23], and salient color features [20]. Although the existing feature learning methods achieve good performance, the cross-view relationships of pedestrian images havent been extensively studied. The proposed approach explicitly models such relationships in different image patch level, and draws strength from them to enhance the re-identification performance.

2.2 Dictionary Learning

As a powerful technique for learning expressive bases in sample space, dictionary learning has attracted lots of attention during the past decades. Some popular dictionary learning methods include K-SVD [1], and projective dictionary pair

learning [5]. Most recently, Liu et al. presented a semi-supervised coupled dictionary learning (SSCDL) method [16], and applied it to person re-identification. The major differences between our approach and SSCDL are two-folds. First, SSCDL is a semi-supervised method, while our approach is supervised. Secondly, SSCDL learns dictionaries from entire gallery and probe human image sets, while the proposed algorithm performs patch cluster based dictionary learning.

3 Method

3.1 Algorithm Overview

Illustration of the framework of the proposed algorithm is shown in Fig. 2. Given N pairs of pedestrian images $\mathcal{T} = \{(G_i, P_i)\}_{i=1}^{N}$, where (G_i, P_i) is the i-th pair of images for the same person captured by camera Cam-G and camera Cam-P. Every images captured by Cam-G $\mathcal{H}_G = \{G_i\}_{i=1}^{N}$ are first down-sized to be the same sizes as their correspondences in $\mathcal{H}_P = \{P_i\}_{i=1}^{N}$ captured by Cam-P, generating a new set of images $\mathcal{H}_G^l = \{G_i^l\}_{i=1}^{N}$. In other words, G_i^l is adjusted from G_i to be the same size as P_i.

Next, images in $\mathcal{H}_G$ and $\mathcal{H}_P$ are divided into patches and patches similar in appearance are clustered together. To cope resolution variations, all images are divided into the same number of equal-sized patches. (Patch sizes keep the same in each image, but vary among different images. The number of patches is the same for all images.) Images in $\mathcal{P}_A$ are divided and clustered based on the clustering result of images in $\mathcal{H}_A^l$, i.e., if two patches of images in $\mathcal{H}_G$ are clustered into a cluster, their corresponding larger-sized patches of images in $\mathcal{H}_G$ are also grouped into a cluster. It is reasonable to assume that image patches in the same cluster share the same dictionary.

In each set of corresponding image patch clusters, latent factor assisted dictionary learning (to be introduced in the next section) is performed. The outcomes after conducting dictionary learning on all sets of corresponding image patch clusters are a set of dictionaries $\hat{\mathcal{D}}_L = \{D_L^i\}_{i=1}^{n}$ for low-resolution image, and a set of dictionaries $\hat{\mathcal{D}}_H = \{D_H^i\}_{i=1}^{n}$ for high-resolution image.

In the testing stage, given a low resolution probe image, P_b, it is firstly divided into n patches. Denote by y_i the feature of i-th patch, we can select best dictionary in $\hat{\mathcal{D}}_L$ that minimizes

$$\min_j \|y_i - D_L^j Z_j\|_F^2. \tag{1}$$

Z_j is then used as the feature representation for y_i. In this way, all patches for P_b can be represented under some dictionaries in $\hat{\mathcal{D}}_L$. All the new feature representations of patches of P_b under $\hat{\mathcal{D}}_L$ are concatenated together to form the new description of P_b, suppose it is F_b. For every pedestrian image in the gallery dataset, a new representation can be generated under $\hat{\mathcal{D}}_H$. Suppose the new representation for the i-th gallery image $\hat{\mathcal{D}}_H$ is F_g^i, the distance between F_g^i and F_b can be used to measure the similarity between P_b and the i-th gallery image. In this way, retrieving the some person shown in F_b can be accomplished.

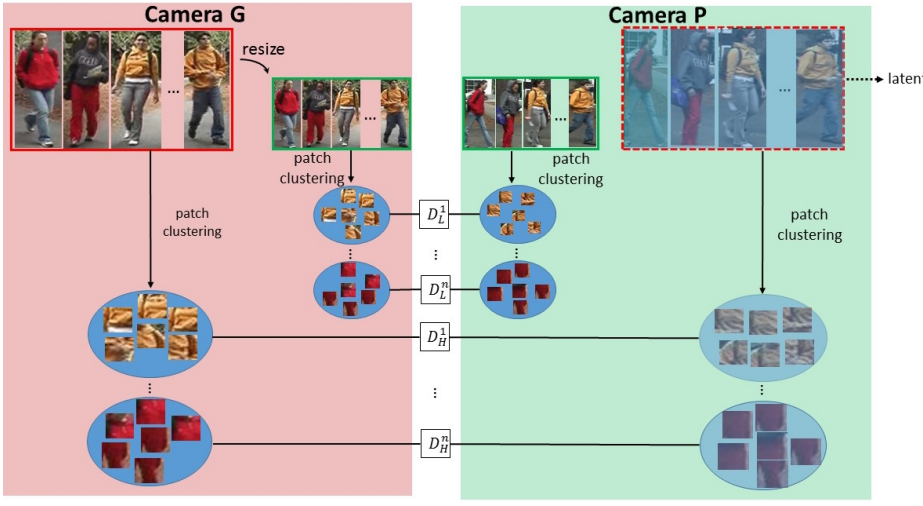

Fig. 2. Illustration of the framework of proposed person Re-ID algorithm.

3.2 Latent Factor Assisted Dictionary Learning

For each set of corresponding image patch clusters, suppose X_{H} and X_{L} denote the high and low resolution image patch features, respectively. The following model is proposed to learn discriminative dictionaries D_{H} and D_{L} from them:

$$\begin{aligned} \min_{A,B,D_{\mathrm{H}},D_{\mathrm{L}},Z_{\mathrm{H}},Z_{\mathrm{L}}} & \|A\|_* + \|B\|_* + \alpha\|X_{\mathrm{L}} - D_{\mathrm{L}}Z_{\mathrm{L}}\|_F^2 + \alpha\|X_{\mathrm{H}} - D_{\mathrm{H}}Z_{\mathrm{H}}\|_F^2 \\ & +\beta\|Z_{\mathrm{L}}\|_1 + \beta\|Z_{\mathrm{H}}\|_1 \\ \text{s.t.} \quad & Z_{\mathrm{L}} = Z_{\mathrm{H}}A + BZ_{\mathrm{L}}, \end{aligned} \tag{2}$$

where Z_{L} and Z_{H} are the new representations of X_{L} and X_{H} under the dictionaries D_{L} and D_{H}, respectively. $Z_{\mathrm{L}} = Z_{\mathrm{H}}A + BZ_{\mathrm{L}}$ is the latent low rank constraint, which requires that the new presentation of low resolution feature Z_{L} can be represented by the new representations of high resolutions feature Z_{H} in a low rank fashion (the nuclear norm of the coefficient matrix $\|A\|_*$ should be minimized). The component BZ_{L} represents the latent factor. A derivation how this component comes and work can be found in [3]. $\|Z_{\mathrm{L}}\|_1$ and $\|Z_{\mathrm{H}}\|_1$ the sparse terms requiring the new representations of original features should be sparse.

3.3 Solving the Optimization Problem

Problem (2) could be solved by off-the-shelf algorithms, e.g., Augmented Lagrange Methods (ALM) [15, 14]. However, extra relax variables in ALM lead to complex matrix operations, e.g., inverse, multiplications, in each iteration. This is essentially caused by the quadratic term in the augmented Lagrangian

function, which includes linear mappings of the target variables. To reduce the computation cost of this part, we propose to use the first order Taylor expansion like approximation to replace the original quadratic term, leading to a simpler solution to the original problem. To make it clear, we first write down the augmented Lagrangian function of problem (2):

$$\begin{aligned}&\min_{D_{\mathrm{H}},D_{\mathrm{L}}} \|A\|_* + \|B\|_* + \alpha\|X_{\mathrm{L}} - D_{\mathrm{L}}Z_{\mathrm{L}}\|_F^2 + \alpha\|X_{\mathrm{H}} - D_{\mathrm{H}}Z_{\mathrm{H}}\|_F^2 \\ &+\beta\|Z_{\mathrm{L}}\|_1 + \beta\|Z_{\mathrm{H}}\|_1 + \langle Y, Z_{\mathrm{L}} - Z_{\mathrm{H}}A - BZ_{\mathrm{L}}\rangle + \frac{\mu}{2}(\|Z_{\mathrm{L}} - Z_{\mathrm{H}}A - BZ_{\mathrm{L}}\|_{\mathrm{F}}^2),\end{aligned} \tag{3}$$

where Y is the Lagrange multiplier and $\mu > 0$ is a penalty parameter. $\langle , \rangle$ is the inner product of matrices and $\langle A, B\rangle = \mathrm{tr}(A^{\mathrm{T}}B)$. We then merge the last two terms into quadratic terms, and formulate it as:

$$\begin{aligned}&\|A\|_* + \|B\|_* + \alpha\|X_{\mathrm{L}} - D_{\mathrm{L}}Z_{\mathrm{L}}\|_F^2 + \alpha\|X_{\mathrm{H}} - D_{\mathrm{H}}Z_{\mathrm{H}}\|_F^2 \\ &+\beta\|Z_{\mathrm{L}}\|_1 + \beta\|Z_{\mathrm{H}}\|_1 + h(A, B, Z_{\mathrm{L}}, Z_{\mathrm{H}}, Y, \mu) - \frac{1}{\mu}\|Y\|_{\mathrm{F}}^2,\end{aligned} \tag{4}$$

where $h(A, B, Z_{\mathrm{L}}, Z_{\mathrm{H}}, Y, \mu) = \frac{\mu}{2}(\|Z_{\mathrm{L}} - Z_{\mathrm{H}}A - BZ_{\mathrm{L}} + Y/\mu\|_{\mathrm{F}}^2)$. Like the conventional ALM, the new formulation is not jointly solvable over A, B, D_{H}, D_{L}, Z_{H}, Z_{L}, but solvable over each of them, by fixing the rest. Therefore, we solve each subproblem at a time, and approximate the quadratic term h with first order expansion at the current point, assuming others are constant. At iteration $t+1$ ($t \geq 0$), we have:

Updating A:

$$\begin{aligned}A^{(t+1)} &= \arg\min_A \|A\|_* + h(A, B, Z_{\mathrm{L}}, Z_{\mathrm{H}}, Y, \mu) \\ &= \arg\min_A \frac{1}{\eta_a\mu}\|A\|_* + \frac{1}{2}\|A - A^{(t)} + \nabla_A h\|_{\mathrm{F}}^2,\end{aligned} \tag{5}$$

where $\nabla_A h = \nabla_A h(A^{(t)}, B^{(t)}, Z_{\mathrm{L}}^{(t)}, Z_{\mathrm{H}}^{(t)}, Y^{(t)}, \mu) = -Z_{\mathrm{H}}^{(t),T}(Z_{\mathrm{L}}^{(t)} - Z_{\mathrm{H}}^{(t)}A^{(t)} - B^{(t)}Z_{\mathrm{L}}^{(t)} + Y^{(t)}/\mu)$ and $\eta_a = \|Z_{\mathrm{H}}^{(t)}\|_2^2$. Problem (5) can be effectively solved by the singular value thresholding (SVT) operator [2]. Similar as updating A, we can

Updating B:

$$\begin{aligned}B^{(t+1)} &= \arg\min_B \|B\|_* + h(A, B, Z_{\mathrm{L}}, Z_{\mathrm{H}}, Y, \mu) \\ &= \arg\min_B \frac{1}{\eta_b\mu}\|B\|_* + \frac{1}{2}\|B - B^{(t)} + \nabla_B h\|_{\mathrm{F}}^2,\end{aligned} \tag{6}$$

where $\nabla_B h = \nabla_A h(A^{(t)}, B^{(t)}, Z_{\mathrm{L}}^{(t)}, Z_{\mathrm{H}}^{(t)}, Y^{(t)}, \mu) = -(Z_{\mathrm{L}}^{(t)} - Z_{\mathrm{H}}^{(t)}A^{(t)} - B^{(t)}Z_{\mathrm{L}}^{(t)} + Y^{(t)}/\mu)Z_{\mathrm{L}}^{(t),T}$ and $\eta_b = \|Z_{\mathrm{L}}^{(t)}\|_2^2$.

Updating Z_{L}, Z_{H}:

$$\begin{aligned}Z_{\mathrm{L}}^{(t+1)} &= \arg\min_{Z_{\mathrm{L}}} \alpha\|X_{\mathrm{L}} - D_{\mathrm{L}}^{(t)}Z_{\mathrm{L}}\|_2^2 + \beta\|Z_{\mathrm{L}}\|_1 \\ &+\frac{\mu}{2}(\|Z_{\mathrm{L}} - Z_{\mathrm{H}}^{(t)}A^{(t)} - B^{(t)}Z_{\mathrm{L}} + Y^{(t)}/\mu\|_{\mathrm{F}}^2),\end{aligned} \tag{7}$$

$$Z_{\mathrm{H}}^{(t+1)} = \arg\min_{Z_{\mathrm{H}}} \alpha\|X_{\mathrm{H}} - D_{\mathrm{H}}^{(t)} Z_{\mathrm{H}}\|_2^2 + \beta\|Z_{\mathrm{H}}\|_1 + \frac{\mu}{2}(\|Z_{\mathrm{L}}^{(t)} - Z_{\mathrm{H}} A^{(t)} - B^{(t)} Z_{\mathrm{L}}^{(t)} + Y^{(t)}/\mu\|_{\mathrm{F}}^2). \tag{8}$$

Since problems (7) and (8) have the same structure, they can be solved in the same way. Take solving problem (8) for example, we rewrite it as:

$$Z_{\mathrm{L}}^{(t+1)} = \arg\min_{Z_{\mathrm{L}}} \alpha\|\hat{X}_{\mathrm{L}} - \hat{D}_{\mathrm{L}}^{(t)} Z_{\mathrm{L}}\|_F^2 + \gamma\|Z_{\mathrm{L}}\|_1, \tag{9}$$

where $\hat{X}_{\mathrm{L}} = \left[X_{\mathrm{L}}, \sqrt{\frac{\mu}{2\alpha}}(Z_H^{(t)} A - Y_1/\mu)\right]$, $\hat{D}^{(t)} = \left[D_L^{(t)}, \sqrt{\frac{\mu}{2\alpha}}(I - B)\right]$, and $\gamma = \beta/\alpha$. Problem (9) can be solved by existing solvers like SPAMS [17].

Updating D_{L}, D_{H}:

$$D_{\mathrm{L}}^{(t+1)} = \arg\min_{D_{\mathrm{L}}} \|X_{\mathrm{L}} - D_{\mathrm{L}} Z_{\mathrm{L}}^{(t)}\|_F^2, \tag{10}$$

$$D_{\mathrm{H}}^{(t+1)} = \arg\min_{D_{\mathrm{H}}} \|X_{\mathrm{H}} - D_{\mathrm{H}} Z_{\mathrm{H}}^{(t)}\|_F^2. \tag{11}$$

Problems 10 and 11 are quadratically constrained quadratic program (QCQP) problems and can be solved using Lagrange dual techniques [11].

Updating Y:

$$Y_1^{(t+1)} = Y_1^{(t)} + \mu(Z_{\mathrm{L}} - Z_{\mathrm{H}} A - B Z_{\mathrm{L}}). \tag{12}$$

Updating μ:

$$\mu = \min(\rho\mu, \max_\mu), \tag{13}$$

where the parameters μ, ρ, ϵ, $\max_\mu$ are set empirically.

The whole procedure of our solutions is outlined in **Algorithm 1**.

4 Experimental Results

Evaluation on the VIPeR Dataset: The VIPeR dataset [4] contains 632 persons with each having a pair of images captured from two outdoor cameras. Similar to [8], down-sampling and smoothing operations are performed on all images from camera B to generate LR images. 632 images from camera A and the generated 632 LR images from camera B form 632 image pairs. All image pairs are randomly split into two sets (316 pairs for each set) with one for training and the other for testing. We take images from camera A in the testing set as the HR gallery image set, and use the LR images from camera B in the testing set to construct the LR probe set. Table 1 report the matching results of all compared methods at sampling rate of 1/8. We can see that the proposed algorithm does outperform some existing ones, despite of not being the best one. This proves some effectiveness of the proposed algorithm.

Algorithm 1 Solving Problem (2)

Input: $X_{\mathrm{H}}, X_{\mathrm{L}}$

Initialize: $A = 0, B = 0, D_{\mathrm{L}} = 0, D_{\mathrm{L}} = 0$

$$\mu = 10^{-6}, \alpha = 1.0, \beta = 0.01, \max_{\mu} = 10^{6}, \mathrm{maxIter} = 50, t = 0.$$

while not converged **or** $t \leq$ maxIter **do**

1. Fix the others and update $Z_{\mathrm{L}}^{(t+1)}$ and $Z_{\mathrm{H}}^{(t+1)}$ according to (9);
2. Fix the others and update $D_{\mathrm{L}}^{(t+1)}$ and $D_{\mathrm{H}}^{(t+1)}$ according to (10) and (11), respectively;
3. Fix the others and update $A^{(t+1)}$ and $B^{(t+1)}$ according to (5) and (6), respectively;
6. Update the multipliers $Y^{(t+1)}$ according to (12);
7. Update the parameter μ according to (13);
8. Check the convergence conditions;
9. $t = t + 1$.

end while

output: $A, B, Z_{\mathrm{L}}, Z_{\mathrm{H}}, D_{\mathrm{L}}, D_{\mathrm{H}}$

5 Conclusions and Future Work

This report introduces a new person re-identification algorithm which targets to solve the resolution change problem in person re-identification. An image patch cluster based dictionary learning framework is proposed to learn dictionaries with discriminative power for low and high resolution image patches. When learning patch based dictionaries, a latent factor corresponding latent high resolution image patches is introduced to help learn better dictionaries. Experiments show that the proposed algorithm outperforms some existing algorithms.

	r=1	r=5	r=10	r=20
RDC [26]	3.48	16.14	26.58	38.29
SSCDL [16]	10.44	31.33	48.42	72.78
RPLM [6]	7.59	26.58	42.72	64.24
KISSME [10]	8.74	28.58	45.02	68.20
SLDDL [9]	16.86	41.22	58.06	79.00
Ours	9.81	27.22	36.08	48.42

Table 1. Top r ranked matching rates (%) on the VIPeR dataset with sampling rate of 1/8.

Due to the limited time, I am unable to refine the proposed algorithm to achieve the state-of-the-art before the end of this semester. I will continue to work on this project and refine the proposed algorithm. Future efforts will be

paid on the following aspect: The relationship among sub-dictionaries should be better exploited, that is, some constraint should be added to required a sub-dictionary has strong representation power for its own image patch cluster, but week representation power for all the other image patch cluster. The current model does not incorporate this constraint, so it is expected the model will produce better person re-identification results once this constraint is introduced in the model.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing 54(11), 4311–4322 (2006)
2. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization 20(4), 1956–1982 (2010)
3. Ding, Z., Shao, M., Fu, Y.: Latent low-rank transfer subspace learning for missing modality recognition. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence (2014)
4. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS) (2007)
5. Gu, S., Zhang, L., Zuo, W., Feng, X.: Projective dictionary pair learning for pattern classification. In: Advances in Neural Information Processing Systems (2014)
6. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: European Conference on Computer Vision (2012)
7. Huang, D.A., Frank Wang, Y.C.: Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. In: Proceedings of the IEEE international conference on computer vision. pp. 2496–2503 (2013)
8. Huang, H., He, H.: Super-resolution method for face recognition using nonlinear mappings on coherent features. IEEE Transactions on Neural Networks 22(1), 121–130 (2011)
9. Jing, X.Y., Zhu, X., Wu, F., You, X., Liu, Q., Yue, D., Hu, R., Xu, B.: Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 695–704 (2015)
10. Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (2012)
11. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Advances in neural information processing systems (2006)
12. Li, S., Shao, M., Fu, Y.: Cross-view projective dictionary learning for person re-identification. In: Proceedings of the 24th International Conference on Artificial Intelligence, AAAI Press (2015)
13. Li, X., Zheng, W.S., Wang, X., Xiang, T., Gong, S.: Multi-scale learning for low-resolution person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3765–3773 (2015)
14. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: Proceedings of the 27th International Conference on Machine Learning. pp. 663–670 (2010)

15. Liu, G., Yan, S.: Latent low-rank representation for subspace segmentation and feature extraction. In: IEEE International Conference on Computer Vision. pp. 1615–1622 (2011)
16. Liu, X., Song, M., Tao, D., Zhou, X., Chen, C., Bu, J.: Semi-supervised coupled dictionary learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014)
17. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning. pp. 689–696 (2009)
18. Peng, P., Xiang, T., Wang, Y., Pontil, M., Gong, S., Huang, T., Tian, Y.: Unsupervised cross-dataset transfer learning for person re-identification. In: 2016 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 543–550. IEEE (2016)
19. Yang, M., Zhang, L., Feng, X., Zhang, D.: Fisher discrimination dictionary learning for sparse representation. In: 2011 International Conference on Computer Vision. pp. 543–550. IEEE (2011)
20. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., Li, S.Z.: Salient color names for person re-identification. In: European Conference on Computer Vision (2014)
21. You, J., Wu, A., Li, X., Zheng, W.S.: Top-push video-based person re-identification. arXiv preprint arXiv:1604.08683 (2016)
22. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2013)
23. Zhao, R., Ouyang, W., Wang, X.: Learning mid-level filters for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014)
24. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: Past, present and future. arXiv preprint arXiv:1610.02984 (2016)
25. Zheng, W.S., Gong, S., Xiang, T.: Transfer re-identification: From person to set-based verification. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (2012)
26. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. IEEE transactions on pattern analysis and machine intelligence 35(3), 653–668 (2013)