PERSON RE-IDENTIFICATION VIA RICH COLOR-GRADIENT FEATURE

Lingxiang Wu[1,2], Jinqiao Wang[2], Guibo Zhu[2], Min Xu[1] and Hanqing Lu[2]

[1]Global Big Data Technologies Centre, University of Technology Sydney, Australia
[2]National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing, China, 100190
Lingxiang.Wu@student.uts.edu.au, Min.Xu@uts.edu.au, {jqwang, gbzhu, luhq}@nlpr.ia.ac.cn
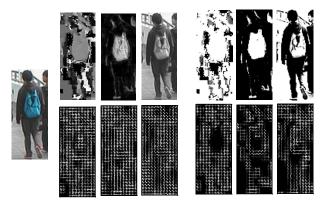
ABSTRACT

Person re-identification refers to match the same pedestrian across disjoint views in non-overlapping camera networks. Lots of local and global features in the literature are put forward to solve the matching problem, where color feature is robust to viewpoint variance and gradient feature provides a rich representation robust to illumination change. However, how to effectively combine the color and gradient features is an open problem. In this paper, to effectively leverage the color-gradient property in multiple color spaces, we propose a novel Second Order Histogram feature (SOH) for person re-identification in large surveillance dataset. Firstly, we utilize discrete encoding to transform commonly used color space into Encoding Color Space (ECS), and calculate the statistical gradient features on each color channel. Then, a second order statistical distribution is calculated on each cell map with a spatial partition. In this way, the proposed SOH feature effectively leverages the statistical property of gradient and color as well as reduces the redundant information. Finally, a metric learned by KISSME [1] with Mahalanobis distance is used for person matching. Experimental results on three public datasets, VIPeR, CAVIAR and CUHK01, show the promise of the proposed approach.

Index Terms— Person re-identification, encoding color space, second order histogram

1. INTRODUCTION

In video surveillance, it is desirable to judge whether or not a pedestrian has appeared across disjoint camera views. That is person re-identification problem. It has received increasing attention recently as it could greatly save human effort on manually browsing and searching persons in a large scale dataset. However, person re-identification is still a challenging topic in computer vision because the same person's appearances often undergo significant variance in illumination, camera viewpoint, pose and clothes, while different pedestrians may look very similar due to the similar dressing style. In addition, complex background clutter, occlusion

Fig. 1. Visualization of HOG features on different color channels. $Top\ row$: Grayscale map of each channel. $Bottom\ row$: HOG model on corresponding channel. The left three channels are form HSV color space, and the right three are from our encoding color space

and low resolution also increase the difficulties of person re-identification.

There are two crucial problems in person re-identification: appearance modeling and person matching. Appearance modeling is required to be representative and discriminative. Some effective descriptors have been proposed, for example: SDALF [2], SCNCD [3], LOMO [4], salMatch [5]. These hand-crafted or learning-based descriptors made great progress in the performance of person re-identification. The other crucial aspect is person matching. Person matching is often formulated as a metric learning problem [6, 1, 7] to maximize the separability of different persons, like KISSME [1], LMNN [8], XQDA [4].

Color and gradient features have shown better performance for appearance modeling, where color feature is robust to viewpoint variance and gradient feature provides a rich representation robust to illumination change. However, traditional approaches usually directly concatenate them together, which ignore the different gradient properties of person appearance in different color spaces. Thus, we extract HOG features over different color spaces to preserve the multi-channel

relationship between color and gradient features. As shown in Fig. 1, it can be seen that HOG features get variant responces on different color channels. It is obvious that the difference of HOG on Hue, Saturation and Value channel is weak, while the gradient distribution in the proposed multi-channel color space shows different representation ability for different parts of persons. Such as the gradient of blue backpack is more obvious in the first channel in multiple-channel color space in Fig. 1. Therefore it is interesting to investigate HOG features in a novel color space to achieve a rich color-gradient representation for appearance modeling. In this way, the noises caused by gradient extraction are inhibited, while subtle gradient changes are stressed or even magnified. The gradient histogram on each cell could look sparse, but robust and discriminative. In this paper, we propose a multiple-channel encoding color space to exploit the color-gradient properties. A Second Order Histogram feature is built to model the appearance of persons. Then a widely used metric learning method– KISSME [1] was employed for person matching.

2. RELATED WORKS

To tackle the problem of person re-identification, many existing approaches have been proposed and can be roughly summarized into two aspects: appearance modeling and person matching.

Appearance Modeling. Currently, many approaches on person re-identification focus on designing a representative feature descriptor. The feature descriptors include hand-craft features [4, 9, 3, 2] and graph-based features [10, 11]. Farenzena *et al.* [2] proposed a symmetry-driven accumulation of local feature (SDALF) by exploiting symmetry property on person to tackle view variance problem. Liu *et al.* [12] proposed an unsupervised approach to learn a bottom-up measurement of feature importance, so that features extracted from different individuals were weighted adaptively driven by their appearance characteristics. Besides vector-based features, graph-based features also attract intention recently because graph has intrinsic advantages on revealing relationship among different parts of an object, and advantages to present feature in a flexible length. Iodice *et al.*[10] formulated person re-id as a graph matching problem and represented person as a graph, which aimed not only at rejecting erroneous matches but also selecting additional useful ones.

Person matching. Metric learning based approaches are often used for person matching [6, 1, 7], where a projection metric is sought so that the projected Mahalanobis-like distance is small for intra-class but large for the inter-class. In [1] Kostinger *et al.* proposed a metric constraint based on a statistical inference perspective. Instead of adopting a fixed metric for all subjects, Wei *et al.* [6] learned transferred metric by re-weighting and selecting training samples according to the similarity with query sample and template set. In [7],

Xiong *et al.* applied multiple matching methods on the kernel space including linear, χ^2 and RBF-χ^2 kernels. Additionally, Prosser *et al.* [13] formulated person re-identification problem as a ranking problem, and employed RankSVM to learn a subspace.

3. SECOND ORDER HISTOGRAM

Overview of the proposed SOH is shown in Fig. 2. In the following, we will introduce each step in details.

3.1. Encoding Color Space

There are several color spaces that are used to present the pixel value of images, *e.g.* RGB, HSV, YUV, Lab, *etc.* Each color space has its advantages and disadvantages with respect to specific application or hardware. Although HSV is widely used in the field of computer vision, it cannot work well in an illumination fusion model, and it weakly responds to natural color. Additionally, all these color spaces are three-channel designed. This inspired us to design a multiple channel color space for a rich representation capability, like 11-dimensional color naming [14]. In order to exploit the rich color-gradient information for a strongly efficient feature representation, we designed an Encoding Color Space (ECS) to maximum the gradient difference in a KISS (*keep it simple and stupid*) way.

After all pixel values in the original color space were normalized to [0, 1], we perform a discretization color encoding, which is to select a fixed discretization scale and to express the pixel value in a discretization method. Let N denote the number of discrete level and $U(n)$ ($n \in Z$) denote the discretization representation of the number n, *e.g.* $U(2) = \{1, 1, 0, 0, 0, 0\}$, $U(5) = \{1, 1, 1, 1, 1, 0\}$ when $N = 6$. To preserve more detailed information, an alternate representation $U'(r)$ is defined for any real number $r \geq 0$ as the unary representation, but substitute for the first zero in the unary representation of $U(\lfloor r \rfloor)$ by $\alpha(r) = r - \lfloor r \rfloor$. For example $U'(2.4) = \{1, 1, 0.4, 0, 0, 0\}$. Then the discretization encoding can be calculated as follows:

$$\phi(x) = U'(Nx) \tag{1}$$

By transferring the discretization encoding from pixel-level to channel-level in different color spaces, we can transform each channel (*e.g.* red-channel, green-channel) into N sub-channels on the basis of Eq.1. Thus a three-channel color space is encoded into a $3 \times$ N-channel color space. N is set to 6 in experiments. We transform original HSV, YCbCr and Lab color space into the encoding color space, and fuse them into a multiple color space with 54 channels.
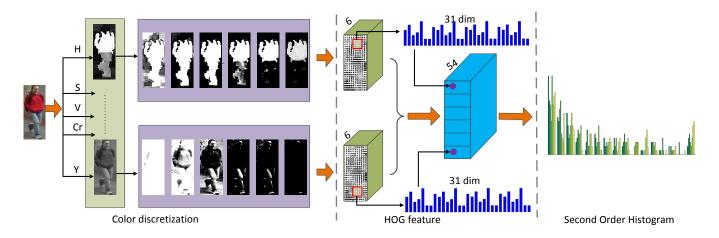
Fig. 2. Overview of the proposed second order histogram.

3.2. Multi-channel HOG

Because of low image resolution and posture variation caused by viewpoint change, color plays a crucial role in appearance modeling. However, variant illumination across disjoint camera view affects the re-identification performance seriously. Changes in the illumination can be modeled by a diagonal-offset model [15] as

$$
\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix} \quad (2)
$$

where u corresponds to the image taken under an unknown illumination, while c corresponds to the same image transformed. Based on the diagonal-offset model, common changes in image value for illumination has been categorized into 5 types, *i.e.* light intensity changes, light intensity shifts, light intensity changes and shifts, light color changes, as well as light color changes and shifts.

In order to overcome illumination challenges, many color descriptors or strategies have been proposed. The most widely used descriptor is the color histogram, which calculates the statistical color distribution with a spatial constraint. But it is sensitive to illumination changes and lack of structure information. In [4], a multiscale Retinex algorithm was used for preprocessing. Although it enhanced details in shadowed regions, it has not dealt with the photometric transformation essentially.

Inspired by HSV-SIFT proposed in [16], we propose to construct a HOG [17, 18, 19] descriptor over encoding color space. HOG is computed on a dense grid of uniformly spaced cell with overlapping local contrast normalization. A 32-bins histogram on each cell is built based on the standard 9 orientations. By removing the last all-zero dimension, we only use the 31-bins information on each cell. If we compute HOG on HSV color space, three cell maps are obtained, where each cell contains 31-bins value. We call it first order histogram.

In this way, we achieve a natural fusion for color and gradient since color feature is robust to viewpoint change, while oriented gradient is robust to illumination variation. In addition, by normalizing the value within each cell, scale-invariance and shift-invariance may be achieved with respect to the light intensity.

3.3. Second Order Histogram

In decades, many descriptors utilized the first order gradient information to characterize the geometric properties of an object, like SIFT, SURF, HOG, LBP, *etc*. However, recent study [20] on human vision have shown that the first order gradient information is far from sufficiency in accurately capturing the perceived visual features by human beings. Thus, we focus on constructing a second order histogram upon the first order histogram, which is the HOG map in this paper.

After we extract HOG features over the multiple encoding color space (54 channels), we get a HOG map with 54×31 values on a cell. Now, the map is regarded as an image and each cell is regarded as a pixel with 1674 "channels". As the discriminability of different color spaces, we add different weights to these "channels" with respect to three color space. These weight is obtained by a grid search with cross-validation. Next, the HOG map is equally divided into 6 horizontal stripes to roughly capture person's head, upper and lower torso, upper and lower legs. Afterwards, a total of 1674 1-D histograms are calculated over each stripe. Then 1674 1-D histograms are also extracted on the whole image to model the global characteristics. In short, our second order histogram feature is represented as:

$$
F = [H_1, H_2, \cdots, H_6, H_G]^T \quad (3)
$$

where

$$
H_i = [h_{i1}, h_{i2}, \cdots, h_{i1674}] \quad (4)
$$

h denotes the 1-D 16-bins histogram.

Although the second order histogram looks simple, it not only captures the high-order color-gradient properties but also greatly reduces the feature dimension. Furthermore, it reflects the statistical characteristics within spatial layout. For instance, without the second order information, the features in first order histogram cannot measure whether those cells with analogue oriented statistics are positioned together or separately.

4. PERSON MATCHING

In general, two sets of person images are given, one set as gallery images, the other set as probe images. For person re-identification, the person from probe set should be matched with its corresponding person image(s) in gallery set. We formulate this problem as follows: Given a gallery set of templates $T = \{T_1, T_2, \cdots, T_n\}$ and a probe Q, find the most similar template $T^* \in T$ with respect to a similarity measure D

$$T^* = \arg\min_{T_i} D(T_i, Q) \qquad (5)$$

In person re-identification, it is supposed to determine whether a pair of samples belong to the same category but not to answer which category they belong to. The multi-class classifiers in traditional machine learning approaches are not suitable here. Therefore, learning a projection metric with metric learning has received great attention. It learns a metric by maximizing the inter-class distances while minimizing the intra-class distances. In this paper, given two SOH features x_i and x_j, we adopt KISSME [1] to learn a metric M in Mahalanobis distance where KISSME views the statistical decision whether a pair is dissimilar or not by the log likelihood ratio test of the two Gaussian distributions.

$$d_M^2(x_i, x_j) = (x_i, x_j)^T M(x_i, x_j) \qquad (6)$$

The metric M is defined as:

$$M = \sum{}_S^{-1} - \sum{}_D^{-1} \qquad (7)$$

where

$$\sum{}_S = \frac{1}{|S|} \sum_{x_i, x_j \in S} (x_i - x_j)(x_i - x_j)^T \qquad (8)$$

$$\sum{}_D = \frac{1}{|D|} \sum_{x_i, x_j \in D} (x_i - x_j)(x_i - x_j)^T \qquad (9)$$

donates the covariance metrics for similarity pairs S and dissimilarity pairs D respectively.

5. EXPERIMENTS

To illustrate the effectiveness of the proposed approach, we carried out exhaustive experiments on three datasets. All experiment results are reported on Cumulative Matching Characteristics (CMC) curves, which represents the chance of the true matching appearing in the top 1, 2,..., N of the ranked list. The first point on the CMC curve is rank-1 accuracy. Our experiments followed the protocol in [7]. We adopt a single-shot setting in experiments, and for fair, we compare method using single-shot setting. Each dataset is divided into two parts, 50% for training and 50% for testing. Specifically, there are 316, 36 and 485 individuals in each test set for VIPeR, CAVIAR and CUHK01. In test, one image for each individual is randomly selected as the gallery set, and the rest compose probe set. This process is implemented 10 times repeatedly, and we report the average matching accuracy.

5.1. Experiments on VIPeR

VIPeR [21] is one of the most challenging and the most widely used person re-id dataset for benchmark evaluation. It contains 632 pairs pedestrians captured by two cameras outdoors. Each pair pedestrians corresponds to two images from two different cameras. This dataset suffers from serious illumination and viewpoint changes.

5.1.1. Encoding Color Space vs. Original Color Space

We compared color histogram and SOH over original color space with those on our encoding color space. The CMC curves and the Rank-1 accuracy are reported in Fig. 3(a).
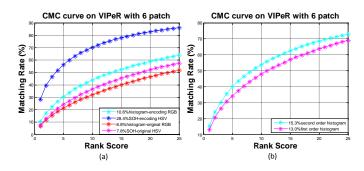
Through rank-1 accuracy, it can be seen obviously that SOH on our encoding color space achieved marginally as higher as almost 20% compared with that descriptor on original HSV. The same situation occurred on histogram on RGB and encoding RGB. It can be illustrated that our encoding color space performs better than traditional color space, especially for SOH descriptor.

5.1.2. Second Order Histogram vs. First Order Histogram

Comparison between second order and first order histogram on encoding RGB color space is reported in Fig. 3(b). Although the Rank-1 accuracy of second order histogram is only 2.3% higher than first order histogram, the second order reduced feature dimension compared to the feature in the first order histogram.

5.1.3. Comparison with the state-of-the-art

Comparison with single feature: As some person re-id papers focus on the machine learning or the ensemble strategies, in this part, we compare the SOH feature with other features. The CMC curves and Rank-1 accuracy could be seen in Fig. 4(a). In KISSME [1], they employed HSV, Lab color histogram, and texture feature extracted by LBP. Using the same metric learning method, our SOH could improve 11.5%. Compared with SalMatch and SDALF, our SOH also outperforms 1.9% and 12.3% respectively. Using the same metric learning method , KISSME, there are only two methods better

Fig. 3. CMC curves on VIPeR: (a) original color space vs. encoding color space on 2 descriptors. (b) first order histogram vs. second order histogram on encoding RGB.

Fig. 4. CMC curves on VIPeR: (a) comparison with published features. (b)The SOH and the ensemble result.

than ours, SCNCD [3] and LOMO [4]. It should be, however, note that their experiment settings are different from ours that they use images from one camera as prob set, others as gallery set.
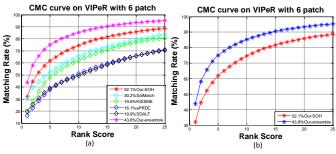
Comparison with existing methods: In this part, we fuse our feature with CNN feature, BoW [22] and LOMO [4] to get a competent feature. The CMC curves of our ensemble feature can be seen in Fig. 4(b) and the comparison results with state-of-the-art approaches are given in Table 1. Our ensemble feature achieves 43.8% at rank-1 accuracy.

Comparison between SOH, our ensemble feature and state-of-the-art approaches is presented in Fig. 5(a) and Table 1. The proposed approach obtains the best performance in CAVIAR dataset. It should be note that we set single-shot in our experiment. If we set multi-shot that each person has 5 images in gallery set, our ensemble feature achieves 68.0%.

Fig. 5. CMC curves: compare with state-of-the-art approaches on (a) CAVIAR (b) CUHK01.

Table 1. Rank-1 Accuracy (%) on 3 datasets.

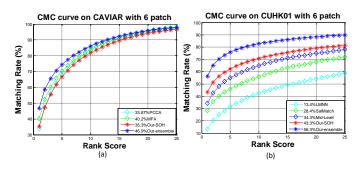|	VIPeR	CAVIAR	CUHK01
aPRDC[12]	16.1	-	-
KISSME[1]	19.6	-	-
SDALF[2]	19.9	-	-
SalMatch[5]	30.2	-	28.5
LOMO[4]	34.1	-	-
SCNCD[3]	37.8	-	-
MidLevel+LADF[23]	43.4	-	-
Ensemble[7]	36.1	-	-
LTR[24]	**45.9**	-	53.4
Semantic[25]	41.6	-	-
MFA[7]	-	40.2	-
PCCA[26]	-	33.9	-
LMNN[6]	-	-	13.4
ImprovedDeep[27]	-	-	47.5
Mid-Level[23]	-	-	34.3
Our-SOH	32.1	35.3	43.3
Our-ensemble	43.8	**46.9**	**56.3**

5.2. Experiments on CAVIAR

CAVIAR [28] is another famous dataset widely used. It contains 1220 images of 72 pedestrians from 2 cameras in a shopping center. The image number of each pedestrian varies from 10 to 20. The major challenge in this dataset arises from pose variance and image resolution.

5.3. Experiments on CUHK01

CUHK01 [5] contains 971 pedestrians captured by two cameras in a campus environment. Each person has two images from each camera. Camera A captures the front view or back view, while camera B captures the side view. Images in this dataset are of high resolution.

The results are shown in Fig. 5(b) and Table 1. Our SOH feature achieves a certain high rank-1 accuracy as 43.3% compared to other hand-craft features. Our ensemble feature achieves 56.3% at rank-1 accuracy, and outperforms all existing methods with single-shot setting.

6. CONCLUSION

In this paper, we propose a novel Second Order Histogram (SOH) feature descriptor over Encoding Color Space (ECS) for person re-identification. As color and gradient feature is

dominant in describing person appearance, our approach is effective in leveraging the color-gradient property to overcome illumination and viewpoint variance. Experiments demonstrate the effectiveness on three popular datasets: VIPeR, CAVIAR and CUHK01. Next step, we will verify the effectiveness of the proposed features on other tasks, such as object detection and tracking.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof, "Large scale metric learning from equivalence constraints," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2288–2295.

[2] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2360–2367.

[3] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z Li, "Salient color names for person re-identification," in *Computer Vision–ECCV 2014*, pp. 536–551. Springer, 2014.

[4] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2197–2206.

[5] Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Person re-identification by salience matching," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2528–2535.

[6] Wei Li, Rui Zhao, and Xiaogang Wang, "Human reidentification with transferred metric learning.," in *ACCV (1)*, 2012, pp. 31–44.

[7] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier, "Person re-identification using kernel-based metric learning methods," in *Computer Vision–ECCV 2014*, pp. 1–16. Springer, 2014.

[8] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid, "Is that you? metric learning approaches for face identification," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 498–505.

[9] Bingpeng Ma, Yu Su, and Frédéric Jurie, "Local descriptors encoded by fisher vectors for person re-identification," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer, 2012, pp. 413–422.

[10] Sara Iodice and Alfredo Petrosino, "Person re-identification based on enriched symmetry salient features and graph matching," in *Pattern Recognition*, pp. 155–164. Springer, 2013.

[11] Luc Brun, Donatello Conte, Pasquale Foggia, and Mario Vento, "People re-identification by graph kernels methods," in *Graph-Based Representations in Pattern Recognition*, pp. 285–294. Springer, 2011.

[12] Chunxiao Liu, Shaogang Gong, Chen Change Loy, and Xinggang Lin, "Person re-identification: What features are important?," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer, 2012, pp. 391–401.

[13] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, Tao Xiang, and Q Mary, "Person re-identification by support vector ranking.," in *BMVC*, 2010, vol. 2, p. 6.

[14] Joost Van De Weijer, Cordelia Schmid, and Jakob Verbeek, "Learning color names from real-world images," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[15] Johannes von Kries, "Influence of adaptation on the effects produced by luminous stimuli," *Sources of color vision*, pp. 109–119, 1970.

[16] Koen EA Van De Sande, Theo Gevers, and Cees GM Snoek, "Evaluating color descriptors for object and scene recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1582–1596, 2010.

[17] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[18] Pedro Felzenszwalb, David McAllester, and Deva Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[19] Guibo Zhu, Jinqiao Wang, Yi Wu, Xiaoyu Zhang, and Hanqing Lu, "Mc-hog correlation tracking with saliency proposal," in *AAAI*, 2016.

[20] Michael J Morgan, "Features and the primal sketch," *Vision research*, vol. 51, no. 7, pp. 738–753, 2011.

[21] 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)*. Citeseer, 2007, vol. 3.

[22] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jiahao Bu, and Qi Tian, "Person re-identification meets image search," *arXiv preprint arXiv:1502.02171*, 2015.

[23] Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Learning mid-level filters for person re-identification," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 144–151.

[24] Sakrapee Paisitkriangkrai, Chunhua Shen, and Anton van den Hengel, "Learning to rank in person re-identification with metric ensembles," *arXiv preprint arXiv:1503.01543*, 2015.

[25] Zhiyuan Shi, Timothy M Hospedales, and Tao Xiang, "Transferring a semantic representation for person re-identification and search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4184–4193.

[26] Alexis Mignon and Frédéric Jurie, "Pcca: A new approach for distance learning from sparse pairwise constraints," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2666–2672.

[27] Ejaz Ahmed, Michael Jones, and Tim K Marks, "An improved deep learning architecture for person re-identification," *Differences*, vol. 5, pp. 25, 2015.

[28] Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino, "Custom pictorial structures for re-identification.," in *BMVC*. Citeseer, 2011, vol. 1, p. 6.