

Person Re-identification with Density-Distance Unsupervised Saliency Learning

Baoliang Zhou¹, Aihua Zheng^{1(✉)}, Bo Jiang¹, Chenglong Li¹, and Jin Tang^{1,2}

¹ Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Computer Science and Technology, Anhui University, Hefei, China
xxzblwc@163.com, {ahzheng214, jiangbo}@ahu.edu.cn
lc1314@foxmail.com

² Key Laboratory of Industry Image Processing and Analysis in Anhui Province, Hefei, China
tj@ahu.edu.cn

Abstract. Human saliency of pedestrians images is distinctive and has been shown importantly in person re-identification (or pedestrians identification) problem. Thus, how to obtain the salient area of pedestrian images is important for this saliency based pedestrians identification problem. In this paper, we first show that this kind of salient area detection can be formulated as a kind of outlier detection problem, and then propose a novel unsupervised saliency learning method using a local outlier-detection technique for person re-identification task. The main feature of the proposed saliency computation method is that it exploits both distance and density information simultaneously. Experimental results on several datasets show the effectiveness of the proposed saliency based person re-identification method.

Keywords: Saliency learning · Person re-identification · Density-Distance · Head-priority

1 Introduction

Person re-identification is an important topic in computer vision area and has been widely used in video surveillance. The aim of person re-identification is to determine associations of the pedestrians in multi-camera networks in video surveillance. Here, we focus on the person re-identification (or pedestrian association problem) in non-overlapping camera networks.

Many works have been proposed for this problem. Generally, these works can be divided into two categories, *i.e.*, supervised and unsupervised methods. For supervised methods, distance metric learning is a classical method towards this problem [3, 4, 9, 12, 24]. Schwartz et al. [10] extracted high-dimensional features for metric learning. Porikli and Divakaran et al. [11] proposed a distance metric method and a non-parametric and non-linear model of color shift function based on color histogram. Li and Wang et al. [12] proposed a method which

learns different metrics for different camera image spaces. Chilgunde et al. [13] applied Kalman Filter to predict the person movement in blind areas of the monitoring cameras, and matched motion trail on Gaussian model. Similarly, Riccardo et al. [14] proposed Multi Social Force Model to predict the tracklets of pedestrians in blind regions. Javed et al. [15] used inter-camera space-time and appearance probabilities to find an object in different cameras by maximizing the conditional probability of the corresponding observations. For unsupervised methods, Ma et al. [20] modified the BiCov descriptor based on Gabor filters for person matching, and used the covariance descriptor to deal with illumination change and image misalignment. Farenzena et al. [17] proposed Symmetry-Driven Accumulation of Local Features to handle the change of camera viewpoint, as pedestrian images are symmetrical. Prosser et al. [1] formulated the person re-identification as a global optimum ranking problem, and utilized an ensemble of RankSVM to learn global feature weights. Liu et al. [9] adopted an attribute-based weighting scheme to find the unique appearance attribute. However, the most existing methods ignored some discriminative and simple features which are actually useful for person pairs matching in human images.

Recently, Zhao et al. [5] proposed a more effective unsupervised salience learning method with some distinct regions which are discriminative, reliable and useful for person matching. The main idea of this method is that it further detects and explores some human salient regions of pedestrian body in computing similarity scores between pedestrian images. Human salience of pedestrian images is distinctive and has been shown importantly in person re-identification problem. Thus, how to obtain the salient area of pedestrian images is important for this salience based pedestrians identification problem.

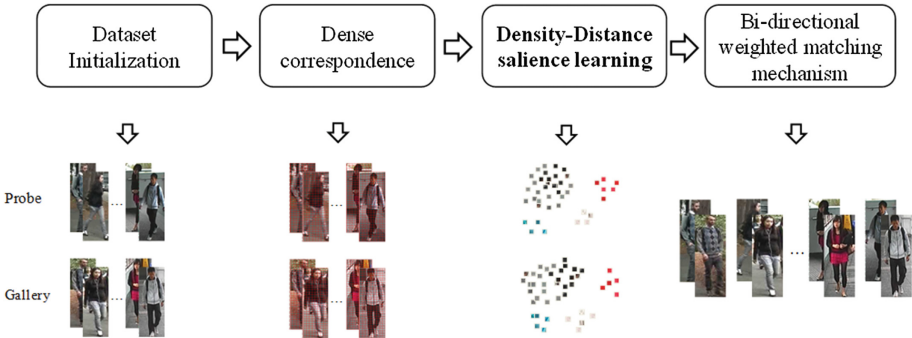


Fig. 1. The whole process of the proposed person re-identification algorithm.

Inspired by the work on salience learning method in [5], in this paper, we first show that this kind of salient area detection can be formulated as a kind of general outlier detection problem. Then, we propose a novel unsupervised salience learning method, called Density-Distance salience learning, which aims to use a local outlier-detection technique for person re-identification problem.

The procedure of the whole algorithm has been shown in Fig. 1. The detailed algorithm is introduced in Sect. 3. One main feature of the proposed salience learning method is that it exploits both distance and density information simultaneously. Compared with traditional methods, the proposed salience detection method is simple and efficient for person re-identification problem. Promising experimental results on the widely used VIPeR and CUHK01 datasets demonstrate that the proposed salience-based person re-identification method is more effective and efficient than some recent methods.

2 Salience Learning with Density-Distance Outlier Detection

In this section, a novel salience learning method is proposed. As mentioned in work [5], the salient patches of pedestrian images generally have the following properties: (1) these patches are deviates markedly from others and thus provide some distinctive information for the pedestrian body image. (2) These patches are robust to the changes of camera viewpoint. Some examples of salient patches are shown in Fig. 2.



Fig. 2. Illustration of person salience

Intuitively, this kind of salience definition can be regarded as the outlier patches in patch dataset. This motivates us to explore some outlier detection methods for salient patch detection. Outlier detection is a fundamental problem in data mining area. In outlier detection, an outlier is defined as an observation which deviates markedly from the other observations [29, 30].

In the following, we first propose a local outlier detection based salience learning method, called Density-Distance salience learning, and then use this method for person re-identification problem. The core idea of the proposed Density-Distance salience learning method is to integrate both distance and density information for salience learning. Each human image is firstly densely segmented into some local patches, then the dLabSift features(details in Sect. 3.1) are extracted for each patch and denoted as $x_{A,i}(m, n)$, where (A, i) indicates the i -th person in camera A , and (m, n) is index denotes the patch located at the m -th row

and the n -th column in the human image. The combined salience value is designed as:

$$\begin{aligned} Svalue_{DD}(x_{A,i}(m, n)) &= \alpha \cdot Svalue_{density}(x_{A,i}(m, n)) \\ &+ (1 - \alpha) \cdot Svalue_{distance}(x_{A,i}(m, n)) \end{aligned} \quad (1)$$

where $Svalue_{density}(x_{A,i}(m, n))$ denotes the density salience values and similarly, $Svalue_{distance}(x_{A,i}(m, n))$ denotes the distance salience values. α ($0 < \alpha < 1$) is a balance parameter. In the following, we introduce the distance salience and density salience, respectively.

2.1 Distance Salience Learning

Each human image is densely segmented into some local patches, as shown in Fig. 3. The patches with discriminative property are called salient patches. Indeed, this kind of salient patches can be regarded as outliers in the patch set, as shown in Fig. 4. Byers et al. [21] proposed a method which uses the KNN distances and been demonstrated to be useful for outlier detection. In this paper, we use the mean value of KNN distance for salience learning.

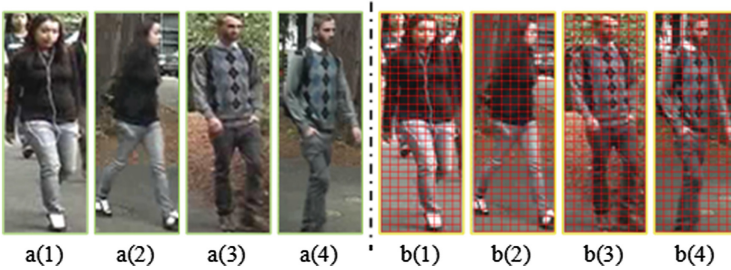


Fig. 3. Examples of Dense Patches. a(1)–a(4) are human images, and b(1)–b(4) are illustrations of dense local patches. Each patch is dense and overlapping.

Before computing the average neighbor distance of each testing patch, the reference set initialization is needed. Let the number of the images in the reference set be R_n . After building the dense correspondences between a testing image and the images in reference set, each patch in the testing image will find an optimal matched patch in every image from the reference set, that is for each testing patch $x_{A,i}(m, n)$ there are R_n matched patches, or called neighbors from the reference set $X_{nn}(x_{A,i}(m, n))$.

$$X_{nn}(x_{A,i}(m, n)) = \{x | \arg \max s(x_{A,i}(m, n), \hat{x}), j = 1, 2, \dots, R_n\}, \hat{x} \in S_{i,j} \quad (2)$$

$$S_{i,j} = S(x_{A,i}(m, n), x_{B,j}) \quad (3)$$

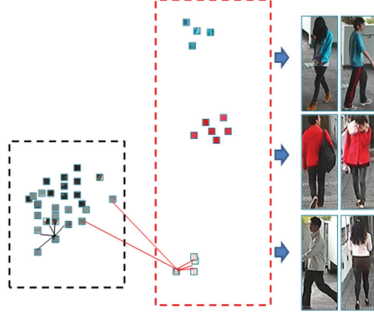


Fig. 4. Illustration of Salient Patches. The patches in the red and the black dotted boxes denote the salient patches and the general patches respectively.

More detailed introduction of Eq. (3) can refer to Sect. 3.1. Byers et al. [21] utilize the K nearest neighbor to find clutters. Instead of the k -th distance used in work [5], the average distance is calculated in our method, which is more reasonable and effective. The salience value can be computed as:

$$Svalue_{distance}(x_{A,i}(m,n)) = D_{average}(X_{nn}(x_{A,i}(m,n))) \quad (4)$$

where $D_{average}$ denotes the average distance to the R_n nearest neighbor. As we can see in Fig. 4, if the patch in the testing image is a salient patch then its $D_{average}$ will be larger than the average distance of general patch. In practice, it is not necessary to take all the R_n neighbors to achieve the global optimum. Therefore, the distance between the testing patch and all R_n neighbors are sorted and the intermediate $k(k = \beta R_n)$ neighbors are selected for computing, where $0 < \beta < 1$. Figure 5 shows the feature weighting map results of salience learning by our average distance strategy, and it is estimated by partial least square (PLS)[10].

2.2 Density Salience Learning

As discussed in works [23, 27, 28], density information is also important for outlier detection, therefore it is also important for salience learning.

Similar to Sect. 2.1, each image in testing set is segmented into $M \times N$ dense patches. As mentioned above, each testing patch has R_n neighbors, *i.e.*, every image in testing set has a set of patches P^i :

$$P^i = \{X_{nn}(x_{A,i}(m,n)) | m = 1, 2, \dots, M; N = 1, 2, \dots, N\} \quad (5)$$

Each neighbor has a matching distance to the testing patch. For each testing image, both maximum and minimum matching distances are used to compute the threshold d_{thr}^i for outliers, *i.e.*,

$$d_{max} = D_{max}(P^i) \quad (6)$$

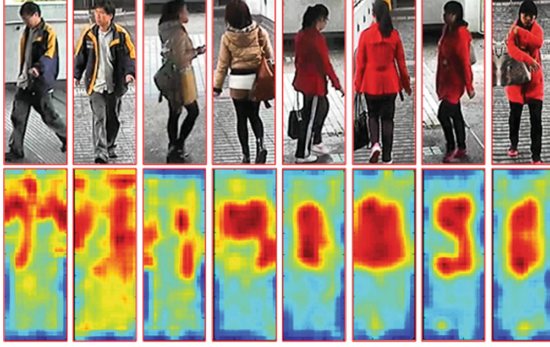


Fig. 5. Illustration of the feature weighting map of our average distance salience learning estimated by partial least square (PLS) [10].

$$d_{min} = D_{min}(P^i) \quad (7)$$

$$d_{thr}^i = \mu(d_{max} - d_{min}) + d_{min} \quad (8)$$

where D_{max} and D_{min} are the maximum and minimum distances respectively in P^i set, $\mu(0 < \mu < 1)$ is a threshold parameter. For each testing patch, $Q^i(m, n)$ denotes the set of matching distances which are larger than d_{thr}^i :

$$Q^i(m, n) = \{D_k(X_{nn}(x_{A,i}(m, n))) | D_k(X_{nn}(x_{A,i}(m, n))) > d_{thr}^i, k = 1, 2, \dots, R_n\} \quad (9)$$

where D_k denotes the distance of the k -th nearest neighbor. The density factor can be defined as:

$$fa_{A,i}(m, n) = (R_n - |Q^i(m, n)|) / R_n \quad (10)$$

where $|Q^i(m, n)|$ is the number of element in $Q^i(m, n)$. Finally, the density salience value is obtained as follows:

$$Svalue_{density}(x_{A,i}(m, n)) = \exp(-fa_{A,i}(m, n) / 2\sigma^2) \quad (11)$$

3 Person Re-identification Algorithm

The whole algorithm of our person re-identification is presented in Algorithm 1.

3.1 Dense Correspondence for Patches

In this section, the dense correspondence proposed in [5, 9, 19] is employed to achieve patch alignment.

Each human image is densely segmented into some local patches. Dense SIFT descriptor and a LAB color histogram, named as dLabSift features are extracted

Algorithm 1. Algorithm of person re-identification

Input:

Probe person $x_{A,i}$ in Camera A
 Gallery person set x_B in Camera B
 Reference set R_n ;

Output:

Find the person x_{B,q^*} mostly matched with $x_{A,i}$
 1: **For** each patch $x_{A,i}(m, n)$ in $x_{A,i}$ **do**
 2: Compute the salience value $Svalue_{DD}(x_{A,i}(m, n))$ using Eq. (3)
 3: **Endfor**
 4: **For** each person $x_{B,j}$ in x_B **do**
 5: **For** each patch $x_{B,j}(m, n)$ in $x_{B,j}$ **do**
 6: Compute the salience value $Svalue_{DD}(x_{B,j}(m, n))$ using Eq. (3)
 7: **Endfor**
 8: **Endfor**
 9: **For** each person $x_{B,j}$ in x_B **do**
 10: **For** each patch $x_{A,i}(m, n)$ in $x_{A,i}$ **do**
 11: Find the corresponding patch $x_{B,j}(m', n')$ in $x_{B,j}$ using Eq.(14)
 12: Compute the similarity score $s(x_{A,i}(m, n), x_{B,j}(m', n'))$ using Eq.(15)
 13: **Endfor**
 14: Compute the similarity score $Sim(x_{A,i}, x_{B,j})$ using Eq.(16) for $x_{A,i}$ and $x_{B,j}$
 15: **Endfor**
 16: Find the person x_{B,q^*} mostly matched with $x_{A,p}$ as

$$q^* = \arg \max_q Sim(x_{A,p}, x_{B,q})$$

17: **return** x_{B,q^*} ;

for each patch with the dimension of $32 \times 3 \times 3 + 128 \times 3 = 672$ (more details can be found in [5]). Similar to work [5], for each image patch $x_{A,i}(m, n)$ in $x_{A,i}$, we first generate candidate patches in $x_{B,i}$ (shown in Fig. 6) as follows.

First, let $T_{A,i}(m)$ be the m -th row patches set of the i -th image of Camera A , *i.e.*,

$$T_{A,i}(m) = \{x_{A,i}(m, n) | n = 1, 2, \dots, N\} \quad (12)$$

Then, we obtain candidate patches in $x_{B,i}$, as

$$S(x_{A,i}(m, n), x_{B,j}) = \{T_{B,j}(a) | a \in \theta(m)\}, \forall x_{A,i}(m, n) \in T_{A,i}(m) \quad (13)$$

where $T_{B,j}(a)$ is the a -th row patches set of the j -th image of Camera B . $\theta(m)$ is the relaxation of adjacency search, as human images may have vertical misalignment.

After that, the optimal corresponding patch $x_{B,j}(m', n')$ is obtained from gallery, as

$$x_{B,j}(m', n') = \arg \max_{\bar{x} \in S_{i,j}} s(x_{A,i}(m, n), \bar{x}) \quad (14)$$

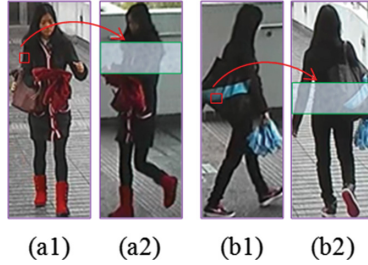


Fig. 6. Illustration of Adjacency Constrained Search. The green box region denotes search region of the patch in red box (Color figure online).

where $S_{ij} = S(x_{A,i}(m, n), x_{B,j})$, and $s(x, y)$ is computed as

$$s(x, y) = \exp\left(-\frac{d(x, y)^2}{2\sigma^2}\right) \quad (15)$$

where $d(x, y)$ is Euclidean distance between x and y .

3.2 Similarity Computation

The testing human images always contain a large part of backgrounds, which are various around pedestrians especially due to the change of camera viewpoint. Therefore, it's unreasonable to take all the testing patches for salience learning.

In this paper, the statistical information of human head position is utilized to remove some background patches before computing the similarity scores. Since the backgrounds around human lower part are normally the similar pavement, only the statistics of human head position are produced to remove the backgrounds around human head.

For each image $x_{A,i}$ in the probe, the matching result is obtained as the image in the gallery with the maximal similarity score $x_{B,j}$, while the similarity score between two images is obtained by accumulating the score of their each patch pair which can be formulated as:

$$Sim(x_{A,i}, x_{B,j}) = \sum_{m,n} \frac{Sa_{A,i}(m, n) \cdot s(x_{A,i}(m, n), x_{B,j}(m', n')) \cdot Sa_{B,j}(m', n')}{\varepsilon + |Sa_{A,i}(m, n) - Sa_{B,j}(m', n')|} \quad (16)$$

where $x_{A,i}$ and $x_{B,j}$ are collection of patch features of the probe image and the gallery image. ε is a parameter controlling the salience difference, and,

$$Sa_{A,i}(m, n) = Svalue_{DD}(x_{A,i}(m, n)) \quad (17)$$

$$Sa_{B,j}(m', n') = Svalue_{DD}(x_{B,j}(m', n')) \quad (18)$$

4 Experiments

The proposed method is evaluated on two publicly available datasets, VIPeR dataset [22] and CUHK01 dataset [8]. Both reflect most challenging problems of person re-identification applications such as human pose variation, cameras viewpoint and illumination change, occlusions between persons.

Evaluation Protocol. Similar to the work [5], each dataset is randomly partitioned into two parts, 50 % for salience learning and 50 % for testing and the standard Cumulated Matching Characteristics (CMC) curve is mainly evaluated during experiments. Images from camera A are used as probe and other images from camera B are used as gallery. For each probe image the rank of matching scores to the images in gallery is obtained. Rank-k recognition rate is the expectation of correct match at rank-k, and the cumulated values of recognition rate at all ranks are recorded as one-trial CMC result. 10 trials of evaluation are executed to achieve stable statistics. Our Density-Distance salience learning method is denoted as DdSal. As mentioned above, before computing the similarity scores the position information of human head is utilized to remove some background patches, this process is denoted as heaPri. The combination of Density-Distance salience learning method and the process of heaPri is denoted as PriDd.

4.1 Evaluations on VIPeR Dataset

VIPeR dataset [22] is one of the most challenging person re-identification dataset and it is captured by two cameras in outdoor with two images for each person shot from different viewpoints. The persons have been accurately detected from

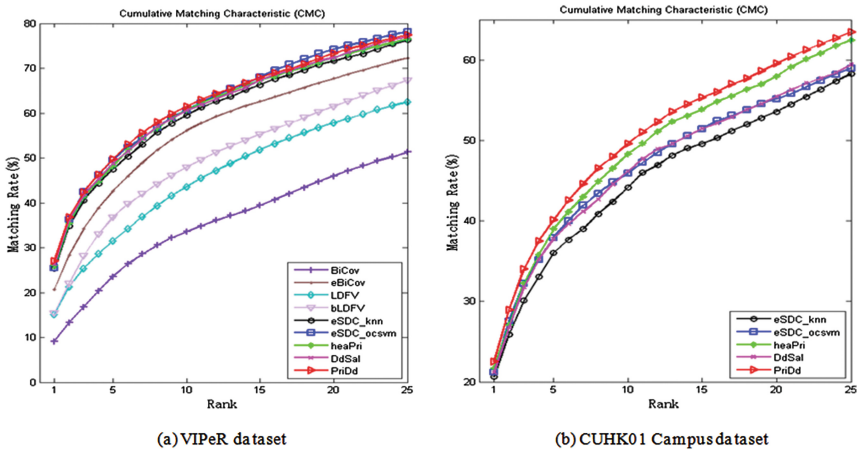


Fig. 7. CMC statistics on the VIPeR dataset and the CUHK01 Campus dataset. (a) On VIPeR dataset, our approach (heaPri, DdSal and PriDd) is compared with BiCov [16], eBiCov [16], LDFV [20], bLDFV [20], eSDC_knn [5] and eSDC_ocsvm [5]. (b) On CUHK01 Campus dataset, our method is compared with eSDC_knn and eSDC_ocsvm.

original videos. It contains 632 pedestrian pairs. Each person has two images from different cameras with different viewpoints, most of which are larger than 90 degree. All images are normalized to the same size of 128×48 [5].

On the VIPeR dataset, the results of DdSal, heaPri and PriDd are reported in Fig. 7(a), compared to several unsupervised methods including eSDC_knn [5], eSDC_ocsvm [5], BiCov [16], eBiCov [16], LDFV [20], bLDFV [20]. Generally speaking, (1) PriDd is slightly better than heaPri, DdSal which means both the Density-Distance salience information and the head background removal improve the matching results. (2) Our approaches and the eSDC-based methods outperform the other four methods. While our PriDd approach is slightly better than eSDC_knn salience learning approach and also slightly better than eSDC_ocsvm salience learning method from rank 1 to rank 15. It's worth mentioning that although eSDC_ocsvm slightly outperforms our approach after rank 15, it requires much higher computing cost.

Table 1. VIPeR dataset: top ranked matching rates in [%] with 316 persons.

Method	r = 1	r = 5	r = 10	r = 20
LMNN [4]	6.23	19.65	32.63	52.25
ITML [4]	11.61	31.39	45.76	63.86
PRDC [4]	15.66	38.42	53.86	70.09
PCCA [7]	19.27	48.89	64.91	80.28
BiCov [16]	9.08	23.58	33.59	45.97
ELF [25]	12.00	31.00	41.00	58.00
LDFV [20]	15.17	31.47	43.60	57.86
bLDFV [20]	15.39	36.77	47.93	61.43
SDALF [17]	19.87	38.89	49.37	65.73
eBiCov [16]	20.66	42.00	56.18	68.00
CPS [26]	21.84	44.00	57.21	71.00
eSDC_knn [5]	25.28	47.50	59.52	71.62
eSDC_ocsvm [5]	25.50	49.62	60.66	74.27
heaPri	25.88	48.22	60.72	72.40
DdSal	26.32	48.73	60.57	72.37
PriDd	27.05	49.68	61.45	73.35

More comparison results on VIPeR dataset with some supervised methods are reported in Table 1, which can also demonstrate the effectiveness of our approaches. eSDC_knn and eSDC_ocsvm seems to have satisfactory performance close to our approaches, however, comprehensively considering Table 2, which reports the running time of our PriDd method compared to eSDC_knn and eSDC_ocsvm on MATLAB (2013a), with 64-bit Win7, Intel core i7, 4.00 GHz CPU, our approaches can achieve better performance with acceptable computing complexity.

Table 2. VIPeR dataset: running time in [seconds] of three salience learning methods with 316 test persons.

Method	Trial1	Trial3	Trial5	Trial7	Trial9
eSDC_knn	0.601	0.578	0.600	0.588	0.597
eSDC_ocsvm	1196.959	1123.837	1260.092	1247.636	1219.456
PriDd	4.694	4.694	4.866	4.862	4.840

4.2 Evaluations on CUHK01 Dataset

The CUHK01 dataset [8] is also captured by two cameras in a campus environment with higher resolution compared to the VIPeR dataset. All the images are normalized to 160×60 for evaluation.

Table 3. CUHK01 dataset: top ranked matching rates in [%] with 316 persons.

Method	r = 1	r = 5	r = 10	r = 20
eSDC_knn	20.66	36.01	44.14	53.54
eSDC_ocsvm	21.20	37.91	45.91	55.16
heaPri	21.62	38.98	48.32	57.94
DdSal	21.13	37.72	46.10	55.41
PriDd	22.53	40.06	49.59	59.52

Table 4. CUHK01 dataset: running time in [seconds] of three salience learning methods with 316 test persons.

Method	Trial2	Trial4	Trial6	Trial8	Trial10
eSDC_knn	0.944	0.947	0.937	0.940	0.938
eSDC_ocsvm	2002.655	2061.217	1991.516	1972.510	2003.010
PriDd	7.213	6.970	7.085	7.005	7.093

For the concise presentation, only two outstanding methods, eSDC_knn and eSDC_ocsvm are implemented on CUHK01 dataset for comparison as shown in Fig. 7(b). From which we can see, our methods, especially PriDd method, have a significant improvement compared to eSDC_knn and eSDC_ocsvm. eSDC_ocsvm has almost identical performance as our DdSal method but much worse performance than our heaPri and PriDd methods, which means the background interference of human head has higher influence on CUHK01 dataset. More detailed comparison results on CUHK01 dataset are shown in Table 3. Same as the evaluation on VIPeR dataset, Table 4 reports the running time of our PriDd method compared to eSDC_knn and eSDC_ocsvm also on MATLAB (2013a),

with 64-bit Win7, Intel core i7, 4.00 GHz CPU, which also can demonstrate that our approaches can achieve better performance with acceptable computing complexity.

5 Conclusion

A novel unsupervised salience learning method is proposed for person re-identification, where a Density-Distance is designed for salience learning. Meanwhile, the statistics of pedestrian's head position is learnt to relieve the background interference during similarity matching. The experimental results on widely used VIPeR dataset and CUHK01 dataset demonstrate the outperformance of our method which improves matching precision with acceptable computing complexity comparing with the state of art methods.

Acknowledgments. Our thanks to the support from 863 (2014AA015104), the National Nature Science Foundation of China (61472002), the Natural Science Foundation of Anhui Higher Education Institutions of China (KJ2014A015), the Natural Science Foundation of Anhui Province (1508085QF127) and the Doctoral Research Foundation of Anhui University (02303203/32030059).

References

1. Prosser, B., Zheng, W., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: BMVC (2010)
2. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 91–102. Springer, Heidelberg (2011)
3. Dikmen, M., Akbas, E., Huang, T.S., Ahuja, N.: Pedestrian recognition with a learned metric. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part IV. LNCS, vol. 6495, pp. 501–512. Springer, Heidelberg (2011)
4. Zheng, W.-S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: CVPR (2011)
5. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: CVPR (2013)
6. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. PAMI **34**(10), 1915–1926 (2012)
7. Mignon, A., Jurie, F.: Pcca.: a new approach for distance learning from sparse pairwise constraints. In: CVPR (2012)
8. 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, Part I. LNCS, vol. 7724, pp. 31–44. Springer, Heidelberg (2013)
9. Liu, C., Gong, S., Loy, C.C., Lin, X.: Person re-identification: what features are important? In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 391–401. Springer, Heidelberg (2012)
10. Schwartz, W., Davis, L.: Learning discriminative appearance-based models using partial least squares. In: XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI) (2009)

11. Porikli F.: Inter-camera color calibration using cross-correlation model function. In: IEEE International Conference on Image Processing, pp. 133–136 (2003)
12. Li, W., Wang, X.: Locally aligned feature transforms across views. In: CVPR (2013)
13. Chilgunde, A., Kumar, P., Ranganath, S.: Multi-camera target tracking in blind regions of cameras with non-overlapping fields of view. In: British Machine Vision Conference, pp. 42.1–42.10 (2004)
14. Mazzon, R., Cavallaro, A.: Multi-camera tracking using a multi-goal social force model. *Neurocomputing* **100**, 41–50 (2013)
15. Javed, O., Rasheed, Z., Shafique, K., Shah, M.: Tracking across multiple cameras with disjoint views. In: Computer Vision (2003)
16. Ma, B., Su, Y., Jurie, F.: Bicov.: a novel image representation for person re-identification and face verification. In: British Machine Vision Conference (2012)
17. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: CVPR (2010)
18. Lu, Y., Lin, L., Zheng, W.-S.: Human re-identification by matching compositional template with cluster sampling. In: ICCV (2013)
19. Ma, K., Ben-Arie, J.: Vector array based multi-view face detection with compound exemplars. In: CVPR (2012)
20. Ma, B., Su, Y., Jurie, F.: Local descriptors encoded by fisher vectors for person re-identification. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 413–422. Springer, Heidelberg (2012)
21. Byers, S., Raftery, A.: Nearest-neighbor clutter removal for estimating features in spatial point processes. *J. Am. Stat. Assoc.* **93**, 577–584 (1998)
22. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS) (2007)
23. Papadimitriou, S., Kitagawa, H., Gibbons, P.B.: Loci: fast outlier detection using the local correlation integral. In: International Conference on Data Engineering, pp. 315–326. IEEE (2003)
24. Zhao, R., Ouyang, W., Wang, X.: Person Re-identification by Saliency Learning. arXiv preprint [arXiv:1412.1908](https://arxiv.org/abs/1412.1908) (2014)
25. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008)
26. Cheng, D., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: BMVC (2011)
27. Hido, S., Tsuboi, Y., Kashima, H.: Statistical outlier detection using direct density ratio estimation. *Knowl. Inf. Syst.* **26**(2), 309–336 (2011)
28. Breunig, M.M., Kriegel, H.P., Ng, R.T.: LOF: identifying density-based local outliers. In: ACM Sigmod International Conference on Management of Data, pp. 93–104 (2000)
29. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. *ACM Comput. Surv. (CSUR)* **41**(3), 15 (2009)
30. Kriegel, H.-P., Krger, P., Zimek, A.: Outlier detection techniques. In: Tutorial at the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (2009)