

RESEARCH

Open Access



Learning deep features from body and parts for person re-identification in camera networks

Zhong Zhang^{1,2*} and Tongzhen Si^{1,2}

Abstract

In this paper, we propose to learn deep features from body and parts (DFBP) in camera networks which combine the advantages of part-based and body-based features. Specifically, we utilize subregion pairs to train the part-based feature learning model and predict whether they belong to positive subregion pairs. Meanwhile, we utilize holistic pedestrian images to train body-based feature learning model and predict the identities of the input images. In order to further improve the discrimination of features, we concatenate the part-based and body-based features to form the final pedestrian representation. We evaluate the proposed DFBP on two large-scale databases, i.e., Market1501 database and CUHK03 database. The results demonstrate that the proposed DFBP outperforms the state-of-the-art methods.

Keywords: Camera networks, Deep feature learning, Person re-identification

1 Introduction

The camera networks, as a kind of wireless sensor networks, have received considerable attention due to the potential value for the practical applications [1–4]. One of the most important applications of camera networks is person re-identification (re-ID) which is an issue of searching the same person from one camera sensor across different camera sensors with a probe image. Although substantial works have been made in this field, it remains some unsolved problems. This is because the appearances of pedestrians from different cameras are easily affected by many environmental factors, such as views, body pose, and lighting.

The existing approaches focus on two fundamental problems, i.e., feature representation and metric learning. On one hand, a lot of efforts expect to develop discriminative feature representations that are robust to the changes in views, pose, and illumination. Gray and Tao [5] proposed the Ensemble of Localized Features (ELF), each of

which consisting of a feature channel, location, and binning information, to overcome the viewpoint changes. Bazzani et al. [6] presented the Symmetry-Driven Accumulation of Local Features (SDALFs) which considers symmetry and asymmetry perceptual principles to handle the environmental variances. Liao et al. [7] introduced the Local Maximal Occurrence (LOMO) features to analyze the horizontal occurrence of local features and maximize the occurrence in order to obtain a stable representation. On the other hand, some methods learn effective metrics. For example, Zheng et al. [8] proposed the Probabilistic Relative Distance Comparison (PRDC) to maximize the matching probability of positive pairs so that the distance between the positive pair is smaller than that of negative pair. Koestinger et al. [9] proposed the KISSME to learn an effective metric using equivalence constraints and had a good effect on the generalization performance. Liao et al. [7] introduced the Cross-view Quadratic Discriminant Analysis (XQDA) to simultaneously learn a discriminant low-dimensional subspace and a metric function on the derived subspace.

Recently, the convolutional neural network (CNN) is applied to person re-ID and achieves attractive performance on the large-scale databases. There are mainly two

*Correspondence: zhong.zhang8848@gmail.com

¹Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin, China

²College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin, China

kinds of CNN models, i.e., verification models and identification models. The verification models [10–12] take pairs of images as the input and determine whether they belong to the same person or not, and the structure is shown in Fig. 1a. The verification models treat the person re-ID as a binary-class classification task and map two images of the same person into the nearby points; otherwise, the points are far apart. However, the verification models are not explicitly utilize re-ID labels. To make full use of the re-ID labels, the identification models [13, 14], which treat the person re-ID as a multi-class recognition task, are proposed, and the structure is shown in Fig. 1b. The identification models usually take holistic images as the input and determine whether images belong to their own labels. The identification models have shown great potentials on current large-scale person re-ID databases. Xiao et al. [15] jointly trained a identification model using multiple datasets and proposed a new dropout function to deal with a large number of classes. Furthermore, Zheng et al. [16] combined the verification and identification models by jointly learning three loss functions.

The above methods utilize the holistic pedestrian images as the input and learn the body-based features, which discards the local characteristics of pedestrians. To overcome this drawback, Yi et al. [17] split a pedestrian image into three horizontal parts and trained three part-CNNs. Cheng et al. [18] used the multi-channel CNN to learn local part features and modified the triplet loss function that requires the intra-class feature distances to be less than the inter-class ones. These existing part-based CNN methods learn the local features which are gradually ignored with the increase of neural network layers.

In this paper, we propose to learn deep features from body and parts (DFBP), which combines the advantages of part-based and body-based features. In the learning process, we simultaneously utilize two kinds of models, i.e.,

verification models and identification models, to learn and extract features. For learning part-based features, we split each pedestrian image into several horizontal subregions. We define the corresponding subregions from the same person as the positive subregion pairs and corresponding subregions from different persons as the negative subregion pairs. We train the verification models using the positive and negative subregion pairs and extract the subregion features from fully connected layer. Then, we utilize the weighted adding to aggregate all the subregion features from one pedestrian image and obtain the part-based feature. Meanwhile, we utilize the holistic images to train an identification model for the body-based features. Since we utilize two kinds of CNNs where they are also feed into different kinds of training samples, the learned part-based and body-based features are complementary. In order to further improve the discrimination of features, we concatenate the part-based and body-based features to form the final pedestrian representation.

The rest of this paper is organized as follows. In Section 2, we introduce the proposed method and the implementation process in detail. In Section 3, we verify the proposed DFBP on two large-scale databases, i.e., Market 1501 [19] and CUHK03 [20], and the experimental results demonstrate that the proposed DFBP outperforms the state-of-the-art methods. Finally, we conclude this paper in Section 4.

2 Method

The proposed DFBP is to learn powerful features by combining the part-based and body-based features in camera networks. The flowchart of the proposed DFBP is shown in Fig. 2. In this section, we introduce our model from three aspects, i.e., the part-based feature learning model, the body-based feature learning model, and the final representation for pedestrian images.

2.1 Part-based feature learning model

Part-based features could provide the effective local information which plays an important role in person re-ID. In order to learn part-based features, we apply the verification model and split each image into several horizontal subregions. We define the subregion pairs from the same person as positive subregion pairs as shown in Fig. 3a, and the subregion pairs from different person as negative subregion pairs as shown in Fig. 3b.

There are two CNN models in the part-based feature learning model as shown in Fig. 2a. The two networks share the same weights which can map subregion pairs into the same feature subspace. Here, we utilize the widely used ResNet-50 [21] pretrained on ImageNet [22] as the CNN model. In the training stage, we resize all the subregions divided from pedestrian images into 256×256 and subtract the mean values computed from all the training

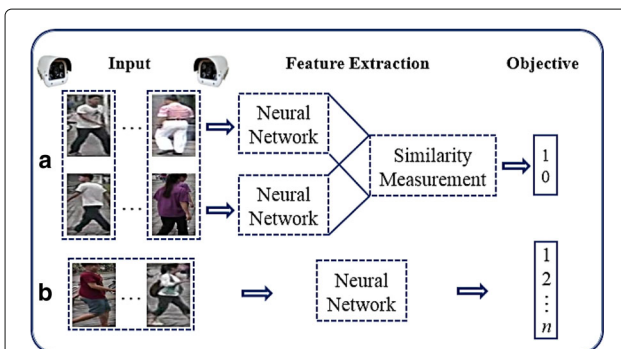


Fig. 1 The structures of **a** verification models and **b** identification models. **a** Verification models take pairs of pedestrian images as the input and predict whether they are the same person or not. **b** Identification models take pedestrian images as the input and predict their identity

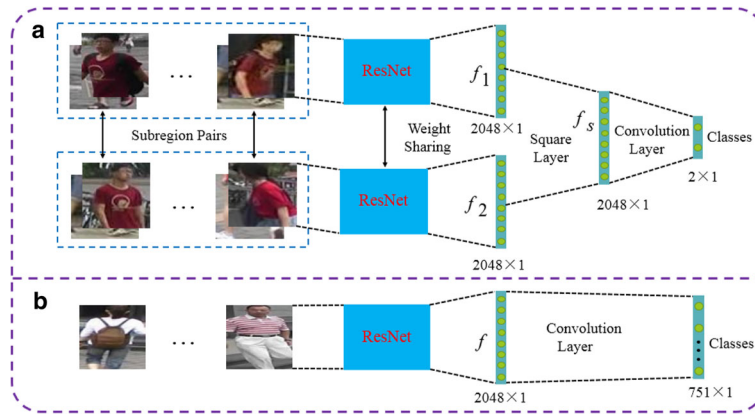


Fig. 2 The structures of **a** part-based and **b** body-based feature learning models

subregions. Then, each subregion is randomly cropped to 224×224 . Given subregion pairs as the input, we obtain two feature vectors f_1 and f_2 which are generated by average pooling at the end of ResNet-50 models as shown in Fig. 2a. In order to connect the two networks, we employ a non-parametric layer called square layer [16] to compare the features f_1, f_2 . The input of square layer is f_1 and f_2 , and the output is formulated as:

$$f_s = (f_1 - f_2)^2 \quad (1)$$

where f_1, f_2 , and f_s are the 2048 dimensional vectors for the ResNet-50 model.

After this layer, we utilize the dropout strategy [23], which introduces zero values at special locations, to regularize the model. Then, we utilize a convolutional layer to turn the resulting vector f_s to a 2-dimensional vector (p_1, p_2) which represents the prediction probability of the input subregion pair belonging to the same person. The convolutional layer takes f_s as the input and has two kernels with the size of $1 \times 1 \times 2048$. We do not add Rectified

Linear Unit (ReLU) after this layer and utilize the softmax to predict the probability:

$$y_i = \frac{e^{a_i}}{\sum_{k=1}^2 e^{a_k}}, \forall i \in 1, 2 \quad (2)$$

where a_i is the activation value of the i -th neuron in the last convolutional layer.

We utilize the cross-entropy as the cost function of part-based feature learning model which is formulated as:

$$L_V = - \sum_{i=1}^2 p_i \log y_i \quad (3)$$

where $y_i \in [0, 1]$ is the prediction probability and p_i is the true probability. If the input is the positive subregion pair, then $p_1=1$ and $p_2=0$; otherwise, $p_1=0$ and $p_2=1$.

Since the two CNN models share the same weights, we can extract subregion features from any ResNet-50 model. We treat f_1 or f_2 as the feature for each subregion.

2.2 Body-based feature learning model

Body-based features could provide holistic information of pedestrian, and they are complementary to part-based

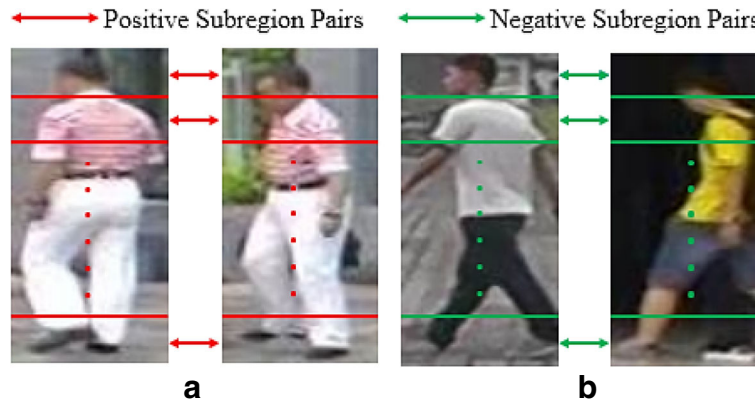


Fig. 3 Illustration of the **a** positive and **b** negative subregion pairs

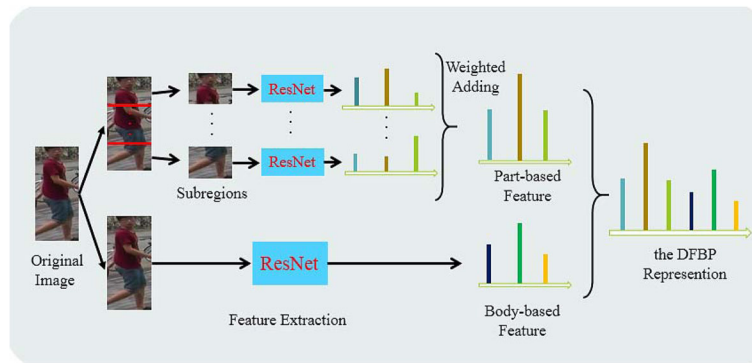


Fig. 4 The flowchart of the proposed DFBP extraction

features. We adopt the identification model for learning body-based features as shown in Fig. 2b and also utilize the ResNet-50 [21] as the CNN model. We take the holistic pedestrian images as the input of ResNet-50 model and utilize the same preprocessing for each pedestrian image as mentioned in Section 2.1. Then, we obtain the feature $f \in \mathbb{R}^{2048 \times 1}$ for each pedestrian image which are generated by the average pooling at the end of ResNet-50 model. Similarly, we employ the dropout strategy to prevent overfitting. Then, we utilize the convolutional layer to turn the resulting vector f into a C dimensional vector where each element represents the prediction probability of the input pedestrian image belonging to one label. The class number of pedestrian is C , so the convolutional layer has C kernels with the size of $1 \times 1 \times 2048$. For example, since the Market1501 database has 751 pedestrian classes, C is set to 751. We do not add ReLU after this layer. We also use the softmax unit to normalize the output and obtain the prediction probability. It is formulated as:

$$y_j = \frac{e^{a_j}}{\sum_{k=1}^C e^{a_k}}, \forall j \in 1, 2, \dots, C \quad (4)$$

where a_j is the activation value of the j -th neuron in the last convolutional layer.

We utilize the cross-entropy loss for body-based feature learning model:

$$L_I = - \sum_{j=1}^K q_j \log y_j \quad (5)$$

where $y_j \in [0, 1]$ is the prediction probability of the input belonging to the j -th class and q_j is the true probability. If n is the true class, then $q_n=1$; otherwise, $q_j=0$ for all j .

After training the body-based feature learning model, we extract f as the body-based feature for each pedestrian image.

2.3 Final representation

In order to obtain a powerful feature representation, we combine the part-based features and the body-based features to form DFBP. The flowchart of the combination strategy is shown in Fig. 4.

We split each pedestrian image into several subregions, and the number of subregions is the same as mentioned in Section 2.1. Then, we extract the feature for each subregion and aggregate them by weighted adding to obtain the part-based feature:

$$P = \alpha_1 p_1 + \alpha_2 p_2 + \dots + \alpha_n p_n \quad (6)$$

Table 1 The rank-1 precious (%) and mAP (%) on the Market-1501 database

Method	Single rank-1	Query mAP	Multi rank-1	Query mAP
DADM [26]	39.40	19.6	49.00	25.8
BoW+KISSME [19]	44.42	20.26	–	–
MST-CNN [27]	45.1	–	55.40	–
MR-CNN [28]	45.58	26.11	56.59	32.26
FisherNet [29]	48.15	29.94	–	–
CAN [30]	48.24	24.43	–	–
SL [31]	51.90	26.11	56.59	32.26
SCSP [31]	51.90	26.35	–	–
DNS [32]	55.43	29.87	71.56	46.03
S-LSTM [33]	–	–	61.60	35.3
Gate Reid [34]	65.88	39.55	76.04	48.45
SOMAnet [35]	73.87	47.89	81.29	56.98
MSCAN [36]	75.45	52.41	83.43	62.03
PIE [37]	78.65	53.87	–	–
Verif-Identif [16]	79.51	59.87	85.84	70.33
Part-based features	36.25	14.47	37.86	13.51
Body-based features	80.29	59.34	86.81	71.18
DFBP	<i>81.71</i>	<i>60.86</i>	<i>87.02</i>	<i>72.21</i>

The data in italics are the best result in each evaluation protocol

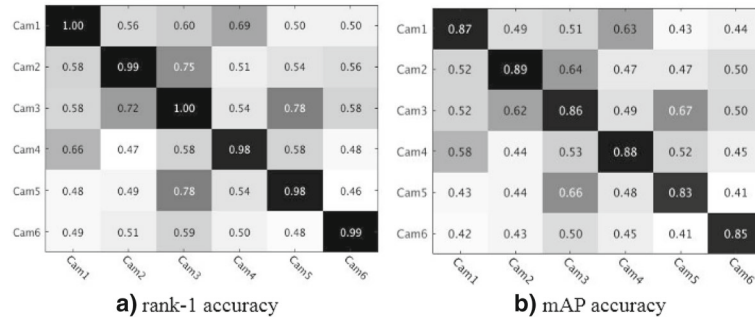


Fig. 5 Re-ID performance between camera pairs on Market1501 database. **a** Rank-1 accuracy. **b** mAP accuracy

where p_i ($i = 1, 2, \dots, n$) represents the feature of the i -th subregion in a pedestrian image, α_i ($i = 1, 2, \dots, n$) is the corresponding coefficient to control the weight of the i -th subregion, and n is the total number of subregions in a pedestrian image.

Finally, we concatenate the part-based feature and body-based feature to form the final DFBP representation:

$$F = [\eta P, \varphi B] \quad (7)$$

where $[\cdot, \cdot]$ indicates directly concatenating the two vectors, B represents the body-based feature, η and φ are the corresponding coefficients to control the weights of P and B respectively, and $F \in \mathbb{R}^{4096 \times 1}$ represents the final feature DFBP. We simply use $\eta = 1$ and $\varphi = 1$. This may not be the best choice for a particular database. But if we do not fore-know the distribution of the database, it is a simple and straight-forward choice.

3 Experiments

In this section, we verify the effectiveness of the proposed DFBP on two large-scale databases. We first introduce the experimental setup in Section 3.1 and then show the

results on the Market1501 database [19] and the CUHK03 database [20] in Sections 3.2 and 3.3 respectively.

3.1 Experimental setup

We implement the proposed model using the MatConvNet package [24]. In the training phase, in order to fine-tune the pre-trained ResNet-50 [21], we modify the last convolution layer to have 751 kernels for the Market1501 database and 1367 kernels for the CUHK03 database. We set the maximum number of training epochs to 75. The learning rate is initialized as 0.1 and decreased to 0.01 after 70 epochs. We utilize stochastic gradient descent (SGD) to update the parameters and set the weight decay to 0.0005. The batch size for training part-based and body-based feature models is 24. In the dropout layer, we set the dropout rate to 0.5. For the part-based feature learning model, we split each pedestrian image into three horizontal subregions and empirically set $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, and $\alpha_3 = 0.2$. As the input of part-based feature learning model, the ratio between positive subregion pairs and negative subregion pairs is 1:1.

3.2 Market1501 database

Market1501 database [19] is a newly released large-scale person re-ID database which is collected in a university campus. The database contains 32,669 annotated bounding boxes of 1501 identities. According to the database setting, it includes 12,936 training images of 751 identities and 19,732 test images of 750 identities. In addition, there

Table 2 Comparison with the state-of-the-art methods on the CUHK03 dataset under the single-shot setting

Method	Rank-1	Rank-5	Rank-10	mAP
KISSME [38]	11.7	33.3	48.0	–
DeepReID [39]	19.9	49.3	64.7	–
BoW+HS [19]	24.3	–	–	–
LOMO+XQDA [7]	46.3	78.9	88.6	–
SI-CI [40]	52.2	84.3	94.8	–
DNS [32]	54.7	80.1	88.3	–
SOMAnet [35]	72.4	92.1	95.8	–
Part-based features	17.8	46.7	64.3	24.8
Body-based features	74.2	92.5	96.8	78.3
DFBP	75.9	93.8	97.9	79.9

The data in italics are the best result in each evaluation protocol

Table 3 Comparison with the state-of-the-art methods on the CUHK03 dataset using the multi-shot setting

Method	Rank-1	Rank-5	Rank-10	mAP
S-LSTM [33]	57.3	80.1	88.3	46.3
Gate-SCNN [34]	68.1	88.1	94.6	58.8
Part-based features	20.6	42.8	55.5	19.8
Body-based features	81.0	90.3	95.0	76.2
DFBP	81.8	91.8	95.7	78.0

The data in italics are the best result in each evaluation protocol

are 3368 query images about the same 750 identities as the test set. Images of each identity are captured by at most six manually set cameras, and each person has 3.6 images on average at each viewpoint. All images are automatically detected by the Deformable Part based Model (DPM) [25]. So the database is close to realistic setting.

In the experiments, we select a pedestrian from the query set and retrieve the truth images in the test set. We utilize the mean average precision (mAP) and rank-1 accuracy to evaluate the performance. Here, rank- i accuracy denotes the probability whether one or more correctly matched images appear in top- i . The evaluation results of the single query setting and multiple query setting are shown in Table 1. From the results, we can see that the proposed DFBP achieves the highest classification accuracy of 81.71% (rank-1) and 60.86% (mAP) in

the single query setting, and 87.01% (rank-1) and 72.21% (mAP) in the multiple query setting. The performance of the proposed DFBP is better than the part-based features and body-based features, because we make full use of the complementarity between the part-based features and body-based features. The proposed DFBP and Verif.-Identif [16] utilize the verification and identification models. Noticeably, the proposed DFBP outperforms the Verif.-Identif in all situations due to considering the structural information of pedestrian by learning part-based features.

We further evaluate the performance of the proposed DFBP between all camera pairs, and the results are shown in Fig. 5. The two matrices are confusion matrices in the single query setting. In the two matrices, the accuracy q_{ij} ($1 \leq i \leq 6$, $1 \leq j \leq 6$) indicates the pedestrian in

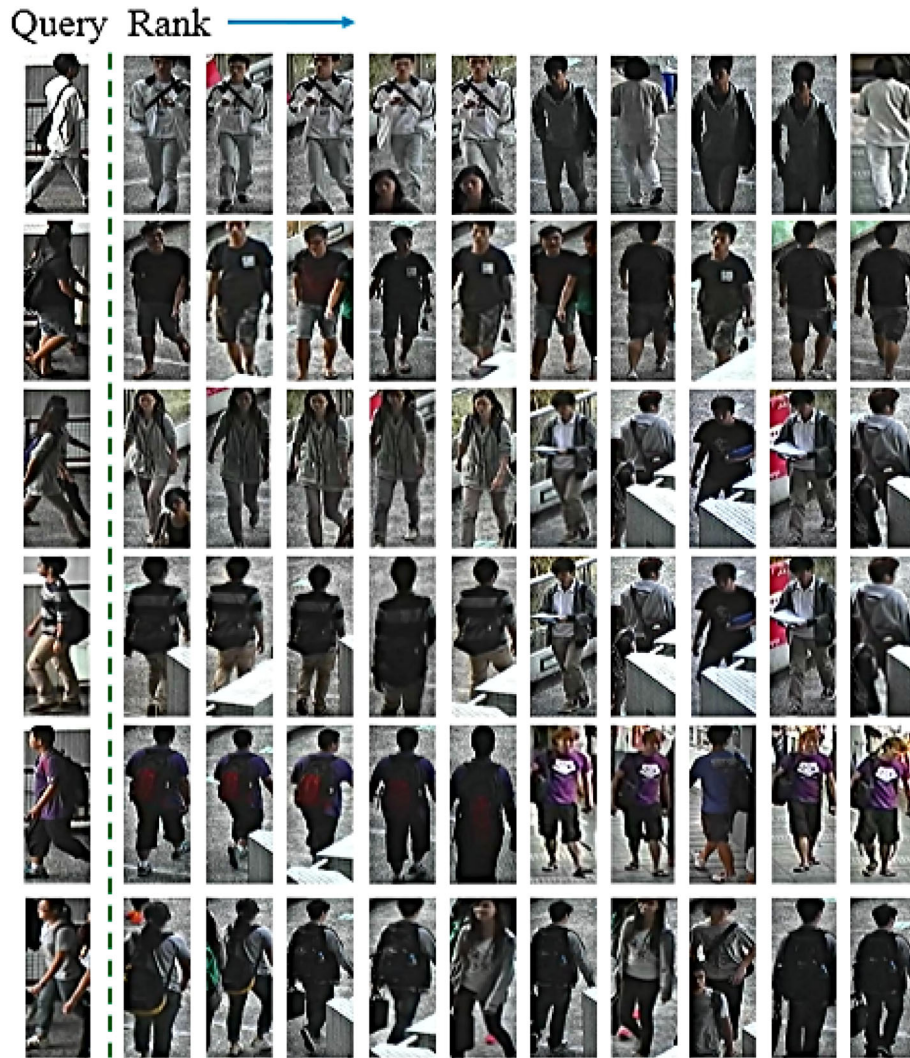


Fig. 6 Samples of pedestrian retrieval on CUHK03 database in the multi-shot setting. The first column indicates query images, and remaining columns are gallery images which are sorted according to the scores from left to right

the i -th camera as the query image and the pedestrian in the j -th camera as the gallery image with the same labels. The camera 6 is a 720×576 low-resolution camera and captures distinct background with the other high-resolution cameras. From the results, we can see that the re-ID accuracies between camera 6 and the others are relatively high. We also compute the average cross-camera accuracy to obtain 50.17% mAP and 56.93% rank-1 accuracy, respectively. The average cross-camera indicates the average of all accuracies in the matrix except the accuracies on the diagonal of the matrix in Fig. 5. The proposed DFBP largely improves the performance compared with the previous reported results, such as 10.51% mAP, 13.72% rank-1 accuracy in [19] and 48.42% mAP, 54.42% rank-1 accuracy in [16]. These results indicate that the proposed DFBP could also learn the discriminative features under different viewpoints.

3.3 CUHK03 database

CUHK03 database [20] includes 14,097 images of 1467 identities which is captured by six cameras over months in the Chinese University of Hong Kong (CUHK) campus. Each pedestrian is observed by two disjoint camera views and has an average of 4.8 images in each view. The database provides two kinds of bounding boxes, i.e., manually annotated bounding boxes and bounding boxes detected by DPM [25]. Since DPM is closer to the realistic setting, we evaluate the proposed DFBP on the bounding box detected by DPM. We separate the database into training data including 1367 pedestrians and test data including 100 pedestrians. This partitioning is independently implemented 10 times, and the average is reported. We evaluate the proposed DFBP by the single-query setting and multiple-query setting.

We utilize the rank- i and mAP to evaluate the proposed DFBP, and the results in the single-shot setting are shown in Table 2. There is only a correct image in the gallery set from the other camera. From the Table 2, we can see that the proposed DFBP achieves the highest classification accuracy of 75.9% rank-1 accuracy, 93.8% rank-5 accuracy, 97.9% rank-10 accuracy, and 79.9% mAP. This demonstrates the effectiveness of the proposed DFBP once again.

In addition, we evaluate the proposed DFBP in the multi-shot setting as shown in Table 3. We use all the images from the other camera as gallery images, and the number of the images is about 500. From Table 3, we can see that the proposed DFBP also achieves the highest accuracy owing to combining the part-based features and body-based features. Furthermore, we represent some re-ID samples in Fig. 6. The first column shows query images and the remaining columns are retrieved images which are sorted according to the similarity scores from left to right. Most retrieved images are correct. Although

there are several incorrect images, such as the last eight images in the sixth row, we think it is a reasonable prediction because the pedestrians are similar to the query in appearance and color.

4 Conclusions

In this paper, we have proposed a novel feature learning method named DFBP in sensor networks, which simultaneously considers the part-based features and body-based features. We utilize subregion pairs and holistic pedestrian images to train part-based and body-based feature learning models respectively. By integrating the two kinds of features, we obtain a discriminative feature. We have proved the effectiveness of the proposed DFBP on two databases which are realistic and challenging, and the proposed DFBP have achieved better performance than the state-of-the-art methods.

Abbreviations

DFBP: Deep feature learning method from body and parts; CNN: Convolutional neural network; DPM: Deformable part-based model; mAP: Mean average precision; SGD: Stochastic gradient descent; ReLU: Rectified linear unit; CUHK: The Chinese University of Hong Kong

Acknowledgements

The authors would like to thank the editor and anonymous reviewers for their helpful comments and suggestions in improving the quality of this paper.

Funding

This work was supported by National Natural Science Foundation of China under Grant No. 61711530240 and No. 61401309, Natural Science Foundation of Tianjin under Grant No. 17JCZDJC30600, the Fund of Tianjin Normal University under Grant No. 135202RC1703, the Open Projects Program of National Laboratory of Pattern Recognition under Grant No. 201700001, and the China Scholarship Council No. 201708120040.

Availability of data and materials

The databases are available online.

Authors' contributions

ZZ contributed to the main idea and algorithm design of the study. TS operated the experiments and ZZ analyzed the results. ZZ and TS drafted and revised the manuscript. Both authors read and approved the final manuscript.

Authors' information

Zhong Zhang received the Ph.D. degree from Institute of Automation, Chinese Academy of Sciences. He is currently an Associate Professor at Tianjin Normal University.

Tongzhen Si is currently pursuing the M.S. degree at Tianjin Normal University.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 February 2018 Accepted: 14 February 2018

Published online: 05 March 2018

References

1. D Deif, Y Gadallah, A comprehensive wireless sensor network reliability metric for critical internet of things applications. *EURASIP J. Wirel. Commun. Netw.* **2017**(1), 145 (2017)

2. F Zhao, X Sun, H Chen, R Bie, Outage performance of relay-assisted primary and secondary transmissions in cognitive relay networks. *EURASIP J. Wirel. Commun. Netw.* **2014**(1), 60 (2014)
3. F Zhao, L Wei, H Chen, Optimal time allocation for wireless information and power transfer in wireless powered communication systems. *IEEE Trans. Veh. Technol.* **65**(3), 1830–1835 (2016)
4. F Zhao, B Li, H Chen, X Lv, Joint beamforming and power allocation for cognitive mimo systems under imperfect csi based on game theory. *Wirel. Pers. Commun.* **73**(3), 679–694 (2013)
5. D Gray, H Tao, in *European Conference on Computer Vision*. Viewpoint invariant pedestrian recognition with an ensemble of localized features (Springer, Marseille, 2008), pp. 262–275
6. L Bazzani, M Cristani, V Murino, Symmetry-driven accumulation of local features for human characterization and re-identification. *Comp. Vision Image Underst.* **117**(2), 130–144 (2013)
7. S Liao, Y Hu, X Zhu, S Li, in *IEEE Conference on Computer Vision and Pattern Recognition*. Person re-identification by local maximal occurrence representation and metric learning (IEEE, Boston, 2015), pp. 2197–2206
8. W Zheng, S Gong, T Xiang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Person re-identification by probabilistic relative distance comparison (IEEE, Colorado Springs, 2011), pp. 649–656
9. M Koestinger, M Hirzer, P Wohlhart, in *IEEE Conference on Computer Vision and Pattern Recognition*. Large scale metric learning from equivalence constraints (IEEE, Providence, 2012), pp. 2288–2295
10. E Ahmed, M Jones, T Marks, in *IEEE Conference on Computer Vision and Pattern Recognition*. An improved deep learning architecture for person re-identification (IEEE, Boston, 2015), pp. 3908–3916
11. S Ding, L Lin, G Wang, H Chao, Deep feature learning with relative distance comparison for person re-identification. *Pattern Recog.* **48**(10), 2993–3003 (2015)
12. F Zhao, W Wang, H Chen, Q Zhang, Interference alignment and game-theoretic power allocation in mimo heterogeneous sensor networks communications. *Signal Process.* **126**, 173–179 (2016)
13. L Zheng, Z Bie, J Wang, C Su, in *European Conference on Computer Vision*. Mars: A video benchmark for large-scale person re-identification (Springer, Amsterdam, 2016), pp. 868–884
14. S Wu, YC Chen, X Li, AC Wu, JJ You, WS Zheng, in *IEEE Winter Conference on Applications of Computer Vision*. An enhanced deep feature representation for person re-identification (IEEE, Lake Placid, 2016), pp. 1–8
15. T Xiao, H Li, W Ouyang, X Wang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Learning deep feature representations with domain guided dropout for person re-identification (IEEE, Seattle, 2016), pp. 1249–1258
16. Z Zheng, L Zheng, Y Yang, A discriminatively learned cnn embedding for person re-identification. *ACM Trans. Multimedia Comput. Commun. Appl.* **14**(1), 1–20 (2017)
17. D Yi, Z Lei, S Li, in *International Conference on Pattern Recognition*. Deep metric learning for person re-identification (IEEE, Stockholm, 2014), pp. 34–39
18. D Cheng, Y Gong, S Zhou, J Wang, N Zheng, in *IEEE Conference on Computer Vision and Pattern Recognition*. Person re-identification by multi-channel parts-based cnn with improved triplet loss function (IEEE, Seattle, 2016), pp. 1335–1344
19. L Zheng, L Shen, L Tian, S Wang, J Wang, Q Tian, in *IEEE International Conference on Computer Vision*. Scalable person re-identification: A benchmark (IEEE, Santiago, 2015), pp. 1116–1124
20. W Li, R Zhao, T Xiao, X Wang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Deepreid: Deep filter pairing neural network for person re-identification (IEEE, Columbus, 2014), pp. 152–159
21. K He, X Zhang, S Ren, J Sun, in *IEEE Conference on Computer Vision and Pattern Recognition*. Deep residual learning for image recognition (IEEE, Seattle, 2016), pp. 770–778
22. O Russakovsky, J Deng, H Su, J Krause, S Satheesh, S Ma, Z Huang, A Karpathy, A Khosla, M Bernstein, Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015)
23. N Srivastava, G Hinton, A Krizhevsky, I Sutskever, R Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
24. A Vedaldi, K Lenc, in *ACM Multimedia*. Matconvnet: Convolutional neural networks for matlab (ACM, Brisbane, 2015), pp. 689–692
25. P Felzenszwalb, R Girshick, D Mcallester, D Ramanan, Object detection with discriminatively trained part-based models. *Trans. Pattern Anal. Mach. Intell.* **32**(9), 1627–1645 (2010)
26. C Su, S Zhang, J Xing, W Gao, Q Tian, in *European Conference on Computer Vision*. Deep attributes driven multi-camera person re-identification (Springer, Amsterdam, 2016), pp. 475–491
27. J Liu, ZJ Zha, Q Tian, D Liu, T Yao, Q Ling, T Mei, in *ACM Multimedia*. Multi-scale triplet cnn for person re-identification (ACM, Amsterdam, 2016), pp. 192–196
28. E Ustinova, Y Ganin, V Lempitsky, in *IEEE International Conference on Advanced Video and Signal Based Surveillance*. Multi-region bilinear convolutional neural networks for person re-identification (IEEE, Lecce, 2017), pp. 1–6
29. L Wu, C Shen, A van den Hengel, Deep linear discriminant analysis on fisher networks. *Pattern Recognit.* **65**, 238–250 (2017)
30. H Liu, J Feng, M Qi, J Jiang, S Yan, End-to-end comparative attention networks for person re-identification. *CoRR*. **abs/1606.04404** (2017). <https://arxiv.org/abs/1606.04404>
31. D Chen, Z Yuan, B Chen, N Zheng, in *IEEE Conference on Computer Vision and Pattern Recognition*. Similarity learning with spatial constraints for person re-identification (IEEE, Seattle, 2016), pp. 1268–1277
32. L Zhang, T Xiang, S Gong, in *IEEE Conference on Computer Vision and Pattern Recognition*. Learning a discriminative null space for person re-identification (IEEE, Seattle, 2016), pp. 1239–1248
33. R Viorior, B Shuai, J Lu, D Xu, G Wang, in *European Conference on Computer Vision*. A siamese long short-term memory architecture for human re-identification (Springer, Amsterdam, 2016), pp. 135–153
34. R Viorior, M Haloi, G Wang, in *European Conference on Computer Vision*. Gated siamese convolutional neural network architecture for human re-identification (Springer, Amsterdam, 2016), pp. 791–808
35. I Barros, M Cristani, B Caputo, A Rognhaugen, T Theoharis, Looking beyond appearances: Synthetic training data for deep cnns in re-identification. *CoRR*. **abs/1701.03153** (2017). <https://arxiv.org/abs/1701.03153>
36. D Li, X Chen, Z Zhang, K Huang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Learning deep context-aware features over body and latent parts for person re-identification (IEEE, Hawaii, 2017), pp. 384–393
37. L Zheng, Y Huang, H Lu, Y Yang, Pose invariant embedding for deep person re-identification. *CoRR*. **abs/1701.07732** (2017). <https://arxiv.org/abs/1701.07732>
38. M Koestinger, M Hirzer, P Wohlhart, P Roth, H Bischof, in *IEEE Conference on Computer Vision and Pattern Recognition*. Large scale metric learning from equivalence constraints (IEEE, Providence, 2012), pp. 2288–2295
39. W Li, R Zhao, T Xiao, X Wang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Deepreid: Deep filter pairing neural network for person re-identification (IEEE, Columbus, 2014), pp. 152–159
40. F Wang, W Zuo, L Lin, D Zhang, L Zhang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Joint learning of single-image and cross-image representations for person re-identification (IEEE, Seattle, 2016), pp. 1288–1296

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)