IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 3, 2018, accepted July 7, 2018, date of publication July 10, 2018, date of current version July 30, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2854830

Pedestrian Retrieval via Part-Based Gradation Regularization in Sensor Networks

SHUANG LIU, (Member, IEEE), XIAOLONG HAO, AND ZHONG ZHANG, (Member, IEEE)
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China

Corresponding author: Shuang Liu (shuangliu.tjnu@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501327, Grant 61711530240, and Grant 61501328, in part by the Natural Science Foundation of Tianjin under Grant 17JCZDJC30600, in part by the Fund of Tianjin Normal University under Grant 135202RC1703, in part by the Open Projects Program of the National Laboratory of Pattern Recognition under Grant 201700001 and Grant 201800002, in part by the China Scholarship Council under Grant 201708120039 and Grant 201708120040, and in part by the Tianjin Higher Education Creative Team Funds Program.

ABSTRACT In this paper, we propose a novel label distribution approach named part-based gradation regularization (PGR) for pedestrian retrieval in sensor networks. Considering different importance of various body parts, we present a gradual function to assign pedestrian label for each horizontal part. In this way, we can conduct part-based supervised learning using the identification network. The proposed PGR not only learns the discriminative local convolutional neural network-based features, but also considers the significance of assigning pedestrian label for different horizontal parts. Experimental results show that the proposed PGR obtains better performance than other approaches on three pedestrian retrieval databases, i.e., Market-1501, CUHK03, and DukeMTMC-reID databases.

INDEX TERMS Sensor networks, pedestrian retrieval, local CNN-based features.

I. INTRODUCTION

Pedestrian retrieval is a significant application of sensor networks due to the potential value for criminal investigation and critical event monitoring [1]–[4]. It aims at searching a specific pedestrian from a large gallery database collected from sensor networks. The key challenges for pedestrian retrieval in sensor networks include pose misalignment, occlusion, illumination changes, etc.

The traditional methods for pedestrian retrieval usually focus on feature representation and metric learning [5], [6]. The feature representation methods [7], [8] employ handcrafted features, such as colors, textures, shapes, etc., to obtain the discriminative features for pedestrian images. The metric learning methods [9]–[11] learn an effective metric measurement to compute the similarity between two pedestrian images. The learned metric measurement could maximize inter-class variations and minimize intra-class variations, and meanwhile could map high dimensional features into a low dimensional space.

With the rise of deep learning, the convolutional neural network (CNN) has attracted more attention in pedestrian retrieval due to its superior performance. One of the most important reasons is that the CNN model is an end-to-end system which jointly optimizes feature representation and distance metric using the back propagation algorithm, while the traditional methods treat feature representation and metric learning as two independent modules leading to sub-optimal solutions. From the aspect of feature representation, there are mainly two types of CNN-based features, global and local CNN-based features. The global CNN-based features are usually learned by feeding holistic pedestrian images into the CNN model [12], [13]. However, the global CNN-based features ignore the structure information which is a key cue for describing pedestrians. Hence, the local CNN-based features [14], [15] are extensively developed to learn the structure information of pedestrians. Some local CNN-based features [16], [17] first divide the original pedestrian images or global features, and then apply them to train the CNN models.

In order to utilize the pedestrian label, Zheng *et al.* [18], [19] employed the Identification Embedding model (IDE), i.e., classification model, to extract the global CNN-based features. The IDE achieves better results than many local CNN-based features, which demonstrates the strong discrimination ability of IDE. Meanwhile, the local CNN-based features could discover the structure information

of pedestrian which is essential for promoting the performance. This inspires us to train the IDE using a local way. We intend to divide the whole image into several horizontal parts so as to learn local features. However, the horizontal part cannot be assigned the same label as the whole pedestrian image as the following reasons. Firstly, the appearance of each horizontal part is pretty different from the corresponding whole image. Secondly, one horizontal part contains a part of identity information and cannot represent the whole pedestrian image as shown in Fig. 1. Hence, it is difficult to assign a pedestrian label to each horizontal part, and this supervised information is essential to train the IDE. It should be noticed that each horizontal part comes from the specific whole image, which encourages us to assign a pedestrian label related to the original label.

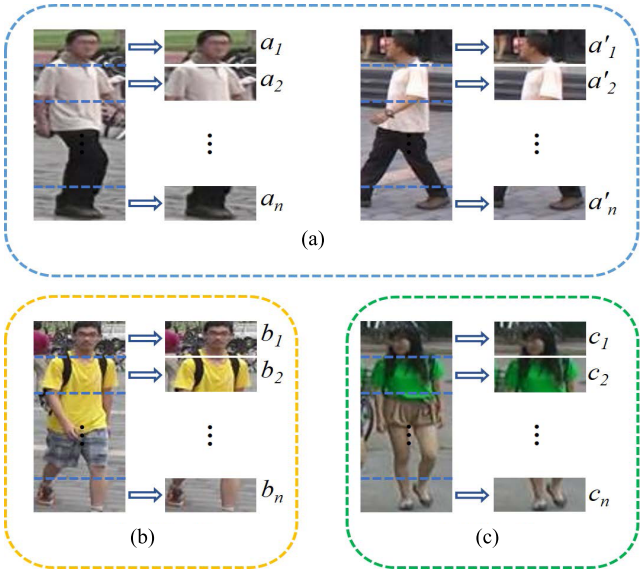

FIGURE 1. Partition a pedestrian image into several horizontal parts, where (a) is the same pedestrian from different camera sensors while (b) and (c) are different pedestrians. a_i and a'_i ($i \in 1, 2, \ldots, n$) express the i-th horizontal part of pedestrian in (a). Similarly, b_i and c_i express the i-th corresponding horizontal part in (b) and (c), respectively.

In this paper, we combine the strengths of the local CNN-based features and the discriminative IDE model, and propose a novel method named Part-based Gradation Regularization (PGR) to assign a pedestrian label for each horizontal part. As shown in Fig. 1, we observe that the upper parts of a pedestrian image contain more identity information than the lower parts, and therefore the proposed PGR utilizes a gradual function to assign pedestrian label for each horizontal part. Meanwhile, the proposed PGR assigns the corresponding parts from the same pedestrian (a_i and a'_i) to the same label distribution, and the corresponding parts from different pedestrians (b_i and c_i) to different label distributions. With the labeled horizontal parts, we fine-tune the IDE model, and treat the fully-connected layer outputs as the horizontal part features. Finally, all the features of horizontal parts in a pedestrian image are concatenated to generate the final representation. We conduct extensive experiments on three large-scale pedestrian retrieval databases, and the results demonstrate the validity of the proposed PGR.

In summary, this paper has the following contributions: 1) the proposed PGR can assign a pedestrian label for each horizontal part, which is essential to training the IDE model; 2) it combines the superiority of local CNN-based features and IDE model to represent pedestrian images; 3) it increases the training samples using the horizontal parts to train the CNN model, which improves the generalization ability of the network and mitigates the risk of overfitting.

The rest of this paper is organized as follows. We present the related work in Section 2. In Section 3, we give a description of the proposed PGR in detail. The experimental results on three large-scale pedestrian retrieval databases, i.e., Market1501 [20], CUHK03 [21] and DukeMTMC-reID, are provided in Section 4. Finally, we conclude this paper in Section 5.

II. RELATED WORK

A. TRADITIONAL METHODS IN PEDESTRIAN RETRIEVAL

The feature representation is vital for traditional pedestrian retrieval methods. To learn the robust features, a lot of approaches are proposed. For example, Ahmed *et al.* [22] presented a novel efficient descriptor named the Hybrid Spatiogram and Covariance Descriptor (HSCD) for pedestrian retrieval where they fuse the spatiogram feature with the covariance feature. To consider the mean and covariance information of pixel, Matsukawa *et al.* [23] proposed the hierarchical Gaussian descriptor to represent the local regions of pedestrian images. Variord *et al.* [24] utilized a data-driven framework to encode pixel values by jointly learning a linear transformation and a dictionary. Moreover, there are other robust feature representations including ELF [6], SDALF [25], LOMO [26], etc.

After acquiring feature vectors of pedestrian images, the similarity scores should be computed for the subsequent matching. Since the Euclidean distance is a computationally efficient metric, it is widely used in pedestrian retrieval for sensor networks [27]–[29]. However, the performance is undesirable because the Euclidean distance is pre-defined and poorly adapts to environmental changes. Hence, many researchers resort to metric learning algorithms which minimize the distance between images from the same pedestrian, and meanwhile maximize the distance between images from different pedestrians. Xiong *et al.* [30] introduced a closed-form approach called the Kernel Local Fisher Discriminant Classifier (KLFDA), which utilizes the kernel trick to handle the high dimensional features and meanwhile maximizes the Fisher criteria. Liao *et al.* [26] presented the Cross-view Quadratic Discriminant Analysis (XQDA) to obtain the discriminative metric in a low dimensional subspace. Jose and Fleuret [31] presented the Weighted Approximate Rank Component Analysis (WARCA), which combines a simple regularizer of orthonormal linear mappings to optimize the precision at top ranks. In addition to the above

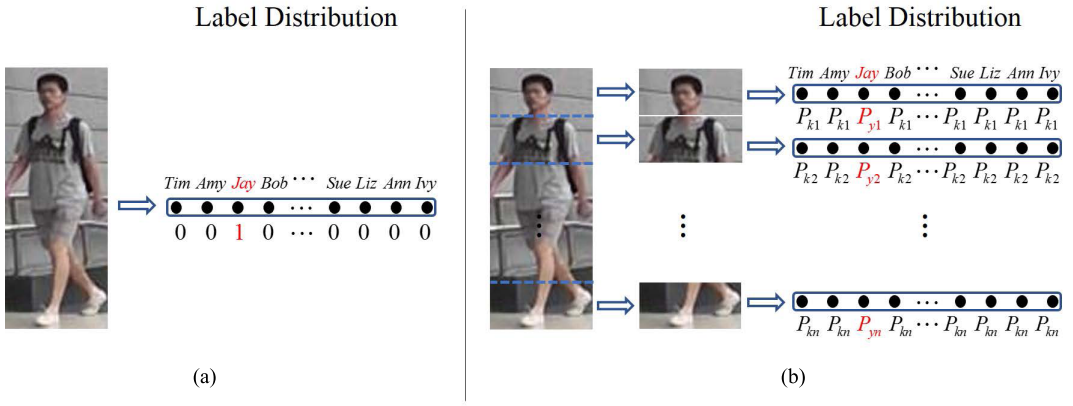

FIGURE 2. The label distributions of (a) the classical label distribution and (b) the proposed PGR. The proposed PGR utilizes a gradual function to assign labels, where $p_{y1} > p_{y2} > \cdots > p_{yn}$.

metric learning methods, there are KISSME [32], LF [33], CFML [34], etc.

B. DEEP LEARNING METHODS IN PEDESTRIAN RETRIEVAL

Presently, CNN has been widely used in pedestrian retrieval for sensor networks [35], [36]. Some methods employ the holistic pedestrian images to train the CNN model. For example, Xiao *et al.* [13] presented the Domain Guided Dropout algorithm to learn global CNN-based features from multiple domains. Wu *et al.* [12] proposed the deep end-to-end neural network to learn high-level global CNN-based features by utilizing smaller convolutional filters. Qian *et al.* [37] introduced a multi-scale deep learning model to mine the global CNN-based features from different scales. Compared to the global CNN-based features, the local CNN-based features could learn the spatial structure of pedestrian. Yi *et al.* [14] divided an images into three overlapped parts and trained three CNNs to capture different statistical properties of pedestrian images. Cheng *et al.* [15] divided the pedestrian images into several regions, where they capture salient local CNN-based features by the body-part convolution layers. Zhao *et al.* [16] considered different importances of various body parts, and merged body-part features using a tree-structured fusion network. Besides, Sun *et al.* [17] utilized a uniform partition strategy to learn discriminative part-informed features.

The classification model could improve the performance of pedestrian retrieval due to rational utilization of pedestrian labels. For instance, in [18], the IDE model was proposed to directly classify the pedestrian images. Wu *et al.* [38] presented a classification model named Feature Fusion Net (FFN) which combines the hand-crafted and CNN-based features. In order to train a deeper classification model, Xiao *et al.* [13] gathered multiple databases and presented the Domain Guided Dropout (DGD) algorithm to process these pedestrian images. In this paper, we utilize a number of horizontal parts labeled by the proposed PGR to train a classification model.

III. METHOD

Firstly, the Label Smoothing Regularization (LSR) is briefly review. Then, we present the proposed PGR in detail. Finally, we introduce details of network.

A. REVIEW OF THE LABEL SMOOTHING REGULARIZATION

Since the IDE model is a supervised learning algorithm, it is necessary to assign the label distribution for each training sample which indicates the probability belonging to each identity. The classical method only assign the ground truth class label to 1, and the others to 0 as shown in Fig. 2(a). Specifically, the classical label distribution is formulated as:

$$q_c(k) = \begin{cases} 0 & k \neq y \\ 1 & k = y \end{cases} \tag{1}$$

where y is the ground truth class label.

The classical label distribution results in the network to be tuned towards the ground truth class, which ignores the contributions of non-ground truth classes. Afterwards, the LSR [39] was proposed to assign small values to non-ground truth classes, which encourages the model to be less confident towards the ground truth class. The label distribution of LSR is formulated as:

$$q_{lsr}(k) = \begin{cases} \dfrac{\varepsilon}{K} & k \neq y \\ 1 - \varepsilon + \dfrac{\varepsilon}{K} & k = y \end{cases} \tag{2}$$

where $\varepsilon \in [0, 1]$ is a parameter to adjust the proportion of non-ground truth labels, and K is the total number of classes.

The LSR achieves better performance than the classical label distribution by considering the non-ground truth classes. However, the LSR is hard to assign a label to each horizontal part because its ground truth label is unknown.

B. PART-BASED GRADATION REGULARIZATION

To address the above mentioned problem, we propose the PGR to assign the label to the horizontal part. Considering different identity information of various body parts, the proposed PGR utilizes a gradual function to assign pedestrian

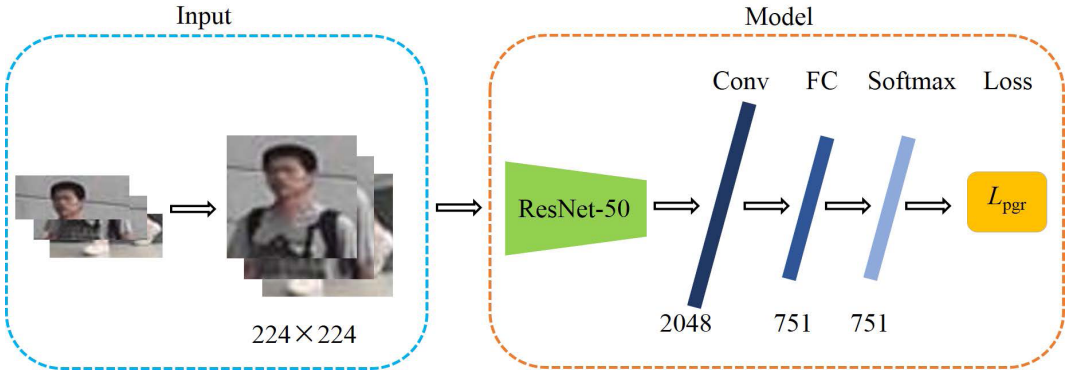

FIGURE 3. The proposed network architecture.

label for each horizontal part as shown in Fig. 2(b). The label distribution of the proposed PGR is defined as:

$$q_{pgr}(k,n) = \begin{cases} \frac{\alpha^n \varepsilon}{K} & k \neq y \\ 1 - \alpha^n \varepsilon + \frac{\alpha^n \varepsilon}{K} & k = y \end{cases} \tag{3}$$

where $n \in \{1, 2, \ldots, N\}$ represents the n-th horizontal part in the pedestrian image, N is the number of horizontal parts in each pedestrian image, ε is a smooth parameter, and $\alpha > 1$ is a hyperparameter which determines the probability of different horizontal parts. It is noticed that the label distribution of the proposed PGR degenerates into the classical method when ε is set to 0. We empirically set ε to 0.1. From Eq. 3, we can see that the proposed PGR assigns the upper parts to larger probability, and gradually reduces to the lower parts.

In this work, the cross-entropy function is employed as the loss measurement. The cross-entropy loss for the n-th horizontal part can be defined as follows:

$$L_{pgr}(n) = -\sum_{k=1}^{K} q_{pgr}(k,n) \log(p(k)) \tag{4}$$

where $k \in \{1, 2, \ldots, K\}$ is the pre-defined classes in the training set, and $p(k)$ represents the predicted probability of the k-th class.

To conveniently expressing Eq. 3, we denote $\frac{\alpha^n \varepsilon}{K}$ as q_k, and denote $1 - \alpha^n \varepsilon + \frac{\alpha^n \varepsilon}{K}$ as q_y. By substituting Eq. 3 into Eq. 4, the proposed loss function can be rewritten as:

$$L_{pgr}(n) = -(q_y - q_k) \log(p(y)) - q_k \sum_{k=1}^{K} \log(p(k)) \tag{5}$$

where $p(y)$ and $p(k)$ are the predicted probability of the y-th and k-th classes, respectively. They are obtained by the softmax function:

$$p(y) = \frac{\exp(x_y)}{\sum_{i=1}^{K} \exp(x_i)} \tag{6}$$

$$p(k) = \frac{\exp(x_k)}{\sum_{i=1}^{K} \exp(x_i)} \tag{7}$$

where x_k represents the k-th output value of the fully-connected layer. By substituting Eq. 6 and Eq. 7 into Eq. 5, our loss function can be simplified as:

$$L_{pgr}(n) = \log(\sum_{k=1}^{K} \exp(x_k)) - q_k \sum_{k=1}^{K} x_k - (q_y - q_k) x_y \tag{8}$$

The Stochastic Gradient Descent (SGD) is employed to optimize neural network parameters. In the back propagation process, we take the derivative of the loss function with respect to x_y:

$$L'_{pgr}(n) = \frac{\exp(x_y)}{\sum_{k=1}^{K} \exp(x_k)} - q_y \tag{9}$$

C. DETAILS OF NETWORK

The architecture of the proposed network is shown in Fig. 3. Given an input image, we first split the image into N horizontal parts and assign a pedestrian label to each horizontal part according to the porposed PGR. Then, each horizontal part is adjusted to 224 × 224 pixels and subtract the mean RGB values computed on the training set. Next, we treat the ResNet-50 [40] as the CNN model, and the architecture of ResNet-50 is shown in Fig. 4. Specifically, the first layer is a convolutional layer followed by the maxpooling layer. After these two layers, the size of input image is reduced by a quarter. The next four parts, Conv2-x, Conv3-x, Conv4-x, Conv5-x, are made up of three blocks, four blocks, six blocks and three blocks, respectively. Each block contains a projection shortcut and three various convolutional layers. The size of filters in three convolutional layers is 1×1, 3×3 and 1×1, respectively, where the first convolutional layer aims at reducing the output dimensions and the third convolutional layer aims at increasing the output dimensions. The last unit is the average pooling which turns the previous output into a tensor (1×1×2,048). Note that the batch normalization (BN) is adopt after each convolutional layer. Finally, we displace the last fully-connected layer with a new one for fine-tuning the network. The dimensionality of new fully-connected layer is 751 for the Market1501 database, 1,367 for the CUHK03 database, and 702 for the DukeMTMC-reID

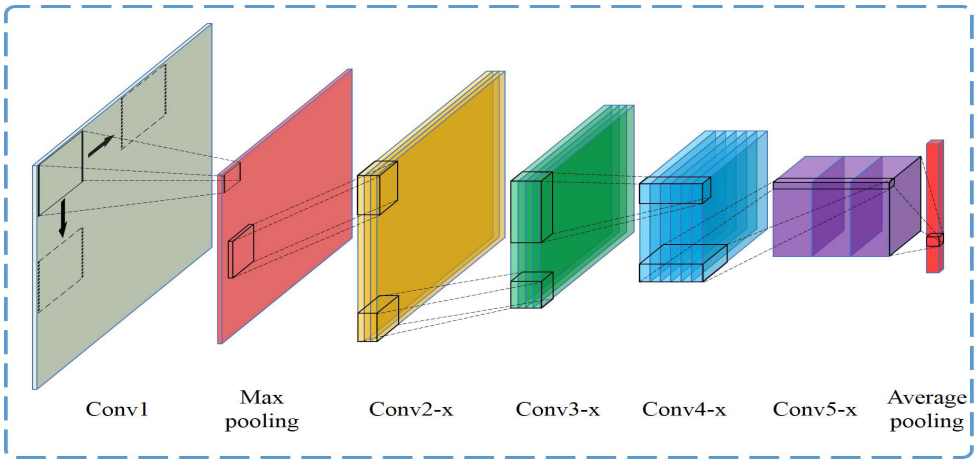

FIGURE 4. The network of the ResNet-50 model.

database. After the fully-connected layer, we then append a softmax unit. During the training stage, the batch is set to 64, and the epochs is set to 50. We treat the output of the last convolutional layer (2,048-dim) as the feature vector for each horizontal part. All the feature vectors of horizontal parts are concatenated to form the final representation for each pedestrian image. Hence, the final feature representation is a $(2{,}048\times N)$-dim vector. The final ranking result is obtained by calculating the Cosine distance of each query image and all gallery images.

IV. EXPERIMENTAL RESULTS

The proposed PGR is evaluated on three large-scale databases (Market-1501 database [20], CUHK03 database [21] and DukeMTMC-reID database [41]). The performance of the proposed PGR is assessed by the standard protocols, rank-1 accuracy and mean average precision (mAP).

A. DATABASES

Market-1501 [20] is a highly popular database for pedestrian retrieval. It contains 32,668 annotated pedestrian images with 1,501 identities, and every identity has an average of 17.2 images. All the images are automatically detected by the deformable part model (DPM) [42]. Moreover, this database is divided into three parts (training set, test set and distractors). There are 12,936 cropped images with 751 identities for training, and 19,732 cropped images with 750 identities for testing (gallery and query). All the images are collected by six camera sensors in a university campus, so this database is close to the real setting. During the evaluation process, we retrieve the real images from the gallery images for 3,368 query images. We utilize two settings, i.e., single query and multiple query, to evaluate the proposed PGR.

CUHK03 [21] is one of the largest pedestrian retrieval databases captured by two disjoint camera sensors in the CUHK campus. It includes 14,097 cropped images of 1,467 pedestrians, and each pedestrian has an average of 9.6 images. There are two image sets for the CUHK03 database, i.e., labeled set and detected set. As for the labeled set, the bounding boxes are produced by hand-drawn. As for the detected set, the bounding boxes are produced by the DPM detector [42]. Experiments are conducted on the detected set, because this set is closer to the real setting. For fair comparison, we utilize the same setting as [26] and [44] that 1,367 pedestrians can be selected as a training set and the rest of 100 pedestrians can be selected as a test set.

DukeMTMC-reID [41] is one of the most challenging pedestrian retrieval databases up to now. All images are captured by eight different camera sensors on the Duke University campus, and are manually annotated by [41]. It involves three parts, i.e., training set, query set and gallery set. There are 16,522 images with 702 identities for training set, 2,228 images with the other 702 identities for query set, and 17,661 images for gallery set. Moreover, single query setting is used to evaluate the proposed approach.

B. SETTINGS

We take ResNet-50 [40] as the CNN model, and then implement the proposed PGR via MatConvNet package [44]. The learning rate is 0.1 and 0.01 for the first 40 epochs and the last 10 epochs, respectively. We empirically set $N = 3$ and $\alpha = 1.1$ in Eq. 3 which are the number of horizontal parts in each pedestrian image and control the probability of different horizontal parts, respectively. We regard the output of the last convolutional layer as a feature vector for each horizontal part resulting in a 2,048-dim vector. Therefore, the feature dimension for each pedestrian image is $2{,}048\times 3{=}6{,}144$. In the evaluation stage, the Cosine distance is used to calculate the similarity between two images. Furthermore, we employ the whole pedestrian images with the classical label distributions to train the ResNet-50 model, and treat this method as the baseline (Basel.). Note that the baseline results are on par with the previous work [18], [19].

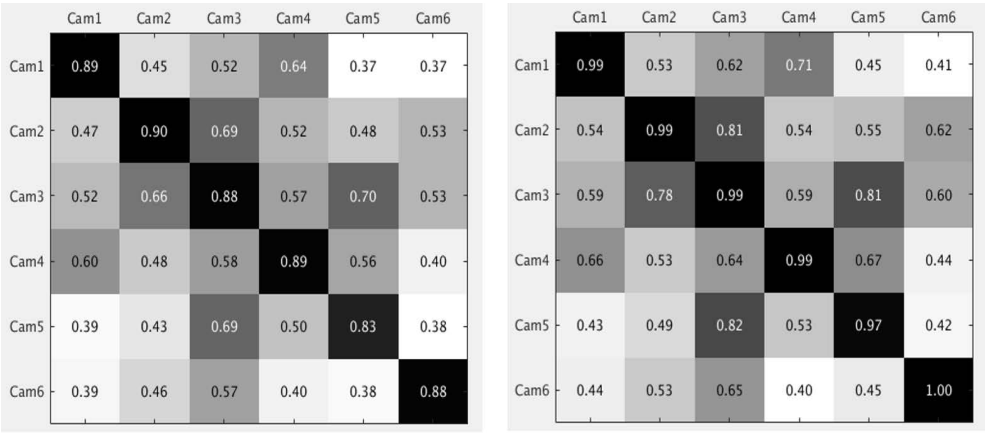

FIGURE 5. Pedestrian retrieval performance between camera sensor pairs on Market-1501. (a) Pairwise mAP. (b) Pairwise rank-1 accuracy.

C. EVALUATION ON MARKET-1501

Table 1 shows the results on the Market-1501 database, where the proposed PGR achieves the best results compared with other methods. Specifically, the proposed PGR achieves the result of 83.31% and 59.69% in the rank-1 accuracy and mAP for the single query, respectively. As for the multiple query, the proposed PGR obtains the result of 87.97% rank-1 accuracy and 71.21% mAP, respectively. Compared with the Basel., we obtain 9.62% rank-1 and 8.21% mAP improvements for the single query, and 10.03% rank-1 and 13.54% mAP improvements for the multiple query. The superiority mainly lies in two reasons. Firstly, the local CNN-based features are utilized to describe the pedestrian images, which plays a vital role in discovering the structural information of pedestrian. Secondly, the proposed PGR assigns different pedestrian label to each horizontal part, which could further improve the feature discrimination. It is worth mentioning that the result of the Basel. exceeds many previous works [26], [45] on this database.

Since the pedestrian images on this database are captured by six different camera sensors, we evaluate the performance between camera sensor pairs as well. The results are shown in Fig. 5. We can see that although the camera sensors have different resolutions, their accuracies are relatively high with each other. Furthermore, we also evaluate the proposed PGR using the cross-camera average rank-1 and the cross-camera average mAP. The cross-camera average rank-1 accuracy is the mean accuracy that computes the average rank-1 accuracy of each camera sensor to other camera sensor. In other word, it computes the mean of non-diagonal elements of matrix in Fig. 5(a). Similarly, the cross-camera average mAP is the mean accuracy by computing the average mAP accuracy of each camera sensor to other camera sensor. The proposed PGR achieves 57.53% cross-camera average rank-1 and 50.73% cross-camera average mAP. These results are considerably better than the previous results in [20], i.e., 13.72% cross-camera average rank-1 accuracy and 10.51% cross-camera average mAP. This indicates the superiority of the proposed PGR in dealing with different camera sensors. Noticeably, these results are also the best results for the published literatures [20], [46].

TABLE 1. Comparison of the proposed PGR with other approaches reported on the Market-1501 database.

Approach	Single Query	(%)	Multi Query	(%)
	rank-1	mAP	rank-1	mAP
BoW+KISSME [20]	44.42	20.26	-	-
MR-CNN [48]	45.58	26.11	56.59	32.26
FisherNet [49]	48.15	29.94	-	-
CAN [50]	48.24	24.43	-	-
SL [10]	51.90	26.11	56.59	32.26
DNS [11]	55.43	29.87	71.56	46.03
Gate Reid [51]	65.88	39.55	76.04	48.45
SOMAnet [52]	73.87	47.89	81.29	56.98
PIE [53]	78.65	53.87	-	-
ResNet-50.+LSRO [44]	78.06	56.23	85.12	68.52
Basel. [19]	73.69	51.48	81.47	63.95
PGR	**83.31**	**59.69**	**87.97**	**71.21**

D. EVALUATION ON CUHK03

On this database, we randomly select one image for each pedestrian from the other camera sensor as gallery, and repeat 10 times to calculate their average rank-1 accuracy and mAP. The results are listed in Table 2 where the proposed PGR achieves the best results once again. Specifically, we achieve the results of 74.9% rank-1, 92.9% rank-5, 96.7% rank-10 and 78.8% mAP for the single shot, which outperforms other reported algorithms. Compared with the Basel., the proposed PGR yields 3.4% and 3.0% improvements for the rank-1 accuracy and the mAP, respectively. It is because the proposed PGR not only learns the local CNN-based features but also possesses the regularization ability. It is noticed that the

TABLE 2. Comparison of the proposed PGR with other approaches reported on the CUHK03 database.

Approach	rank-1	rank-5	rank-10	mAP
KISSME [54]	11.7	33.3	48.0	-
DeepReID [21]	19.9	49.3	64.7	-
Bow+HS [20]	24.3	-	-	-
LOMO+XQDA [26]	46.3	78.9	88.6	-
SI-CI [55]	52.2	84.3	94.8	-
DNS [11]	54.7	80.1	88.3	-
SOMAnet [52]	72.4	92.1	95.8	-
ResNet-50.+LSRO [44]	73.1	92.7	96.7	77.4
Basel. [19]	71.5	91.5	95.9	75.8
PGR	**74.9**	**92.9**	**96.7**	**78.8**

Basel. itself is higher than some previous works [26], [45], because it utilizes the discriminative IDE model.

E. EVALUATION ON DUKEMTMC-REID

As shown in Table 3, the proposed PGR achieves the best results of 72.29% rank-1 and 52.20% mAP. These results yield 7.07% and 7.21% improvements for the rank-1 accuracy and the mAP compared with the Basel. There are two main reasons for improvements. On the one hand, the proposed PGR can mine the structural information to represent pedestrians. On the other hand, the proposed PGR produces a lot of extra labeled training samples, which improves the generalization ability of the model and mitigates the risk of overfitting.

TABLE 3. Comparison of the proposed PGR with other approaches reported on the DukeMTMC-reID database.

Approach	rank-1	mAP
BoW+KISSME [20]	25.13	12.17
LOMO+XQDA [26]	30.75	17.04
ResNet-50.+LSRO [44]	67.68	47.13
OIM [56]	68.1	47.4
APR [57]	70.69	51.88
PAN [58]	71.59	51.51
Basel. [19]	65.22	44.99
PGR	**72.29**	**52.20**

V. CONCLUSION

In this paper, we have proposed a novel pedestrian retrieval method named PGR, which assigns pedestrian labels for horizontal parts. Concretely, we start from dividing the pedestrian image into several horizontal parts, and then allocate a pedestrian label to each horizontal part based on the proposed PGR. Finally, we utilize horizontal parts with the pedestrian labels to train the identification network in a supervised manner. Experiments on three challenging pedestrian retrieval databases, Market-1501, CUHK03 and DukeMTMC-reID, show that the proposed PGR outperforms the other methods.

REFERENCES

[1] N. Aparajita, P. K. Sa, S. K. Choudhury, S. Bakshi, and B. Majhi, "A neuromorphic person re-identification framework for video surveillance," *IEEE Access*, vol. 5, pp. 6471–6482, 2017.

[2] Z. Zhang, C. Wang, B. Xiao, W. Zhou, and S. Liu, "Action recognition using context-constrained linear coding," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 439–442, Jul. 2012.

[3] Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, and C. Shi, "Cross-view action recognition via a continuous virtual path," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2690–2697.

[4] Z. Zhang, C. Wang, B. Xiao, W. Zhou, and S. Liu, "Robust relative attributes for human action recognition," *Pattern Anal. Appl.*, vol. 18, no. 1, pp. 157–171, Feb. 2015.

[5] J. Guo, Y. Zhang, Z. Huang, and W. Qiu, "Person re-identification by weighted integration of sparse and collaborative representation," *IEEE Access*, vol. 5, pp. 21632–21639, 2017.

[6] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2008, pp. 262–275.

[7] L. Lin, R. Huang, X. Li, F. Zhang, and M. Ye, "Person re-identification by optimally organizing multiple similarity measures," *IEEE Access*, vol. 5, pp. 26034–26045, Nov. 2017.

[8] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Person re-identification using spatial covariance regions of human body parts," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Aug. 2010, pp. 435–440.

[9] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 649–656.

[10] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1268–1277.

[11] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1239–1248.

[12] L. Wu, C. Shen, and A. van den Hengel. (2016). "PersonNet: Person re-identification with deep convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1601.07255

[13] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1249–1258.

[14] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Aug. 2014, pp. 34–39.

[15] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based CNN with improved triplet loss function," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1335–1344.

[16] H. Zhao *et al.*, "Spindle Net: Person re-identification with human body region guided feature decomposition and fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1077–1085.

[17] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang. (2017). "Beyond part models: Person retrieval with refined part pooling." [Online]. Available: https://arxiv.org/abs/1711.09349

[18] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, "Person re-identification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1367–1376.

[19] L. Zheng, Y. Yang, and A. G. Hauptmann. (2016). "Person re-identification: Past, present and future." [Online]. Available: https://arxiv.org/abs/1610.02984

[20] L. Zheng, L. Sheng, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1116–1124.

[21] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. CVPR*, Jun. 2014, pp. 152–159.

[22] M. Zeng, Z. Wu, C. Tian, L. Zhang, and L. Hu, "Efficient person re-identification by hybrid spatiogram and covariance descriptor," in *Proc. CVPRW*, Jun. 2015, pp. 48–56.

[23] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical gaussian descriptor for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1363–1372.

[24] R. R. Varior, G. Wang, J. Lu, and T. Liu, "Learning invariant color features for person reidentification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3395–3410, Jul. 2016.
[25] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Comput. Vis. Image Understand.*, vol. 117, no. 2, pp. 130–144, Feb. 2013.
[26] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. CVPR*, Jun. 2015, pp. 2197–2206.
[27] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Nov. 2010, pp. 501–512.
[28] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by Fisher vectors for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 413–422.
[29] F. Zhao, B. Li, H. Chen, and X. Lv, "Joint beamforming and power allocation for cognitive MIMO systems under imperfect CSI based on game theory," *Wireless Pers. Commun.*, vol. 73, no. 3, pp. 679–694, 2013.
[30] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 1–16.
[31] C. Jose and F. Fleuret, "Scalable metric learning via weighted approximate rank component analysis," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 875–890.
[32] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2288–2295.
[33] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. IEEE Conf. CVPR*, Jun. 2013, pp. 3318–3325.
[34] B. Alipanahi, M. Biggs, and A. Ghodsi, "Distance metric learning vs. Fisher discriminant analysis," in *Proc. IEEE Int. Conf. Artif. Intell.*, Jul. 2013, pp. 598–603.
[35] F. Zhao, W. Wang, H. Chen, and Q. Zhang, "Interference alignment and game-theoretic power allocation in MIMO heterogeneous sensor networks communications," *Signal Process.*, vol. 126, pp. 173–179, May 2015.
[36] F. Zhao, H. Nie, and H. Chen, "Group buying spectrum auction algorithm for fractional frequency reuse cognitive cellular systems," *Ad Hoc Netw.*, vol. 58, pp. 239–246, Apr. 2017.
[37] X. Qian, Y. Fu, Y. G. Jiang, T. Xiang, and X. Xue, "Multi-scale deep learning architectures for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 5399–5408.
[38] S. Wu, Y.-C. Chen, X. Li, A.-C. Wu, J.-J. You, and W.-S. Zheng, "An enhanced deep feature representation for person re-identification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–8.
[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[41] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. Eur. Conf. Comput. Vis. Workshop*, Nov. 2016, pp. 17–35.
[42] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[43] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3754–3762.
[44] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. ACM Multimedia*, Sep. 2015, pp. 689–692.
[45] R. R. Varior, B. Shuai, J. Lu, D. Xu, and G. Wang, "A siamese long short-term memory architecture for human re-identification," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 135–153.
[46] Z. Zheng, L. Zheng, and Y. Yang, "A discriminatively learned CNN embedding for person reidentification," *ACM Trans. Multimedia Comput. Appl.*, vol. 14, no. 1, pp. 1–20, Jan. 2018.
[47] E. Ustinova, Y. Ganin, and V. Lempitsky, "Multi-region bilinear convolutional neural networks for person re-identification," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2017, pp. 1–6.
[48] L. Wu, C. Shen, and A. van den Hengel, "Deep linear discriminant analysis on Fisher networks: A hybrid architecture for person re-identification," *Pattern Recognit.*, vol. 65, pp. 238–250, May 2017.
[49] H. Liu, J. Feng, M. Qi, J. Jiang, and S. Yan, "End-to-end comparative attention networks for person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3492–3506, Jul. 2017.
[50] R. R. Varior, M. Haloi, and G. Wang, "Gated Siamese convolutional neural network architecture for human re-identification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2016, pp. 791–808.
[51] I. B. Barbosa, M. Cristani, B. Caputo, A. Rognhaugen, and T. Theoharis. (2017). "Looking beyond appearances: Synthetic training data for deep CNNs in re-identification." [Online]. Available: https://arxiv.org/abs/1701.03153
[52] L. Zheng, Y. Huang, H. Lu, and Y. Yang. (2017). "Pose invariant embedding for deep person re-identification." [Online]. Available: https://arxiv.org/abs/1701.07732
[53] F.-Q. Wang, W.-M. Zuo, L. Lin, D. Zhang, and L. Zhang, "Joint learning of single-image and cross-image representations for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1288–1296.
[54] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "Joint detection and identification feature learning for person search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3376–3385.
[55] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, and Y. Yang. (2017). "Improving person re-identification by attribute and identity learning." [Online]. Available: https://arxiv.org/abs/1703.07220
[56] Z. Zheng, L. Zheng, and Y. Yang. (2017). "Pedestrian alignment network for large-scale person re-identification." [Online]. Available: https://arxiv.org/abs/1707.00408

SHUANG LIU (M'18) received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. She is currently an Associate Professor with Tianjin Normal University, Tianjin, China.

XIAOLONG HAO is currently pursuing the master's degree with Tianjin Normal University, Tianjin, China. His research interests include sensor networks, person re-identification, and deep learning.

ZHONG ZHANG (M'14) received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently an Associate Professor with Tianjin Normal University, Tianjin, China. He has published about 80 papers in international journals and conferences, such as *Pattern Recognition*, the IEEE TRANSACTIONS ON CIRCUITS SYSTEMS VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, *Signal Processing* (Elsevier), CVPR, ICPR, and ICIP.