2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Person Re-identification via Person DPM based Partition

Shaomei Li[1*], Chao Gao[1], Hongtao Yu[1], Jianpeng Zhang[1,2]
1 National Digital Switching System Engineering and Technological Research and Development Center
Zhengzhou, China
2 Eindhoven University of Technology, 5600 MB Eindhoven, the Netherlands
lishaomei_may@126.com

***Abstract*—In surveillance videos, the pictures of a same person often present significant variation which makes person re-identification difficult. Though the globe appearances may present great difference, some local patches still have great similarities, and human eyes can be used to distinguish the identity of each person via these local patches. Inspired from it, patch matching is introduced in person re-identification and has been shown to be an efficient method to solve these problems caused by different viewpoints, poses, camera settings, illumination and occlusion. But until now there is no guide for how to decide the size of patch, and most researches got these patches either by dense sampling or coarse partition. In both case, the person structure information is missing. To improve re-identification accuracy, we propose a method via person DPM (Deformable Parts Model) partition. First, both compared appearances are partitioned into several body parts by pre-trained person DPM and these parts are grouped according to their positions in the body; Second, part matching is conducted between two appearances' parts in each group based on deep learning features; Finally, fusing the similarities of each group are to decide whether these two appearances are from the same person or not. Experiments on VIPeR dataset illustrate that without supervised training, the proposed method can obtain good re-identification performances compared with state-of-the-art methods.**

Keywords—Person re-identification; Deformable Part Model; Deep learning feature; Part matching

I. INTRODUCTION

Person re-identification is to identify whether people from different surveillance videos are the same one[1]. It's a fundamental work for surveillance video analysis, robotics, automatic image annotation, human-computer interaction and so on.

The visual features can be used in person re-identification including face, height, shape, clothing, hair color, hair style, gait, etc. Since most surveillance videos are captured under uncontrolled conditions and with low definition, person re-identification based on the biologic features, such as face and gait is always infeasible. The appearance is the most widely used feature, but due to the differences in viewpoints, poses, camera settings, illumination, occlusion and background, the appearances of a same person in different surveillance videos often undergo great variation which makes person re-identification a challenging task. Meanwhile, researches have shown that traditional distance measurements could not accurately evaluate the appearance changes of the same person, and this is a main reason that person re-identification is still an unsolved problem[2]. Based on this point, researchers recently pay more attention to metric learning based on person re-identification[2-6].

Considering that some local appearance patches from different views of the same person still possess great similarity and human eyes can be used to distinguish the identity of each person according to these local patches, patch matching is proposed to person re-identification and recent researches have shown that patch matching can improve re-identification accuracy. But analyzing these patch matching methods we can find that each method either partitions person into patches by dense sampling or by coarse partition, such as three parts (head, body and leg) just according to height. Also, there are few methods to use person structure as prior knowledge to partition person body. In fact, human body possesses a rigid structure and this geometric information can be helpful for re-identification[7]. To make use of such structure information for person re-identification, [8] first proposes to employ pre-trained person DPM (Deformable Part Model) to extract body parts. Since person DPM integrates body structure information, such partitioned parts have semantic meaning and it's easy to get the correspondence between two parts from two appearances. As shown in [8], person DPM based partition is effective for re-identification.

However, the method proposed in [8] still has some problems. First, body part localization using general DPM is time-consuming. Second, the features used in [8] are HSV histogram and MSCR (Maximally Stable Colour Region), these traditional hand-craft features have proven to be weaker than deep learning features in many computer visual tasks. Third, to compute the distance between two images, [8] simply uses the linear combination of the distances based on the HSV histogram and MSCR features.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

To solve these problems, we present a new person re-identification method via person DPM based partition. The main contributions include as follows:

(1) To speed up person DPM based partition, we improve the traditional DPM process by extracting the HOG feature pyramid in a faster way;

(2) To effectively describe each body part, SPP-Net is used to extract deep learning features;

(3) Since different body part has different importance for re-identification, the final distance between two images is the weighted combination of the distances based on each part.

II. RELATED WORK

Feature extraction for appearance is the most important process for person re-identification[9]. The most widely used features are various color features including RGB, LAB and HSV color histogram[7,10,11], texture features including LBP histogram[10] and Gabor features[10], local features including SIFT[11], and their fusion. Color feature is the most discriminative one among them. Meanwhile, recent researches have shown that symmetry structure and silhouette of person[9], color invariant signature[12] and salience regions[13] also can improve re-identification accuracy.

Based on above features, many feature matching methods also proposed and supervised methods have better performance. These supervised methods include Boosting[14], Rank SVM[15], PLS[15] and many metric learning methods[2,5,10]. While most metric learning methods take person appearance as a whole in evaluating the similarity of two appearances, [10] proposes to divide the appearance into several groups based on its poses and respectively learn different metrics for each group. Moreover, with the development of deep learning, metric learning based on siamese convolutional neural network is proposed in [4].

While most work treat the person appearance as a whole in person re-identification, recent researches have shown that patch matching is an efficient way to improve re-identification performance[4,11]. In these methods, person appearances are partitioned into patches, and similarity is derived by patch matching.

III. FAST BODY PARTITION BASED ON PERSON DPM

The pictures of a same person observed in surveillance videos often present great variation due to viewpoints, poses, camera settings, illumination, occlusion and background. Though the similarities of the holistic person regions maybe small, there exist some local patches to keep stable which provide valuable information for re-identification[13]. As shown in Fig. 1, though the pose, illumination and background have great changes between these two images, the white hat and horizontal stripe T-shirt still have great similarities and can be used as evidence for re-identification. Based on this observation, we propose a novel method to partition person body into parts, and conduct person re-identification by part matching.

Fig. 1. Two pictures of a same person from different camera views

Considering that geometry structure of person can be helpful for person partition, we use a pre-trained person DPM to complete this process. DPM is a pictorial structure based on HOG feature, it represents an object as a group of many deformable parts, and there are elastic connections among these parts [16]. A DPM is comprised of a root filter and n part filters. The root filter is used to detect the whole silhouette of the object, and the n parts have a displacement from the root filter, which reflect the detailed and deformable feature of the object. The ith part is parameterized by filter w_i and deformation term d_i $(i = 1,...,n)$. A proposed object location is defined by $\{p_0, p_1, ..., p_n\}$, where p_0 is the location of root, and p_i is the location of ith part. The root filter and part filters are connected by pictorial structure, and the deformable feature between them is described by deformable model. In our work, the part filters of the person DPM are used to find the locations of the main body parts $\{p_1, ..., p_n\}$, through which person partition is accomplished.

The person DPM used in this paper is shown in Fig .2. It contains two person structures, and each is composed of 8 parts. When it used for person partition, it successively employs each person structure to inference the best location of each part and calculate the score. The structure with a higher score is considered as the partition result. We can see that the partition parts have obvious semantic meanings, which are approximately corresponding to four body regions, head, chest, waist and leg. Based on such partition, we propose to do part matching within each body region to improve re-identification accuracy.

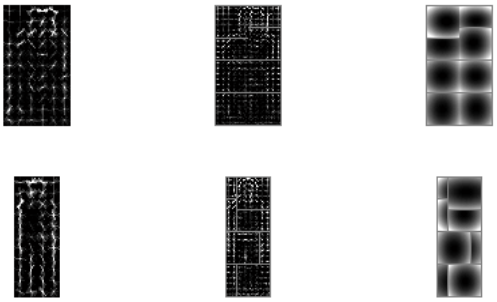

Fig. 2. The person DPM used in this paper

The main two steps involved in person partition via DPM are the extraction of HOG feature and parts location. The computational complexity of these two steps is large, and the implementation of DPM takes about 5 seconds per VGA image on a single thread. Obviously, such implementation cost is not suitable for person partition in this paper, which is used as a pre-process step for person re-identification. To make it possible, we firstly optimize the traditional DPM detection flow.

A. Fast HOG Feature Pyramid Construction based on Power Law

In traditional DPM based person detection, considering mismatch may exist between the size of person used in training and the size of testing person. It needs to downsample and smooth the testing image many times to get image pyramid at every scale. Then for each scale image pyramid, calculate HOG feature at every position to generate fine feature pyramids. Each HOG feature pyramid represents the computation of the image at a scale s which is evenly sampled in log-space and started from 1. Typically each octave includes 4 to 12 scales and an octave is the interval between one scale and another with half or double its value. In the standard DPM, it needs to compute the HOG for every s to construct a feature pyramid. By this way, fine HOG feature pyramid can be provided but the computational cost is huge.

Recently, research on statistics of multi-scale features found that the features of neighborhood scale in feature pyramid have some relation and follows a power law [17,18]. Suppose the size of image I is $h \times w$, I_s denote I sampled at scale s and its size is $h_s \times w_s$. Suppose the reflection function from image to HOG is $f_\Omega(\cdot)$, then the HOG of image with size $s1$ and the HOG of image with size $s2$ satisfies the following power law:

$$f_\Omega(I_{s1}) / f_\Omega(I_{s2}) \approx (s1/s2)^{-\lambda_\Omega} \quad (1)$$

Where Ω denotes channel type and λ_Ω is a Ω related parameter. λ_Ω is set as 0.34 by experiments in this paper.

Based on Eqn. (1), we introduce an efficient approximation scheme for constructing feature pyramid. We begin by just computing HOG of a base scale $f_\Omega(I_{s_{base}})$ per octave ($s \in \{1,1/2,1/4,...\}$). And for the intermediate scales, their HOGs are computed using $f_\Omega(I_s) \approx f_\Omega(I_{s_{base}}) \times (s/s_{base})^{-\lambda_\Omega}$. Such method can get a good tradeoff between speed and accuracy, since the cost of approximating is less than 33% of computing at the original scale. The difference of the standard computation flow and our fast flow is illustrated in Fig .3.

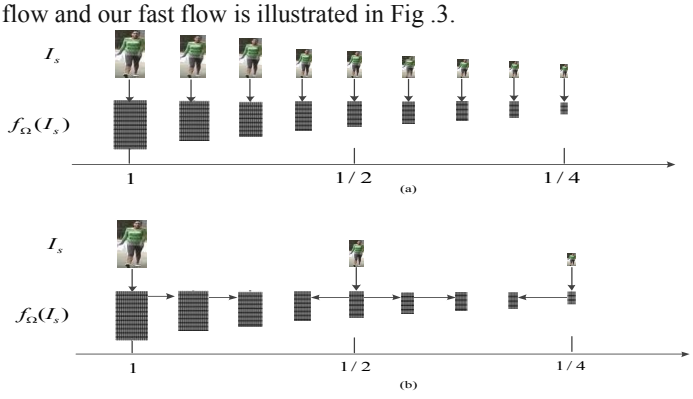

Fig. 3. (a) Standard construction of HOG feature pyramid; (b) Fast construction of HOG feature pyramid.

B. Person DPM based Parts Location

In person DPM based person partition, firstly we need to find the location of the root filter, and decide the location of each part according to the pictorial structure between the root filter and part filters. In generic application, the location of the person is unknown and it may be anywhere in the image, then we need to do exhaustive search on image of each scale by sliding widow. In person re-identification, since person region is known in advance (cropped in dataset), we just need to treat the person region as a whole and only search once.

Every group of locations of the body parts $\{p_1,...,p_n\}$ is a person hypothesis, which specifies a configuration of parts. To exactly explore the location of each body part in person region, we calculate the score of each hypothsis：

$$score(p_0) = max_{p_1,...,p_n} score(p_0,...,p_n) \quad (2)$$

where

$$score(p_0,...,p_n) = w_0^T \phi_\alpha(p_0,H) + \sum_{i=1}^{n} w_i^T \phi_\alpha(p_i,H) - d_i^T \phi_d(p_i,p_0) + b \quad (3)$$

H denotes HOG feature pyramid calculated in Section III.A, ϕ_α denotes the HOG feature vector at corresponding location, and ϕ_d is separable quadratic function for deformation. b is a real-valued bias term which reflects the different pose and deformation of the object.

The location of part p_i is determined by maximizing the score of appearance minus the deformation cost:

$$p_i = argmax_p w_i^T \phi_\alpha(p,H) - d_i^T \phi_d(p,p_0) \quad (4)$$

where p denotes the possible location of part. Then we can partition person according to the location of each part.

IV. Person Re-identification Based on Part Matching

Based on the above person partition, we design the following person re-identification flow as shown in Fig. 4.

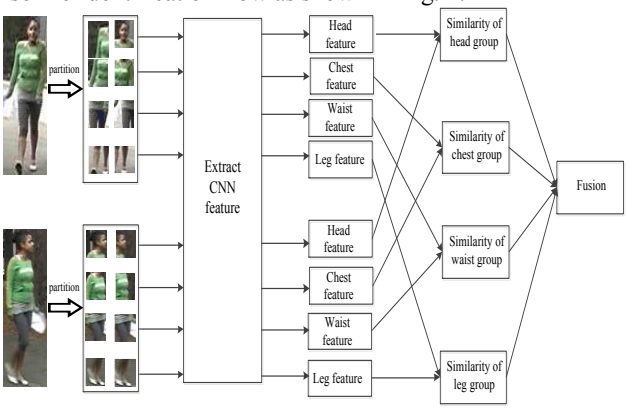

Fig. 4. Person re-identification based on part matching, the person in the top row is the image(**024_0**,cam_a) from VIPeR dataset, and the person in the bottom row is the (**024_45**,cam_b) from VIPeR dataset.

First, each person is partitioned by the person DPM as shown in Fig 2, then we get the 8 parts of each person. Second, we group these 8 parts into 4 groups (head\chest\waist\leg) according to their heights in the body. Third, we extract the deep learning feature of each part and compare the deep learning features of parts from different persons in the same group. Finally, make the decision based on the fusion of the four groups.

Considering that the same parts from different persons may have different sizes, we use SPP-Net to extract the deep learning

features of each part. Compared with other CNN models, between the convolutional layers and the fully-connected layers, SPP-Net inserts a spatial pyramid pooling layer[19], which makes it not demand that the input region must have the same size. The SPP-Net used in this paper is the one provided in [19] which is trained on PASCAL VOC 2007[20] (https://github.com/Shaoqing Ren/SPP_net.git). This model is originally trained for object detection, so its output is the object category. Different from [19], we directly extract the high-dimensional features after its spatial pyramid pooling layer as features which is used for further matching. The original process flow is shown in the red dotted region in Fig 5, while the process flow used by this paper is marked in red solid region.

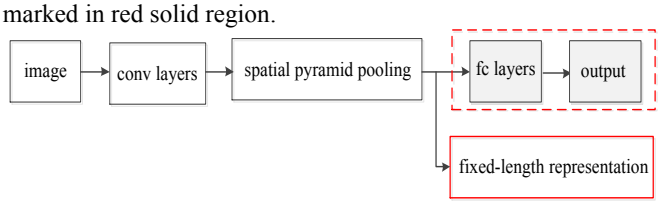

Fig. 5. Extraction of deep learning feature via SPP-Net.

The similarity of two body parts is determined by calculating the cosine distance of their deep learning features. Suppose the deep learning features of two parts are $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$, their cosine distance is:

$$dist(\boldsymbol{f}_1,\boldsymbol{f}_2)=\frac{\boldsymbol{f}_1^T\boldsymbol{f}_2}{\sqrt{\boldsymbol{f}_1^T\boldsymbol{f}_1}\sqrt{\boldsymbol{f}_2^T\boldsymbol{f}_2}} \tag{5}$$

The bigger *dist* is, the more similar these two parts.

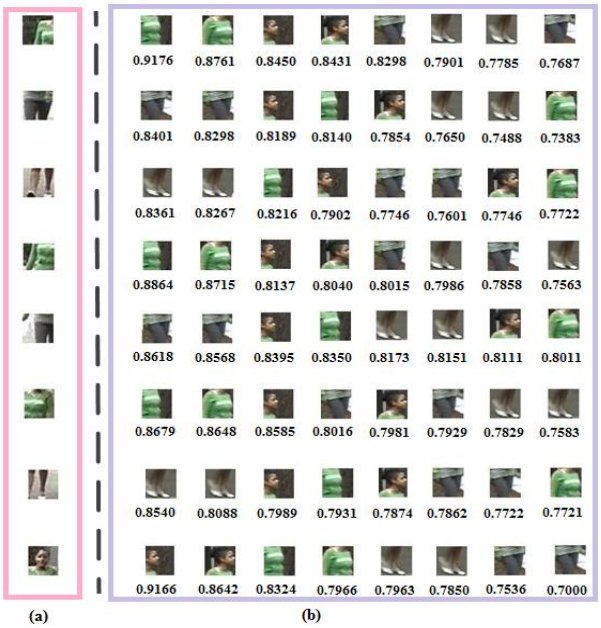

Fig. 6. Examples of cosine distance based on deep learning features extracted via SPP-Net. (a)The eight parts of the image(**024_0**,cam_a) from VIPeR dataset.(b) The top are the eight parts of the other image(**024_45**,cam_b) from VIPeR dataset, the bottom are the cosine distances between these parts and the one in the same row of (a).

The distances of the eight parts of person 024_0 and the eight parts of person 024_45 from dataset VIPeR[21] and the ranking results are shown in Fig. 6. It can be seen that similar parts have higher distance, and such deep learning features based cosine distance can well reflect the differences between parts.

After partition, a body part group of a person may be composed of several parts. Suppose a body part group of person A is composed of M parts, $A_{p_1},A_{p_2},...,A_{p_M}$, the deep learning features of $A_{p_1},A_{p_2},...,A_{p_M}$ are $\boldsymbol{f}_{A_{p_1}},\boldsymbol{f}_{A_{p2}},...,\boldsymbol{f}_{A_{pM}}$,the same body part group of pedestrian B has N parts, $B_{p_1},B_{p_2},...,B_{p_N}$, the deep learning features of $B_{p_1},B_{p_2},...,B_{p_N}$ are $\boldsymbol{f}_{B_{p_1}},\boldsymbol{f}_{B_{p_2}},...,\boldsymbol{f}_{B_{pN}}$, then the distance of this body group between these two persons is:

$$dist_i=\sum_{m=1}^{M}\sum_{n=1}^{N}\min dist(\boldsymbol{f}_{A_{pm}},\boldsymbol{f}_{B_{pn}}) \tag{6}$$

Where $M=2$ and $N=2$ for the person DPM we used in this paper.

Moreover, considering the difference of two different persons may be reflected with different degrees between different body part groups. We calculate the final distance by summing the cosine distances of the four groups with their weights. Suppose the cosine distance of the head part group is $dist_1$, the cosine distance of the chest part group is $dist_2$, the cosine distance of the waist part group is $dist_3$, the cosine distance of the leg part group is $dist_4$ and the weights of these four part groups are w_1, w_2, w_3 and w_4, respectively, the whole distance between these two persons is:

$$D=\sum_{i=1}^{4}w_i\times dist_i \tag{7}$$

The weights in Equation (7) are related with the importance of each part group, and the larger weight value is set to the part group with more discriminability. The value of each weight is set in Section V by experiments.

V. Experiments

A. Datasets and Settings

The most popular datasets used in person re-identification include ETHZ[22], VIPeR[21], i-LIDS[23], CIVAR[24], CUHK Campus[10]. Since VIPeR has the clearest evaluation protocol, we compare our method with other methods on VIPeR. The images in the VIPeR[21] dataset are from 632 people with each person having two images. These images are captured in outdoor (academic environment) by two cameras from different viewpoints. All the images in this dataset are normalized to 128×64 pixels. View angle change and illumination change are the main challenges of this dataset. In our experiment, the image of each person from CAM A is used as gallery and the other one from CAM B is used as probe.

The experimental results are measured by CMC (Cumulative match characteristic)[2] curve. CMC means the percentage that the correct match is included in the top-n best matches.

Our experiments are conducted under single-threaded implementations on a 3.3 GHz Intel Xeon(R) E5-2690 CPU computer running Linux, and all the codes are programmed by matlab2013a.

B. Person DPM based Person Partition

The person DPM used for pedestrian partition is trained based on the 4192 person images from PASCAL VOC 2007[20]. As described in Section III.A, to speed up partition, in the calculation of HOG feature pyramid, we only calculate the HOG features of a set of base scales, and the HOG features of other scales are approximated based on the HOG features of these base scales according to the power law.

Obviously, the smaller the number of base scales, the less time is needed to construct the feature pyramid, but two less base scales may affect accuracy. To balance between the speed and the accuracy, we analyze how the number of scales approximated by a base scale to affect speed and accuracy. If one scale is approximated by a base scale, we label it as DPM-1, and if n scales are approximated by a base scale, we label it as DPM-n. When $n=0$, it's the standard calculation of fine HOG feature pyramid. We compare DPM-0, DPM-1, DPM-2 and DPM-3 in this section.

In the testing, we firstly partition each image in VIPeR by DPM-0 and record the region of each part, marked as groundtruth; Then partition each image by DPM-1, DPM-2 and DPM-3, record the region of each part and compared with the groundtruth. If each of the 8 parts of a person has more than 90% overlap rate with the groundtruth, this person is treated as successful partition. Suppose the region of the groundtruth is A_{truth}, and the partitioned region is A_{test}, the overlap rate(OR) is calculated as:

$$OR = \frac{A_{test} \cap A_{truth}}{A_{test} \cup A_{truth}} \tag{8}$$

Successful partition rate is the number of successfully partitioned images versus the number of total images.

TABLE I. THE RESULTS OF DPM-n

	DPM-0	DPM-1	DPM-2	DPM-3
successful partition rate	100%	100%	99.68%	95.96%
partition time(s)	0.1735	0.0882	0.0438	0.0326

Considering successful partition rate and partition time simultaneously, the number of approximated scales is fixed to 2 for person partition in the following experiments. So for each image, the extra partition time needed is less than 0.05s.

C. Setting of Weigths

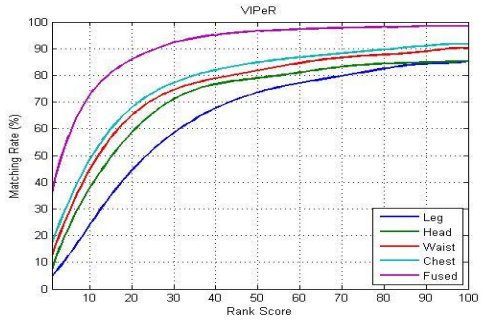

Fig. 7. The rank curves of the 4 body part groups and the whole image on VIPeR.

After partition, the size of each part is about 30×30. To analyze the importance of each body part group for person re-identification, we conduct person re-identification on VIPeR just based on one body part group respectively. And the result is shown in Fig. 7.

As shown in Fig. 7, the order of discrimination is chest, waist, head and leg, so the weights of these four body part groups are set as 0.4, 0.3, 0.2 and 0.1 to satisfy that the sum of these four weights is 1. Then the fused result with these weights is also shown in Fig 7.

D. Compared with the Method Proposed in [8]

Since person DPM based partition is the basis of our work, and this idea was firstly proposed in [8], to validate the effectiveness of our method, we compared it with the method proposed in [8].

The same person DPM described in V.B is used for person partition for these two methods. And since the settings of weights for the two features, HSV histogram and MSCR in linear combination are not stated in [8], we test nine sets of weights with one is from 0.1 to 0.9 and the other is from 0.9 to 0.1. Finally, the best result achieved with $\{0.7,0.3\}$ is used to compare with our method.

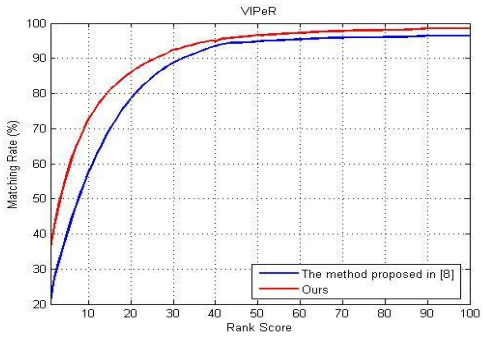

Fig. 8. Comparison with the method proposed in [8].

As shown in Fig. 8, our method outperforms the method proposed in [8], which means that it's useful to use deep feature and our fusion strategy.

E. Compared with Othe Method

Some state-of-art person reidentificaiton methods, SDALF[9], RDC[2], PPCA[27], Salience[11], RPML[28], LAFT[10], DML[4], whose results on VIPeR have published are used for comparison.

For precise comparison, the results of compared methods are the values reported in their original papers. And the unavailable results are labelled by "-".

As seen in Table II, our method outperforms most of these methods. And it is on a par with LAFT and DML. It must be noted that since the person DPM is trained in advance with person images from other dataset, which means no data from testing dataset is used for training, our method is unsupervised. Most compared methods, except ELF, need about half of the images in the dataset for training. And DML even use all the images in the VIPeR to tune the model parameters. Such unsupervised way makes our method more applicable.

TABLE II. THE ACCURACY OF THE ALGORITHMS ON VIPER(%)

	Rank -1	Rank -5	Rank -10	Rank -20	Rank -25	Rank -50
ELF[13]	12.00	31.00	41.00	58.00	-	-
RDC[2]	15.66	38.42	53.86	70.09	-	-
PPCA[37]	19.27	48.89	64.91	80.28	-	-
Salience[10]	26.74	50.70	62.37	76.36	-	-
RPML[38]	27.00	-	69.00	83.00	-	95.00
LAFT[9]	**29.60**	-	69.31	-	88.70	**96.80**
DML[4]	28.23	59.27	**73.45**	**86.39**	**89.53**	96.68
Ours	28.48	58.86	72.55	85.84	89.00	96.68

VI. CONCLUSION

This paper proposes a new method to partition pedestrian based person DPM, and re-identification pedestrian by part matching approach. To speed up person DPM based partition, we improve traditional DPM detection process. Meanwhile, in order to improve the accuracy of part matching, we use deep learning features to capture both the color and texture characteristics of each part. Experiments show that such partition improves re-identification accuracy than other methods. Since the SPP-Net model used in the paper is trained on generic image dataset, we intend to use person parts to train specific SPP-Net model in the future work.

ACKNOWLEDGMENT

This work was supported by National Science and Technology Support Plan of China (No.2014BAH30B01), and National Natural Science Foundation of China(No.61521003).

REFERENCES

[1] Gong, S., Cristani, M., Yan, S., Loy, C.C. Person Re-Identification[M]. London:*Springer*-Verlag. 2014:1-30.

[2] Zheng, W., Gong, S., Xiang, T. Re-identification by relative distance comparison[J]. *PAMI* 2013, 35(3):653-668.

[3] Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.. Learning locally-adaptive decision functions for person verification[C]. Proceedings of *Computer Vision and Pattern Recognition*(2013), 2013: 3610-3617.

[4] Dong Yi, Zhen Lei, Shengcai Liao. Deep Metric Learning for Person Re-Identification. *2014 22nd International Conference on Pattern Recognition (ICPR)*, pp: 34-39.

[5] Kostinger, M., Hirzer, M.,Wohlhart, P., Roth, P.M., Bischof, H. Large scale metric learning from equivalence constraints. [C]. Proceedings of *Computer Vision and Pattern Recognition*(2012), 2012:2288-2295.

[6] Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.. Learning locally-adaptive decision functions for person verification[C]. Proceedings of *Computer Vision and Pattern Recognition*(2013), 2013: 3610-3617.

[7] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani,"Person re-identification by symmetry-driven accumulation of localfeatures," in *Computer Vision and Pattern Recognition* (CVPR), 2010 IEEE Conference on, 2010, pp. 2360-2367.

[8] VH. Nguyen, K. Nguyen, DD. Le, DA. Duong, S. Satoh. Person Re-identification Using Deformable Part Models[M]. Berlin Heidelberg: *Springer*, 2013, 8228:616-623.

[9] O. Javed, K. Shafique, and M. Shah, "Appearance modeling for tracking in multiple non-overlapping cameras," in Computer Vision and Pattern Recognition, 2005. *CVPR* 2005. IEEE Computer Society Conference on, vol. 2, 2005, pp. 26-33.

[10] W. Li and X. Wang, "Locally aligned feature transforms across views,"in *Computer Vision and Pattern Recognition* (CVPR), 2013 IEEE Conference on, 2013, pp. 3594-3601.

[11] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Computer Vision and Pattern Recognition* (CVPR), 2013 IEEE Conference on, 2013, pp. 3586-3593.

[12] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person re-identification," *Pattern Analysis and Machine Intelligence,* IEEE Transactions on, vol. 35, no. 7, pp. 1622-1634, 2013.

[13] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in Proceedings of IEEE *International Conference on Computer Vision*, 2013, pp. 2528-2535.

[14] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in Computer Vision *C ECCV* 2008, ser. Lecture Notes in Computer Science, D. Forsyth, P. Torr, andA. Zisserman, Eds. Springer Berlin Heidelberg, 2008, vol. 5302, pp.262-275.

[15] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *BMVC*, 2010, pp. 1-11.

[16] Felzenszwalb P F, Huttenlocher D P. Pictorial Structures for Object Recognition[C]. *International Journal of Computer Vision*.2005, 61(1): 55-79.

[17] Ruderm D L. The statistics of natural images[J]. *Network Computation in Neural Systems*, 2009, 5(4):517-548.

[18] Dollar P, Appel R, Belongie S, et al. Fast Feature Pyramids for Object Detection[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 36(8):1-14.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *arXiv:1406.4729v4*.

[20] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. The PASCAL Visual Object Classes (VOC) Challenge[J], *International Journal of Computer Vision*, 2010, 88(2): 303-338.

[21] D. Gray, S. Brennan, and H. Tao. Evaluating Appearance Models for Recognition Reacquisition, and Tracking[C]. Proceedings of IEEE *International Workshop Performance Evaluation of Tracking and Surveillance*, 2007:151-157.

[22] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance based models using partial least squares," in Proceedings of the *XXII Brazilian Symposium on Computer Graphics and Image Processing*,2009.

[23] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by prob-abilistic relative distance comparison," in *Computer Vision and Pattern Recognition* (CVPR), 2011 IEEE Conference on, 2011, pp. 649−656.

[24] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, Custom pictorial structures for re-identification[C], in *British Machine Vision Conference* (BMVC), 2011, pp. 68.1-68.11.

[25] A. Ess, B. Leibe, and L. Van Gool. Depth and Appearance for Mobile Scene Analysis[J]. Proceedings of *Computer Vision and Pattern Recognition*(2007),2007:1-8.

[26] UK Home Office. i-LIDS Multiple Camera Tracking Scenario[OL]. *http://scienceandre-search.homeoffice.gov.uk/hosdb/cctv-imaging-tech-nology/i-lids/.*

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

[28] M. Hirzer, P. Roth, M. K ¨ ostinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in Computer Vision *C ECCV* 2012, ser. Lecture Notes in Computer Science, A. Fitzgibbon,S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Springer Berlin Heidelberg, 2012, vol. 7577, pp. 780−793.