VP-ReID: Vehicle and Person Re-Identification System

Longhui Wei, Xiaobin Liu, Jianing Li, Shiliang Zhang
School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China
{longhuiwei,liu-xb,ljn-vmc,slzhang.jdl}@pku.edu.cn

ABSTRACT

With the capability of locating and tracking specific suspects or vehicles in a large camera network, person Re-Identification (ReID) and vehicle ReID show potential to be a key technology in smart surveillance system. They have been drawing lots of attentions from both academia and industry. To demonstrate our recent research progresses on those two tasks, we develop a robust and efficient person and video ReID system named as VP-ReID. This system is build based on our recent works including Deep Convolutional Neural Network design for discriminative feature extraction, efficient off-line indexing, as well as distance metric optimization for deep feature learning. Constructed upon those algorithms, VP-ReID identifies query vehicle and person efficiently and accurately from a large gallery set.

KEYWORDS

Person Re-Identification, Vehicle Re-Identification

1 INTRODUCTION

Person Re-Identification (ReID) [11] and vehicle ReID [3] can be regarded as two special instance retrieval tasks. Both of them target to match images of a given instance, *i.e.*, a person or vehicle from a large-scale surveillance video dataset. With the capability of locating and tracking specific suspects or cars in a large camera network, person and vehicle ReID are important for applications on public security. In recent years, those two tasks have been attracting more and more attentions from both the academia and industry.

Person and vehicle ReID face many challenging issues. For instance, the appearance of a vehicle or person image can be easily affected by various factors like lighting conditions, camera viewpoints, misalignment error in detected bounding boxes, *etc.* They also need to cope with the issue of large intraclass variance and small interclass difference, as well as the massive amount of surveillance data. In recent years, we have conducted many works toward conquering those challenges [2, 4–7, 10–12]. Those research progresses have significantly boosted the performance of person ReID and vehicle ReID. Based on those works, we develop a robust and efficient person and video ReID system named as VP-ReID.

VP-ReID is a web-based demo system composed of two modules, *i.e.*, person ReID module and vehicle ReID module, respectively. Those two models are implemented with different algorithms but has similar idea and user interface. Users can upload and choose a person or vehicle image. The system would return the matched person or vehicle images from a gallery set. Targeting to achieve high accuracy and efficiency, VP-ReID is designed from two aspect, *i.e.*, discriminative descriptor extraction and efficient retrieval strategy, respectively. VP-ReID trains Deep Convolutional Neural Networks (DCNNs) for feature extraction and fuses regional and global descriptors to achieve high discriminative power and robustness. To improve the ReID efficiency, VP-ReID introduces an offline indexing algorithm, which builds the connections among gallery images of the same instance during offline indexing stage. The resulting image connections thus help to improve the online accuracy, recall, and efficiency. There, VP-ReID has the potential to retrieve a given person or vehicle image from a large-scale dataset with high accuracy and efficiency.

Table 1: Comparison on Market1501 in single query mode.

Methods	mAP	Top-1
Spindle [15]	-	76.9
SVDNet [8]	62.1	82.3
PAN [17]	63.4	82.8
DML [14]	68.8	87.7
Proposed Method	**73.9**	**89.9**

2 ALGORITHMS

2.1 Person Re-Identification

The person ReID module is mainly based on our recent work GLAD [11]. It includes the following components for human parts detection, descriptors extraction, and indexing and retrieval, respectively.

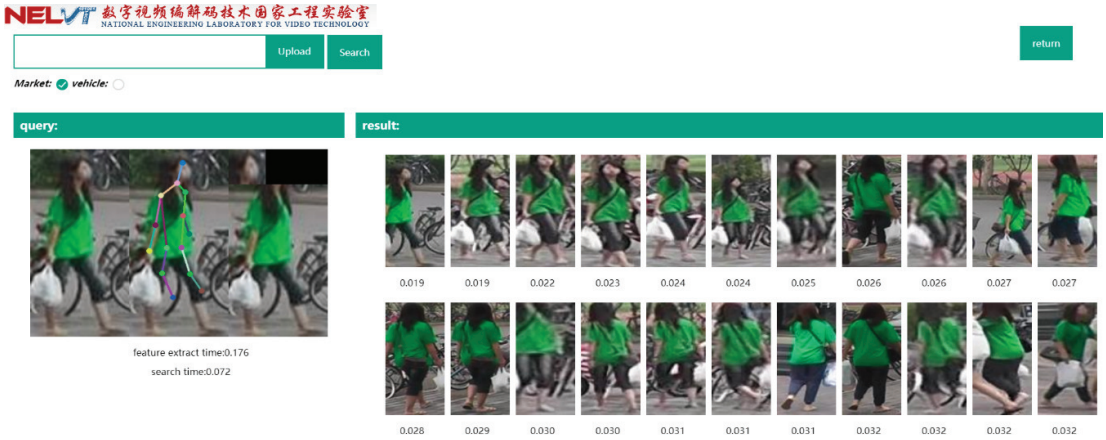
Part detection: Human parts convey detailed and distinctive visual cues [4]. We thus detect human parts for feature extraction. Originally detected person bounding boxes are commonly not well-aligned. To overcome the misalignment error, we detect human parts using Deepercut [1], instead of directly dividing images into fixed strips. Deepercut [1] outputs human joint points, based on which, three body regions can be cropped, *i.e.*, the head region, upper-body region, and lower-body region, respectively. Our experiments show coarse part segmentation is robust to pose variations and makes the feature extraction model compact and efficient [11].

Descriptor extraction: The descriptor extraction is conducted by a CNN network composed of four sub-networks, that correspond to the global image and the three body regions, respectively. Those sub-networks share parameters in the first several layers, then have their own parameters for feature learning. In the testing stage, the global image and three detected part regions are inputted into those four sub-networks for global and regional feature extraction, respectively. The resulting four descriptors are concatenated as the final descriptor.

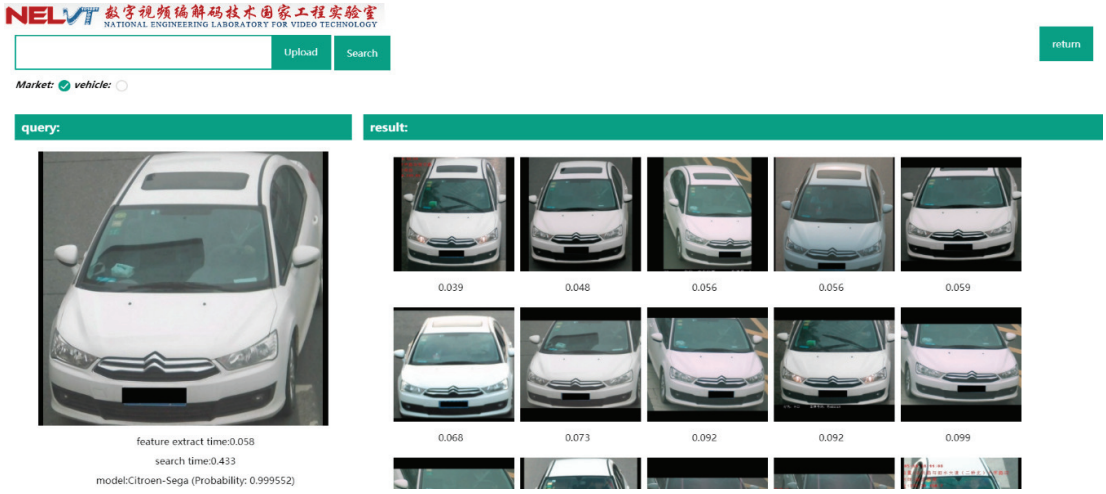
Retrieval procedure: To accelerate the retrieval procedure, offline indexing and coarse-to-fine retrieval strategy are utilized. In offline

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn



(a) Illustration of person ReID result



(b) Illustration of vehicle ReID result

Figure 1: Illustration of VP-ReID System.

stage, a hierarchical clustering method, *e.g.*, TDC [11], is utilized to group the images belong to the same person together. In online retrieval stage, coarse retrieval are conducted to retrieve the relevant image groups, then fine retrieval is conducted to get a precise image rank list. As shown in our experiments [11], this strategy substantially accelerates the online retrieval without degrading the accuracy.

Performance of Person ReID We use *Market1501* [16] dataset to evaluate our person Re-ID methods. *Market1501* [3] contains 32,668 images of 1,501 person identities. 12,936 images of 751 identities are selected as training set and others are used for testing. Following previous work [16], we report the Top-1 accuracies and mAP in Table 1. As shown in Table 1, our proposed methods achieve the best performance compared with state-of-the-art methods.

2.2 Vehicle Re-Identification

The vehicle ReID model is designed with similar idea with the one of person ReID. The CNN model have several sub-networks to extract complementary global and local features. A weights prediction sub-network is also designed to predict weights for different features to chase a more reasonable feature fusion. We also propose a new distance loss to accelerate the feature learning procedure. We briefly introduce each component in the following parts.

Global feature extraction: We use two different sub-networks to extract complementary global features, *i.e.*, the CN branch and BN branch, respectively. CN branch uses two Fully Connected (FC) layers to extract a global feature. The BN branch has similar structure with CN branch, but embeds a Batch Normalization (BN) layer [13] between feature maps and the pooling layer. We observed that BN and CN branches have different focuses on the input image, *e.g.*, local discriminative regions vs. larger contextual regions, respectively [13]. CN and BN branches thus extract different global features.

Regional feature extraction: Regional differences could be important for differentiating similar vehicles sharing the identical model

Table 2: Performance comparison on *VehicleID*.

Method	*VehicleID*					
	Small		Medium		Large	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
VGG+CCL [3]	0.436	0.642	0.370	0.571	0.329	0.533
Mixed Diff+CCL [3]	0.490	0.735	0.428	0.668	0.382	0.616
OIF$^+$ [9]	-	-	-	-	0.670	0.829
Proposed	**0.771**	**0.928**	**0.727**	**0.892**	**0.700**	**0.871**

and color. We use regional sub-networks to extract regional features. Regional sub-networks first divide feature maps extracted from global image into overlapped regions. Then regional features are extracted from those regions, respectively. Adjacent regions are set to be overlapped to take alignment error and variations of angle of views into account.

Weights prediction: To chase a more reasonable feature fusion and enhance the feature robustness against vehicle detection errors, viewpoint changes, and other noises, we design a weight prediction sub-network to predict feature weights. During the training stage, weights prediction sub-network takes the global and local features as inputs and outputs feature weights to maximize the vehicle classification accuracy. This sub-network hence dynamically predicts feature weights during the testing stage.

Feature Learning: A new distance loss is proposed to accelerate the procedure of feature learning and network training. Different from the sample-wise triplet loss, it considers samples of the identical vehicle as an image set, then guides the network training procedure to optimize the distance between and within image sets. Specifically, it pulls samples in the same set close to each other and pushes different sets away from each other. It exhibits better efficiency than the commonly used sample-wise triplet loss.

Performance of Vehicle ReID: We use *VehicleID* [3] dataset to evaluate our vehicle Re-ID methods. *VehicleID* [3] contains 221,763 images of 26,267 vehicles. 13,134 vehicles and 110,178 images are selected for training and others are used for testing. Three subsets containing 800, 1,600, and 2,400 vehicles are extracted from the testing set as *small*, *medium*, and *large* sets. Following previous work [3], we report the Top-1, Top-5 accuracies in Table 2. On the three testsets, proposed methods achieve promising performance compared with state-of-the-art methods.

3 THE VP-REID SYSTEM

VP-ReID is a web-based demo system build with the above two models. It provides two web interfaces: one allows users to upload a person or vehicle image, and another one shows the retrieved person or vehicle images from gallery set. User can choose any returned image as a new query by clicking it.

We show a person ReID result page in Fig. 1-(a). The three images in the left of Fig. 1 are query person image, detected 14 joint points of human body, and the cropped part regions. Under the query, the system shows the feature extraction time and search time, which are computed on a GTX-1080 GPU and an intel i7 CPU. With our off-line indexing strategy, VP-ReID can finish person ReID in a large-scale dataset in real time [11, 12]. The ReID results are shown on the right side of the webpage. The number under each returned image is the feature distance between it and the query. We rank the results by the feature distance.

We also show a vehicle ReID result in Fig. 1-(b). Similar to person ReID, left side of the webpage shows the query image, feature extraction time, and search time. We additionally show the fine-grained category of the query predicted by our model, *i.e.*, the model and maker of query vehicle, as well as the prediction probability. The ReID results are shown on the right side of the webpage, and the feature distance is also shown under each returned image.

4 CONCLUSIONS

This demo presents a robust and efficient person and video ReID system named VP-ReID. VP-ReID is build upon our recent research progresses on person ReID and vehicle ReID [2, 4–7, 10–12]. It consists of two models, *i.e.*, person ReID model and vehicle ReID model, respectively. By learning and fusing complementary regional and global features with multi-branch DCNNs, VP-ReID extracts robust visual descriptors. VP-ReID also optimizes the offline indexing to ensure high online ReID efficiency. Therefore, VP-ReID has the potential to identify a given person or vehicle image from a large-scale dataset with high accuracy and efficiency.

REFERENCES

[1] Eldar Insafutdinov, Leonid Pishchulin, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. 2016. Deepercut: A deeper, stronger, and faster multi-person pose estimation model. In *ECCV*.

[2] Jianing Li, Shiliang Zhang, Jingdong Wang, Wen Gao, and Qi Tian. 2017. LVreID: Person Re-Identification with Long Sequence Videos. *arXiv preprint arXiv:1712.07286* (2017).

[3] Hongye Liu, Yonghong Tian, Yaowei Yang, Lu Pang, and Tiejun Huang. 2016. Deep relative distance learning: Tell the difference between similar vehicles. In *CVPR*.

[4] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. 2017. Pose-driven Deep Convolutional Model for Person Re-identification. In *ICCV*.

[5] Chi Su, Fan Yang, Shiliang Zhang, Qi Tian, Larry S Davis, and Wen Gao. 2017. Multi-Task Learning with Low Rank Attribute Embedding for Multi-Camera Person Re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).

[6] Chi Su, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. 2018. Multi-type attributes driven multi-camera person re-identification. *Pattern Recognition* 75 (2018), 77–89.

[7] Chi Su, Shiliang Zhang, Fan Yang, Guangxiao Zhang, Qi Tian, Wen Gao, and Larry S Davis. 2017. Attributes driven tracklet-to-tracklet person re-identification using latent prototypes space mapping. *Pattern Recognition* 66 (2017), 4–15.

[8] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. 2017. SVDNet for Pedestrian Retrieval. In *ICCV*.

[9] Zhongdao Wang, Luming Tang, Xihui Liu, Zhuliang Yao, Shuai Yi, Jing Shao, Junjie Yan, Shengjin Wang, Hongsheng Li, and Xiaogang Wang. 2017. Orientation Invariant Feature Embedding and Spatial Temporal Regularization for Vehicle Re-Identification. In *ICCV*.

[10] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. 2017. Person Transfer GAN to Bridge Domain Gap for Person Re-Identification. *arXiv preprint arXiv:1711.08565* (2017).

[11] Longhui Wei, Shiliang Zhang, Hantao Yao, Wen Gao, and Qi Tian. 2017. GLAD: Global-Local-Alignment Descriptor for Pedestrian Retrieval. In *ACM MM*.

[12] Hantao Yao, Shiliang Zhang, Dongming Zhang, Yongdong Zhang, Jintao Li, Yu Wang, and Qi Tian. 2017. Large-scale person re-identification as retrieval. In *ICME*.

[13] Hantao Yao, Shiliang Zhang, Yongdong Zhang, Jintao Li, and Qi Tian. 2017. One-Shot Fine-Grained Instance Retrieval. In *ACM Multimedia*.

[14] Ying Zhang, Tao Xiang, Timothy M Hospedales, and Huchuan Lu. 2017. Deep Mutual Learning. *arXiv preprint arXiv:1706.00384* (2017).

[15] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. 2017. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*.

[16] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable person re-identification: A benchmark. In *ICCV*.

[17] Zhedong Zheng, Liang Zheng, and Yi Yang. 2017. Pedestrian Alignment Network for Large-scale Person Re-identification. *arXiv preprint arXiv:1707.00408* (2017).