Vehicle Re-Identification Based on the Authenticity of Orthographic Projection

Qiang Lu[1], Fengwei Quan[2], Mingkai Qiu[3], Xiying Li[4*]

[1,2,3,4]Guangdong Provincial Key Laboratory of Intelligent Transportation Systems
[1,2,3,4]School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou Guangdong 510006, China
[*]Corresponding author at: Research Center of Intelligent Transportation System, School of Engineering, Sun Yat-sen University, Guangzhou, Guangdong, 510006, China

Abstract— *Vehicle re-identification is still a problem do not receive much attention in the multimedia and vision communities. Since most existing approaches mainly focus on the overall vehicle appearance for re-identification and do not consider the visual appearance changes of sides of vehicle, called local deformation. In this paper, we propose a vehicle re-identification method based on the authenticity of orthographic projection, in which three sides of vehicle are extracted, and the local deformation is explicitly minimized by scaling each pair of corresponding side to uniform size before computing similarity. To compute the similarity between two vehicle images, we 1) construct 3D bounding boxes around the vehicles, 2) extract sub-images of the three sides of each vehicle like a three-view drawing, 3) compute the similarity between each pair of corresponding side sub-images, and 4) use their weighted mean as the final measure of similarity. After computing the similarity between the query vehicle and all candidate vehicles, we rank these similarities and take the vehicle with the maximum similarity as the best match. To evaluate this approach, we use a dataset with 240 pairs of vehicle images extracted from surveillance videos shot at seven locations in different directions. The experimental results show that our proposed method can achieve 75.83% matching accuracy for the top-1 ranked vehicle and 91.25% accuracy for the top-5 vehicles.*

Keywords— ***3D bounding boxes, local deformation, vehicle re-identification, weighted mean.***

I. INTRODUCTION

Vehicle re-identification refers to the problem of identifying a query vehicle in a gallery of candidates captured from non-overlapping cameras. As the development of smart city, how to research a given vehicle in a large-scale surveillance video data is an emerging and important problem that should pay more attention. Unlike person re-identification [1,2,3] which attract widespread attention, researches on vehicle re-identification are still limited. In vehicle-related research, the major researches in the multimedia and computer vision fields are focus on vehicle detection [4], fine-grained recognition [5] and driver behavior modeling [6]. Different with vehicle fine-grained recognition, which aims at recognizing the model of a given vehicle, vehicle re-identification is a more challenging task since there exist many vehicles share the same model and they should be identified as different classes.

As each vehicle's license plate number is unique, vehicle re-identification may not difficult if the license plate number canbe distinguished [7,8,9]. However, in real-world applications, especially in most surveillance videos, license plates cannot be clear enough to identify a vehicle due to the influences of camera distance and resolution. Therefore, license plate number matching is not a reliable method of re-identification. Instead, achieving high re-identification accuracy based on vehicle appearance is desired.

Existing re-identification approaches [10,11,12,13] focus on learning an embedding space in which similar images are pulled closer and dissimilar images are pushed far away, and the embedding spaces are optimized by a triplet loss [14], circle loss [15] or other improved triplet losses function. These methods all reduce the intra-class variance between images of same vehicles implicitly, and here we aim to construct a method that could explicit reduce the intra-class variance.

Due to the variations in viewing angle, shooting distance and background clutter, one of the challenges in vehicle re-identification is same vehicle capture from different camera via different views may have significantly different appearance, as shown in Fig. 1. Even after scaling the vehicle images to uniform size, differences in appearance still exist since visual appearance changes of each side of vehicle, called local deformation, is not uniform. So, to minimize the influence of vehicle deformation, we aim to find a method that can deal with local deformation rather than overall deformation, here, overall deformation means visual appearance changes of the whole vehicle.

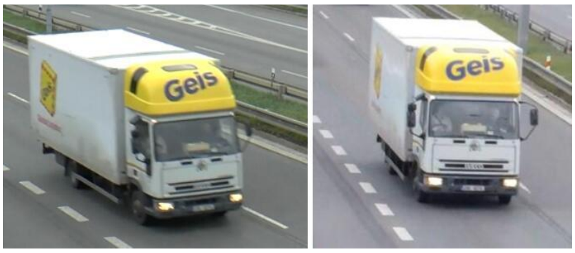

FIGURE 1: A vehicle captured from different camera viewpoints.

Images captured by orthographic projection can be used for scale- and rotation-invariant detection [16]. By making three-view drawings based on orthographic projection, the real shape of all three directions (front, top and side) can be obtained, which is called the authenticity of orthographic projection. As vehicles can be approximated as rectangular solids, it is possible to construct an approximate three-view drawing of a vehicle by geometric transformation. So, if we can obtain each vehicle's three views as a three-view drawing and unify the size of each pair of corresponding views (front-to-front, top-to-top and side-to-side), then the deformation of each view can be reduced, and the influence of vehicle deformation minimized. Based on this idea, we propose a vehicle re-identification method based on the authenticity of orthographic projection.

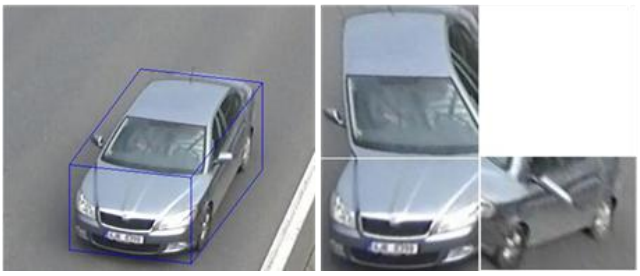

FIGURE 2: 3D bounding box constructed around vehicle, and then split so the three views form a three-view drawing.

Given two images of vehicles, we first construct 3D bounding boxes [17] around them, split each vehicle's three views like a three-view drawing, as shown in Fig. 2. Then, we unify the size of each pair of corresponding views, compute the three similarities between them and apply a weighting strategy to obtain the final similarity. After computing the similarities between the query vehicle and all candidate vehicles, we sort these similarities and take the vehicle with the maximum similarity to be the query vehicle. The re-identification procedure is as shown in Fig. 3.

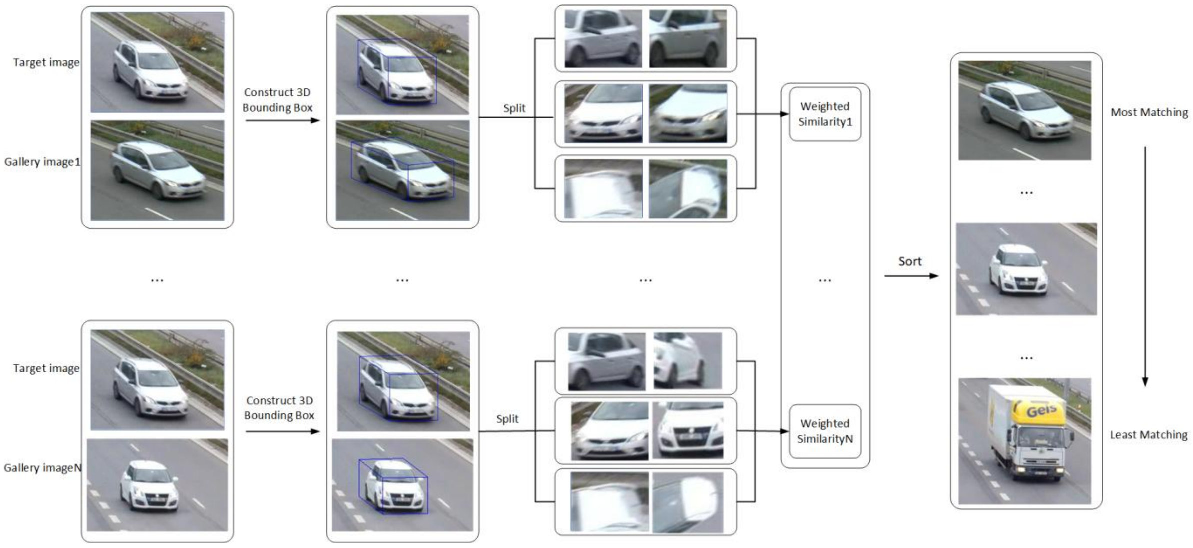

FIGURE 3: Vehicle re-identification method proposed in this paper

The main contribution of this work to the field of vehicle re-identification is that, unlike previous studies, we propose a method that can deal with the deformation of local sides rather than overall deformation. The proposed model can be used to effectively improve the performance of some existing methods in vehicle re-identification or vehicle retrieval.

The remainder of this paper is organized as follows: In Section II, early and recent algorithms for target re-identification are reviewed. In Section III, we explain our proposed method for vehicle re-identification. In Section IV, we describe the datasets used, the evaluation metrics and the experimental results. Finally, in Section V, we conclude our work.

II. RELATED WORK

Like person re-identification, vehicle re-identification is a type of target re-identification. This involves a computer being given an image of a target object from one camera view, then using it to identify an image of a target object from a set of candidates captured from other camera views. As the principle of these two types of target re-identification is basically the same, some methods of person re-identification can be applied to vehicle re-identification.

Person re-identification: The work of person re-identification mainly focus on two parts: 1) developing a robust descriptor to handle variations in appearance, 2) designing an effective distance metric to measure the similarity between images.

The first category of research mainly focuses on developing a discriminative representation that is robust to cross-view appearance variations. Sunderrajan et al. [1] propose a clothing context-aware color extraction method to learn color drift patterns in a non-parametric manner using the random forest distance (RFD) function. Wei et al. [18] develop a pedestrian image descriptor named Global-Local-Alignment Descriptor; this descriptor explicitly leverages the local and global cues in human body to generate a discriminative and robust representation. To promote the results of designing a descriptor, Wang et al [19] design a data-specific adaptive metric method to conquer the zero-shot and fine-grained difficulties in the re-id problem.

The second category of research aims to find a mapping function from the feature space to another distance space where feature vectors from the same target are more similar than those from different ones. Triplet loss or other improved triplet loss is used to model the intra-class variance and inter-class variance. Ding et al. [2] considered the re-identification problem as a ranking issue and used triplet loss to obtain the relative distance between images. Chen et al. [20] designed a quadruplet loss process, which can lead to model outputs with larger inter-class variation and smaller intra-class variation compared with the triplet loss method. Sun et al. [15] propose a circle loss which offers a more flexible optimization approach towards a more definite convergence target.

Except for these two categories, there are also some other researches focus on re-ranking. Leng et al. [3] proposed an automatic bidirectional ranking method based on content and context similarity, the gallery images are treated as new probes to requery in the original gallery set. Ye et al. [21] propose a method exploring both similarity and dissimilarity relationship for ranking optimization, the method improve the quasi-similar galleries' ranking orders and penalize the quasi-dissimilar galleries. Sarfraz et al. [22] introduce an expanded cross neighborhood re-ranking method by integrating the cross neighborhood distance. A local blurring re-ranking [23] employs the clustering structure to improve neighborhood similarity measurement, refining the ranking list.

Vehicle re-identification: Some researches focus on doing vehicle re-identification not only by vehicle image, but also with license plate information or spatial-temporal information of the vehicle, Liu et al. [24,25] utilizes vehicle image and license plate feature to do coarse-to-fine vehicle research, then do re-ranking by the spatiotemporal information of vehicle to get a better result. Shen et al. [26] and Jiang et al. [27] also train a network with multi task to learn discriminative representations of vehicle, then use the spatiotemporal information of vehicle to do re-ranking.

However, license plate information and the spatial-temporal information always cannot obtain in unconstrained surveillance environment, so, develop a general model only based on visual appearance is more desired.

Similar to person re-identification, vision-based vehicle re-identification methods focus on learning discriminative and robust feature representations. These methods train a multi-task network to extract features, use triplet loss, contrastive loss or other improved triplet loss to model inter-class or intra-class variance. Li et al. [10] designed a multi-task training network including identification, attribute recognition, verification and triplet tasks, and used the element-wise absolute difference of extracted features as the similarity score. Zhu et al. [11, 28] design a hybrid similarity learning function to compute the similarity score, this hybrid similarity is computed by simultaneously projecting the element-wise absolute difference and multiplication of the corresponding deep learning feature pair with a group of learned weight coefficients. Liu et al. [12] design a coupled cluster loss to make the training phase more stable and accelerate the convergence speed. Bai et al. [29] design a group-sensitive-triplet embedding to model intra-class and inter-class variance in learning representation, images of vehicle are divided into groups, and images of each group are supposed to share similar attributes, in this way, the intra-class variance are well modeled. He et al. [30] proposed a part-regularized discriminative feature preserving method which enhances the perceptive ability of subtle discrepancies.

Some other researches focus on constructing a descriptive representation containing all-view information to solve the multi-view vehicle re-identification problem. Zhou et al. [13] use GAN to generate all-view features from the visible view's feature, then fuse all the inferred features in different views and adopt the final representation for distance metric learning. Zhou et al [31] propose two architecture: one use only CNN and the other use CNN and LSTM to generate all -view features from single view image.

III. VEHICLE RE-IDENTIFICATION BASED ON THE AUTHENTICITY OF ORTHOGRAPHIC PROJECTION

3.1 3D Bounding Boxes Construction

We construct 3D bounding boxes around vehicles from a single image based on the method proposed in [17]. The vehicle's contour C and three directions, as shown in Fig. 4, are needed for the construction. The first direction V_1 (marked in yellow) is the direction of the traffic. The second direction V_2 (marked in red) is perpendicular to the first direction and parallel to the road. The third direction V_3 (marked in blue) is orthogonal to V_1 and V_2.

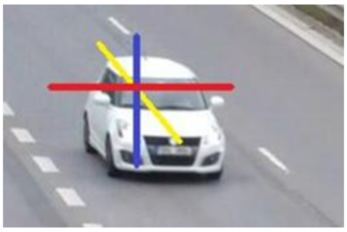

FIGURE 4: The three directions needed for 3D bounding box construction

By quantizing the angle space by bins of 3° from −90° to 90°, which separates the angle space into 60 bins per direction, we transform the regression problem of direction prediction into a classification task. We use Resnet50 [32] with three separate outputs to conduct multi-task prediction of the three directions (see Fig. 5), and use eq. (1) to compute the sum of the cross-entropy of the three predictions as the total loss function for network prediction:

$$Loss = -\sum_{i=0}^{2}\sum_{j=0}^{59} y_{i,j}^{'}\log(y_{i,j}) \tag{1}$$

Where $y_{i,j}^{'}$ is defined as:

$$y_{i,j}^{'} = \begin{cases} 1, if\ truth\ value\ of\ i^{th}\ direction\ lies\ in\ j^{th}\ bin \\ 0, otherwise \end{cases} \tag{2}$$

$y_{i,j}$ is similar to $y_{i,j}^{'}$, but represents the predicted value.

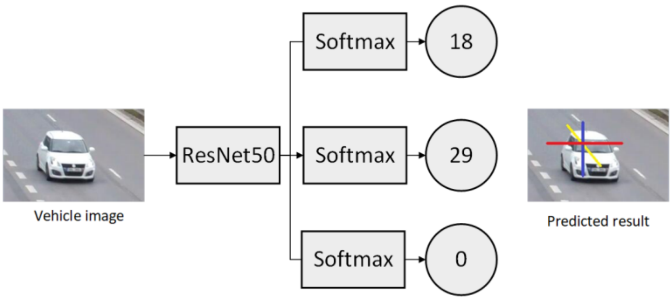

FIGURE 5: Using Resnet50 for prediction of directions. The vehicle image is fed to the network with three separate outputs that predict the angle spaces that its three directions belong to.

For vehicle contour detection, we use the fully convolutional encoder-decoder network designed in [33], which masks with probabilities of vehicle contours for each image pixel (see Fig. 6). To obtain the final contour, we use a different approach to [17], which searches for global maxima along the line segment from the center to the edge points of the 2D bounding box. Here, we only employ binarization to the probability map and use the binary image as the final contour. The reasons we do this simplification are: 1) It will accelerate the whole process. The search process used in [17] will be time-consuming when the 2D bounding boxes are large, as there will be hundreds of lines needed to replicate the search process. In addition, the extra work needed for 2D bounding box estimation is also time-consuming; 2) The difference has a little influence on the re-identification result. Compared with [17], the final contour will be just a little different in the edge region. In the similarity measure we use, this small difference only has a small effect on the computed similarity and can be ignored.

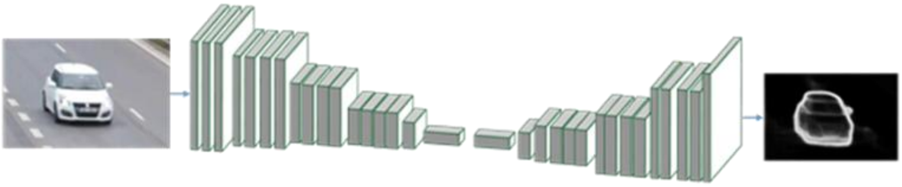

FIGURE 6: Using a fully convolutional encoder-decoder network for estimating vehicle contours

The process of using the vehicle's contour and three directions to construct 3D bounding boxes is as shown in Fig. 7. We put the intersections of each tangent in order, then a 3D bounding box can be obtained by connecting, in order, the seven points: $[(x_A, y_A), (x_B, y_B), (x_C, y_C), (x_D, y_D), (x_E, y_E), (x_F, y_F), (x_G, y_G)]$.

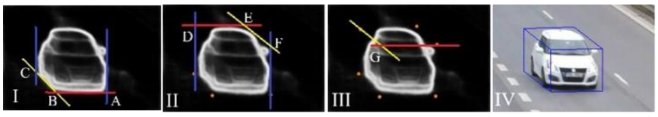

FIGURE 7: Use of a vehicle's three directions and contours to construct a 3D bounding box. (I) Tangent lines and their relevant intersections A, B, C. (II) Derived lines and their intersections E, D, F. (III) Derived lines and intersection G. (IV) Constructed bounding box

3.2 Similarity Measure

The global similarity between two images is calculated after we construct each vehicle's 3D bounding box. All three views of the vehicle—the front, top, and side—are split and the sub-region similarity between each pair of corresponding sides of these two vehicles is computed. Here, we use a simplified deformable diversity similarity (DDIS) method to compute the sub-region similarity.

3.2.1 Deformable Diversity Similarity

DDIS [34] is an algorithm for template matching, which can be used for computing the similarity between two images, and no prior information of the images is needed.

Given two images I_1 and I_2, let points $p_i, q_j \in \mathbb{R}^d$ represent patches of I_1 and I_2, respectively. DDIS is used to compute thesimilarity between two sets of points, $P = \{p_i\}_{i=1}^{N}$ and $Q = \{q_j\}_{j=1}^{M}$. Let p^a denote the appearance and p^l the location of patch p (and similarity with q). Find the appearance-based nearest-neighbor (NN) patch p_i for every point q_j s.t. $p_i = \mathrm{NN}^a(q_j, P) = argmin_{p\in P} d(q_j^a, p_a)$ for a given distance $d(q^a, p^a)$. Let $r_j = d(q_j^l, p_i^l)$ denote the location distance between point q_j and its NN^a. Then, define $k(p_i)$ as the number of patches $q \in Q$whose NN^a is p_i:

$$k(p_i) = |\{q \in Q: \mathrm{NN}^a(q, P) = p_i\}| \tag{3}$$

Finally, DDIS can be computed as:

$$\mathrm{DDIS}_{Q\to P} = c \sum_{q_j \in Q} \frac{1}{r_j + 1} \cdot \exp\left(1 - k\left(\mathrm{NN}^a(q_j, P)\right)\right) \tag{4}$$

where c=1/min $\{M, N\}$ is a normalization factor.

3.2.2 Simplified Deformable Diversity Similarity (SDDIS)

In our method, we simplify the DDIS according to the actual situation, which is reflected in two ways. 1) We set M as equal to N, where M and N are the numbers of patches of images I_1 and I_2, respectively. Because scale normalization to the corresponding views is performed, the two corresponding views are of the same size, which means that M equals N. 2) We set r_j = 0. Because $1/(r_j + 1)$ is used to quantify deformation in Equation (4), large deformation is given a penalty as r_j is big, so it finds the best and plausible match of a small image within a large image. In our work, we assume that the corresponding views are the best matching sides to each other.

In addition, as we cannot split all three views of the vehicle precisely, the images of some sides will include background orparts of other sides. This means that the location distance between a patch and its nearest-neighbor patch sometimes will be large, even if two images are of the same vehicle. So, the SDDIS similarity measure we use is:

$$\mathrm{SDDIS}_{Q\rightarrow P} = \frac{1}{M}\sum_{q_j \in Q} \exp\left(1 - k\left(\mathrm{NN}^{a}\left(q_j, P\right)\right)\right) \tag{5}$$

3.2.3 Global Similarity

Given two images—a query image T and candidate image S—let T_f, T_t, T_s be the three views (front, top and side) that are split from the 3D bounding box of T, and similarity for S. After computing all three sub-region similarities between each pair of corresponding views, the global similarity between T and S is calculated as:

$$\mathrm{F}(T, S) = W_f \cdot \mathrm{SDDIS}(T_f, S_f) + W_t \cdot \mathrm{SDDIS}(T_t, S_t) + W_s \cdot \mathrm{SDDIS}(T_s, S_s) \tag{6}$$

where $W_f, W_t\ and\ W_s$ denote the different weights given to each similarity. Different weights are given because each side has a different influence on the vehicle re-identification process. For example, the front side contains the lights, radiator grille and other discriminative features, which are more useful for distinguishing vehicles than the features of the other sides.

3.3 Vehicle Re-identification (VRID)

In summary, given a query vehicle T, we find the best-matching vehicle amongst a set of gallery candidates $S = \{S_1, \cdots, S_n\}$. For n pairs of vehicles $\{T, S_i\}$, we compute their global similarities with the process shown in Fig. 8, sort these n similarities $\{\mathrm{F}(T, S_1), \cdots, \mathrm{F}(T, S_i), \cdots, \mathrm{F}(T, S_n)\}$, and then the vehicle with the maximum similarity is the best match.

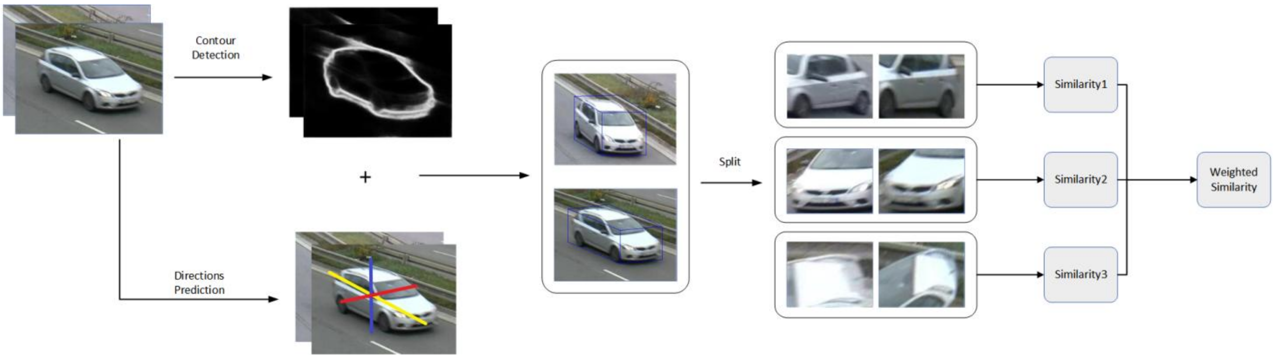

FIGURE 8: Complete process of similarity computing

IV. EXPERIMENTS

4.1 Experiment Dataset

4.1.1 Direction Prediction Dataset

For direction prediction, we used the BOXCAR116K dataset [17] to train the network. BOXCAR116K contains 116,287 images of vehicles shot by 137 cameras. It has annotations for each vehicles' 2D and 3D bounding boxes so we can obtain the required directions from each image by computing the angle between the edges of the 3D bounding boxes and the horizontal line, then take the result to be the ground truth.

4.1.2 Vehicle Re-identification Dataset

For vehicle re-identification, we used the BrnoCompSpeed dataset [35]. It contains 18 full-HD videos, each around 1 h long, captured at six different locations and in three directions for each location. We captured vehicle images manually and paired images of vehicles captured at two different directions. 240 pairs of images were obtained, with examples shown in Fig. 9.

4.2 Experiment Setting and Results

4.2.1 Direction Prediction

After analyzing the BOXCAR116K dataset, we found that direction V_3 was always approximately90°. Accordingly, we set V_3 to 90° and predicted the other two directions in the experiments.

We trained the Resnet50 network in TensorFlow, added two separate fully-connected layers with *softmax* activation (one for each direction) after the last pooling layer to perform multi-task prediction. We used 96,000 images for training and the other 20,286 images for testing. The prediction accuracy was 89.75% for V_1 and 91.99% for V_2.

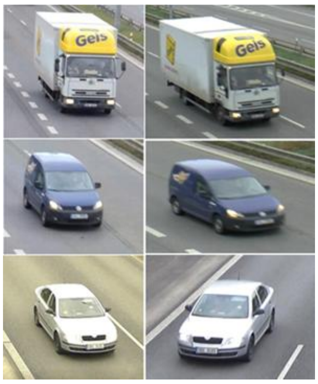

FIGURE 9: Pairs of vehicle images from a vehicle re-identification dataset.

4.2.2 Contour Detection:

For contour detection, we used the model directly pre-trained with the Pascal VOC2007 dataset. Some of the detection results are shown in Fig. 10.

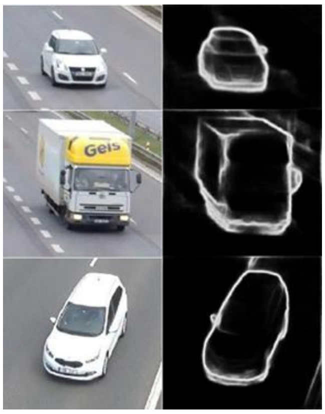

FIGURE 10: Examples of contour detection using the pre-trained model

The images show that the pre-trained contour detection model can acceptably predict a result that completely extracts the vehicle's contour with only a few background pixels incorrectly predicted as contour pixels.

4.2.3 3D Bounding Box Construction

After obtaining the vehicle directions and contours, we constructed 3D bounding boxes following the process shown in Fig. 7. Some of the results are shown in Fig. 11.The method of 3D bounding box construction can create a good result that captures the vehicle's 3D shape. This allows us to split the vehicle's three sides confidently based on the 3D bounding boxes.

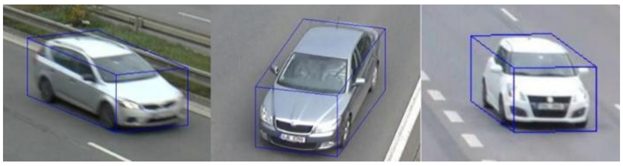

FIGURE 11: Examples of 3D bounding box construction

4.2.4 Vehicle Re-identification

As there are few studies on vehicle re-identification, we only compared our method with two others. 1) HOG+LLC, which is a typical method in the field of target re-identification. It first extracts an image's histograms of oriented gradients (HOG) [36], then uses locality-constrained linear coding (LLC) [37] to encode the HOG features, obtains a higher level of image semantic description and, finally, uses the correlation coefficient between the two images' feature vectors as a measure of their similarity. 2) DDIS. To prove that our method can effectively improve the performance of DDIS, we compared it with the method that directly computes the DDIS between two whole images. Furthermore, to compute the DDIS or SDDIS of two images, we first need to divide the images into patches according to a preset patch size. So, to make comparisons in different patch sizes, the patch size was set to 5, 7, and 9 in the experiment.

In this paper, the evaluation metric is the cumulative matching characteristic curve (CMC) [38], in which the accuracy in *k* is the probability of observing the correct identity within the top *k* ranks. The result is shown in Fig. 12 and Table I.

Using the surveillance video dataset, the proposed method achieved 75.83% re-identification accuracy in the top-1 rank and 91.25% accuracy in the top-5 ranks. This proves the feasibility of our approach.

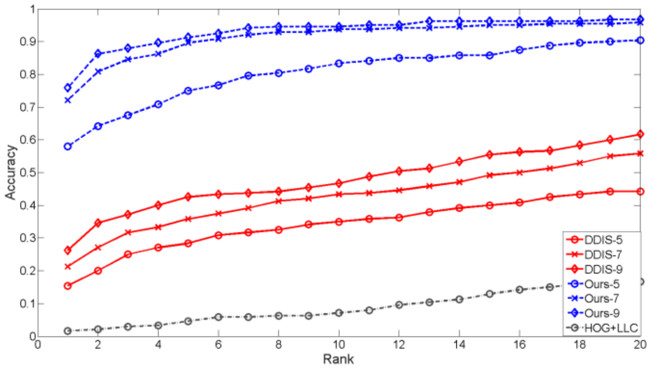

Figure 12: Results of vehicle re-identification of 240 pairs of vehicles using our method, the direct DDIS method, and HOG+LLC. The numbers 5, 7, and 9 represent the patch sizes used in SDDIS and DDIS to divide images into patches

Table I
Comparison of Different Vehicle Re-Identification Methods. Top-K Represents the Probability of Observing the Correct Identity Within the Top K Ranks

Method	Top-1	Top-2	Top-3	Top-4	Top-5
HOG+LLC	0.0167	0.0208	0.0292	0.0333	0.0458
DDIS-5	0.1542	0.2	0.25	0.2708	0.2833
DDIS-7	0.2125	0.2708	0.3167	0.3333	0.3583
DDIS-9	0.2625	0.3458	0.3708	0.4	0.425
Ours-5	0.5792	0.6417	0.675	0.7083	0.75
Ours-7	0.7208	0.8083	0.8458	0.8625	0.8958
Ours-9	**0.7583**	**0.8625**	**0.8792**	**0.8958**	**0.9125**

4.3 Analysis

From the experimental results, we can see that our method has much better performance than the HOG+LLC and direct DDIS methods. With the patch size set to 9, our method obtains the best result: the top-1 accuracy is 75.83%, and the top-5 accuracy is 91.25%. Compared with using DDIS directly, our method was 188.9% and 114.7% more accurate, respectively, which proves that it can effectively improve DDIS performance. Both the overall performance and the comparison using DDIS directly prove the feasibility of our method.

The reason why our method can achieve higher accuracy is that it can deal with the deformation of each view rather than just the overall deformation. Just as shown in Fig. 13, the images of a vehicle captured from different camera views have obvious differences in appearance, especially when the images are shot from different sides of the vehicle. Even when we resize the images so that the vehicles are the same size, deformation also exists, so that the differences between images are obvious. In our method, three views of the vehicle are split and scale-normalized to ensure that the vehicle sub-images are the same size. For images shot from different directions, we flipped them. As we can see in Fig. 13, differences in appearances between images can be minimized by using the process above.

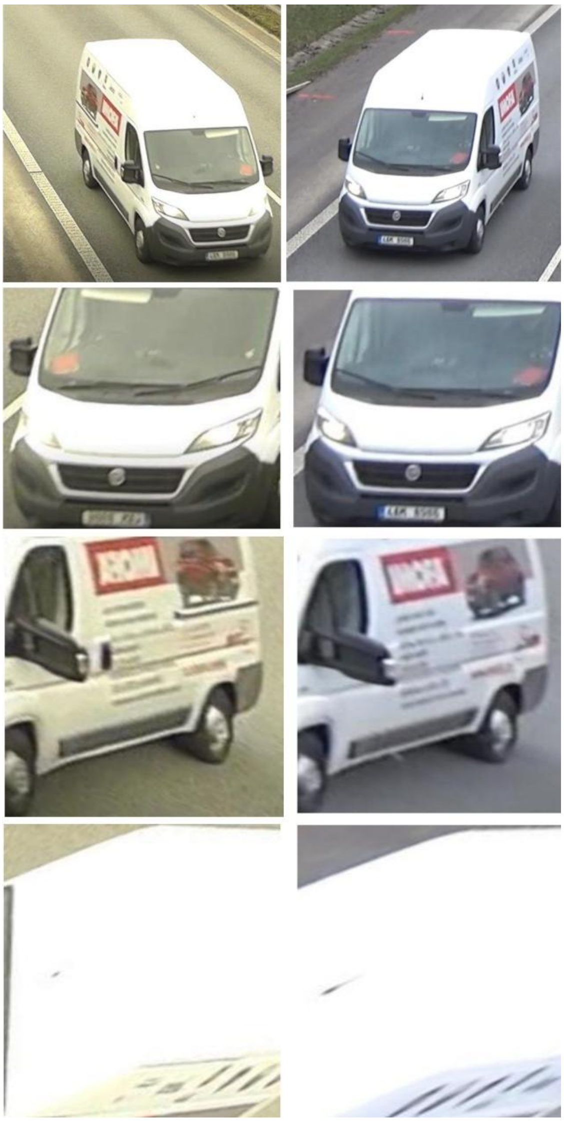

FIGURE 13: Images of a vehicle captured from different camera views, with their three sides split according to 3D bounding boxes. After performing scale-normalization on each corresponding side, differences in the appearance of image pairs can be minimized.

The results indicate that the HOG+LLC method achieved low accuracy. Variation in vehicle orientation and the exclusion of color information may account for this issue.

From the experimental results, the bigger the path size set, the higher the accuracy. This may be because the texture of the vehicle is not complicated. When the patch size is small, many patches in the candidate image may share the same nearest-neighbor patch in the query image as their appearances are similar, which results in a relatively small DDIS value. When the patch size increases, this problem will appear less, so the accuracy will be higher.

V. CONCLUSION

In actual situations, the main differences between images captured from different camera views are differences in scale and image deformation. In order to eliminate these differences, we need a method that can deal with the local deformation of each side of a vehicle, rather than its overall deformation. In this paper, we proposed a method for vehicle re-identification based on the authenticity of orthogonal projection. This means that by making three-view drawings according to orthographic projection, we can obtain the real shape of all three sides of an object. This minimizes scale differences and image deformation.

By splitting vehicle information into three views and flipping the images as required, we can solve the problem presented by images being shot from different sides. This may provide a feasible way to perform vehicle re-identification when images are shot from different directions, including anterior and posterior views.

A comparison between our method and the direct DDIS method demonstrates the potential of our approach as an auxiliary method for improving the performance of some existing vehicle re-identification methods. Future work will expand this approach to other vehicle re-identification methods, ultimately aiming to build a system suitable for application to expressways.

ACKNOWLEDGEMENTS

This work is supported by Key-Area Research and Development Program of Guangdong Province, No.2020B090921003.

REFERENCES

[1] S. Sunderrajan and B. Manjunath, “Context-aware hypergraph modeling for re-identification and summarization,” IEEE Transactions on Multimedia, vol. 18, no. 1, pp. 51–63, 2016.

[2] S. Ding, L. Lin, G. Wang, and H. Chao. “Deep feature learning with relative distance comparison for person reidentification”. Pattern Recognition, vol. 48, no. 10, pp. 2993–3003,2015.

[3] Q. Leng, R. Hu, C. Liang, Y. Wang and J. Chen, "Bidirectional ranking for person re-identification," 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, 2013, pp. 1-6.

[4] R. S. Feris et al., "Large-Scale Vehicle Detection, Indexing, and Search in Urban Surveillance Videos," IEEE Transactions on Multimedia, vol. 14, no. 1, pp. 28-42, Feb. 2012.

[5] W. Ge, X. Lin, Y. Yu and J. Sochor, “Weakly Supervised Complementary Parts Models for Fine-Grained Image Classification from the Bottom Up,” 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3029-3038.

[6] N. Li, J. J. Jain, and C. Busso, “Modeling of driver behavior in real world scenarios using multiple noninvasive sensors,” IEEE Transactions on Multimedia, vol. 15, no. 5, pp. 1213–1225, 2013.

[7] A. R. Selokar and S. Jain, “Automatic number plate recognition system using a fast stroke-based method,” IEEE Transactions on Multimedia, vol. 1, no. 7, pp. 1-5, Apr. 2014.

[8] S. Du, M. Ibrahim, M. Shehata and W. Badawy, "Automatic License Plate Recognition (ALPR): A State-of-the-Art Review," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 2, pp. 311-325, Feb. 2013.

[9] C. Gou, K. Wang, Y. Yao and Z. Li, "Vehicle License Plate Recognition Based on Extremal Regions and Restricted Boltzmann Machines," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 4, pp. 1096-1107, April 2016.

[10] Y. Li, Y. Li, H. Yan and J. Liu, "Deep joint discriminative learning for vehicle re-identification and retrieval," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 395-399.

[11] J. Zhu, H. Zeng, Y. Du, Z. Lei, L. Zheng and C. Cai, "Joint Feature and Similarity Deep Learning for Vehicle Re-identification," IEEE Access, vol. 6, pp. 43724-43731, 2018.

[12] H. Liu, Y. Tian, Y. Wang, L. Pang and T. Huang, "Deep Relative Distance Learning: Tell the Difference between Similar Vehicles," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2167-2175.

[13] Y. Zhou and L. Shao, "Vehicle Re-Identification by Adversarial Bi-Directional LSTM Network," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, 2018, pp. 653-662.

[14] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 815-823.

[15] Y. Sun et al., "Circle Loss: A Unified Perspective of Pair Similarity Optimization," 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 6398-6407.

[16] Jae-Hyeung Park, Joohwan Kim, and Byoungho Lee, "Three-dimensional optical correlator using a sub-image array," Optics express, vol 13, no. 13, pp:5116-5126, 2005.

[17] J. Sochor, J. Spanhel and A. Herout, "BoxCars: Improving Fine-Grained Recognition of Vehicles Using 3-D Bounding Boxes in Traffic Surveillance," IEEE Transactions on Intelligent Transportation Systems. DOI: 10.1109/TITS.2018.2799228.

[18] L. Wei, S. Zhang, H. Yao, W. Gao and Q. Tian, "GLAD: Global-Local-Alignment Descriptor for Scalable Person Re-Identification," IEEE Transactions on Multimedia. DOI: 10.1109/TMM.2018.2870522.

[19] Z. Wang et al., "Zero-Shot Person Re-identification via Cross-View Consistency," IEEE Transactions on Multimedia, vol. 18, no. 2, pp. 260-272, Feb. 2016.

[20] W. Chen, X. Chen, J. Zhang and K. Huang, "Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-identification," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 2017, pp. 1320-1329.

[21] M. Ye et al., "Person Reidentification via Ranking Aggregation of Similarity Pulling and Dissimilarity Pushing," IEEE Transactions on Multimedia, vol. 18, no. 12, pp. 2553-2566, Dec. 2016.

[22] M. S. Sarfraz, A. Schumann, A. Eberle, and R. Stiefelhagen, "A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking," 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 420–429.

[23] C. Luo, Y. Chen, N. Wang, and Z. Zhang, "Spectral feature transformation for person re-identification," 2019 IEEE International Conference on Computer Vision (ICCV), 2019, pp. 4976–4985.

[24] X. Liu, W. Liu, H. Ma and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, 2016, pp. 1-6.

[25] X. Liu, W. Liu, T. Mei and H. Ma, "PROVID: Progressive and Multimodal Vehicle Reidentification for Large-Scale Urban Surveillance," IEEE Transactions on Multimedia, vol. 20, no. 3, pp. 645-658, March 2018.

[26] Y. Shen, T. Xiao, H. Li, S. Yi and X. Wang, "Learning Deep Neural Networks for Vehicle Re-ID with Visual-spatio-Temporal Path Proposals," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 1918-1927.

[27] N. Jiang, Y. Xu, Z. Zhou and W. Wu, "Multi-Attribute Driven Vehicle Re-Identification with Spatial-Temporal Re-Ranking," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018, pp. 858-862.

[28] Zhu, J., Du, Y., Hu, Y., Zheng, L., and Cai, C. "VRSDNet: vehicle re-identification with a shortly and densely connected convolutional neural network". Multimedia Tools and Applications, 2018, pp. 1-15.

[29] Y. Bai, Y. Lou, F. Gao, S. Wang, Y. Wu and L. Duan, "Group-Sensitive Triplet Embedding for Vehicle Reidentification," IEEE Transactions on Multimedia, vol. 20, no. 9, pp. 2385-2399, Sept. 2018.

[30] B. He, J. Li, Y. Zhao and Y. Tian, "Part-Regularized Near-Duplicate Vehicle Re-Identification," 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3992-4000.

[31] Y. Zhou, L. Liu and L. Shao, "Vehicle Re-Identification by Deep Hidden Multi-View Inference," IEEE Transactions on Image Processing, vol. 27, no. 7, pp. 3275-3287, July 2018.

[32] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[33] J. Yang, B. Price, S. Cohen, H. Lee and M. Yang, "Object Contour Detection with a Fully Convolutional Encoder-Decoder Network," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 193-202.

[34] Talmi, R. Mechrez and L. Zelnik-Manor, "Template Matching with Deformable Diversity Similarity," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 2017, pp. 1311-1319.

[35] J. Sochor et al., "Comprehensive Data Set for Automatic Single Camera Visual Speed Measurement," in IEEE Transactions on Intelligent Transportation Systems. DOI: 10.1109/TITS.2018.2825609.

[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1.

[37] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang and Y. Gong, "Locality-constrained Linear Coding for image classification," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2010, pp. 3360-3367.

[38] C. Arth, C. Leistner and H. Bischof, "Object Reacquisition and Tracking in Large-Scale Smart Camera Networks," 2007 First ACM/IEEE International Conference on Distributed Smart Cameras, Vienna, 2007, pp. 156-163.