Vehicle Re-identification with Spatio-temporal Information

Hye-Geun Kim[1], YouKyoung Na[2], Hae-Won Joe[2], Yong-Hyuk Moon[3,4], Yeong-Jun Cho[1]
[1]Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju, Republic of Korea
[2]Department of Artificial Intelligence, Chonnam National University, Gwangju, Republic of Korea
[3]Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea
[4]Department of AI, University of Science and Technology, Daejeon, Republic of Korea
Email: hyegeunkim@jnu.ac.kr, youkyoung823@gmail.com, 214583@jnu.ac.kr, yhmoon@etri.re.kr, yj.cho@jnu.ac.kr

***Abstract*—Vehicle re-identification (re-id) in a large-scale camera network is important in public safety, traffic control, and security. However, due to the appearance ambiguities of vehicle, the previous appearance-based re-id models often fail to track vehicle across multiple cameras. To overcome the challenge, in this work, we utilize spatial and temporal information between cameras and individual vehicle. To this end, we build camera network topology that consists of transition time probabilities between camera pairs. Then, we merge the appearance similarity and spatio-temporal probability scores for the final re-id similarity. Based on our methods, we improved vehicle re-id performance on public dataset (`VeRi776`). The experiment results support that utilizing spatial and temporal information in vehicle re-id can leverage the baseline re-id performance and handle the vehicle appearance ambiguities.**

***Index Terms*—Vehicle re-identification, Spatial and temporal information, Large-scale camera networks**

I. Introduction

Vehicle re-identification (re-id) aims to identify the same vehicles across non-overlapping multiple cameras. Recently, it is increasingly important in public safety, traffic control, and security. Performing vehicle re-id in the large-scale multi-camera network presents several challenges. First, there is a very large number of vehicles appearing in the camera network. It occurs high computational complexity in performing re-id. Second, there are vehicles that are exactly the same in appearance (shape, model and color) making an appearance ambiguity problem in vehicle re-id. Many studies have been focused on feature learning [1]–[3] and metric learning [4]–[6] to represent appearances of objects to perform re-id. However, those appearance-based re-id models still

struggle with appearance ambiguity.

To handle the challenges in vehicle re-id, we first analyze the characteristics of vehicles in camera networks. Compared to person re-id [1], [4], [7]–[9], vehicles have unique characteristics as follows: first people have distinctive appearance features, such as different face, clothing and body shapes, whereas vehicles can have the exactly same appearance (e.g., model, shape and color) according to their model types. Second, vehicle movements across cameras are predictable than people because vehicles only move along roads and highways.

Based on the characteristics of vehicles, in this work, we additionally utilize spatial and temporal information of camera networks and individual vehicles to overcome the appearance ambiguity problem. To this end, we first build camera network topology using ground-truth vehicle matching data between each camera pair. The camera network topology consists of camera transition time distributions that involve spatial and temporal relationships between camera pairs. Then, we aggregate appearance similarity and spatio-temporal probability scores for the final re-id similarity.

We tested a public vehicle re-id dataset `VeRi776` [10]. Due to the proposed spatial and temporal re-id framework, Rank-1 score was improved by 3.64% and mAP was improved by 2.89%. The experimental results support that using spatial and temporal information can significantly leverage the re-id performance of the baseline appearance-based re-id model. The framework of the proposed methods is simple but very effective.

II. Proposed Methods

Figure. 1 illustrates overall framework of the proposed methods. For the appearance-based re-identification (re-id) model, we utilized FastReID [11]. Note that we can use any kind of appearance-based models as the baseline of our framework. In the training stage, transition time distributions are built by the spatial-temporal information in the training data. Finally, we combine the appearance similarity with the transition time distribution for the final re-id similarity.

A. Transition time distribution

We estimate transition time distributions by using the spatial-temporal information of the given data. First, we extract transition time data $D = \{d_1, d_2, \ldots, d_N\}$ with spatial-temporal information and convert it to a normalized histogram. N is a number of transition data. The transition time distribution is fitted with a Gaussian distribution based on a normalized histogram, with the μ and σ deviation calculated as follows

$$\mu = \frac{\sum_i x_i \times h_i}{\sum_i h_i}, \quad \sigma = \sqrt{\sum_i h_i \times (x_i - \mu)^2} \tag{1}$$

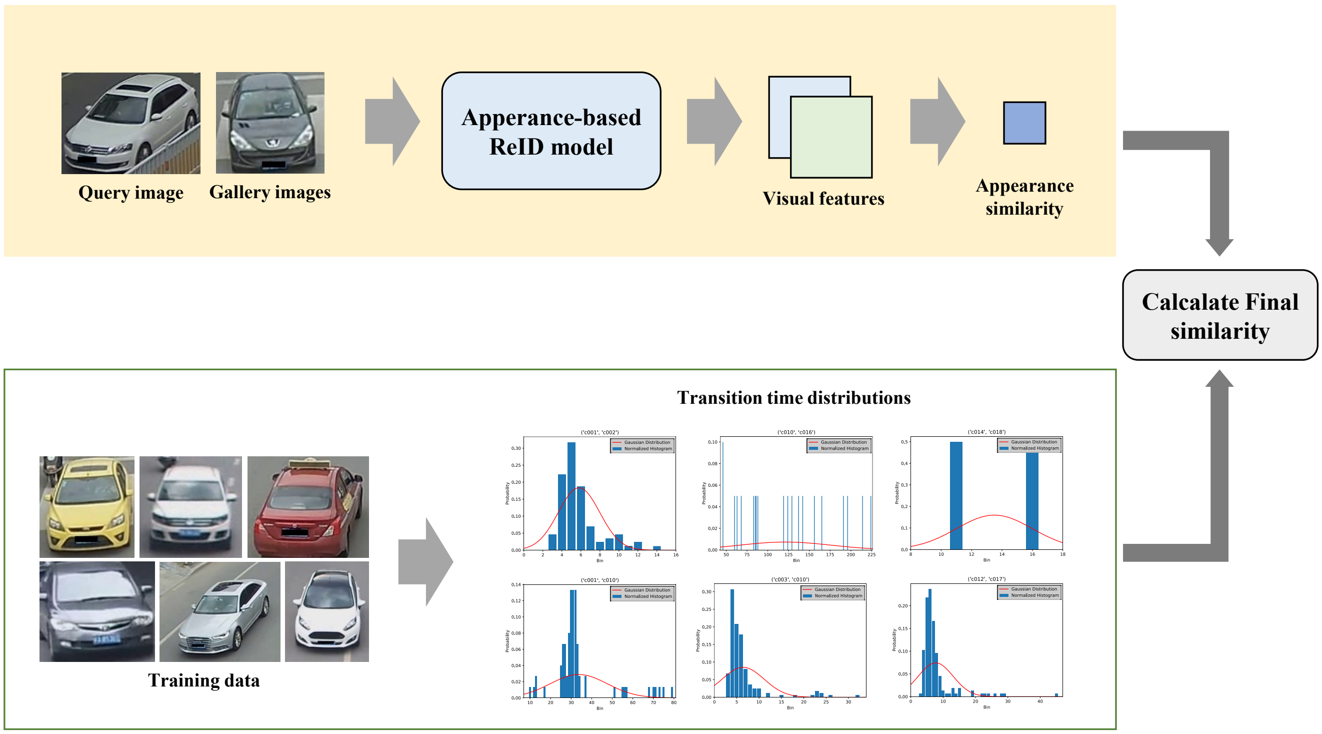

Fig. 1: The overall proposed spatio-temporal vehicle re-identification framework

We can calculate the spatial-temporal probability $p(d_i|\mu, \sigma^2)$ for transition time data d_i from a transition time distribution fitted with a Gaussian distribution as follows:

$$p(d_i|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(k_{d_i} - \mu)^2}{2\sigma^2} \tag{2}$$

where k_{d_i} is the bin index of the histogram to the transition time data d_i.

B. ReID with spatial-temporal information

We convert the cosine similarity S_c to the appearance similarity S_A in the range [0, 1] calculated by the appearance-based ReID model. We can calculate the final similarity of images $\mathbf{x}_i$ and $\mathbf{x}_j$ by simply multiple the appearance similarity and spatial-temporal probability as follows

$$S_P(\mathbf{x}_i, \mathbf{x}_j) = S_{A_{i,j}} p(\Delta f_{i,j}|\mu, \sigma^2) \tag{3}$$

where $S_P(\mathbf{x}_i, \mathbf{x}_j)$ be the final similarity for images $\mathbf{x}_i$ and $\mathbf{x}_j$, $S_{A_{i,j}}$ be the appearance similarity for the two images, and $p(\Delta f_{i,j}|\mu, \sigma^2)$ be the space-time probability for the two images. The $\Delta f_{i,j}$ is the transition time difference between the two images. However, this can remove the effect of appearance similarity because it applies spatial-temporal probabilities directly to appearance similarity. We utilize a balanced factor λ to combine appearance similarity and spatial-temporal probability optimally.

$$S_P(\mathbf{x}_i, \mathbf{x}_j) = \lambda S_{A_{i,j}} + (1 - \lambda) p(\Delta f_{i,j}|\mu, \sigma^2) \tag{4}$$

We set λ to 0.5. Following eq.4, we calculate the final similarity S_P by combining the appearance similarity and the spatial-temporal probability.

III. Experimental Results

A. Dataset and Settings

For experiments, we used `VeRi776` [10] vehicle re-identification (re-id) dataset. It has over 49,000 images of 776 different vehicles captured by 20 non-overlapping synchronized cameras. Among them, we selected six camera pairs as summarized in Table. I. To evaluate the re-id performance, we calculated Rank-1, 5, 10 accuracies and mean average precision (mAP) score. For the baseline appearance model, we trained FastReID [11] to extract appearance similarity. ResNet-50 is set to its backbone network structure and the training parameters as follows: epoch – 60, batch size – 64. Note the we can use other state-of-the-art appearance-based re-id model for our framework.

B. Experiments

In the experiments, we tested six different camera pairs in `Veri776` dataset. 1–4 pairs have sufficient numbers of transition vehicles. On the other hand, 5–6 pairs have a few numbers of transition vehicles between cameras. For the camera pairs (1–4), Rank-1 performance is increased significantly by 6.46%, 6.98%, 1.19%, and 8.33%, and mAP increased by 6.49%, 4.06%, 0.91%, and 6.49%. In comparison, we observed a slight decrease and maintained performance in

Index	Camera pair	Method	Rank-1	Rank-5	Rank-10	mAP
1	Cam1–Cam2	baseline	83.06	88.71	95.16	83.10
		+ st info.	89.52	93.55	96.77	89.59
2	Cam1–Cam10	baseline	80.23	89.53	96.51	83.60
		+ st info.	87.21	90.70	96.51	87.66
3	Cam3–Cam10	baseline	99.45	100.0	100.0	98.01
		+ st info.	98.35	100.0	100.0	97.65
4	Cam12–Cam17	baseline	98.21	99.40	100.0	97.10
		+ st info.	99.40	99.40	100.0	98.01
5	Cam10–Cam16	baseline	98.33	98.33	98.33	96.35
		+ st info.	98.33	98.33	98.33	96.13
6	Cam14–Cam18	baseline	91.67	91.67	100.0	91.53
		+ st info.	100.0	100.0	100.0	98.02

TABLE I: Performance comparison of the baseline (FastReID) and the proposed spatial-temporal re-id method (+st info.)

Rank-1 of 1.1% and 0%, and mAP of 0.36% and 0.22% for the two camera pairs (5–6). Although the re-id performances of those camera pairs containing a few transition vehicles are not improved, on average, re-id performance has improved significantly. Especially, the performance of camera pair (Cam14–Cam18) is improved Rank-1 from 91.67% to 100.0% significantly, alleviating the appearance ambiguity perfectly. The results support that using spatial-temporal information alleviates appearance ambiguity in the appearance-based re-id model.

IV. Conclusion

In this work, we proposed a vehicle re-identification (re-id) framework that combines appearance similarity and spatio-temporal similarity to handle appearance ambiguity problem in vehicle re-id. In our experiments with in `VeRi776` data, we achieved significant performance improvements. The results confirm that the proposed framework alleviates appearance ambiguity and helps accurate vehicle re-id in the large-scale camera network. We fit the transition time distribution using the Gaussian distribution in this work. In future work, we will give the histogram a proper distribution so that every camera pair has a proper spatial-temporal probability.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2021-0-00907, Development of Adaptive and Lightweight Edge-Collaborative Analysis Technology for Enabling Proactively Immediate Response and Rapid Learning).

References

[1] S. Khamis, C.-H. Kuo, V. K. Singh, V. D. Shet, and L. S. Davis, “Joint learning for attribute-consistent person re-identification,” in *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III 13*. Springer, 2015, pp. 134–146.

[2] N. Peri, P. Khorramshahi, S. S. Rambhatla, V. Shenoy, S. Rawat, J.-C. Chen, and R. Chellappa, “Towards real-time systems for vehicle re-identification, multi-camera tracking, and anomaly detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 622–623.

[3] J. Zhao, Y. Zhao, J. Li, K. Yan, and Y. Tian, “Heterogeneous relational complement for vehicle re-identification,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 205–214.

[4] A. Hermans, L. Beyer, and B. Leibe, “In defense of the triplet loss for person re-identification,” *arXiv preprint arXiv:1703.07737*, 2017.

[5] Y. Bai, Y. Lou, F. Gao, S. Wang, Y. Wu, and L.-Y. Duan, “Group-sensitive triplet embedding for vehicle reidentification,” *IEEE Transactions on Multimedia*, vol. 20, no. 9, pp. 2385–2399, 2018.

[6] A. Ghosh, K. Shanmugalingam, and W.-Y. Lin, “Relation preserving triplet mining for stabilising the triplet loss in re-identification systems,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4840–4849.

[7] D. Chen, Z. Yuan, B. Chen, and N. Zheng, “Similarity learning with spatial constraints for person re-identification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1268–1277.

[8] 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 *Proceedings of the iEEE conference on computer vision and pattern recognition*, 2016, pp. 1335–1344.

[9] G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, G. Yu, E. Zhou, and J. Sun, “High-order information matters: Learning relation and topology for occluded person re-identification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6449–6458.

[10] X. Liu, W. Liu, T. Mei, and H. Ma, “A deep learning-based approach to progressive vehicle re-identification for urban surveillance,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 869–884.

[11] L. He, X. Liao, W. Liu, X. Liu, P. Cheng, and T. Mei, “Fastreid: A pytorch toolbox for general instance re-identification,” *arXiv preprint arXiv:2006.02631*, 2020.