Tracklet-based vessel re-identification for multicamera vessel-speed enforcement

Tracklet-based Vessel Re-identification for Multi-Camera Vessel-Speed Enforcement

H.G.J. Groot[1] M.H. Zwemer[1,2] E. Bondarev[1]
R.G.J. Wijnhoven[2] P.H.N. de With[1]

[1]Eindhoven University of Technology, Dept. of EE, VCA Research Group
P.O. Box 513, 5600 MB Eindhoven, the Netherlands
[2]ViNotion B.V., Daalakkersweg 2-58, 5641 JA Eindhoven, the Netherlands
emails: {H.G.J.Groot, M.Zwemer, E.Bondarev}@tue.nl,
rob.wijnhoven@vinotion.nl, P.H.N.de.With@tue.nl

Abstract

In crowded waterways, maritime traffic is bound to speed regulations for safety reasons. Although several speed measurement techniques exist for road traffic, such systems are not available for vessels. This paper proposes a new approach for tracklet-based re-identification (re-ID) as a solution for vessel-speed enforcement. For evaluation, the Vessel-reID dataset is used that we introduced in previous work [2]. The core of the tracklet re-ID approach is based on a novel Tracklet-based Querying Procedure as a more effective alternative to the Common Querying Procedure (CQP) found in popular re-ID datasets [7, 8]. The existing procedure randomly selects a single image from the whole query-vessel trajectory (in one camera view). This is improved by (1) detecting a set of most representative images per tracklet of a query-vessel, and by (2) raising the matching accuracy based on accumulating the gallery similarity scores for all images in the set. In the experimental validation, we adopt two well-known person re-ID algorithms, TriNet [3] and MGN [6], since most re-ID literature focuses on person re-ID. Results show a significant increase in performance by applying the tracklet-based approach instead of CQP: a gain of 5.6% and 8.1% Rank-1 for MGN and TriNet, respectively.

1 Introduction

For maritime traffic, speed regulations are important, since speeding vessels can produce waves with a high water displacement that can be dangerous to other waterway users, especially small boats or swimmers. Road-traffic speed-control systems are typically implemented by applying a radar system to detect the speeding vehicles in combination with a camera system to record their license plates. Systems that measure the speed of road users over a longer trajectory also exist and are typically implemented by applying license-plate detection in two sufficiently distant cameras. However, for maritime traffic, none of these systems are suitable, since (1) vessels do not have well-defined license plates or other standardized visual registration markers, and (2) radar systems are too expensive and have difficulties with irregularly maneuvering vessels.

The overall visual vessel appearance of maritime traffic is often unique, because most vessels are of different vessel type, have a different bow or cabin, or can have distinguishing details such as flags or buoys. Therefore, visual re-identification of vessels between various camera locations is theoretically possible. To this end, we introduce a novel speed-enforcement system of vessels by applying video-based vessel re-identification (re-ID). Since this system measures the speed of vessels over a long

Figure 1: Visual example of several vessel trajectories in Camera 1 (left) and Camera 2 (right) of the Vessel-reID dataset, where each row shows a vessel instance in our dataset. Some images are skipped for visualization (denoted by dotted red line). Image from [2].

Figure 2: Example images of the same vessel appearing in Camera 1 (left) and 2 (right). Image adopted from [2]

trajectory, this will inherently lead to an accurate speed measurement. However, visual vessel re-ID poses several challenges, such as fluctuating weather and lighting conditions. Moreover, depending on the camera angle and position constraints, it should deal with varying poses of the object with respect to the camera.

This work builds further on our previous research [2], where we have introduced a novel Vessel-reID dataset to investigate whether visual-based vessel re-ID is feasible in practice. This Vessel-reID dataset is captured with two cameras placed several kilometers apart and contains image cutouts of all the detected vessels. In total, there are 2,474 unique vessels, where several example cutouts are shown in Figure 1 for both cameras. These image cutouts are automatically determined by cropping and selecting the bounding-box detections that a Single Shot Multibox Detector (SSD) [5, 9] produces on the full-frame camera footage, see Figure 2. Since all popular re-ID datasets use this image cutout format [7, 8], our Vessel-reID dataset is suitable to evaluate re-ID techniques on vessels. Contrary to these datasets though, we will test our technique on new unseen data during other weather conditions, instead of mixing the data of all captured conditions.

In this paper, we improve the querying procedure commonly applied in re-ID literature. Furthermore, as compared to our previous work [2], we incorporate an improved re-ID model. These two contributions further enhance re-ID performance and application feasibility. We refer to our initial work [2] for the details of creating the Vessel-reID dataset and focus on re-ID in this paper.

The main contributions of this paper are as follows. First, a novel Tracklet-based

Querying Procedure is introduced as an alternative to the querying procedure commonly applied in nearly all re-ID literature. Second, it is shown that the Tracklet-based Querying Procedure significantly improves baseline performance, independent of the re-ID model used.

The remainder of the paper is structured as follows. Section 2, explains important related work on re-identification. Section 3 introduces the system overview for the proposed Tracklet-based Querying Procedure. Section 4 discusses the experimental validation of the procedure. The paper finalizes in Section 5 with the conclusions.

2 Related work

Most re-ID literature focuses on person re-ID, where persons are tracked when they travel from one camera view to another. Recently, re-ID algorithms have also been evaluated on the road vehicle object class, while achieving high performance [1]. This shows that re-ID learning networks do generalize well to other domains, such as vessel detection. These techniques often adopt the use of Convolutional Neural Networks (CNNs), since their introduction clearly made re-ID algorithms more mature [4]. Initially, pairwise verification CNNs were popular, where the networks trained using image pairs of samples from either similar or different persons [4]. By applying the contrastive loss, these networks learn to embed the visual properties of the persons into meaningful feature vector descriptions. Verification CNNs achieve this by increasing the feature distance between images of different persons, while decreasing the distance between images of the same person. However, at present nearly all re-ID algorithms are metric-embedding CNNs [3, 1, 6]. These are trained by sets of 3 images, so-called triplets. A triplet consists of an anchor image, an image of the same person (as in the anchor), and an image of a different person. The training is guided by applying the corresponding triplet loss. Together with a more-effective mining strategy, this type of CNN is popular and offers state-of-the-art performance [3]. The TriNet network [3] proves this concept and still continues to perform well.

Another often exploited technique is to focus on image partitions, for instance on a person's head, body and legs. Here, the Multiple Granularity Network (MGN) [6] devotes 3 separate ResNet-50 branches to different partitionings. Furthermore, MGN also introduces a training strategy that involves both the triplet loss and the cross-entropy loss. The combination of both contributions results in a significant performance gain. The majority of related work, including [3, 1, 6], utilize the ResNet-50 network as a backbone architecture.

3 System overview

As depicted in Figure 3, the re-ID system aims to find a vessel of the same identity as the query-vessel in the available gallery. The Query-vessel tracklet shown in the figure can originate from any camera and represents a newly detected vessel when it appears. The gallery consists of all images of all vessels previously seen by the camera network. All images from both the Query-vessel tracklet and the gallery, are embedded before matching commences in the depicted Tracklet-based matching component. The two Feature embedding components are responsible for creating those embeddings. We experimentally validate our Tracklet-based matching component with two well-known person re-ID algorithms, TriNet [3] and MGN [6]. The Feature embedding component is thus either TriNet or MGN.

The core of the proposed re-ID system, i.e. the Tracklet-based matching component, is presented in Figure 4. From now on, this component is referred to as the Tracklet-based Querying Procedure. The key property of this procedure is that it utilizes several images from the query tracklet instead of just one. More exactly, it selects and

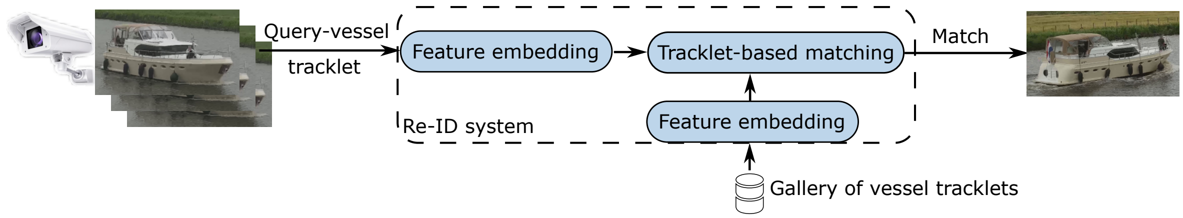

Figure 3: System overview of the re-ID system. The Query-vessel tracklet may originate from any camera and the Tracklet-based matching component is illustrated in Figure 4.

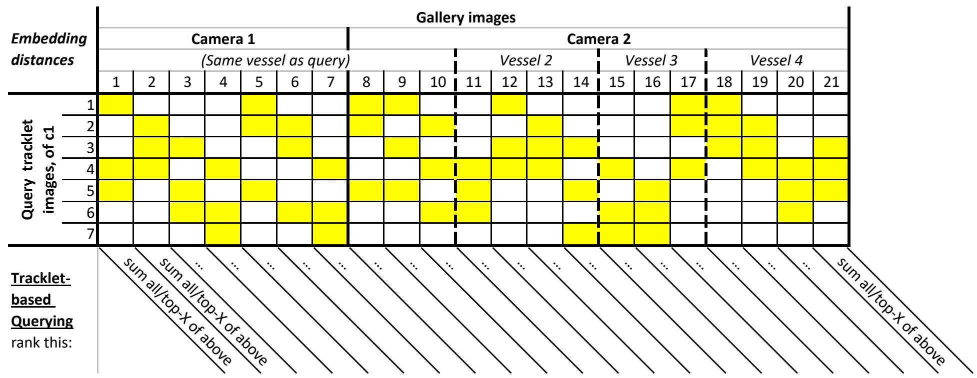

Figure 4: Visualization of our novel Tracklet-based Querying Procedure. The yellow boxes indicate the top-3 query tracklet images per gallery image (i.e. per column). For conciseness, there is only a single vessel visualized in the gallery of Camera 1 (c1).

exploits only the most representative images of the query tracklet. To achieve this, the procedure contains the following processing steps. First, the Euclidian distance between every embedded image in the query tracklet and every embedded image in the gallery is determined. This results in the matrix as depicted in Figure 4. Note that this Euclidian distance is effectively a similarity score, where smaller means more similar. Second, for every gallery image (for every column shown), all similarity scores are accumulated into a single (accumulated) similarity score. The accumulation happens by summing only the Top-X scores in that column, i.e. the X smallest distances in the column. These correspond to the most representative images of the query tracklet. The end result of this step is the horizontal vector illustrated underneath the matrix in Figure 4. Third, the vessel ID corresponding to the gallery image with the smallest accumulated similarity score defines the match. This process is called ranking. Since the gallery per definition also contains the query tracklet, the accumulated similarity scores corresponding to the query tracklet are excluded during ranking. Otherwise, the re-ID system would always be correct, since the embedding distance towards the same image is zero.

In line with the most popular (person) re-ID datasets [7, 8], we adopt the so-called closed-set approach. This means that each vessel tracklet in the test set is alternatingly used as the query, while all vessel tracklets are in the gallery. Several examples of the applied images are shown in Figure 1. During matching, re-ID algorithms typically apply the Common Querying Procedure, which is also used here. This procedure would replace the Tracklet-based matching component in Figure 3 and is substantially different from the proposed Tracklet-based Querying Procedure as explained above. In the Common Querying Procedure, a single image from the whole query-vessel tracklet (in one camera view) is selected randomly as the query-vessel representative. During ranking, the gallery image closest to this single representing image (in embedding

Table 1: Re-ID performance on our Vessel-reID dataset, when applying the novel querying procedure. The best performance a re-ID model achieves is indicated in bold. Results reported as mean (std.dev.) over 10 cycles.

	Accuracy on test set [%]			
	TriNet		**MGN**	
Description	**Rank-1**	**mAP**	**Rank-1**	**mAP**
Common Querying Procedure	55.9 (±1.4)	49.7 (±1.0)	68.9 (±0.9)	62.6 (±0.6)
Tracklet-based Querying				
All	61.9 (±1.4)	55.0 (±1.0)	73.9 (±0.7)	68.1 (±0.5)
Top-10	63.7 (±1.1)	56.3 (±0.8)	73.8 (±0.9)	68.0 (±0.6)
Top-15	**64.0** (±**1.2**)	**56.5** (±**0.9**)	74.3 (±0.9)	68.2 (±0.6)
Top-25	63.8 (±1.2)	56.4 (±0.9)	**74.5** (±**0.9**)	**68.3** (±**0.5**)

space) defines the most likely match. Hence, in Figure 4, only a single random row of the matrix would be incorporated during ranking. However, this single random image is likely not always a good representative of the query vessel. For instance, it occurs often that the query-vessel is temporarily occluded by another vessel, e.g. a vessel that travels in the opposite direction. Therefore, we propose to apply the Tracklet-based Querying Procedure as a more effective alternative. It should be noted that the Common Querying Procedure is implied by the popular datasets and their definition of the query set.

4 Experiments

To evaluate the proposed querying procedure, we use Rank-1 and mAP (mean Average Precision). The Rank-1 is most important because this metric indicates best how the overall system will perform in practice, i.e. in an open-set case. The adopted TriNet and MGN re-ID models are trained 10 times, to determine the resulting re-ID performance on the test set of each training cycle and to report mean and standard deviations on the results. These results are presented in Table 1. For the Tracklet-based Querying Procedure, the per-gallery-image similarity scores are accumulated by summing only the Top-X scores of that gallery image, as described in Section 3. For the experimental validation, we choose X=*all* (all query tracklet images are summed), and X=10, 15, and 25 (also indicated in Table 1).

Evidently, the proposed Tracklet-based Querying Procedure is significantly beneficial for re-ID, starting with a performance gain of 5.0% Rank-1 (73.9%), when summing all query-tracklet images for MGN. Further improvement is observed with the Top-15 (74.3%) and Top-25 (74.5%) matching methods. Clearly, using a multi-image query is preferred over a single-image query. Moreover, when using multiple but not all images, the matching method can filter out the least representative images of the query vessel. This explains why the *Top-X* approach shows higher performance than the *All* approach. Intuitively, when a query-vessel is occluded in some of its tracklet images, those samples can be removed with the *Top-X* approach.

Finally, when comparing TriNet with MGN, it is clear that the gain in performance is substantial and independent of the applied re-ID model. Ultimately, the improvements are 8.1% and 5.6% Rank-1, for TriNet and MGN, respectively. The gain is somewhat lower for MGN and this can be explained by the baseline performance of MGN. As this is significantly higher, +13.0% Rank-1 compared to TriNet, it is also harder to improve this performance. Furthermore, although the difference is well within the observed spread between training cycles, it is remarkable that Top-15 performs better than Top-25 for TriNet, while for MGN this is vice versa.

5 Conclusions

This paper has presented a novel tracklet-based approach for vessel re-ID. In contrast to common re-ID datasets, our Vessel-reID dataset provides the full tracklet for each query-vessel. This helps to assess the impact of the tracklet-based procedure on re-ID performance. In comparison with the querying procedure of existing re-ID literature, the proposed tracklet-based alternative exploits most images from the query-vessel tracklet instead of only a single image. Moreover, since the tracklet-based alternative solely uses the most representative images of the tracklet, it is less susceptible to temporal occlusions. Experimental evaluation reveals that the Tracklet-based Querying Procedure significantly outperforms the Common Querying Procedure, elevating performance from 68.9% Rank-1 to 74.5% Rank-1 on our Vessel-reID dataset. For the intended vessel-speed enforcement application, re-ID is applied to measure the vessel speeds. For this, the final re-ID result is attractive in two ways. First, the obtained Rank-1 score presents a feasible performance for a practical system, since the performance level supports direct law enforcement. Second, the result is still conservative, because we have tested on new unseen data during other weather conditions, instead of using data with mixed conditions.

References

[1] H. Chen, B. Lagadec, and F. Bremond. Partition and reunion: A two-branch neural network for vehicle re-identification. In *CVPR Workshops*, 2019.

[2] H. G. J. Groot., M. H. Zwemer., R. G. J. Wijnhoven., Y. Bondarev., and P. H. N. de With. Vessel-speed enforcement system by multi-camera detection and re-identification. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*, pp. 268–277. INSTICC, SciTePress, 2020. doi: 10.5220/0008911202680277

[3] A. Hermans*, L. Beyer*, and B. Leibe. In Defense of the Triplet Loss for Person Re-Identification. *arXiv preprint arXiv:1703.07737*, 2017.

[4] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, and R. J. Radke. A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3):523–536, March 2019. doi: 10.1109/TPAMI.2018.2807450

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *ECCV*, pp. 21–37. Springer, 2016.

[6] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou. Learning discriminative features with multiple granularities for person re-identification. In *2018 ACM Multimedia Conference on Multimedia Conference*, pp. 274–282. ACM, 2018.

[7] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1116–1124, 2015.

[8] Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3754–3762, 2017.

[9] M. H. Zwemer, R. G. Wijnhoven, and P. H. N. de With. Ship detection in harbour surveillance based on large-scale data and cnns. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*. INSTICC, INSTICC, Funchal, Madeira, Portugal, 2018. doi: 10.5220/0006541501530160