Vessel-speed Enforcement System by Multi-camera Detection and Re-identification

H. G. J. Groot[1,a], M. H. Zwemer[1,2,a], R. G. J. Wijnhoven[2], Y. Bondarev[1] and P. H. N. de With[1]
[1] *Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
[2] *ViNotion B.V., Eindhoven, The Netherlands*

Keywords: Maritime Traffic Management, Vessel Detection, Vessel Re-identification, Speed Enforcement Application.

Abstract: In crowded waterways, maritime traffic is bound to speed regulations for safety reasons. Although several speed measurement techniques exist for road traffic, no known systems are available for maritime traffic. In this paper, we introduce a novel vessel speed enforcement system, based on visual detection and re-identification between two cameras along a waterway. We introduce a newly captured Vessel-reID dataset, containing 2,474 unique vessels. Our vessel detector is based on the Single Shot Multibox Detector and localizes vessels in each camera individually. Our re-identification algorithm, based on the TriNet model, matches vessels between the cameras. In general, vessels are detected over a large range of their in-view trajectory (over 92% and 95%, for Camera 1 and 2, respectively), which makes the re-identification experiments reliable. For re-identification, application specific techniques, i.e. trajectory matching and time filtering, improve our baseline re-identification model (49.5% mAP) with over 20% mAP. In the final evaluation, we show that 77% (Rank-1 score) of the vessels are correctly re-identified in the other camera. This final result presents a feasible score for our novel vessel re-identification application. Moreover, our result could be further improved, as we have tested on new unseen data during other weather conditions.

1 INTRODUCTION

Maritime traffic is bound to speed regulations for safety, especially in crowded waterways where commercial and tourist vessels are mixed. Speed regulation is important, since waves generated by speeding vessels can cause safety risks for other waterway users. In general, the higher the speed of a vessel, the more water it displaces (though exceptions exist for planing speedboats). In addition to damage to the shore-line which causes increased maintenance costs, the water displacement may cause dangerous currents for other waterway users, like swimmers or small boats. Moreover, the motor sound of speeding vessels generates noise disturbance.

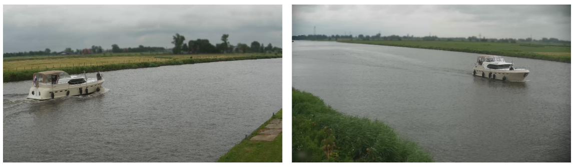

Figure 1: Example images of the same vessel appearing in Camera 1 (left) and 2 (right).

Continuous measurement of vessel speed enables active monitoring and law enforcement. To this end, we introduce a novel system for the application for vessel speed enforcement on waterways. For road vehicles, speed enforcement is a well-known subject and is typically implemented for single-location speed measurements using magnetic loops and radar systems. However, magnetic loops cannot be used on waterways and radar systems are expensive and have difficulties with irregularly manoeuvring vessels. Measuring the average speed of road users more robustly over a longer trajectory is typically implemented using re-identification of vehicles between two camera locations based on automatic license plate recognition. In contrast to vehicles, vessels do not have well-defined licence plates or other common visual registration markers. However, the overall vessel appearance is often unique because most vessels have different vessel type, bow, cabin or different details such as flags or buoys. Therefore, the vessel image theoretically allows for re-identification of vessels between different camera locations. However, the application poses several challenges, such as fluctuating weather conditions and as a consequence of these, the highly dynamic lighting conditions at the constantly

[a] equal contributions

Groot, H., Zwemer, M., Wijnhoven, R., Bondarev, Y. and de With, P.
Vessel-speed Enforcement System by Multi-camera Detection and Re-identification.
DOI: 10.5220/0008911202680277
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 5: VISAPP*, pages 268-277
ISBN: 978-989-758-402-2; ISSN: 2184-4321

moving water surface. To this end, we develop a novel system to measure the speed of vessels over a long trajectory, inherently obtaining accurate speed measurements, using two cameras with visual re-identification from raw vessel images.

In this paper, we propose the application of accurate vessel speed measurements by visual re-identification based on two surveillance cameras, placed several kilometers apart. To our knowledge, this is the first time that such a system is proposed, tested and evaluated. Vessels are detected and tracked within each individual camera view. Then, a re-identification (re-ID) algorithm links vessels from one camera to the most similar historic vessel image in the other camera. When a vessel is recognized and linked in both cameras, its speed is determined using the travel time compared to the distance between the two cameras. Our system is developed using video data collected along a canal in the Netherlands. Figure 1 shows example images of both cameras containing the same vessel when entering and leaving the captured area covered by the cameras. We specifically focus on three aspects: video-based vessel detection and tracking, re-ID of vessels based on their visual appearance in two cameras, and creation of an image dataset used to train the detection- and re-ID algorithm. The main contributions of this paper are as follows. First, we present a combination of state-of-the-art detection, tracking and re-ID algorithm for vessel speed computation. Second, the system setup is discussed of the new application of vessel speed measurement. Third, we experiment with the different detection and re-identification sub-systems and show that adding application understanding to the re-identification problem significantly improves recognition performance.

The remainder of the paper is divided as follows. Section 2 introduces related work for all components of our system. Next, Section 3 describes the proposed system. The dataset used for the experiment is presented in Section 4. In Section 5, the experimental validation of our system is divided in the evaluation of the vessel detection performance and the re-identification performance. The paper ends with conclusions in Section 6.

2 RELATED WORK

The proposed application consists of a complete pipeline for visual detection and re-identification of vessels. Limited work is available for vessel-speed enforcement with surveillance cameras. One particular work presents ARGOS (Bloisi and Iocchi, 2009), a vessel-traffic monitoring system in the city of Venice. This system employs surveillance cameras, with slightly overlapping views, mounted high above the waterway. The authors employ background modeling for detection and tracking of all vessels. Although background modeling achieves good performance, in our work we focus on surveillance cameras with more dynamic backgrounds and our system only utilizes two cameras several kilometers apart. Other work concentrates on unrestricted detection, tracking and re-ID from moving vessels, to recognize other vessels in their surroundings (Qiao et al., 2019; Bibby and Reid, 2005). In the following paragraphs, we discuss related work for detection and re-identification. Besides this, we discuss publicly available datasets.

A. Detection Techniques: In the generic field of visual object detection, Convolutional Neural Networks (CNNs) achieve state-of-the-art performance. Currently, there are two dominant kinds of CNN detection techniques. The first technique splits the problem into two stages: region proposal and refinement. These stages are combined into a single CNN (Girshick, 2015; Ren et al., 2015) and may even perform instance segmentation in the refinement step (He et al., 2017). The other common CNN detection technique is to skip the region proposal step altogether and estimate bounding boxes directly from the input image such as YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017) and Single Shot Multibox Detector (Liu et al., 2016). YOLO uses the topmost feature map to predict bounding boxes directly for each cell in a fixed grid. The SSD detector extends this concept by using multiple default bounding boxes with various aspect ratios at several scaled versions of the topmost feature map. Prior work (Zwemer et al., 2018) shows that the SSD detector is robust against large-scale variations of vessels. We select the SSD detector for our application because of the relatively low computational requirements and high accuracy, proven for the vessel detection problem. Moreover, since the detector can operate at a high frame rate and its related requirements for the visual tracking method are limited (only one object type and mostly a clear view of the object), which enables the use of a computationally efficient tracking algorithm.

B. Re-identification Algorithms: For re-ID, all related work is mainly focused on the person re-identification domain. However, we can experiment and apply the same techniques to our vessel application. Based on road vehicle applications, the performance of state-of-the-art re-ID algorithms has been recently evaluated (Chen et al., 2019), showing that re-ID networks generalize to other domains. In the last decade, visual-based re-ID has become more mature due to the general developments in CNNs (Karanam et al., 2019) and can be divided in two main techniques: pairwise verification and metric embedding. Pairwise verification networks usually apply the contrastive loss during training, while for metric embedding the triplet loss is popular (Hermans* et al., 2017). Pairwise verification networks are commonly Siamese networks (Ahmed et al., 2015) and are trained on image pairs. These networks then learn to differentiate different persons by increasing the feature distance, while decreasing the distance for two images of the same person. Alternatively, metric embedding networks consider image triplets, considering two images from the same person and one from a different person. As a result, metric embedding networks utilize inter-class variance in the resulting embedding space more efficiently. The TriNet model (Hermans* et al., 2017) builds upon a customized ResNet-50 architecture and is suitable for embedded applications because of its low computational complexity. Another commonly exploited technique is to force the network to separately focus on e.g. head, body and legs of persons (Wang et al., 2018). However, vessels are much more diverse in appearance. Cargo vessels are for instance very long, while sailing yachts often have height exceeding their length. In this paper, we propose to employ a state-of-the-art TriNet model (Hermans* et al., 2017) with its attractively low complexity.

C. Datasets: For our specific vessel application, both the CNN detector and re-ID network require a vessel image dataset for training. Regarding generic visual object detection, popular datasets are MS-COCO (Lin et al., 2014), ImageNet (Russakovsky et al., 2015) and PASCAL-VOC (Everingham et al., 2012). Although these datasets do contain vessel images, they are only taken from very different camera viewpoints not matching with our surveillance scenario. The dataset proposed by (Zwemer et al., 2018) contains vessels in surveillance scenarios with multiple camera viewpoints and can be used for training a vessel detector. As this dataset does not contain vessel trajectories and identifications of the vessels, it cannot be used for training the re-ID network. Regarding re-identification, the most commonly used datasets focus on persons, such as DukeMTMC-reID (Zheng et al., 2017) and Market-1501 (Zheng et al., 2015), so that they cannot be applied for vessel re-ID. Consequently, we introduce a novel large vessel dataset containing trajectories and identifications, which is used to train our vessel re-ID CNN.

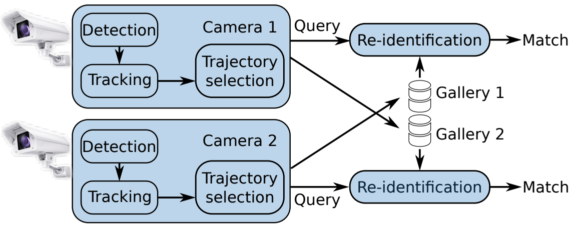

Figure 2: System overview.

3 SYSTEM OVERVIEW

The proposed system is divided in two components (see Figure 2): vessel detection and vessel re-identification to link vessels between the two cameras. The first component performs real-time detection and tracking of vessels for each camera individually. For the second component, a set of vessel images is extracted and stored for re-identification, for each detected vessel. Each vessel image is used both as a query image and as a reference image, where the reference images are stored in the so-called gallery for future queries from the other camera. The query image is thus needed when matching it with vessels that have previously appeared in the other camera. Since our application focuses on speed enforcement, gallery images of vessels are accumulated until the application constraints invalidate those vessel images that do not relate to a speed violation (travel time too long). For every detected vessel, the re-identification query results in the best matching object from the gallery set of the other camera. After finding a match for a vessel query image, the travel time of that vessel is determined by comparing the times of appearance within both cameras. Since the distance between the two cameras is known and fixed, the travel time leads directly to the average vessel speed. In the following subsections, the implementations of both components are discussed in more detail.

3.1 Detection, Tracking and Trajectory Selection

Vessel detection is performed using the Single Shot Multibox Detector (SSD) (Liu et al., 2016). This detector consists of a base network and a detection head. The input of the detector is an image of 512×512 pixels from which the base network computes features. The detection head consists of several convolutional layers which create downscaled versions of these features. The detection head predicts an object confidence and bounding box on each of these feature layers. The box is predicted using offsets to a

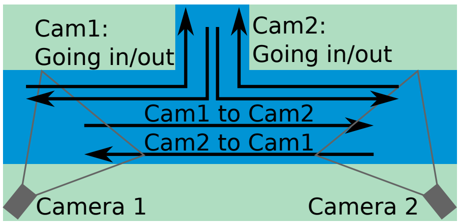

Figure 3: Schematic overview of the two cameras positioned along the main waterway (top-view).

set of fixed prior boxes. The output of the detection head is created by combining the class confidences, location offsets and prior boxes of all feature layers. These combinations are then filtered by thresholding the class confidences followed by non-maximum suppression, based on the Jaccard overlap (IOU) of the boxes, leading to the final set of detections.

To match the input resolution of the SSD detector network, a square region in the camera image is selected, cropped and scaled to 512 × 512 pixels. Vessel detection by the SSD network is performed at 5 frames per second. To create trajectories over time, visual tracking is performed. Tracking is carried out using feature point tracking (Shi and Tomasi, 1994) with optical flow. Within each detection box, a uniform grid of feature points is created and new positions of each point are estimated in the next frame based on their optical flow. The median displacement of the feature points determines the new position estimate of the vessel box. This process is repeated for every consecutive video frame.

The result after detection and tracking for each vessel is a set of bounding-box coordinates over time, representing the vessel trajectory. These trajectories are temporally sampled (every second) and the resulting selected vessel images are then used for re-ID between the two cameras.

3.2 Re-identification (re-ID)

The re-ID system receives vessel images of a detected vessel from one camera and is responsible for finding the corresponding vessel in the images from the other camera. The re-identification task is performed using the TriNet model (Hermans* et al., 2017). Each vessel image is converted into a low-dimensional feature representation. To this end, each vessel image is first scaled to a resolution of 288 × 144 pixels and then randomly cropped to 256 × 128 pixels to match the input of the ResNet-50 base network of the TriNet model. The last layer of the ResNet-50 model is replaced by two fully connected layers of 1024 and 128 units. Each of the vessel images in a vessel trajectory is converted to this 128-dimensional feature representation and stored in the gallery of the respective camera. The system is trained end-to-end and uses the triplet loss function to perform deep metric learning. Each training sample is a combination of two feature vectors of a similar vessel and a single feature vector of a different vessel. After training the re-ID model, the model is used to determine the similarity between one known gallery image and a previously unseen vessel image (query). During typical system operation (inference only), matching is implemented by computing the Euclidean distance between the query feature vector and each gallery feature vector. This results in a similarity score for each gallery image. The most similar gallery image defines the matching object.

We now propose to use three application aspects of understanding to improve the performance of our re-identification system. Firstly, we utilize multiple query images per vessel and propose to accumulate the similarity scores for all query images to improve accuracy. Each query image is matched individually and the per-gallery-image similarity scores are combined over the different query images (by summation). The gallery image with the highest summed similarity score defines the matching object. Secondly, based on the minimum and maximum travel time of vessels, a time selection limits the size of the galleries of both cameras. This aspect always increases the accuracy because false matches are reduced from the search (gallery) set. Lastly, vessels from Camera 1 only need to be matched with vessels from Camera 2, and vice versa. Hence, we ensure that the gallery only contains images from a single camera. The effectiveness of each aspect will be evaluated later in the experiments section.

4 VESSEL DATASET

We introduce our novel Vessel-reID dataset containing vessels in two cameras, spaced about 6 kilometers apart along a canal in the Netherlands. The set was recorded over 4 days and contains 2,474 vessels, each sampled with multiple images, and linked between the two cameras. Figure 3 shows a schematic overview of the positions of the cameras. Note that the connecting canal is not covered by any camera (at the top in the figure). This causes some vessels to appear in one camera, but not in the other camera. However, this is a side canal and the majority of the vessels use the main canal and pass both cameras.

The Vessel-reID dataset is created semi-automatically using an existing ship detector and tracker. For each day of video, we first process each

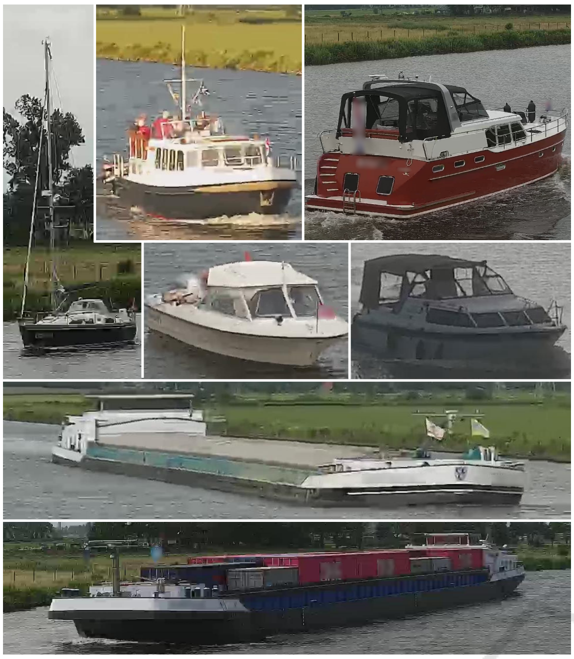

Figure 4: Visual examples of several types of vessels in our Vessel-reID dataset.

camera stream individually to extract per-camera trajectories of vessels. The video is processed with an existing vessel detector (Zwemer et al., 2018) to identify vessels and track them over time through the camera view, as is proposed in the final system (see Figure 2). The obtained vessel trajectories are temporally sub-sampled (every second). Manual validation of detection and tracking are performed to enforce accurate localization and full coverage of all vessels. To increase the annotation accuracy of our semi-automatic creation of the initial dataset, the vessel detector is retrained with the annotations of the first day of video for both cameras prior to applying it on the data of the other days. After creating trajectories of the two individual cameras, the vessels in the two cameras are manually linked together to define the final ground truth.

In total, we have annotated 4 days of video from 6.00 AM until 9.00 PM from both cameras. This resulted in 2,474 trajectories of vessels moving through both cameras (1,237 per camera) of which we have one representative sample per second. On average, there are 44 samples of a vessel in Camera 1 and 66 samples in Camera 2. The different numbers can be explained by the different viewing angles of both cameras, resulting in differently covered canal lengths (see Figure 1). Figure 5 shows example trajectories of vessels in Camera 1 and 2. Table 1 gives an overview of the vessel trajectories per day. Note that about 38% of the vessels are moving into the side canal, mooring at a local harbour or appear in one camera only.

Table 1: Number of vessels going in each direction per day.

Direction	**Day 1**	**2**	**3**	**4 (Test)**
Cam1 to Cam2	178	104	167	181
Cam2 to Cam1	155	134	176	142
Cam1 going in	34	23	45	31
Cam2 going in	31	30	50	33
Cam1 going out	25	28	41	33
Cam2 going out	24	27	40	44
Total Cam1	392	289	429	387
Total Cam2	388	295	433	400

5 EXPERIMENTS

Our experiments focus on the evaluation of the visual detection and vessel re-identification components. The first two experiments focus on the vessel detector, while the remainder of the evaluations measure the performance of our re-identification application. For re-identification, we first validate our dataset and then incrementally apply and evaluate our three specific application aspects of understanding. We conclude the experiments with a final application-oriented evaluation.

5.1 Detection Performance

For measuring detection performance, we compare four SSD detectors with the same CNN network architecture but trained with different datasets. The first detector is the original SSD512 detector (Liu et al., 2016) trained on the PASCAL-VOC 2007 and 2012 sets (Everingham et al., 2012). The second detector is proposed by Zwemer *et al.* and is trained on their harbour surveillance dataset (Zwemer et al., 2018). The third detector is trained on a combination of the harbour surveillance dataset and our novel Vessel-reID dataset. The fourth detector is trained on our novel dataset only.

Training is performed using Stochastic Gradient Decent (SGD) for 120k iterations, starting at a base learning rate of 0.001, decaying with a factor of 10 at 80k and 100k iterations. The SSD model weights are initialized with the default pre-trained VGG network (Liu et al., 2016). Weight decay is set to 0.0005, gamma is 0.1 and we use a batch size of 32. Days 1, 2 and 3 are used for training. Day 4 is used for testing.

All four detectors are evaluated both on the harbour surveillance set from (Zwemer et al., 2018) and on the novel Vessel-reID dataset. The comparison is based on the recall-precision curve and the Area under the recall-precision Curve (AuC) metric. Figure 6 shows the results on our Vessel-reID set and Fig-

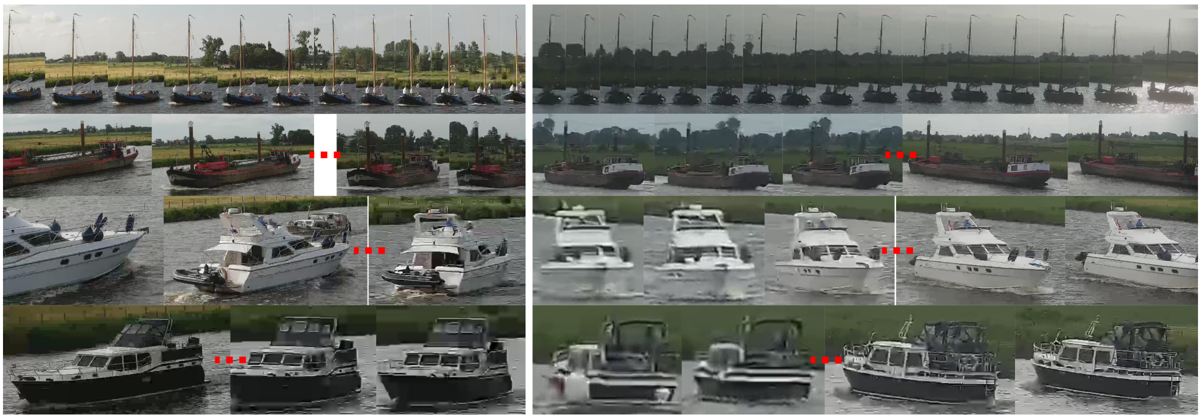

Figure 5: Visual example of several vessel trajectories in Camera 1 (left) and Camera 2 (right), where each row shows a unique vessel in our dataset. Some images are skipped for visualization (denoted by dotted red line).

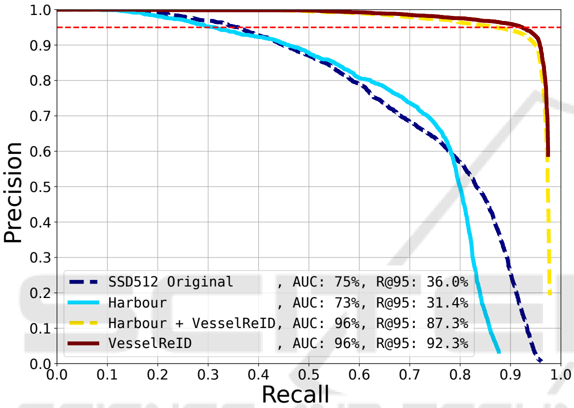

Figure 6: Vessel detection performance on the Vessel-reID set.

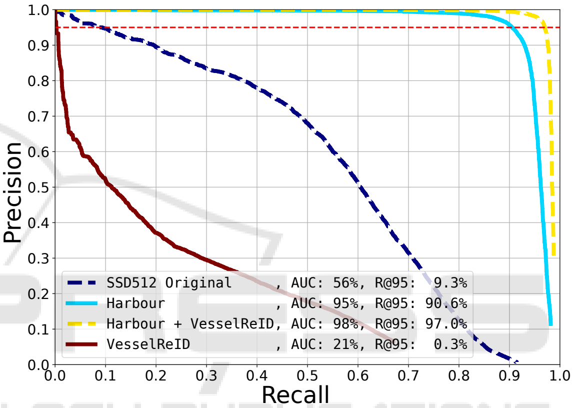

Figure 7: Vessel detection performance on the harbour surveillance set (Zwemer et al., 2018).

ure 7 depicts the results on the harbour surveillance set (Zwemer et al., 2018). The default SSD detector (dotted blue line) performs poorly on both datasets. Interestingly, the harbour (solid cyan line) and Vessel-reID (solid red line) detectors perform well on their own set, but have poor detection performance on the other set. This indicates that the sets are complementary, which is motivated by the information that the harbour set contains mostly large inland and sea-going cargo vessels, while the Vessel-reID contains mostly small pleasure craft. Training with the combined sets (dotted yellow) leads to the highest detection performance on both datasets. Therefore, we select the detector trained with the combined harbour and Vessel-reID sets for our next experiments.

5.2 Detection for Re-identification

In our application of re-ID, it is important that each vessel is detected at least once in both cameras. Our testing dataset contains several images of the trajectory of each vessel, allowing us to measure the amount of detections per vessel over its trajectory. In this experiment, for each vessel, we measure the detection ratio, i.e. the amount of correct detections with respect to the amount of ground-truth annotations. This detection ratio is reported over all vessel trajectories per camera location in our dataset using a histogram representation. For the detector, we have selected the threshold at 95% precision to limit the amount of false detections, while still having a high recall of 87.3%.

Figure 8 shows the results per camera. In general, many vessels are detected over more than 90% of their trajectory and almost all vessels have a trajectory coverage of more than 60%. Unfortunately, there are few vessels (10 in camera 1 and 7 in camera 2) that are not detected at all. Visual inspection of these missed vessels shows that they are typically very small boats (such as ‘dinghies’ behind larger boats) or boats moving close together (see Figure 9).

This experiment shows that our detector has a high

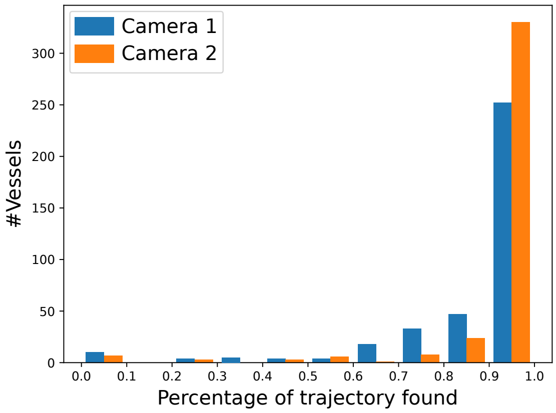

Figure 8: Percentage of trajectory detected.

detection accuracy and delivers a dense set of image samples for our re-identification sub-system. For re-identification, it is important to detect each vessel at least once. Although we have a few vessels that are never detected, they are mostly attached to a larger boat ('dinghies'), which does not influence the re-identification performance, since the main vessel is detected. Missed vessels that are close to each-other, require the combined vessel to be detected for correct re-identification in both cameras.

5.3 Re-identification Hyper-parameters

In a first re-identification experiment, we evaluate the effect of the main hyper-parameters for training. As motivated by (Hermans* et al., 2017; Groot et al., 2019), we evaluate the learning rate, the number of iterations and the start iteration of the learning rate decay. A total of eight different parameter combinations is evaluated five times and averages and standard deviations are reported. This experiment is carried out on a subset of the total dataset (only our test set, Day 4). Evaluation is performed analogous to the most popular person re-ID datasets (Zheng et al., 2015; Zheng et al., 2017) on 10% of our test set, while the remaining 90% of the vessels are used for training. For each vessel trajectory in the test set, a single random sample is moved to our query set and the remaining samples are added to the gallery.

The results are shown in Table 2. The bold settings are the values as used in (Hermans* et al., 2017) and act as reference. We can conclude that the effect of the different parameter combinations is limited. The learning rate of 0.001 does not lead to stable training and should be avoided. The effect of the number of training iterations is negligible. Therefore, we select the following parameters for the remainder of our experiments: learning rate 3×10^{-4}, 25k iterations and a learning rate decay of 0.001 after 15k iterations.

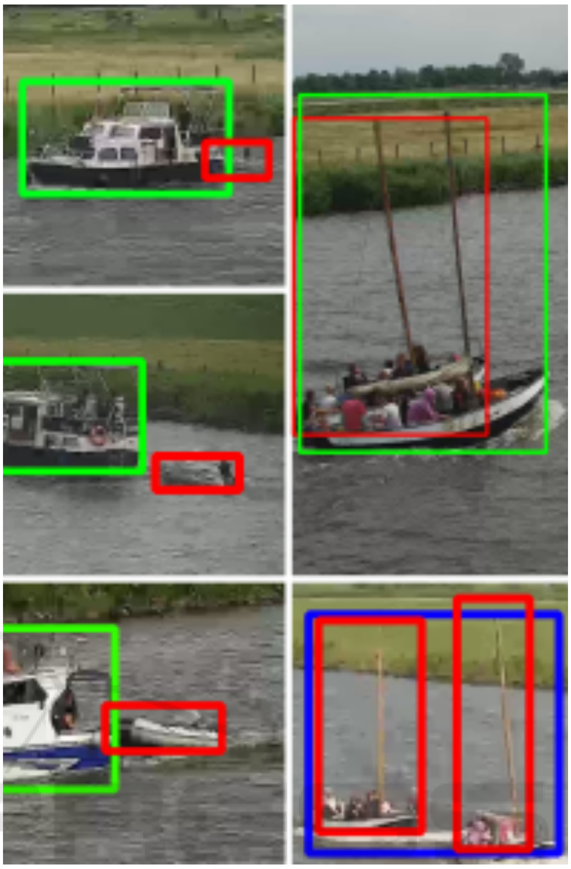

Figure 9: Correct detections (green), Missed (red) and false (blue) detections of our vessel detector.

Table 2: Effect of the learning rate (Lr), number of iterations (Iter) and start of the learning rate decay (Dec) on re-identification accuracy. Reported numbers are averages (stddev) over 5 runs. Reference and selected settings in bold.

Lr	Iter	Dec	Accuracy [%]	
			mAP	**Rank-1**
$1 \cdot 10^{-4}$	25k	15k	71.4 (±2.8)	73.3 (±2.8)
	35k	20k	72.7 (±2.6)	74.3 (±3.2)
$\mathbf{3 \cdot 10^{-4}}$	**25k**	**15k**	**69.2** (±**1.0**)	**72.7** (±**1.5**)
	35k	20k	69.5 (±2.7)	72.4 (±4.8)
$5 \cdot 10^{-4}$	25k	15k	69.6 (±2.7)	73.0 (±2.0)
	35k	20k	69.8 (±3.7)	76.4 (±4.0)
$1 \cdot 10^{-3}$	25k	15k	57.0 (±22.3)	59.4 (±22.6)
	35k	20k	66.1 (±2.0)	71.2 (±2.4)

5.4 Re-identification Training Data

We will now investigate the effect of the amount of training data on the re-identification performance. To this end, we first train using a single day of training

Table 3: Re-identification performance on our Vessel-reID dataset, when incrementally applying application aspects (see Section 3.2). Results are averages (stddev) over 5 runs.

	Scores on validation set [%]	
Description	**mAP**	**Rank-1**
VesselReId (Day 3)	36.2 (±0.9)	42.9 (±1.5)
VesselReId (full)	49.5 (±1.0)	55.6 (±1.7)
+ Tracklet		
Sum of ALL	54.8 (±0.6)	61.5 (±0.8)
Sum of Top-10	56.2 (±0.6)	63.6 (±0.9)
Sum of Top-15	56.3 (±0.6)	63.7 (±1.0)
+ Time filtering	59.6 (±0.8)	65.7 (±1.0)
+ Only cross-camera	71.0 (±0.9)	77.3 (±1.3)

data (Day 3), resulting in a 50/50 train/test distribution which is common in re-ID. Secondly, we train utilizing the full three days of our Vessel-reID set. The test set is as defined in our Vessel-reID dataset (Day 4), where the query and gallery sets are constructed in a similar way as in our previous experiment (one random sample per vessel trajectory in the query set). The performance is reported by the mean average precision (mAP) and Rank-1 score, including standard deviation, measured over 5 training runs.

The results are presented in Table 3. It can be observed that when training on all data (Vessel-reID (full)), the performance is significantly higher than when training on only a single day of data (Vessel-reID (Day 3)). The results on our dataset show a large performance gap when comparing to the performance of the TriNet model on public datasets. The TriNet model performs better on popular public person re-ID datasets, DukeMTMC (rank-1 75.4%, see (Groot et al., 2019)) and Market-1501 (mAP 69.1%, Rank-1 84.9%, see (Hermans* et al., 2017; Groot et al., 2019)). We expect that this originates from the fact that our 50/50 train/test distribution results in a training set containing half the amount of unique objects with respect to these datasets. Furthermore, even when training with more data (3 days in our train set), the re-ID performance is still relatively low. This is explained by the fact that the test set was recorded during a completely different day with changed conditions, while the validation set of the public person re-ID datasets are recorded under the same conditions as the training. In the remainder of the paper, we train using all training data (row Vessel-reID (full)).

5.5 Re-identification Tracklet-based Querying

In our tracklet-based query approach, we no longer consider a single image as a query. Instead, we apply inference on every image appearing in a single trajectory of a vessel. By applying inference on the whole tracklet, we can effectively combine the results of all images in the tracklet, instead of just one (randomly selected) image, which is common in re-ID. However, this approach has an impact on the final evaluation. If we would still adopt the common query/gallery division approach, we have to move each query from the gallery into the query set. This would now lead to an empty gallery set, because we use all images of a tracklet as a single query, instead of just one image. Therefore, we propose to keep all images in the gallery and consider one tracklet at a time as the query. Furthermore, we have carefully validated that this changed gallery alone has a negligible impact on performance.

Matching of a tracklet query is performed by computing the Euclidean distance of each individual image in a tracklet to those in the gallery. Consequently, for each gallery image, the distance to all tracklet images is known, which we then combine in three different ways. First, in our most basic version, we take the sum of all these distances. Once this is done for all gallery images, we rank the gallery accordingly to obtain *all* most likely matches for the tracklet query. Second, we only take the *top-10* most likely matches for the tracklet query. Third, we only take the *top-15* matches. Ranking is done similarly for all methods. The results are shown in Table 3 and applied on top of our baseline model. Evidently, tracklet-based matching is significantly beneficial for re-ID, achieving a performance gain of 5.3% mAP for matching all tracklet images (54.8%). Further improvement can be seen for the top-10 (56.2%) and top-15 (56.3%) matching methods. Overall, using multiple images as query request, results in a significant performance gain as compared to a single-image query. In our application of vessel re-ID, the Top-15 matching method works best and is therefore selected.

5.6 Re-identification Time Filtering

In this experiment, we evaluate our time-filtering extension, which is applied on top of our tracklet-based querying extension. For this extension, we were inspired by our application, but it can be directly applied to other object classes such as persons or vehicles.

Inspection of our Vessel-reID dataset shows that the vessels take an average travel time of around 32 minutes between cameras. Furthermore, we have found that none of the ships require longer than 3 hours to pass. Hence, we exclude any possible matches where the implied transition time is longer than 3 hours. This filtering is applied on top of the top-15 variant of our tracklet-based querying method.

The result is presented in Table 3 and shows that including timing information significantly improves the mAP score by 3.3% and the Rank-1 by 2.0%. Our result indicates that incorporating timing information in re-ID yields higher performance, and can be directly applied to other re-ID applications.

5.7 Application-based Evaluation

In our previous experiments, we have used the evaluation methods commonly adopted in re-ID. In these methods, each query image is matched to a combined gallery, consisting of images from both cameras. Given our application, we specifically require matching a query from one camera to the gallery of another camera, and vice versa. In this experiment, we change this generic method of evaluation such that it only considers images of the other camera in the gallery (thus only cross-camera). Evaluating with these constraints will provide insight in the final performance of our specific application in practice.

We evaluate the performance of re-ID for our application with the top-15 tracklet-based querying approach and additional time filtering. The bottom row of Table 3 shows the result of combining these approaches with our cross-camera evaluation method. For our application, a re-ID performance score of 71.0% mAP and 77.3% Rank-1 is achieved. Hence, 77% of the vessels are correctly re-identified in the other camera. This final result is attractive in two ways. First, the obtained Rank-1 score is comparable with systems for person re-ID, meaning that we have obtained the same level of quality in re-ID in an otherwise novel application. Second, our result is conservative, because we have tested on new unseen data during other conditions, instead of mixing the conditions of training and testing (typical for person re-ID).

6 CONCLUSIONS

We have proposed the novel application of vessel speed measurement using visual re-identification. To our knowledge, this is the first time that an automated trajectory speed measurement for vessels has been proposed. In addition, it is the first time that re-identification has been applied to vessels, as re-identification literature typically considers the person class. The proposed system uses a setup with two cameras, spaced several kilometers apart. Each camera system applies detection and tracking to localize all vessels in its own camera view. Then, multiple images are collected for each vessel and stored in a database for visual vessel matching between the two cameras. The proposed re-identification system compares newly detected vessels in one camera as query to the gallery set of all vessels detected in the other camera, and vice versa. For our implementation, we use the Single Shot Multibox Detector (SSD) with a VGG base network and train it specifically to localize vessels. Re-identification is implemented using the TriNet model with a ResNet-50 base network, trained with the Triplet loss function on our Vessel-reID dataset.

For this purpose, we have introduced a novel Vessel-reID dataset. This extensive vessel dataset was constructed from two camera positions mounted 6 kilometers apart at a canal in the Netherlands. During four days, a total of 2,474 different vessels were captured. Each vessel is represented by multiple images captured along its trajectory in the camera view. The set contains a large variation in vessel appearance, where most vessels are pleasure crafts.

The performance of the detection and re-identification systems are experimentally validated. We have compared the detection performance of four different SSD detectors and conclude that training on a combination of a harbour dataset and our new Vessel-reID dataset results in the best performance on both sets. This can be explained by the harbour set containing mostly inland and sea-going vessels, whereas the Vessel-reID set is complementary with mostly pleasure crafts. Of the total set of 787 vessels in the test set, only 14 vessels are missed and therefore not considered for re-identification. In general, vessels are detected over a large range of their in-view trajectory (over 92% and 95%, for Camera 1 and 2, respectively), making the re-ID experiment reliable.

For re-identification, after tuning the system hyper-parameters, we have added application-specific techniques for understanding vessel re-identification. First, combining re-identification scores over multiple images of the same vessel increases the performance with 6.8% mAP (8.0% Rank-1). An additional time-filtering stage adds another 3.3% mAP (2.0% Rank-1), leading to a combined performance of 59.6% mAP and a Rank-1 score of 65.7%, without re-ranking. Finally, when we evaluate the re-identification performance specifically for our application, we show that 77% (Rank-1) of the vessels are correctly re-identified in the other camera. This final result is attractive in two ways. First, the obtained Rank-1 score presents a feasible score for our novel application of vessel re-identification. Second, our result is still conservative, because we have tested on new unseen data during other weather conditions, instead of mixing data of all captured conditions.

The proposed system enables the automatic speed measurement of vessels over a large-distance trajectory. Despite the obtained high performance of the system, the current process of law enforcement still requires the intervention of a human operator. However, the performance of our automated system is approaching the level of directly supporting law enforcement.

ACKNOWLEDGEMENTS

We want to thank the Dutch Ministry of Infrastructure and Water Management for supporting the experiments and providing access to the test location. Parts of this work were funded by the European ITEA PS-Crimson project on surveillance.

REFERENCES

Ahmed, E., Jones, M., and Marks, T. K. (2015). An improved deep learning architecture for person re-identification. In *CVPR*, pages 3908–3916.

Bibby, C. and Reid, I. (2005). Visual tracking at sea. In *Proceedings of the 2005 IEEE Int. Conference on Robotics and Automation*, pages 1841–1846. IEEE.

Bloisi, D. and Iocchi, L. (2009). Argos—a video surveillance system for boat traffic monitoring in venice. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(07):1477–1502.

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

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2012). The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results.

Girshick, R. (2015). Fast R-CNN. In *Proceedings of the International Conference on Computer Vision*.

Groot, H., Bondarau, E., et al. (2019). Improving person re-identification performance by customized dataset and person detection. In *IS&T International Symposium on Electronic Imaging 2019, Image Processing: Algorithms and Systems XVII*.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In *Proceedings of the International Conference on Computer Vision*.

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

Karanam, S., Gou, M., Wu, Z., Rates-Borras, A., Camps, O., and Radke, R. J. (2019). 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.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer.

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

Qiao, D., Liu, G., Zhang, J., Zhang, Q., Wu, G., and Dong, F. (2019). M3c: Multimodel-and-multicue-based tracking by detection of surrounding vessels in maritime environment for usv. *Electronics*, 8(7):723.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE CVPR*, pages 779–788.

Redmon, J. and Farhadi, A. (2017). Yolo9000: Better, faster, stronger. In *IEEE CVPR*, pages 6517–6525.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

Shi, J. and Tomasi, C. (1994). Good features to track. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600.

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

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

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

Zwemer, M. H., Wijnhoven, R. G., and de With, P. H. N. (2018). 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,*, Funchal, Madeira, Portugal. INSTICC, INSTICC.