Detecting Road Users at Intersections Through Changing Weather Using RGB-Thermal Videos

Bahnsen, Chris; Moeslund, Thomas B.

Published in:
Advances in Visual Computing

DOI (link to publication from Publisher):
10.1007/978-3-319-27857-5_66

Publication date:
2015

Document Version
Accepted author manuscript, peer reviewed version

Detecting Road Users at Intersections Through Changing Weather Using RGB-Thermal Video

Chris Bahnsen and Thomas B. Moeslund

Visual Analysis of People Laboratory, Aalborg University, Aalborg, Denmark.

Abstract. This paper compares the performance of a watch-dog system that detects road user actions in urban intersections to a KLT-based tracking system used in traffic surveillance. The two approaches are evaluated on 16 hours of video data captured by RGB and thermal cameras under challenging light and weather conditions. On this dataset, the detection performance of right turning vehicles, left turning vehicles, and straight going cyclists are evaluated. Results from both systems show good performance when detecting turning vehicles with a precision of 0.90 and above depending on environmental conditions. The detection performance of cyclists shows that further work on both systems is needed in order to obtain acceptable recall rates.

1 Introduction

Road safety is a subject of high interest amongst governments and the research community. In the European Union, for instance, it is the goal of the European Commission to halve the number of road deaths by 2020 [1]. One of the ways to increase the passive safety of a road is to improve the layout of the road based on historical data of traffic events such as police and hospital records. However, not all accidents are reported and when conflict data is sparse, it might be difficult to assess the safety of a particular road. Video data, on the other hand, allows the generation of detailed information about the road users such as trajectories, speed profiles, and road user types.

However, when conflict data is sparse, one must look through thousands of hours of video to extract events of interest. Another approach is *surrogate safety analysis* which studies potential conflicts as a surrogate for real conflicts. The foundation of surrogate safety analysis is the existence of a continuous relationship between the severity of the interactions and the volume which forms a so-called conflict pyramid [2]. The fatal injuries reside on the top of the pyramid and resembles a low volume of the traffic data while normal traffic with no conflicts make up the majority of the traffic and resides in the lower part of the pyramid. Surrogate safety analysis builds on the claim that the analysis of near-conflicts gives a surrogate measure of the number of fatal interactions between road users [3].

The long-time goal of this work is to obtain a surrogate measure of the safety of bicyclists, pedestrians, and other vulnerable road users at urban intersections.

In order to do this, one must reliably detect and track all road users in intersections by automatic analysis of video data. This paper presents a thorough evaluation of the block-based road user action detection technique presented in [4] on 16 hours of thermal and RGB video in three urban intersections. The data set includes a variety of sub-optimal conditions for visual algorithms, including rain, hard shadows, reflections, low lighting, and occlusion. We compare the mentioned approach to a feature-based Kanade-Lucas-Tomasi (KLT) tracker for traffic analysis presented by Saunier et al. [5] which is made available through the open-source *trafficintelligence* project [6]. To the best of our knowledge, this is the first cross-evaluation of tracking algorithms for infrastructure-side monitoring on real-world, non-optimal thermal-visible video data.

The following section of this paper contains an overview of related work in infrastructure-side traffic surveillance. Section 3 outlines the main methods used for the block-based road user action detection system used in [4] and Section 4 explains the KLT-based feature tracker used for comparison. In Section 5, the thermal-visible dataset is described, including context and weather information of the data. Section 6 contains the experimental results of using the mentioned algorithms for cyclist and vehicle detection and Section 7 concludes the work.

2 Related Work

Traditional computer-vision based methods on traffic surveillance concerns the monitoring of motorized vehicles at highways [7]. In highways, the detection and tracking of vehicles is easier because they are usually well separated, run in separate lanes, and follow certain routes. A comprehensive survey of traffic surveillance in highway applications is found in [8].

In the past decade, researchers have explored the more complex task of monitoring vehicles at urban areas which includes monitoring of intersections [9], [5] and pedestrians and two-wheelers such as mopeds and cyclists [10], [11], [12]. Monitoring urban traffic is challenging due to the density of the traffic, variable types of road users, and lower camera orientations which aggravates occlusion. Due to the vast amount of challenges, the field of traffic analysis in computer vision is very diverse and includes a broad range of approaches. In their extensive review of urban traffic analysis with computer vision, Buch et al. divides the field into two main approaches; top-down and bottom-up surveillance systems which eventually are combined with a tracking system [13].

The foundation of *top-down surveillance* is the segmentation of the foreground which is accomplished by using a variety of classic techniques, including frame differencing, background averaging, Kalman filtering, and the Gaussian mixture model (GMM). Foreground segmentation is followed by grouping and vehicle classification which includes region- and contour-based features, and advanced machine learning. Examples of top-down approaches are found in [14], [9], and [15] where the authors use a background model to detect vehicles and [16] which is based on frame differencing.

In *bottom-up surveillance*, the foreground segmentation is replaced by patch detectors and classifiers. Examples in traffic surveillance include Hessian corners [5], SIFT [17], and boosting [18].

Tracking is used to connect observations of road users in consecutive frames into spatio-temporal trajectories. The classic Kalman filter is used in a variety of applications, including [10]. Trackers based on the Kalman filter assumes a Gaussian process and measurement noise, which is not fulfilled in the general case of tracking urban traffic. The Particle Filter removes these assumptions at the cost of computational simplicity and is used for tracking motorcycles in [19]. Saunier and Sayed use the KLT tracker to track keypoints of vehicles in intersections [5]. Tracked features are grouped over time according to the spatial distance of the tracks. The work of Saunier and Sayed has been used in [20] to predict collision amongst vehicles in intersections and extended in [12] to include the classification of road users, including pedestrians and bicycles.

3 Watch-Dog Detection of Road User Actions

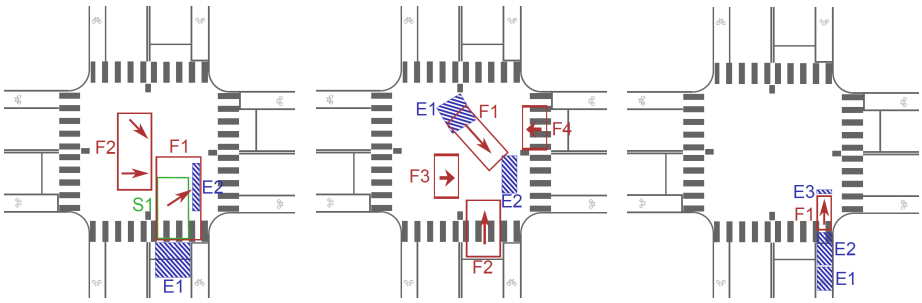
Detection, tracking, and classification of urban traffic in unconstrained scenarios pose a substantial challenge to computer vision algorithms. Existing methods are typically evaluated either at short intervals or under ideal conditions. An automated system which is able to accomplish these tasks in unconstrained scenarios does currently not exist [13]. On the other hand, manual monitoring of vast amounts of video is an expensive and tedious task which indeed does not scale well to analysis of complex transport networks. In the recent work of [4], the authors take steps to close the gap between automated and manual analysis by introducing a semi-automated watch-dog system, whose aim is to reduce the amount of video data for inspection by the traffic analysts. This semi-automated system is specialized for the detection of interactions between Right Turning Vehicles (RTV), Left Turning Vehicles (LTV), and Straight Going Cyclists (SGC) at intersections. The goal of the watch-dog is not to perform perfect tracking of road users but to obtain a reasonable data reduction.

The watch-dog system contains a cascade of two fundamental detector types that registers either presence or movement in a region of interest (ROI). Each fundamental detector is laid out in a predefined ROI where the road users of interest may be observed. In the watch-dog system, a detector is either triggered or non-triggered, and it is a combination of this binary logic that lays out the RTV, LTV, and SGC detectors.

The *presence detector* uses a background subtraction technique based on reference images. The reference images are compared to the current frame by computing the Canny edges [21] in the ROI of the frame. If the difference of the two edge images is greater than a specified threshold, the detector is marked as triggered. If the difference of the edge images is below 80 % of the threshold, and the background has not been updated for τ consecutive frames, the background image is updated with the current image.

The *movement detector* estimates the movement in a certain ROI of the intersection by computing the dense optical flow [22] between two frames. Only the flow vectors within a desired direction of movement and above a certain magnitude are kept. If the number of remaining flow vectors surpass a threshold, the detector is triggered. A detailed description of the movement and presence detectors is found in [4].

The movement and presence detectors are overlaid on specific parts of the intersection and several detectors are chained to detect RTV, LTV, or SGC. For instance, if we want to detect RTV, we know that vehicles of interest must enter the intersection, perform a right turn in a designated area, and eventually exit the intersection. In the watch-dog framework, this translates to three sequential detections; detecting presence, detecting movement, and detecting presence. Prototype layouts of the RTV, LTV and SGC detectors are shown in Figure 1. The presence detector is abbreviated E (edge), the movement detector F (flow), and a new S (stationary) detector is introduced which detects if something is present, but not moving. The stationary detector is a combination of the presence and movement detectors configured on the same ROI. Activity diagrams which describe the sequential logic of the RTV and SGC detectors, are shown in Figure 2. The LTV and RTV detectors contain one or more modules (F2, F3, F4) which are used to prevent the detection of road users from other directions.

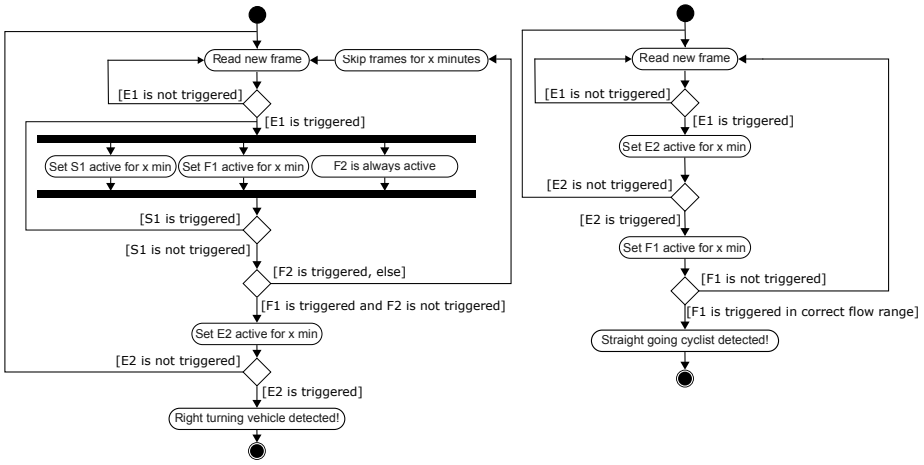


(a) RTV (b) LTV (c) SGC

Fig. 1: Prototype layouts for detecting road user actions. The arrows of the F detectors indicates the direction of movement for which the detector is configured

3.1 Fusing Modalities

The watch-dog operates on RGB, thermal, and combined RGBT video data. In RGBT mode, the fundamental detectors of the watch-dog run in parallel on the synchronized RGB and thermal video data. The fundamental detectors output a confidence value between 0 and 1 and a value ≥ 0.5 indicates that the detector



(a) RTV detector (b) SGC detector

Fig. 2: Activity diagrams of RTV and SGC watch-dog detectors. The LTV detector is similar to the RTV detector. In the LTV detector, F3 and F4 shares the behaviour of F2 in the RTV detector

is triggered. If the averaged confidence value of the RGB and thermal detectors is above 0.5, the multi-modal detector is triggered.

4 Feature-Based Tracking of Road Users

We use the feature-based tracker of Saunier et al. [5] to compare the results of the watch-dog and assess the method in less-than-ideal weather conditions. The feature-based tracking algorithm is an extension of the method by Beymer et al. [7] which was used to track vehicles at highways. The cornerstone of the tracking algorithm is the KLT-tracker [23]. The extracted features are tracked over time and in order to mimic the physical constraints of the movement of the road users, the features are kept only if the spatio-temporal displacement is small and the averaged motion of the features is smooth. The displacement and motion constraints also entail that only objects in motion are tracked, i.e. if road users stop at any point in the intersection, the tracking of the road user is suspended and will be initiated under a different ID once the road user starts moving.

The features from the tracking stage are grouped in a subsequent, offline step. The grouping algorithm operates in world coordinates and groups feature tracks with similar motion. A feature is grouped to nearby features if the distance is below the maximum distance threshold, $D_{\text{connection}}$. A new feature is easily grouped to several feature groups through this connection stage. However, new

features are only added to existing groups if the feature and the group are similar for a minimum number of frames.

For each frame, it is checked if the available pairs of connected features are still belonging together. The distance $d_{i,j}$ between the connected features is computed and it is checked if the relative motion between the two feature tracks is within a segmentation threshold, $D_{\text{segmentation}}$. If their relative motion is above $d_{\text{segmentation}}$, the features are disconnected.

The $D_{\text{connection}}$ and $D_{\text{segmentation}}$ thresholds are tuned to obtain a balance between overgrouping and oversegmentation. If the thresholds are set too low, road users will be oversegmented, i.e. a single road user will be represented as multiple tracks. If the thresholds are set too high, adjacent road users are detected as one. Finding the right thresholds is a challenge, however, and one has to choose the road user type for which the thresholds should be optimized. For instance, the algorithm might correctly detect car-sized objects while smaller road user types, such as cyclists, are prone to overgrouping and larger objects, such as lorries, are oversegmented.

4.1 Fusing modalities

Although the feature-based tracker of [5] is built to operate on RGB video, it also translates well to video in the thermal domain. As described in the following section, objects in thermal video generally contain less information than their RGB representation. This is taken into account by adjusting the KLT-tracking parameters and allowing the formation of trajectory groups with fewer trajectories. Additional grouping parameters need not be changed because the grouping is performed in world coordinates in both domains. In RGBT mode, features are extracted separately in the RGB and thermal modality and mapped to a common world coordinate system. Grouping is then performed on the combined RGB and thermal trajectories to produce one single output.

In order to compare the performance of the watch-dog and feature-based tracking, we need to obtain measures of RTV, LTV, and SGC from the latter. Because the entry and exit points of these road users are well-defined in intersections, we define entry and exit masks for each road user action type. We thus define a RTV from an object trajectory if the trajectory passes through entry and exit masks defined for the intersection in question.

5 Thermal-Visible Intersection Data Set

In order to track road users in a diverse range of weather and environmental conditions, the road users themselves must be visible to the algorithms. This is difficult during the night if artificial illumination is sparse. Vehicles might be detected by their headlights - but what about pedestrians and bicyclists?

Thermal cameras are independent of the availability of visible light and only depends on the emitted radiation from objects. Thus, thermal cameras allows to see objects through the night, as long as the temperature of the objects is

different from the temperature of the surroundings. Contrary to RGB cameras, thermal cameras are not susceptible to shadows. However, features are sparse in thermal images, and it might thus be more difficult to determine identities or distinguish between objects. An extensive review of thermal cameras in computer vision is found in [24]. When combined, RGB and thermal cameras extend the visibility of the road users and improves the robustness of traffic surveillance algorithms.

The proposed data set is an extension of the data set used in [4]. We extend the original four hours of video data to 16 hours and include a broader variety of weather and lighting conditions such as rain, wind, twilight, overcast, and full daylight. Further details regarding the contextual information of the data set is found in Table 1. Samples of each of the three locations are shown in Figure 3. For each of the three locations, RTV, LTV, and SGC have been manually annotated and assigned a time stamp whenever the desired road user type enters the area of the intersection corresponding to the E2 or E3 module of the watch-dog detectors shown in Figure 1.

Table 1: Environmental conditions of the proposed data set. The sequences are distributed over three days for each intersection

Location	Seq.	Time of day	Weather	Temp.	Lighting
A	1	07:00 - 08:00	Partly cloudy	12 °C	Full daylight
A	2	07:00 - 08:00	Light rain	17 °C	Overcast
A	3	07:00 - 08:00	Mostly Cloudy	15 °C	Overcast
A	4	12:00 - 13:00	Clear	19 °C	Full daylight
A	5	15:00 - 16:00	Light rain	19 °C	Overcast
A	6	16:00 - 17:00	Mostly Cloudy	19 °C	Full daylight
B	1	06:00 - 07:00	Rain	12 °C	Twilight
B	2	07:00 - 08:00	Rain	12 °C	Overcast
B	3	07:00 - 08:00	Shallow Fog, Partly Cloudy	6 °C	Overcast
B	4	12:00 - 13:00	Mostly Cloudy	13 °C	Full daylight
B	5	16:00 - 17:00	Partly Cloudy	17 °C	Full daylight
C	1	07:00 - 08:00	Light Rain Showers	13 °C	Overcast
C	2	07:00 - 08:00	Mostly Cloudy	13 °C	Overcast
C	3	12:00 - 13:00	Mostly Cloudy	16 °C	Overcast
C	4	16:00 - 17:00	Light Rain Showers	14 °C	Overcast
C	5	21:00 - 22:00	Rain Showers	12 °C	Deep twilight

6 Experimental Results

The watch-dog and the feature-based tracker are applied on the 16 hours of video described in Table 1. In order to mitigate the oversegmentation of vehicles caused

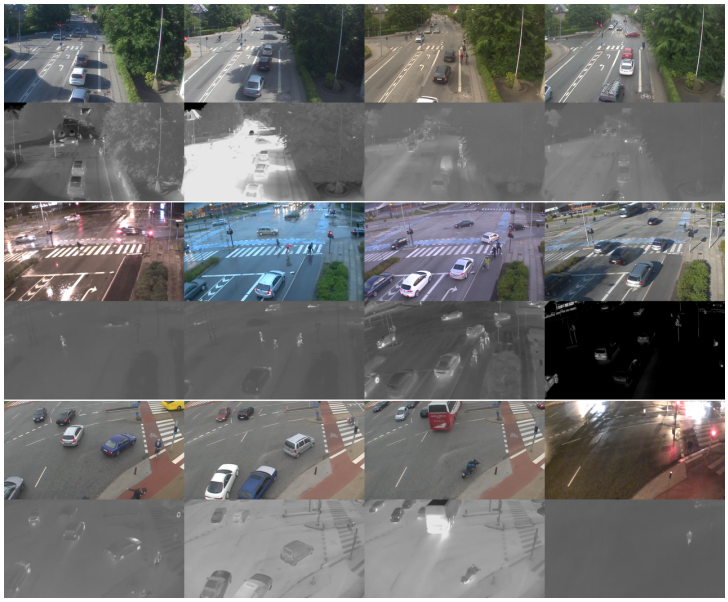


Fig. 3: Sample images of intersection A (top), B (middle), and C (bottom). The appearance of the intersections in both modalities vary greatly due to changing environmental conditions

by the feature-based tracker, duplicate tracks are filtered in a post-processing step. Only one object trajectory per second is allowed to pass through what corresponds to the E2 and E3 module of the watch-dog detectors and other tracks within the second are filtered out. Because oversegmentation is not a problem amongst cyclists, the 1 second filter is only applied to the detection of RTV and LTV. For detection of SGC, the threshold is relaxed to 0.3 seconds. A detection is marked as a true positive if is within $\pm$ 2 seconds of the ground truth.

The results of the test are shown in Table 2. It is seen that the watch-dog based detection of RTV and LTV generally is robust with precision and recall rates above 0.90 in several sequences. The RGBT mode of the watch-dog is shown to be stable whenever each individual modality gives acceptable results. Whenever the watch-dog fail to detect road users in a single modality, the RGBT results will suffer accordingly. Whenever results are stable, the RGBT mode performs best or trails behind the best performing modality by few percentage points. The results of the feature-based tracker shows remarkable precision with few or none false positives. Recall of RTV and LTV shows to be comparable or slightly better than the watch-dog performance. However, the detection of SGC by either the watch-dog or feature-based approaches shows considerable room for improvement. The smaller size and the irregular motion of the SGC are still challenges that need to be solved.

Table 2: RTV, LTV, and SGC detection performance of the watch-dog and feature-based trackers. The number of manually annotated SGC, RTV, and LTV road user actions is marked in italics in the right side of the table

Seq.		Block-based Watch-dog						Feature-based tracker						GT
		Precision			Recall			Precision			Recall			
		SGC	RTV	LTV	SGC	RTV	LTV	SGC	RTV	LTV	SGC	RTV	LTV	
A1	RGB	0.20	0.52	0.93	**0.71**	0.50	0.84	0.48	0.72	**0.91**	0.27	**0.80**	0.62	*146*
	T	0.77	**0.98**	0.94	0.70	0.75	0.65	**0.96**	**0.96**	0.77	**0.32**	0.79	**0.97**	*122*
	RGBT	**0.80**	0.97	**0.96**	**0.71**	**0.80**	**0.92**	0.57	0.74	0.86	0.24	0.75	0.92	*77*
A2	RGB	0.29	0.98	**0.99**	**0.64**	0.78	**0.99**	0.95	0.96	0.88	0.43	0.81	0.96	*131*
	T	0.57	**0.99**	0.91	0.56	0.61	0.55	**1.00**	**0.99**	0.93	0.19	0.69	0.96	*130*
	RGBT	**0.76**	0.97	0.97	0.60	**0.84**	0.89	0.95	0.95	**0.94**	**0.44**	**0.82**	**0.97**	*74*
A3	RGB	0.68	0.98	0.91	0.65	**0.90**	**0.97**	0.93	**0.99**	**0.87**	0.47	**0.88**	0.87	*141*
	T	0.85	0.98	0.94	0.65	0.85	0.16	**0.96**	**0.99**	0.83	0.33	0.70	**0.95**	*120*
	RGBT	**0.90**	**1.00**	**0.97**	**0.69**	**0.90**	0.89	0.95	0.97	0.85	**0.52**	0.87	0.94	*93*
A4	RGB	0.11	**1.00**	**0.91**	**0.90**	0.50	0.87	0.75	0.98	**0.93**	0.10	**0.88**	0.68	*29*
	T	0.73	0.97	0.90	0.55	**0.83**	0.87	**1.00**	**1.00**	0.84	0.10	0.77	**0.99**	*101*
	RGBT	**0.77**	0.99	0.90	0.79	0.79	**0.96**	0.83	0.99	0.91	**0.17**	0.87	**0.99**	*82*
A5	RGB	0.54	**0.98**	0.95	**0.70**	0.78	**0.91**	0.95	**0.98**	**0.93**	**0.41**	**0.88**	0.85	*44*
	T	0.63	0.94	0.90	**0.70**	0.78	0.27	**1.00**	0.96	0.91	0.07	0.51	**0.90**	*156*
	RGBT	**0.66**	0.97	**0.99**	**0.70**	**0.91**	0.73	0.92	0.96	0.91	0.27	0.77	0.88	*139*
A6	RGB	**1.00**	**0.96**	**0.98**	0.08	**0.80**	0.88	**1.00**	0.97	**0.91**	**0.18**	**0.90**	0.86	*39*
	T	0.27	0.73	0.96	**0.90**	0.27	0.88	**1.00**	**0.98**	0.89	0.15	0.82	**0.99**	*137*
	RGBT	**1.00**	0.95	0.97	0.00	0.42	**0.96**	**1.00**	**0.98**	0.89	0.15	0.82	**0.99**	*155*
B1	RGB	0.19	**0.97**	**0.98**	0.82	0.55	0.48	**1.00**	**0.94**	0.92	0.71	**0.85**	**1.00**	*28*
	T	**0.78**	0.91	0.97	**0.89**	0.71	0.69	**1.00**	0.93	**0.95**	0.64	0.67	0.70	*195*
	RGBT	0.63	0.94	0.96	0.86	**0.78**	**0.82**	0.96	0.92	0.62	**0.79**	0.81	0.65	*89*
B2	RGB	0.34	**0.97**	**0.99**	**0.73**	0.67	0.83	**1.00**	**0.97**	**0.93**	**0.72**	**0.88**	**0.97**	*71*
	T	0.65	0.84	0.98	0.68	0.78	0.69	0.98	**0.97**	0.92	0.61	0.75	0.91	*353*
	RGBT	**0.67**	0.95	0.98	0.68	**0.85**	**0.95**	**1.00**	**0.97**	0.83	0.69	0.83	0.79	*210*
B3	RGB	0.43	**0.99**	0.92	**0.65**	0.89	0.93	**1.00**	**1.00**	0.90	0.63	0.84	**0.98**	*92*
	T	0.19	0.94	0.92	0.64	0.80	0.83	0.98	0.99	0.71	**0.68**	**0.87**	0.81	*377*
	RGBT	**0.49**	**0.99**	**0.93**	0.63	**0.90**	**0.99**	**1.00**	**1.00**	**0.91**	0.61	0.81	0.88	*177*
B4	RGB	**0.08**	0.98	0.96	**0.82**	**0.67**	**0.89**	**1.00**	0.96	0.93	**0.71**	**0.81**	**0.99**	*28*
	T	0.03	**1.00**	**1.00**	0.79	0.00	0.02	**1.00**	**0.99**	0.86	0.64	0.78	**0.99**	*205*
	RGBT	0.04	**1.00**	**1.00**	0.79	0.00	0.05	**1.00**	0.98	**0.95**	**0.71**	0.75	**0.99**	*87*
B5	RGB	0.09	**0.99**	0.97	**0.83**	0.88	**0.94**	0.83	**0.99**	0.95	0.60	0.83	**0.99**	*48*
	T	0.26	**0.99**	**1.00**	0.60	0.91	0.87	**0.95**	0.96	0.87	**0.77**	**0.89**	0.91	*347*
	RGBT	0.07	**0.99**	0.99	0.81	**0.93**	**0.94**	0.86	**0.99**	**0.98**	0.63	0.80	0.95	*102*
C1	RGB	0.90	**0.94**	**0.97**	**0.73**	**0.94**	**0.95**	**1.00**	0.96	0.96	0.51	0.94	0.69	*74*
	T	0.89	0.88	0.96	0.66	0.91	0.81	**1.00**	0.93	0.93	**0.54**	**0.97**	**0.96**	*116*
	RGBT	**0.96**	**0.94**	0.96	0.70	0.93	0.94	**1.00**	**0.97**	**0.97**	0.53	**0.97**	0.70	*99*
C2	RGB	0.80	0.97	0.94	**0.75**	0.94	0.81	**0.98**	0.97	0.98	**0.52**	0.97	0.72	*88*
	T	0.93	0.42	**0.97**	0.74	0.96	0.88	**0.98**	0.94	0.94	0.51	**1.00**	**0.96**	*120*
	RGBT	**0.94**	**0.98**	0.95	**0.75**	**0.97**	**0.92**	**0.98**	**0.99**	**0.99**	**0.52**	0.97	0.73	*113*
C3	RGB	**0.79**	**0.99**	**0.94**	0.92	**0.94**	**0.94**	0.91	**1.00**	0.98	**0.83**	0.97	0.82	*12*
	T	0.03	0.55	0.81	**1.00**	0.46	0.48	**1.00**	0.98	0.98	**0.83**	**1.00**	**0.91**	*128*
	RGBT	0.14	0.98	**0.94**	0.75	0.48	0.61	0.91	**1.00**	**0.99**	**0.83**	0.99	0.77	*109*
C4	RGB	**0.84**	**0.98**	**0.99**	**0.79**	0.35	**0.93**	**1.00**	**1.00**	**0.99**	0.56	0.96	0.82	*34*
	T	0.31	0.93	0.88	**0.79**	**0.74**	0.81	**1.00**	0.97	0.98	0.65	**0.98**	**0.93**	*155*
	RGBT	0.33	0.96	**0.99**	0.76	0.61	0.81	**1.00**	0.99	**0.99**	**0.62**	0.97	0.80	*103*
C5	RGB	0.01	0.53	**1.00**	**0.67**	0.24	0.16	**1.00**	0.97	**1.00**	**0.67**	0.94	0.84	*3*
	T	**0.67**	0.94	**1.00**	**0.67**	0.88	0.68	0.67	0.92	**1.00**	**0.67**	**1.00**	**0.95**	*33*
	RGBT	**0.67**	**1.00**	0.93	**0.67**	**1.00**	**0.74**	**1.00**	**1.00**	**1.00**	**0.67**	**1.00**	**0.95**	*19*

7 Conclusion

This work evaluates the detection performance of left and right turning vehicles and straight going cyclists at urban intersections. Two detection approaches are evaluated at 16 hours of RGB and thermal video data featuring challenging weather and light levels. The first approach, the watch-dog, detects road user actions by using a chained set of basic detectors and spatial constraints of the intersection. The second approach, the feature-based detector, uses a KLT-tracker and additional grouping to track moving objects in the intersection. Both approaches show promising results when detecting vehicles while the detection of cyclists shows room for further improvement. The use of RGB and thermal modalities generally results in more stable performance for both detection approaches. However, more sophisticated weighting of modalities is needed to filter out false negatives whenever a detection algorithm breaks down in one modality.

Future work includes more persistent tracking of road users at all speeds in the intersection and further road user classification. Once full trajectories are found, trajectory classification techniques will be investigated to gather more detailed information of the road user actions [25].

Acknowledgements

The authors thank Tanja Kidmann Osmann Madsen for acquiring the data as well as assistance on the ground truth. This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 635895. This publication reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

References

1. European Commission: White paper roadmap to a single european transport area towards a competitive and resource efficient transport system. COM (2011) **144** (2011)
2. Hydén, C.: The development of a method for traffic safety evaluation: The swedish traffic conflicts technique. BULLETIN LUND INSTITUTE OF TECHNOLOGY, DEPARTMENT (1987)
3. Svensson, Å., Hydén, C.: Estimating the severity of safety related behaviour. Accident Analysis & Prevention **38** (2006) 379–385
4. Bahnsen, C., Moeslund, T.B.: Detecting road user actions in traffic intersections using rgb and thermal video. In: Advanced Video and Signal Based Surveillance, 2015. AVSS 2015. IEEE Conference on, IEEE (2015)
5. Saunier, N., Sayed, T.: A feature-based tracking algorithm for vehicles in intersections. In: Computer and Robot Vision, 2006. The 3rd Canadian Conference on, IEEE (2006) 59–59
6. Saunier, N.: "trafficintelligence". https://bitbucket.org/Nicolas/trafficintelligence (2015) Accessed: 2015-07-29.

7. Beymer, D., McLauchlan, P., Coifman, B., Malik, J.: A real-time computer vision system for measuring traffic parameters. In: Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, IEEE (1997) 495–501
8. Kastrinaki, V., Zervakis, M., Kalaitzakis, K.: A survey of video processing techniques for traffic applications. Image and vision computing **21** (2003) 359–381
9. Veeraraghavan, H., Masoud, O., Papanikolopoulos, N.P.: Computer vision algorithms for intersection monitoring. Intelligent Transportation Systems, IEEE Transactions on **4** (2003) 78–89
10. Maurin, B., Masoud, O., Papanikolopoulos, N.P.: Tracking all traffic: computer vision algorithms for monitoring vehicles, individuals, and crowds. Robotics & Automation Magazine, IEEE **12** (2005) 29–36
11. Messelodi, S., Modena, C.M., Zanin, M.: A computer vision system for the detection and classification of vehicles at urban road intersections. Pattern analysis and applications **8** (2005) 17–31
12. Zangenehpour, S., Miranda-Moreno, L.F., Saunier, N.: Automated classification in traffic video at intersections with heavy pedestrian and bicycle traffic. In: Transportation Research Board 93rd Annual Meeting. Number 14-4337 (2014)
13. Buch, N., Velastin, S., Orwell, J., et al.: A review of computer vision techniques for the analysis of urban traffic. Intelligent Transportation Systems, IEEE Transactions on **12** (2011) 920–939
14. Aköz, Ö., Karsligil, M.E.: Traffic event classification at intersections based on the severity of abnormality. Machine vision and applications (2014) 1–20
15. Kamijo, S., Matsushita, Y., Ikeuchi, K., Sakauchi, M.: Traffic monitoring and accident detection at intersections. Intelligent Transportation Systems, IEEE Transactions on **1** (2000) 108–118
16. Ki, Y.K., Lee, D.Y.: A traffic accident recording and reporting model at intersections. Intelligent Transportation Systems, IEEE Transactions on **8** (2007) 188–194
17. Zhang, W., Yu, B., Zelinsky, G.J., Samaras, D.: Object class recognition using multiple layer boosting with heterogeneous features. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 2., IEEE (2005) 323–330
18. Khammari, A., Nashashibi, F., Abramson, Y., Laurgeau, C.: Vehicle detection combining gradient analysis and adaboost classification. In: Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, IEEE (2005) 66–71
19. Nguyen, P.V., Le, H.B.: A multi-modal particle filter based motorcycle tracking system. In: PRICAI 2008: Trends in Artificial Intelligence. Springer (2008) 819–828
20. Saunier, N., Sayed, T., Ismail, K.: Large-scale automated analysis of vehicle interactions and collisions. Transportation Research Record: Journal of the Transportation Research Board **2147** (2010) 42–50
21. Canny, J.: A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on (1986) 679–698
22. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Image Analysis. Springer (2003) 363–370
23. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. International Journal of Computer Vision **56** (2004) 221–255
24. Gade, R., Moeslund, T.B.: Thermal cameras and applications: A survey. Machine vision and applications **25** (2014) 245–262
25. Morris, B.T., Trivedi, M.M.: Learning, modeling, and classification of vehicle track patterns from live video. Intelligent Transportation Systems, IEEE Transactions on **9** (2008) 425–437