

APPLICATION OF AUTOMATED VIDEO ANALYSIS FOR BEHAVIOURAL STUDIES: CONCEPT AND EXPERIENCE

Aliaksei Laureshyn¹, Håkan Ardö²,
Thomas Jonsson³, Åse Svensson⁴

ABSTRACT. Lund University, Sweden, is developing a video analysis system for making long-term behavioural studies, primarily in complex urban environments. The road users are detected using KLT interest point tracker. The trajectories are estimated using foreground-background segmentation, while speed is estimated using the shape analysis of interest point. The extracted trajectories are further used for behavioural analysis. In this paper we present the experience from an on-going study in Stockholm city where the task is to find out if allowing two-way bicycle traffic on one-way streets has negative effects on safety. The video analysis system was applied to detect biking in the “wrong” direction and analyse traffic conflicts between cyclists and other road users. The manual observations done parallel allowed validating the accuracy of the system performance.

INTRODUCTION

Video recording is commonly used when making behavioural studies in road traffic. It allows collecting data over long time periods and gives possibility to look through the material later and in more comfortable conditions. An installed video camera has less distractive effect on the road users compared to a road-side observer. The other advantage is that video is very illustrative and once an event of interest has been detected it can be looked over again, and additional relevant information can be retrieved. It is also possible to make more exact measurements from the video data, e.g. extract road users' position and speed.

¹ PhD Candidate, Department of Technology and Society, Faculty of Engineering LTH, University of Lund, Sweden, e-mail: aliaksei.laureshyn@tft.lth.se

² PhD Candidate, Centre for Mathematical Sciences, Faculty of Engineering LTH, University of Lund, Sweden, e-mail: ardo@maths.lth.se

³ Postdoctoral Fellow, Department of Technology and Society, Faculty of Engineering LTH, University of Lund, Sweden, e-mail: thomas.jonsson@tft.lth.se

⁴ Senior Researcher, Department of Technology and Society, Faculty of Engineering LTH, University of Lund, Sweden, e-mail: ase.svensson@tft.lth.se

However, as the frequency of the searched events goes down and the observational period increases, the problem of event detection becomes crucial. It is resource-consuming to use an observer watching through the entire film and the results might still be quite inaccurate as the attentiveness decreases with time. The extraction of the position and speed data manually is also a very time-consuming process (e.g. Archer, 2005, mentions the ratio between the time spent on manual “clicking” the vehicle position and the length of the video film processed as 10:1).

A research group at Lund University, Faculty of Engineering LTH is developing a system which can process the video data automatically. The system is aimed at studying the behaviour of road users in complex traffic environment (primarily urban conditions and mixed traffic modes). Now, at the final stage of a four-year project, the system can detect and track vehicles, cyclists and pedestrians and measure their speed and acceleration with relatively high accuracy. Special procedures are developed for describing the interaction between road users, e.g. detection of encounters (simultaneous presence in a certain area), presence of a collision course, detection of conflict situations, analysis of the speed profile shapes, etc.

This paper describes the methodology implemented in the system on an example of a study where it was practically applied. The study concerns the effects of introduction of two-way cycling on streets with one-way traffic in Stockholm city and includes observations done in two stages, before and after the introduction. So far only the before-observations have been completed (the second stage is planned for the summer 2008) and therefore it is not possible to make conclusions about the effects of the measure. The main focus of this paper lies on the performance of the system, problems encountered and the perspectives for further improvements.

STOCKHOLM STUDY: BACKGROUND AND SCOPE

The city of Stockholm has traditionally had a rather small mode share of bicycle trips compared to many other Swedish cities. However, there are initiatives trying to promote cycling. One possibility is to generally allow biking against one-way traffic. This would extend the available network for bicycle trips and lead to shorter travel routes and times. The backside would be that it might also lead to new dangerous situations and conflicts between the cyclists against one-way traffic and other road users.

The project aims at investigating the total safety effect of allowing cycling against one-way traffic, not only estimating the risk at a specific street. Therefore the design of the study also includes studying changes in the total bicycle flow and in the route choice, i.e. from which streets bicycle traffic transfers to the one-way streets. To be able to both establish risk at specific sites and route changes the sites were in some cases chosen so as to be able to count cyclists at several alternative routes.

Initially, 32 places were selected as potentially interesting for observations. However, finding a good place for camera installations turned out to be a complicated task. The cameras were normally attached to railings on balconies of apartments or offices, but in some places there were no buildings with balconies located near enough. At other sites there were potential camera positions but no electrical power available. Some owners of the buildings did not want to cooperate, or it was impossible to get in contact with them. Finally, only 22 places were filmed of which 18 were further analysed. Three of the excluded sites did not have any one way streets

entering or exiting the intersection (only being selected for counting bicyclists), and the fourth was excluded since the camera turned out to be too far away from the intersection to allow for proper analysis.



Figure 1. A camera unit.

Eight camera units⁵ were used for the study (Figure 1), moved between sites just before or after the weekend, resulting in three to four weekdays of recordings at each site. Further, the video material was processed and the objects moving in the “wrong” direction detected. To ensure the quality and validate the work of the video analysis system much work was still done manually. This included: a) calculation of the vehicle, pedestrian and cyclist flows for short periods at each site; b) visual control and sorting of the system detections, detection among them the situations which potentially might be conflicts.

AUTOMATED ANALYSES OF THE VIDEO DATA

Analysis of the recorded video films employs several techniques, that vary in degree of automation, complexity and computation intensity. Generally, the more advanced technique, the more sensitive it is for eventual errors, quality of input data and calibration procedures and the more validation it requires.

A “wrong-way” detector is a relatively simple analysis performed to find situations where something (e.g. pedestrian, vehicle, bicycle or just noise) is moving within the road but in the “wrong” direction. For each of those events a small video clip is saved to allow the events to be analysed in further details or manually sorted afterwards. Further the tracks of road users can be extracted and their speeds estimated. Finally, the tracks can be analysed to find the specific situations, for example encounters or traffic conflicts.

“Wrong-way” detector

The wrong way detection is a fast filter that is capable of removing a lot of the uninteresting parts of the video sequences. Typically the parameters are chosen to make sure that no interesting

⁵ A camera unit consists of a camera house which contains the camera itself and a mini server with a high capacity hard disk (750 Gb). The disk capacity allows for continuous filming for approximately 2 weeks with 320*240 resolution and 30 fps frame rate. The unit needs a power supply (12 V) and is connected to a portable computer via USB port for adjustment of the camera focus and starting/ending the recordings. Video is stored as 0.5-hour files in .mjpeg format.

events are removed, even if this means that quite a few uninteresting events are included among the detections.

The filter was implemented using the KLT interest point tracker (Shi & Tomasi, 1994). It finds points in the image that are expected to be easy to identify within the following frames. Typically this consists of points with a lot of structure such as corner points or edge junctions. Then it tracks those points over the entire video sequence. As points are lost new points are chosen to replace the lost ones. Typically a single road user contains several interest points and large road users contain more interest points than small road users. Some results from this tracker are shown in Figure 2.

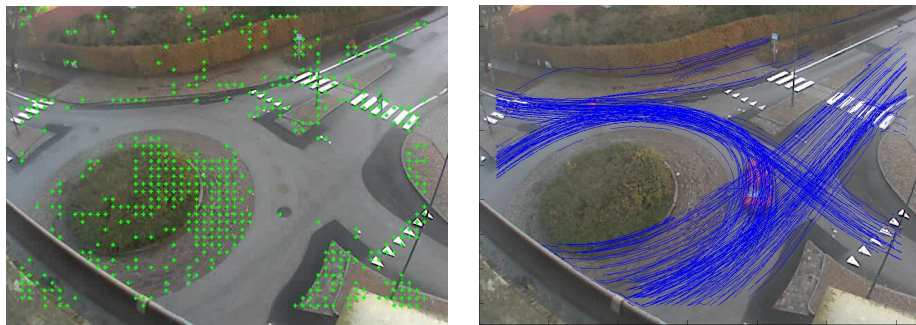


Figure 2. Results from the KLT interest point tracker (Shi & Tomasi, 1994). The green points are the interest points chosen and the blue lines are the tracks generated from a 30 min sequence. Typically each road user generates several tracks.

The tracks from the KLT tracker are then filtered based on a set of heuristics. The street is manually marked in the sequence, and all tracks outside it are removed. Also, the “wrong”-direction is manually specified, and tracks with a direction significantly different from it are removed. This discards road users travelling in the allowed directions and crossing the street. Finally all tracks moving too slowly are removed. This discards some slow moving pedestrians crossing the street at small angles as well as a lot of noise from interest points detected in the background.

The remaining tracks are counted and if this count exceeds some threshold the situation is considered interesting and a video cut is extracted and stored for later processing.

Road user tracking

The tracking of the road users can be performed either on the video clips produced by the detector described in the previous section or on the entire video film (usually, a video clip has to be at least 1-2 minutes long to ensure the quality of the background model, therefore if the number of detection is high and the clips “overlap”, it is easier to process the entire video material directly).

First a probabilistic background/foreground segmentation similar to described in Ardö & Åström, 2008, is performed. Background/foreground segmentation is a generalization of the background subtraction method. Several such methods exist and they all are based on the same principle of

estimating the background and then deciding what parts of the image currently show the background and what parts show something else, the foreground.

The method used has a background model that only consists of the temporal mean and variance for each pixel. Then each 8x8 block of pixels in the background model is compared to the corresponding pixels in the current frame using the correlation coefficient. As it is independent of intensity-level scaling and translation, the results become fairly independent on the lighting conditions as long as the lighting is constant over the entire 8x8 patch, i.e. it fails at the borders of sharp shadows. The output of this step is a probability for each pixel which is close to 1 if the pixel belongs to the foreground and close 0 if it belongs to the background. For nearly uniform blocks with very little structure, such as a car roof or road pavement, this probability becomes close to 0.5. This is because in those cases the correlation coefficient becomes unreliable and thus the algorithm becomes more uncertain. Figure 3 shows some example of the analysis.

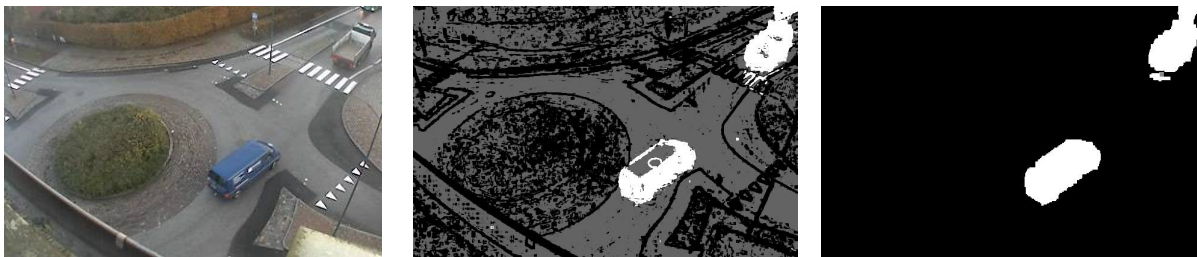


Figure 3. Background/foreground segmentation example. Left: Input frame. Middle: foreground probability per pixel (white pixels have high probability of being a foreground while the black pixels have a very low probability). Right: Resulting segmentation.

After the foreground probability is calculated for each pixel, the surrounding pixels will be used to decide on the final classification. A Markov random field is formulated where in addition to the per pixel foreground probabilities also the pair-wise probabilities of two neighbouring pixels either belonging to the same segment (foreground or background) or to different segments are considered. Typically there is much higher likelihood for pixels to belong to the same segment than to different segments. By using this information the unknown parts of the image can be filled in as either background or foreground. Also errors such as small noise segments, small holes or shadow borders are removed.

The problem of solving the Markov random field, e.g. finding the segmentation that maximises all the probabilities, can be formulated as a graph-cut problem, which can be solved fast and gives a globally optimal solution. That means that among all $2^{640 \times 480} \approx 10^{92476}$ possible segmentations for a 640x480 image, the one found is guaranteed to be the one with highest likelihood. In the case of video it is also possible to speed up the calculations even more by utilizing the fact that the result for adjacent frames looks very similar, typically the road users have only moved a few pixels each. This is described in Kohli & Torr, 2005, where they need on average only 33.6 ms to process a 368x256 frame and 78.0 ms to process a 720x578 frame on a 2.8 GHz P4.

The detected foreground pixels, e.g. those painted white in Figure 3 right, are within each frame clustered together into a few connected components, where two foreground pixels are considered connected if they are neighbours. The neighbours of a pixel are defined as the pixels directly

above, below, left and right of it, i.e. each pixel has 4 neighbours. These connected segments typically correspond to one or more road users. It is only in rare occasions that a single road user is split into several connected components, while it is quite common for several road users to belong to the same segments. The later happens each time one road user is occluded by another, but it is also fairly common for two road users that are very close to become a single connected component, even there is no occlusion.

Those connected components are then used to cluster the interest point tracks from the previous sections into clusters of tracks belonging to the same road user. If there are more interest point tracks connecting two connected components than there are interest point tracks separating them, they are grouped together. This is a very easy operation as there can be no ambiguities, e.g. either two connected components are in the same group or they are not. There is no optimization needed. But each group will contain several road users, as whenever there are occlusions road users will be clustered together. To resolve this, the groups have to be split again. As it is fairly uncommon for a single road user to be split into several connected component, event of two sets of tracks belonging to different connected components for more than a few adjacent frames can be used to convert those two sets into two different road users. Allowing a single road user to consist of more than one connected component for a few frames improves the robustness of the system as this does happen when for example a car passes behind a lamppost. Figure 4 shows the tracks from a single car after the clustering is done.

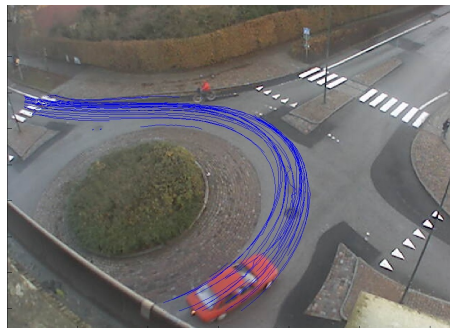


Figure 4. The interest point tracks belonging to a single road user.

After the clustering of the interest point tracks, each road user is represented as a set of interest points in each frame. By calculating the mean over all those points a point close to centre of the road user is found, which is used as the resulting position of the generated track. Unfortunately, this position jumps a bit back and forth as interest points are lost and new ones are picked up. This means that it cannot be used to estimate velocities.

The tracks going in the "wrong" way can be easily detected by using, for example, some entrance and exit gates which a track has to pass. The size of the connected segments representing a road user in the image can give a rough estimation of its real size, which allows to make estimations to which type a road user belongs (vehicle, pedestrian or cyclist).

Velocity estimation

To get a more precise estimation of the velocity of a road user a shape analysis of the interest points, similar to Åström et al., 2007, is performed. The transformation between each set of

interest points in terms of rotation translation and scaling is estimated as illustrated in Figure 5. A mean shape for the total set of interest points is estimated at the same time as the transformations. Then the position in each frame can be expressed as a transformation in terms of rotation translations and scaling from this mean shape into the shape observed in the image. The centre point is then estimated once for the entire track as the mean over all the points in the mean shape. Its position in the different frames is then found by applying the transformation for that frame.



Figure 5. Transformation between two sets of interesting points.

This gives a much smoother estimate of the position as it is no longer assumed that the same set of interest points is available in each frame, so the events of losing interest points and picking up new ones are handled much better. Unfortunately, it is assumed that the shape of the interest points does not change, e.g. the only transformations it accounts for are rotation translation and scaling. This is not the case when the interest points comes from a 3-dimensional object viewed with a projective camera, as they can be transformed with a full projective transformation. This means that this measurement is not very good as a position estimation, especially at the beginning and end of the track, where it might be well outside the entire road user. But differentiation of this position gives a very good estimation of the velocity of a road user, while the measurements from the previous section can be used as a fairly good position estimation.

Systematic error in speed and position estimation

An image observed in a camera view is a 2-dimensional representation of the 3-dimensional reality, therefore it is not possible to calculate the distances between the objects in reality using the measurements in the image only. However, having some prior knowledge about the reality some approximation can be done. For example, if the image is transformed as if taken from straight above the intersection (i.e. **rectified**), it would remind an intersection map and the distances can be re-calculated using simple scaling.

The rectified image provides accurate distances between the points in a certain plane in reality, usually the road plane. However, the distances between the object's parts which are elevated would be distorted, and the higher the elevation the higher the error is. Approximation that the objects are flat and lie in the road plane introduces a systematic error in position estimation as, seen from aside, an object appear to take more place on the road than it actually does. This error depends on the object's height, orientation and distance from the camera, the angle at which it is seen, i.e. theoretically the error is not constant as the object passes through the camera view. In practice, the error varies between 1-2 meters for smaller road users (cars, cyclists and pedestrians), but for high vehicles, such as busses or lorries, it is much higher.

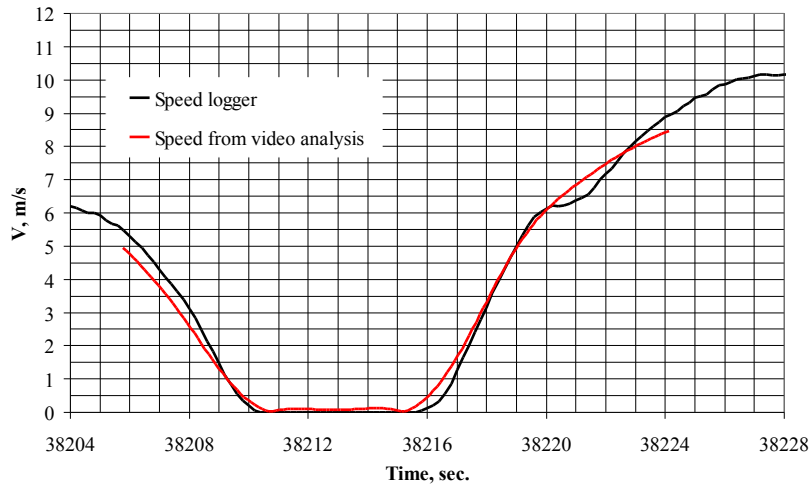


Figure 6. Speed estimated from video data vs. speed log in a car.

Despite the error in position the estimation of speed remains quite accurate. If the position error size does not change much during the object passage, the speed (i.e. the **change** of the position) is nearly free from the error and thus the speed estimations would have higher accuracy even without special corrections. Figure 6 illustrates some results from a test where a car with an installed speed logger was filmed and the logged speed was compared with estimations from the video data.

Detection of conflict situations

Traffic conflicts are situations close to accidents by nature (a break-down in the interaction between road users) but having less degree of severity. There are several Traffic Conflict Techniques developed, where a conflict is defined using different indicators. The most common indicators are Time-to-Collision and Post-Encroachment Time or some variation of these parameters (Hydén, 1987; Kraay & van_der_Horst, 1985; Gettman & Head, 2003).

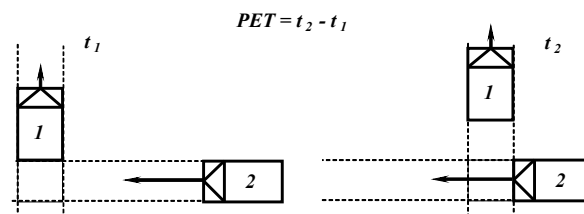


Figure 7. The conventional definition of Post-Encroachment Time (PET).

Time-to-Collision (TTC) is “the time required for two vehicles to collide if they continue at their present speed and on the same path” (Hydén, 1987). If the two road users pass a common spatial zone but at different time calculation of TTC is not possible. Such encounters can be described by Post-Encroachment Time (PET), defined as the time difference between the first road user leaving the common spatial zone and the second arriving at it (Archer, 2005, Figure 7). A similar, but continuous, parameter saying what the PET value will be if the road users continue at there

present speed and path is called Time Advantage (TAdv) in Hansson, 1975, and Gap Time in Gettman & Head, 2003.

The mentioned indicators are quite simple to calculate if the two vehicles are on a parallel or perpendicular course (the calculation procedures are, for instance, described in Vogel, 2003; van_der_Horst, 1990). However, in most general case two vehicles can approach each other from any side and at any angle. Moreover, even for the same approaching angle there are several possible ways for two vehicles to collide (Figure 8).

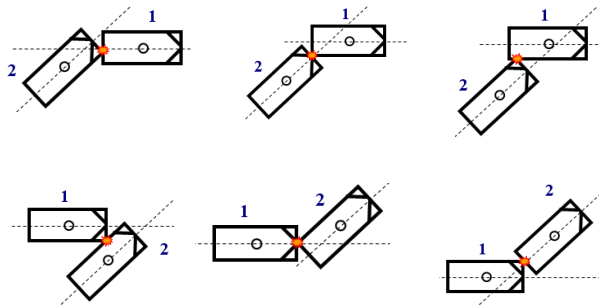


Figure 8. Possible collision types for the same approach angle (adopted from van_der_Horst, 1990)

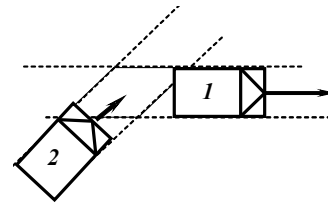


Figure 9. The problem of calculation of Time Advantage in case of non-perpendicular approaching angle: the zone become larger and the second vehicle can enter it while the first one has not left it and still the collision is avoided.

It can be seen from Figure 8 that in all the cases it is a corner of one vehicle that hits a side of another. In the most general case one have to separately calculate TTC for a side and a corner of two vehicles in all possible combinations and find the minimal value which will be the TTC for the entire vehicles.

The problem in calculation of TAdv is the definition of the common spatial zone, which can be quite large if the crossing angle is not perpendicular (see Figure 9). In some cases both vehicles can appear in the common zone but still avoid a collision. To overcome this problem the indicators have to be re-defined in other terms excluding the uncertain geometrical criteria. Here we used the following definition of Time Advantage: “the minimal delay of the first vehicle which, if applied, will result in a collision (assuming that otherwise the vehicles preserve the same speed and path)”. Practically, this also implies that the Time Advantage has to be calculated all corner-side combinations of the two vehicles and the minimal value will be the TAdv for the entire vehicles.

Another problem is the definition of the planned path for each road user. A human observer can easily project the planned trajectory, but it is quite difficult to explain how exactly this projection is done. A possible approximation is to assume that the road user actually follows the planned path, i.e. to calculate the indicators along the trajectory which is known. This can be misleading in case the road user avoided a conflict by changing the path, for example taking a larger radius in a turn or changing a lane.

Figure 10 shows an example of calculated TTC and TAdv profiles for an encounter between two road users. At the beginning (time from T_1 to T_2) they are on a collision course and the TTC is decreasing as they approach each other. However, at the same time vehicle 1 starts braking and vehicle 2 accelerates. Because of this from the moment T_2 they are no longer on a collision

course and the Time Advantage starts going up from zero. From the moment T_3 vehicle 2 is no longer on the way of vehicle 1 and none of the parameters can be calculated.

PET can here be calculated as the time interval between T_3 and the moment when the vehicle 1 arrives at the area which has been occupied by vehicle 2 at T_3 . However, we consider the Time Advantage value at T_3 as a more relevant indicator, since PET can be affected by changes in speed of vehicle 1 after T_3 which might not be really related to the interaction with vehicle 1.

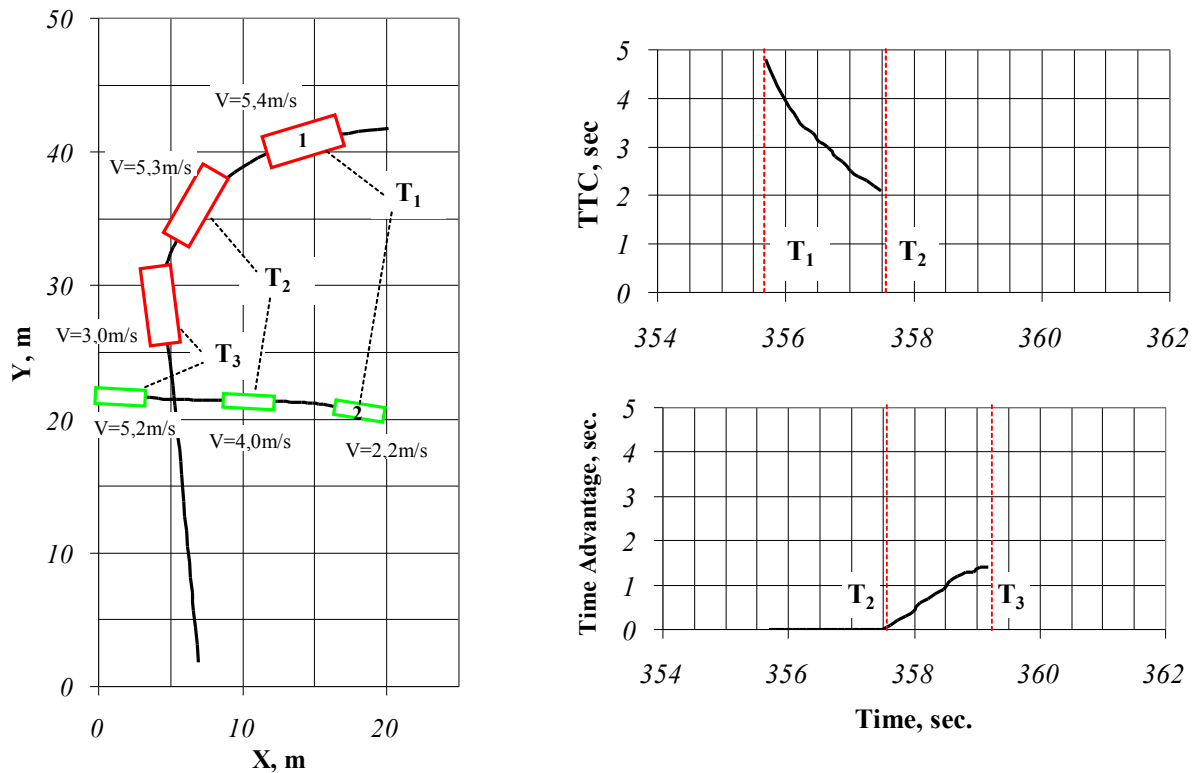


Figure 10. TTC and Time Advantage profiles for an interaction – an example.

In this way the interactions between the detected road users can be analysed. By setting certain threshold values for TTC and Time Advantage potential conflict situations can be detected.

RESULTS

The recording at 18 sites for 3-4 days resulted in 2.5 Tb of data and 900 hours of daytime video material. After the first stage of video analysis, the detection of objects moving in “wrong” direction, this amount was decreased to approximately 27000 short video clips, with a total length of approximately 115 hours. It was decided that this material would be looked through by two observers who sorted all the detections in 4 categories: bicyclists, pedestrians, cars and other (wrong detection or odd situations). This work took approximately one month of full-time work for the observers. The results are presented in Table 1. Since the observational periods were not the same at each site, the numbers are given as an average per day.

Table 1. The results of manual sorting of the detections at each site (average per day).

Site	Bicyclists	Pedestrians	Cars	Other	Sum
2	147	894	11	12	1063
4	100	44	19	7	170
5	110	54	9	14	187
6	63	938	26	126	1153
7	8	13	4	1	26
9	42	367	4	159	572
11	35	104	29	63	230
12	13	312	5	5	334
14	31	140	16	48	235
15	35	426	12	26	497
16	55	667	7	17	745
23	208	347	35	54	645
27	52	163	61	61	337
29	13	11	4	14	42
33	55	50	9	6	120
34	28	491	11	18	548
36	30	15	20	10	74
37	12	1	3	4	19
Total	1037	5037	285	645	6997
Percentage	15%	72%	4%	9%	100%

To estimate the accuracy of the automatic detection, manual cyclist counts were also done at each site for one or two 0.5-hour periods (from the video films). The comparison between manual counts and automatic detection is given in Table 2.

Another task performed by the observers was to detect situations that involve “wrong-way” cyclists and potentially might be conflicts. Totally, only 43 such situations were found, none of them was classified as a serious conflict according to the definition of the Swedish Traffic Conflict Technique (Hydén, 1987).

The information available after the video clips had been manually sorted was sufficient for the purpose of the study, therefore the extraction of road users’ tracks was not done on a large scale, but only for a test purpose. Site 33 was chosen for this test as it had relatively high number of potential conflicts (6) concentrated during four 0.5-hour periods (i.e. totally 2 hours of video). The tracks were extracted for all the road users during this period (even though there was a possibility to analyse only short sequences detected in the first stage, it was interesting to compare the performance of these two techniques in detection of “wrong-way” cyclists, too). As it was known that there are no serious conflicts to be found, the conflict criteria were set quite loose: one of the road user in an encounter had to move in the “wrong” direction with $TTC < 2$ sec. or $TAdv < 1$ sec. Table 3 shows the results of this test.

It turns out that the studied site was in a shade of a large tree during the most part of the day. This resulted in many false track detections located on the shade border (as the leaves were moving in the wind, the shades were detected as separate objects). These tracks were, however, very easy to sort out as they were abnormally long in time while the travel length did not exceed 1-2 meters.

Table 2. Manually observed vs. automatically detected “wrong-way” cyclists.

Site	Time	Manually observed	Automatically detected			
		Bicyclists	Bicyclists	Pedestrians	Cars	Other
2	07.00-07.30	2	3	7	2	0
4	07.00-07.30	3	3	0	0	0
	15.00-15.30	4	3	0	1	1
5	07.00-07.30	2	2	1	0	0
	15.00-15.30	2	4	3	1	0
6	07.00-07.30	1	1	15	3	0
	15.00-15.30	3	1	29	0	0
7	07.00-07.30	2	0	0	1	0
	15.00-15.30	1	0	0	0	0
9	07.00-07.30	2	1	1	0	0
	15.00-15.30	7	4	10	0	0
11	07.00-07.30	1	1	0	1	0
12	07.00-07.30	0	0	2	0	0
14	07.00-07.30	0	0	2	1	0
15	07.00-07.30	1	1	7	1	1
16	07.00-07.30	3	1	5	0	1
	15.00-15.30	4	0	22	0	0
23	07.30-08.00	7	6	12	0	1
	15.00-15.30	7	12	18	3	2
27	07.00-07.30	1	1	4	1	0
	15.00-15.30	0	0	7	0	1
29	07.00-07.30	0	0	0	0	0
	15.00-15.30	0	0	0	0	0
33	07.00-07.30	8	5	0	1	0
	15.00-15.30	4	4	3	0	0
34	07.00-07.30	1	1	9	0	0
	15.00-15.30	5	3	24	0	1
36	07.00-07.30	1	1	0	1	1
	15.00-15.30	3	2	0	0	1
37	07.00-07.30	3	0	0	0	0
SUM		78	60	181	17	10

DISCUSSION

One of the main benefits of the automated video analysis is that it condenses the video material when the events of interest are rare. The amount of raw video data collected in this study was hardly feasible to process employing only human observers. However, the amount of manual work done was still significant, therefore there is a high need for further automation of the process (e.g. use of the track detection and analysis technique on a large scale).

Table 3. Detection of “wrong-way” cyclists and conflicts by 2 techniques.

Period	"Wrong-way" detector (manually classified)			Track-based detection		
	Cyclists	Pedestrians	Conflicts	Cyclists (manually classified)	Pedestrians (manually classified)	Conflicts (automatically detected)
<i>1</i>	<i>8</i>	<i>1</i>	<i>2</i>	<i>9</i>	<i>12</i>	<i>2</i>
<i>2</i>	<i>4</i>	<i>0</i>	<i>1</i>	<i>3</i>	<i>10</i>	<i>0</i>
<i>3</i>	<i>2</i>	<i>3</i>	<i>2</i>	<i>3</i>	<i>5</i>	<i>2</i>
<i>4</i>	<i>3</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>0</i>
Total	<i>17</i>	<i>4</i>	<i>6</i>	<i>17</i>	<i>30</i>	<i>4</i>

The detection rate of the “wrong-way” cyclists is quite high, most of the manually counted “wrong-way” cyclists were also detected by the system (60 out of 78 in Table 2, i.e 77%). However, configuration of the system for a high detection rate results in many false detections. Only 15% of all the detections were cyclist while the main part (72%) were pedestrians walking on the street. This problem can be partly mitigated if a more advanced filter to distinguish between cyclists and pedestrians is introduced. The filter should include the threshold values for both the size and the average speed of a moving object.

Generally, the automated detection has lower detection rate compared to human observers. However, in some cases (sites 2, 5 and 23 in Table 2) the video detector found cyclists that were missed by the observers. All three sites are very lively with pedestrians and cyclists mixed, crossing or moving on the street in all possible directions. Such environment might be quite distracting for a human observer, while the automated detector is not much affected as long as the space between road users is large enough to detect them as separate objects.

Only 4 of the 6 known conflicts were detected by the track-based detector (Table 3). Analysis of the “misses” showed that in both cases the reason was that the “wrong-way” cyclists were not detected at all. However, the general detection rate of the simple “wrong-way” detector and a track-based detector is quite the same, so it might be just a coincidence that the missed cyclists were involved in conflicts. Further tests with more conflicts used will help to get more reliable figures.

The systematic bias in position estimation can contribute to the errors in detection. For example, cyclists moving close to the street side might appear as moving on a sidewalk. Extension of the detection zone on a sidewalk, on the other hand, results in many pedestrian detections which are not relevant (the difference in pedestrian detection can be seen in Table 3 where the detection zone for the track-based detector was extended 0.5 meter on a sidewalk compared to the simple “wrong-way” detector). The quality of the position estimation is also crucial if exact measurements of the conflict parameters (TTC, TAdv) are necessary.

Despite the high number of the cyclist passages observed, only few of them were classified as potential conflicts. It is expected, however, that after the introduction of two-way cycling on one-way streets the cyclist flow will increase, and, possibly, the number of conflicts, too. Even if the amount of serious conflicts will remain low, the analysis of the distribution of TTC and Time Advantage parameters can give some idea about the changes in safety after the measure introduction.

The installation of the cameras turned out to be a complicated task. It is quite difficult to find a place which is close, provides a good view and camera angle and has power supply. The problem of power supply can be partially solved if mobile sources (batteries) are used. The total power consumption of a camera unit is relatively small and a set of batteries can provide power for approximately one day of observations. However, this introduces new problems of changing the batteries during longer observation periods and securing them from sabotage if left unattended in the open.

A possible way to compensate for a poor camera view and occlusions is to use several cameras looking at a site from different angles. This would allow covering larger areas but also increases the amount of calculations and thus the time for analysis significantly.

At the moment the program code is written in Matlab and partly Visual Basic environments and not optimised for the fastest performance. Therefore there is still a potential to speed up the calculations by using C++ language and employing several parallel processors.

Calculation of the conflict indicators relies highly on the accuracy of the position and speed data. The quality of data can be improved by using higher resolution of the video but also by employing the data from several cameras which allows restoring the 3D-information and thus avoiding the described systematic error. If this problem is solved, the system opens completely new possibilities for validation and enhancement of the conflict techniques, which at the present are limited and simplified to match the capacity of human observers used for conflict detection.

CONCLUSIONS

The developed automated video analysis system has a great potential for use in behavioural studies, especially when the studied events are rare. However, at the moment, the amount of false detections is still very high and more advanced filtering algorithms are needed.

It is rarely possible to find a very good location to install a camera. Therefore more efforts should be put on developing techniques to compensate for a poor view by using data from several cameras.

There is still a need to improve the accuracy of position and speed estimation, which are important parameters for calculation of safety-related indicators. Since video analysis provides continuous description of road users' trajectories and speed profiles, it provides data for more comprehensive analysis of the behaviour and interactions and can be used, among other, for validation and enhancement of the conflict techniques.

REFERENCES

Archer, J. (2005) Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling. Doctoral thesis, Royal Institute of Technology, Stockholm, Department of Infrastructure.

- Ardö, H., K. Åström (2008) Bayesian Formulation of Image Patch Matching Using Cross-correlation. *Computer Vision and Pattern Recognition. Proceedings of IEEE Computer Society Conference on CVPR 08'* (submitted), pp. xxx-xxx.
- Gettman, D., L. Head (2003) *Surrogate Safety Measures from Traffic Simulation Models*. U.S. Department of Transportation, Federal Highway Administration, Report No. FHWA-RD-03-050.
- Hansson, A (1975) *Studies in driver behaviour, with applications in traffic design and planning. Two examples (In Swedish)*. Lund Institute of Technology, Department of Traffic Planning and Engineering, University of Lund, Bulletin 11.
- Hydén, C. (1987) *The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflict Technique*. Department of Traffic Planning and Engineering, Lund University, Bulletin 70.
- Kohli, P., P. H. S. Torr (2005) Efficiently Solving Dynamic Markov Random Fields Using Graph Cuts. *Computer Vision, 2005. 10th IEEE International Conference on ICCV 2005*, pp. 922-929.
- Kraay, J. H., A. R. A. van_der_Horst (1985) *The Trautenfels study: a diagnosis of road safety using Dutch conflict observation technique DOCTOR*. Institute for Road Safety Research SWOV, The Netherlands, R-85-53.
- Shi, J., C. Tomasi (1994) Good features to track. *Computer Vision and Pattern Recognition. Proceedings of IEEE Computer Society Conference on CVPR 94'*, pp. 593-600.
- van_der_Horst, R. (1990) *A time-based analysis of road user behaviour in normal and critical encounters*. Doctoral thesis, Delft University of Technology.
- Vogel, K. (2003) A comparison of headway and time to collision as safety indicators. *Accident Analysis and Prevention* (38), pp. 427-433.
- Åström, K., J. Karlsson, O. Enquist, A. Ericsson, F. Kahl (2007) Automatic Feature Point Correspondences and Shape Analysis with Missing Data and Outliers. In: B. K. Ersböll and K. S. Stenstrup (Eds.) *Lecture Notes in Computer Science*. Springer.