Collection of microlevel Safety and efficiency Indicators with Automated Video Analysis

Svensson, Åse; Laureshyn, Aliaksei; Jonsson, Thomas; Ardö, Håkan

2011

Document Version:
Publisher's PDF, also known as Version of record

COLLECTION OF MICRO-LEVEL SAFETY AND EFFICIENCY INDICATORS WITH AUTOMATED VIDEO ANALYSIS

Åse Svensson, PhD[1]
Aliaksei Laureshyn, PhD
Thomas Jonsson, PhD
Traffic and Roads, Department of Technology and Society, Faculty of Engineering, LTH, Lund University, Sweden

Håkan Ardö, PhD
Centre for Mathematical Sciences, Faculty of Engineering, LTH, Lund University, Sweden

Anna Persson, MSc student
Civil Engineering program, Faculty of Engineering, LTH, Lund University, Sweden

ABSTRACT
This paper revises the theoretical grounds for relation of the micro-level behaviour indicators to the safety and efficiency (level-of service) of the local traffic system. Automated video analysis is suggested as a tool for collection of the individual data about road users driving trajectories and speeds and interaction with other road users. We describe the current state of the video analysis system developed at Lund University, Sweden, and illustrate its performance as a watch-dog for the interaction situations that can be further used for safety and level-of-service analysis.

Keywords: road safety, level-of-service, indicator, severity hierarchy, interaction process, automated video analysis

INTRODUCTION

Safety and indicators on safety

Traffic safety is about reducing injuries in traffic. For a specific location accidents are however too rare events to provide a sufficient basis for safety analyses and suggestions of suitable measures. It is the accident and the sometimes devastating consequences that is the aim of traffic safety work but in order to get there we need to explore the traffic safety processes; we need to better understand the relationship between road user behaviour and injury accidents. A lot of research is therefore put into the work of developing methods for analysing surrogates to accidents; examples of safety indicators are Time-to-Collision, TTC, (e.g. Hayward, 1971; van der Horst,1990; Minderhoud & Bovy, 2001; Kiefer et al., 2005), Post-Encroachment-Time, PET,

[1] Corresponding author, e-mail: ase.svensson@tft.lth.se, tel +46 46 222 91 42.

(Cooper, 1983) and Deceleration Rate, DR, (Gettman & Head, 2003, Malkhamah, 2005). The Swedish Traffic Conflicts Technique (TCT) is such a method (Hydén, 1987), see Figure 1. Validation studies show that events above a specific severity, Serious Conflicts according to the Swedish TCT, have an established relationship with police reported injury accidents (Svensson, 1992).

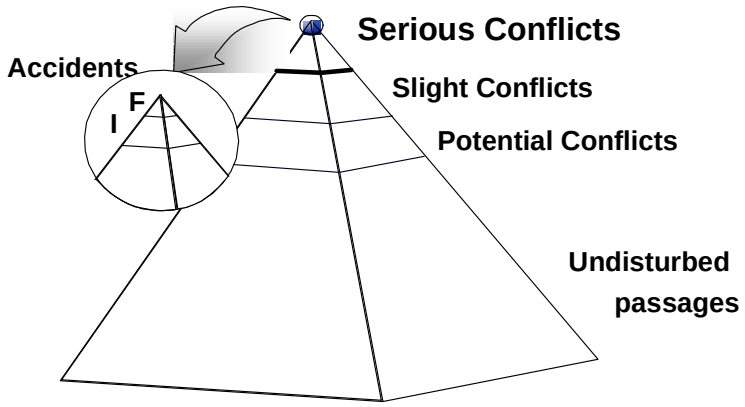

Figure 1. The pyramid - the interaction between road users as a continuum of events (Hydén, 1987)

With "collision course" as a prerequisite the TCT scope was later extended (Svensson, 1998) to besides injury accidents and serious conflicts also include normal interactive behaviour – severity hierarchies were constructed and analysed. A severity hierarchy is a continuum where events can be ordered with regard to their injury accident potential. The aim is to better understand the traffic safety process by describing the relationship between interactions, serious conflicts and injury accidents. The results showed that the shape of the hierarchy revealed information about the traffic safety process and that the shape differed with regard to type of manoeuvre, geometrical design, etc.

In Laureshyn et al. (2010) the severity hierarchy concept was broadened even more. Here safety related indicators like PET and Time Advantage (TAdv) + T_2 were introduced in order to handle all encounters thus not only those with collision course. TAdv is an indicator that describes situations where two road users pass a common spatial zone but at different times (Hansson, 1975). These are indeed situations without collision course but TAdvcan be used to describe *closeness* to a collision course. TAdv can also be looked upon as a continuous extension of the single value indicator PET. T_2 is a supplementary parameter to TAdv as TAdv itself is not sufficient to describe collision risk as it is also important to know how soon the encroachment will occur. T_2 is measured as the time of the second road users arriving at the common spatial zone and thus describes the nearness of the encroachment. In Laureshyn et al. (2010) the advantage of using continuous indicators like TTC or TAdv profiles instead of indicator values at

a certain moment like TTC_{min}, Time-to-Accident (TA) or PET was discussed. Compared to narrowing down a whole process to a single value, continuous indicators provide the possibility of analysing the development of an interaction as a process. Here the theoretical framework was established and continuous safety related indicators on micro-level developed.

Indicators on safety and level of service

When analysing qualities and effects of traffic and traffic planning, level of service is besides safety regarded as one of the most important ones. An interesting aspect is that level of service and safety often are presented as to be diametrical opposites. It is often assumed that if level of service is to be improved then safety will have to become worse and vice versa. A reason for this position may be found in the fact that both qualities are related to speed. For level of service high speed is an advantage while for safety high speed is a disadvantage. In a literature study Hagring (2000) starts to question these "facts" and points at empirical studies on macro level that show a relationship between flow and safety and between level of service and safety for motor vehicles. The relationships between accident rate and motor vehicle flow and between accident rate and delay are either linear and increasing or quadratic. For intersections without signal control Hagring assumes that this pattern may depend on a relationship between delay and risk taking on the individual level, an interpretation based on studies showing that shorter time-gaps are accepted as delay increases. For signal controlled intersections and road segments a corresponding relationship may be argued, that the propensity to take risks increases when delay increases. This can be seen as a strong indication on the prevalence of a compromise between delay and safety on an individual level. There are however no known studies on the relationship between level of service and safety on an individual level. Hagring concludes that simulation studies may give some guidance.

Another way of approaching this issue may be found in theories that include psychological and social aspects. Here many different theories to explain why road users (with emphasis on drivers) behave as they do in traffic have been proposed. Wilde's risk homeostasis theory (Wilde, 1994) is based on the assumption that a driver adapts his/her driving behaviour in such a way that a balance between what happens on the road (the perceived risk) and the level of risk that the driver can accept (accepted risk) is achieved. Thus, Wilde presumes that safety is the driving (and only?) motivational module when moving around in traffic and interacting with other road users and the road and vehicle environment. In contrast to Wilde's homeostasis theory the zero risk theory (Näätänen&Summala, 1976) presumes that risk is not the only motivational module. Other motives to satisfy are for instance hurry, maintaining speed and conservation of effort. The zero risk theory says that risks normally are avoided by staying within certain safety-margin thresholds and that behaviour is corrected when the safety-margin threshold is violated. Here Summala (1996) emphasizes the necessity of getting knowledge about this unknown mechanism that sees to it that the safety margins are not exceeded.

By studying interactive behaviour based on assumptions of suitable parameters regarding safety and time-efficiency it would be possible construct severity hierarchies (Svensson, 1998; Svensson & Hydén, 2006; Laureshyn et al., 2010). The shape of these severity hierarchies could then be analysed with regard to "optimal" behaviour and even pinpoint elements in the road and vehicle environment that hinders respectively enhances these behaviour. In Svensson's first

attempts at analysing such severity hierarchies based on TA/Speed the shapes were primarily analysed with regard to safety (Svensson, 1998). It was however concluded that the convexity of the severity hierarchy also could be interpreted as a distribution of individual safety margins - safety margins that differ due to each individual's unique acceptance of comfortable margins in time and space at interactions, and due to time of detection. It would be possible to argue that these margins also depend on considerations other than safety, such as the wish to maintain a certain speed or the wish to conserve energy and comfort. Different convexities at different types of locations would then be due to the individual road user's perception of the prevailing circumstances regarding safety, comfort etc. which might vary with regard to type of intersection, type of manoeuvre etc., if we have no reason to consider the sample of road users at the different locations as differing.

In these first attempts (Svensson, 1998; Svensson & Hydén, 2006) it was possible to relate the shape of the hierarchy to safety outcomes. The analyses of two intersections were performed with regard to a border in the severity hierarchy, above which injury accidents and serious conflicts were located. The results indicated considerable differences between the intersections regarding the distribution of events both considering above and below the border but also among the events below the border. At the signalised intersection there were injury accidents and serious conflicts, there were also many events at the lowest severities but hardly any events just below the border. At the not signalised intersections the distribution was completely different. Here there were no injury accidents or serious conflicts, hardly any events at the lowest severities but an accumulation of events just below the border. Thus, the latter distribution was associated with a safer location with more safe interactive behaviour. This shape of the hierarchy could also indicate an efficient way to proceed at the location as also indicated by Näätänen's and Summala's (1996) zero-risk theory. A reasonable interpretation was that a high interaction frequency below "the border" in the severity hierarchy could be an indication of an approved severity that is in accordance with the road users preferences regarding risk, time efficiency, degree of control etc. If in addition the probability of a serious injury accident is fairly low at this location in the severity hierarchy, this shape of a severity hierarchy seems to satisfy both the traffic engineers' and the road users' notions of a safe type of interaction. Thus, it is possible to connect Näätänen's and Summala's safety-margin threshold and the safety margins implied in the severity hierarchy. The first is a subjective, individual, safety margin and the latter probably an objective one, measured by physical parameters. There is, however, a possibility that the two might meet.

Continuous indicators on safety and level of service

As mentioned earlier the severity hierarchy concept was extended further by Laureshyn et al. (2010) to also open up for the option to use continuous indicators instead of indicator values at a certain moment. Compared to narrowing down a whole process to a single value, continuous indicators provide the possibility of analysing the development of an interaction as a process. To be able to study road users' interactional behaviour from a safety and level of service perspective it would therefore be possible to use a combination of TTC, TAdv, T_2 and relative speed. TTC is a very strong indicator on safety. Together with speed it might possess the same qualities as TA/Speed i.e. being an indication of where approved individual time-margins are in accordance with the individual's preference regarding perceived safety, time efficiency, degree of control,

etc., and when they are not. TAdv together with T_2 and relative speed could perhaps indicate something similar for situations without collision course. Here it is very interesting to note that the interpretation of T_2 is very similar to TTC, i.e. it describes how soon the road users will collide *if* they come on collision course. Another important property of T_2 is that it provides "smooth" transfer between the "collision course"- and "crossing course"-situations. At the moment of transfer from "collision course" to "crossing course" the TTC ceases to exist, but TAdv still equals zero. This makes T_2 equal to TTC, and if both TTC and T_2 are plotted on the same graph, they will make a continuous curve. Similarly to TTC, T_2 "jumps" into infinity if the second road user comes to a complete stop.

Automated video analysis

To be able to study safety related indicators on a micro-level the data collection has to be automated as these data are not possible for a human observer to detect. At Lund University in Sweden an automated video analysis system is being developed with the aim of analysing road user behaviour. A great advantage of having video recordings as a base in the system as compared to e.g. pure logged data, is that it is possible to link analyses to video clips of the situations. The system consists of one or several video cameras that simultaneously record an intersection from different positions and angles. The main advantages of using several cameras is that a larger area may be covered perhaps also including one or two of the entrances and exits and an increased accuracy in position. The main disadvantage is the significantly increased processing time. After the video processing, including foreground-background segmentation, etc., a data set is produced showing each road user's movement in terms of time and space coordinates, so called trajectories. Based on these trajectories algorithms for detection of relevant events are developed and applied. How the system is improved and adapted for the tasks in this paper is further explained in section Video Analysis below.

AIM

The aim of this paper is twofold: 1) To report on enhanced algorithms for improved tracking and accuracy in position for mixed manoeuvres at an urban intersection, 2) To analyse the effectiveness of a watch-dog system selecting relevant interactions.

DESCRIPTION OF STUDY LOCATION

In this study a signalized intersection was recorded during 6 weeks in the summer of 2010. The site was simultaneously recorded with 6 cameras, viewing the intersection and one of the approaches from different angles. Two cameras covered two different parts of one of the main approaches. Together they covered a stretch of approximately 100 m. Four cameras covered the intersection as such from different angles and with different zooming. The status of the study is the following: The transformation between road and image coordinates is completed. The cameras have been connected so that the objects (cars) are identified and tracked as they leave the view of one camera and enter the view of another camera.

Right now accidents and serious conflicts are manually selected from the recordings. The next step will be to find out with what accuracy the system is able to detect these events based on different safety indicators like TTC, PET andT_{adv}.

VIDEO ANALYSIS

The task of the image processing is to extract trajectories for all the road users in the intersection. A region of interest is defined and only the road users within this region are considered in the analyses. It is a process consisting of multiple steps illustrated in Figure 2 and described in more detail below.

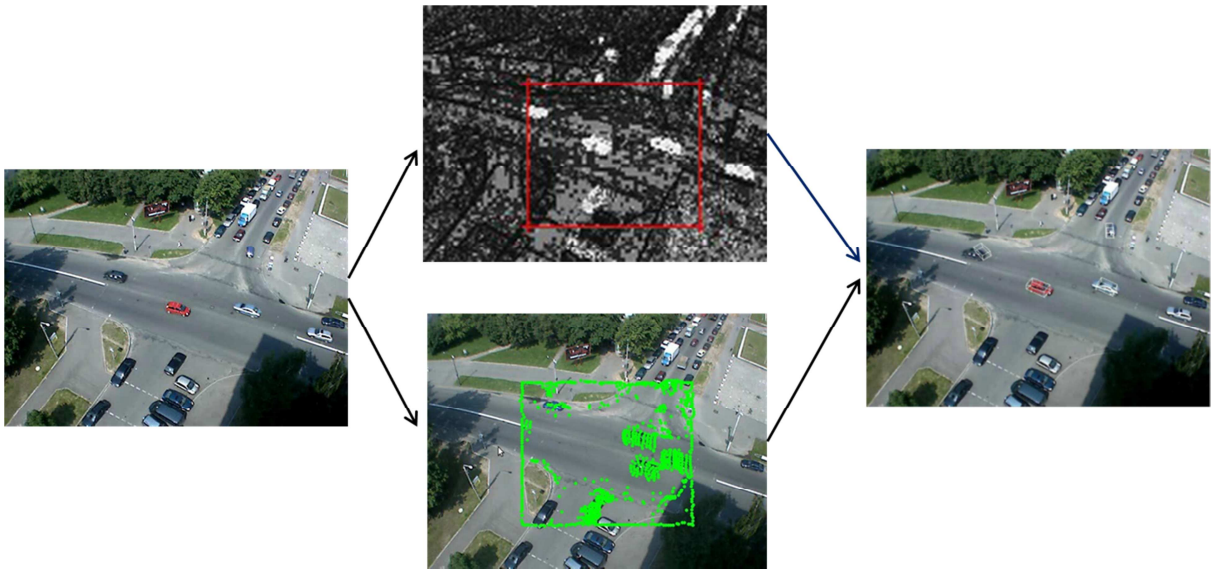

Figure 2. The video processing is performed in several steps. Information are extracted from the input video (left image) using background/foreground segmentation (middle top image) and point tracking (middle bottom image). A simple 3D-box model describing the road user shape is then fitted to this data (right). A region of interest is defined (marked in middle images) and only road users within this area are considered in the analyses.

Background/Foreground segmentation

The pixels of the video can be segmented into background pixels and foreground pixels. Background pixels are those that show the static background, which in these kinds of scenes are mostly pavements. Foreground pixels are those that show the moving foreground objects, i.e. the road users. This segmentation is tricky because the static background is not really static. It varies not only due to noise in the video but also due to lighting variations. On a sunny day those variations are slow, but on a cloudy day a cloud passing by the sun will cast large shadows across the entire scene moving faster than the road users. To handle this problem the pixels of the input video is divided into very small blocks (4x4 pixels) and then the blocks are normalized with respect the lighting conditions (Ardö&Åström, 2009).From these normalized blocks a

background image is estimated as a sliding temporal median. Also the expected variation of this background image is estimated as a sliding 25/75 % quartile.

Each input frame is then compared to the background image and by using the expected variation a probability can be estimated of how likely each block in the frame is to be showing the background. Figure 1, middle top, shows the results of this background/foreground segmentation. White pixels are likely to be showing foreground while black pixels are likely to be showing background. Grey pixels are uncertain. This typically happens for blocks that have very little structure, i.e. are close to be uniformly coloured. In those cases it is not possible to distinguish a lighting variation from a physical change. The result of this step gives information about where the road users are likely to be located in each frame.

Point tracking

To also get information about how the road user likely will move a point tracker is used. It is initiated by placing a set of points uniformly spaced along the border of the region of interest. These points are tracked using the KLT optical flow tracker (Shi&Tomasi,1994). As soon as one of them moves more than a few pixels from its starting position a new point is initialized there. Once a point has moved outside the region of interest it will be removed from the set of points being tracked and its trajectory through the intersection will be saved and passed on to the processing steps below. This will generate several point trajectories for each road user (Figure 1 middle bottom) and all of them will both start and stop at the border of the region of interest. Unfortunately there will also be quite a lot of point trajectories that jump from one road user to another and also point trajectories that get stuck on the background for long periods of time.

Object tracking

An object tracker combines the information from the point tracker with the information from the background foreground segmentation and tries to produce object tracks that agrees with both of them. A two dimensional rectangle is used here to model the road users, and a uniform grid of possible locations covering the intersection is defined. For each of those locations, 8 different orientations uniformly spaced over the entire circle are considered. This defines a discrete set of possible states, each described by its position orientation and frame number. For each of those states (position, orientation and frame) the likelihood of an object with that state being present is defined as the likelihood of all pixels within the rectangle being foreground. This likelihood is converted into a cost with the property that if it's more likely that there is an object present at a specific state, the cost for that state will be negative, otherwise positive.

The states are connected into a directed graph by using the information from the point tracker. The point trajectories produced are considered one by one. The positions along the trajectory are replaced by their closest state forming a sequence of states. The traveling direction of the point trajectory is used to define the orientation of the state. The states in this sequence are connected together by placing a directed edge from one to the other in the graph. This indicates that it is possible for an object to move from one state to the next in the same order as the point trajectory did. This is performed for each of the point tracks forming a large graph containing all states as nodes. Each state will have several outgoing edges (as there are several point tracks passing

through each of them). These edges restrict how objects are allowed to change their state over time to motions close to what was observed by the point tracker.

This graph now contains all the information needed. The background foreground segmentation gives the cost for an object to be located in any give state, and the possible directions to move between the different states are restricted to the motions observed by the point tracker. Road users traveling through the intersection can now be found as paths through this graph. The object tracking algorithm will try to find such paths with as low cost as possible, where the cost of a path is the sum of the costs of all the states along it. This search is restricted to not allow two objects to have the exact same state at the same time. However, even if objects are located at neighbouring states they will be overlapping a lot. This is not physically possible and it would be desirable to also impose the restriction that this is not allowed. However, doing that can increase the runtime of this optimization step several hundred times. That makes the resulting system more or less unusable.

Instead objects are allowed to overlap, which means that the same object will be detected several times as slightly different trajectories. The objects along those trajectories will overlap quite a lot. This can be used to cluster the produced trajectories into one cluster per actual road user.

Position refinement

So far only a single camera was used to detect the road-users. To make higher precision estimates of their position, several cameras viewing the intersection from different angles can be used. For this a three dimensional box model is needed. All the cameras are calibrated with respect to a single three dimensional representation of the intersection. By placing a box representing a road user within itin this representation of the intersection, the calibration can be used to project this box into each of the cameras. This will produce a polygon within each of the cameras that is supposed to contain foreground if there is a roadusers located where the box was placed. The likelihood of each three dimensional location is defined as the likelihood of all the pixels within in all those polygons to be foreground.

The two dimensional rectangle trajectories are converted into three dimensional box trajectories by projecting them onto the ground plane of the intersection. This will generate a systematic error in the position estimate. It can however be refined by doing a local search over nearby positions and orientations and choosing the one with the highest likelihood, considering all cameras simultaneously. This search can be performed with much higher resolution than the grid used above without losing much in total execution time.

INTERACTION ANALYSIS

At the moment, the raw data produced by video analysis does not provide the necessary accuracy to completely automate analysis of the interactions between the road users. We believe that, at least for some time in the future, this analysis will be done in two steps: first, a video analysis system selects the situations that might be of interest and then human observers look through and sort the detections, possibly manually correct trajectories and extract the necessary indicator values. Generally, the need to look at the situation should not be seen as “an evil”, but, on the

contrary, as a chance to really understand what is happening at the traffic site. The variety of situations that really take place in traffic is far beyond what we can imagine having seen the dry technical documentation available at the beginning of a study. Understanding and deriving common regularities of how road users interact is impossible without looking at the each individual interaction at the lowest micro-level. This can reveal the factors that are not present or obvious in the trajectory data, such as communication between the road users, presence of a third road user that affects the situation, odd manoeuvres (e.g. cars making U-turns or backing-up, pedestrians crossing diagonally, etc.) that occur sufficiently often and have too much effect on the traffic in general to be ignored.

The performance of the video analysis system is influenced by the traffic conditions. A few vehicles appearing simultaneously in the scene are easily detected and tracked and the errors are rare. On the contrary, when the intersection is congested, vehicles occlude the road surface and each other for long periods of time, making the background model less reliable and separation between the single road users more difficult. Thus errors like missed vehicles, false detections and trajectories that jumps between two vehicles close to each other become more common. For this test we chose evening time after the main afternoon peak but still with sufficient traffic at the intersection. Half-an-hour periods between 19.00 and 19.30 during four consecutive days were analysed.

The accident history at the intersection indicates that collisions between the left-turning and on-going vehicles are quite common. For this reason the interactions between these two traffic flows were chosen for the test.

First, the video analysis system was used to produce the trajectories of all the vehicles passing the intersection. Then, three criteria for selection of the relevant situation were tested:

1. **“Left turn”**. Special “gates” were set defining which area a trajectory has to pass to be counted as a left turn. According to this criterion, each left-turning vehicle was treated as separate detection.
2. **“Left turn + meeting”**. The previous criterion, completed with a requirement for an on-going vehicle to be present simultaneously at some moment.
3. **“Left turn + meeting + time”**. The previous criterion further restricted by a minimal TTC and Time Advantage or PET values for the left-turning and on-going vehicles.

The same video sequences were watched by an observer who counted all the left-turning vehicles and also situations where such vehicles interacted with on-coming traffic.

The results of the automated and manual detection are shown in Table 1.

Table 1. Detections of the relevant interactions by automated video analysis system and a human observer.

	Automated video analysis			Human observer	
	Detection criterion	**Detections, total**	**Relevant interaction**	**Left-turning vehicles**	**Relevant interactions**
Day 1	Left turn	18	6	18	6
	Left turn + meeting	22	6		
	Left turn + meeting + time	8	4		
Day 2	Left turn	31	16	34	17
	Left turn + meeting	43	16		
	Left turn + meeting + time	22	8		
Day 3	Left turn	21	10	26	11
	Left turn + meeting	13	7		
	Left turn + meeting + time	7	4		
Day 4	Left turn	24	8	24	8
	Left turn + meeting	23	8		
	Left turn + meeting + time	10	5		
Total	Left turn	94	40	102	42
	Left turn + meeting	101	37		
	Left turn + meeting + time	47	21		

The simple detector of all left-turning vehicles performs quite well. However, a great part of the detections are not really relevant as no interaction took place. The second criterion results in even higher number of detections. The reason for that is that while a left-turning vehicle is waiting for a suitable gap, several on-coming vehicles might pass and each of them generates a separate detection. Theoretically, this can be eliminated by adding one more criterion that limits the time nearness between the road users (since both TTC and Time Advantage values for a standing vehicle are very high). In reality however the third criterion resulted in that too many of the relevant situations missed (21 of 42, i.e. 50%). The reason for that, most probably, is the imperfection of the trajectory data that affected the calculation of the time-related indicators.

The test results clearly illustrate the problem of finding a trade-off between the degree of automation of the detection process and the detection rate. The more complex is the detection criteria, the more possibilities there is for the calculations to go wrong due to inaccuracy in the original trajectory. On the other hand, the price of missing some relevant situation can be worth the opportunity to analyse long time periods with low manual input and in this way collect large datasets – something that has so far been unfeasible impossible due to extreme costs of the manual labour.

CONCLUSIONS

We argue that detailed examination of the individual interactions might help to better understand the motives that produce certain road user behaviour. This analysis has to include both safety and

efficiency (level of service) perspectives in an integrated way, in the best case using the same indicators that reflect both aspects.

Traffic interaction is a process and thus should be described with continuous indicators. Collection of continuous data type is very laborious and is hardly feasible without automated techniques such as video analysis.

Our test shows that the video analysis system at Lund University can be used for initial filtering of the situations that are relevant, however, the final classification and data extraction still requires manual input.

It is still an urgent problem to improve the accuracy of the trajectory data produced by the video analysis algorithms. A possible way to go is to use data from several cameras to refine the road user positions. Hopefully, this will also allow us to analyse even congested traffic since the objects occluded in one of the cameras can be seen in some other one.

REFERENCES

Ardö, H., Åström, K. (2009) *Bayesian Formulation of Image Patch Matching Using Cross-correlation*. Third ACM/IEEE International Conference on Distributed Smart Cameras, 2009

Hagring, O. (2000) Sambandmellanframkomlighet och trafiksäkerhet – En förstudie."*Relationship between level of service and traffic safety – A before study*".Bulletin 192.Department of Traffic Engineering, Lund University, Sweden.

Hayward, J.C. (1971) *Near misses as a measure of safety at urban intersections.*Doctoral thesis.The Pensilvania State University, Department of Civil Engineering.

Hydén, C. (1987) *The development of a method for traffic safety evaluation: the Swedish traffic conflict technique*. Doctoral thesis.Lund University, Department of Traffic Planning and Engineering, Sweden.

Laureshyn, A., Svensson, Å.,Hydén, C. (2010) *Evaluation of traffic safety based on micro-level behavioural data: Theoretical framework and first implementation*. Accident Analysis & Prevention 42 (2010) 1637–1646. Elsevier

Laureshyn, A., Ardö, H., Svensson, Å., Jonsson, T. (2009) *Application of automated video analysis for behavioural studies: concept and experience*. IET Intelligent Transport Systems, 2009, Vol. 3, Iss. 3, pp. 345–357

Näätänen, R. and Summala, H. (1976) *Road user behaviour and traffic accidents*. North-Holland / American Elsevier, Amsterdam and New York.

Shi, J., Tomasi, C. (1994) *Good features to track.* Computer Vision Pattern Recognition. Proc. IEEE Computer Society Conf. CVPR, 1994, pp. 593– 600

Summala, H. (1996) *Accident risk and driver behaviour*. Safety Science, Vol. 22, No 1-3, pp. 103-117

Svensson, Å. and Hydén, C. (2006) *Estimating the severity of safety related behaviour*. Accident Analysis & Prevention 38 (2006) 379-385. Elsevier

Svensson, Å (1992) *Vidareutveckling och validering av den svenska konfliktteknikens.* Department of Traffic Planning and Engineering, Lund University, Sweden.

Svensson, Å. (1998) *A method for analysing the traffic process in a safety perspective*. Doctoral Thesis.Bulletin 166.Department of Traffic Engineering, Lund University, Sweden.

Wilde, G. (1994) *Target Risk*. ISBN 0-9699124-0-4