SmartCountplus – Towards Automated Counting and Modelling of Non -Motorised Traffic with a Stand-Alone Sensor Device

Norbert Brändle, Ahme Nabil Belbachir, Stephan Schraml

(Norbert Brändle, Austrian Insititue of Technology, Dynamic Transportation Systems, Norbert.Braendle@ait.ac.at)
(Ahmed Nabil Belbachir, Austrian Institute of Technology, Neuroinformatics, Address, Nabil.Belbachir@ait.ac.at)
(Stephan Schraml, Austrian Institute of Technology, Neuroinformatics, Stephan.Schraml@ait.ac.at)

1 ABSTRACT

We introduce a novel visual counting device being able to automatically discriminate between participants of non-motorised traffic (pedestrians, bicyclists). The sensor elements (pixels) respond to relative light intensity changes, thus avoiding conventional imaging and privacy issues usually raised by the public when it comes to visual surveillance. Three-dimensional depth information is computed with the stereo principle, and the set of light intensity change events is grouped together with a clustering algorithm to discriminate between moving objects. A classification algorithm based on descriptive features then identifies indivual participants of non-motorised traffic. A preliminary evaluation on a dataset with 128 passages shows a classification rate of 92% for riding cyclists and 100% for pedestrians for 2+1 classification, and 43-96% for 4+1 classification distinguishing between riding cyclists, pedestrians, walking cyclists, umbrellas and other objects.

2 INTRODUCTION

Volumes of non-motorised traffic are defined by the number of pedestrians and bicyclists per unit time. They are a key performance measure necessary to evaluate the impact of pedestrian and bicycle infrastructure improvements, to develop estimates of pedestrian and bicyclist risks, and to understand the environmental correlates of walking and cycling (SCHWARTZ et al., 2000). One of the most promising strategies for improving the amount and quality of non-motorised traffic volume data is to employ automated counting devices. Automated devices have the potential to reduce costs associated with traditional manual counting methods, including the cost of data input and storage, and to produce long-term continuous counts of non motorised traffic activity. Without automated devices, the manual collection of counts of more than a few days in length is highly impractical. (GREEN-ROESEL et al., 2008).

Ideally, rather than seperately counting pedestrians and bicylcles with dedicated automated devices, it is desirable that a single self-contained device be able to discriminate between the participants and provide at the interface the various traffic counts. Fig. 1a) shows a typical setup, where a park has two separate lanes for pedestrians and cyclicts, respectively. Fig. 1b) shows a similar setup with a bicyle and pedestrian lane. Fig. 1c) shows a mixed scenario comprising riding cyclists, pedestrians and pedestrians with umbrellas (a feature often not taken into account) captured from the bridge shown Fig. 1b).

The main objective of the SmartCountPlus project is to implement a stand-alone sensor device being able to deliver separate counts for pedestrians and bicycles and their velocities. After a brief review of the state of the art of automated pedestrian and bicyle counting in Section 3, this paper introduces the SmartCountPlus sensor device in Section 4. Section 5 sketches the main principles of individually counting non motorised traffic participants on data captured with the SmartCountPlus sensor device. Section 5 provides preliminarly experimental results performed on data captured at the scenario of Fig. 1b) and c).

Figure 1 a, b and c

Fig. 1: Many non-motorised traffic scenarios are mixed (a) bicycle and pedestrian lane in a park (b) bicycle and pedestrian lane under a bridge (c) mixed scenario involving bicyclists, pedestrians and pedestrians with umbrellas as viewed from the bridge in (b)

3 STATE OF THE ART AND CONTRIBUTION

Technologies for obtaining automatic pedestrian counts have been mainly developed for indoor environments (e.g. shopping malls) or low-density outdoor environments (e.g. trails). The study in (GREEN-ROESEL et al., 2008) provides an overview of existing pedestrian counting technologies. Due to strongly varying environmental conditions such as rainfall, snow and lighting, existing technologies are often not suited for counting pedestrians in urban outdoor environments. For example, (CLARK, 2009) reports findings from monitoring success and failure of walking investment in London, where laser based counters wer reported not to work as desired. Instead, (CLARK, 2009) reports 'CCTV' (Closed Ciruit Television) as a sucessful technology, without specifying the technology or product which actually performs automated analysis of the captured video data for pedestrian counting. The same holds for the study of pedestrian quality standards in New York City (NG, 2009). Indeed, reliable automated video analysis for pedestrian counting is still a challenging scientific top in the field of computer vision, especially for crowded scenarios involving dense groups of people, see e.g. the proceedings of the Performance Evaluation of Tracking and Surveillance (PETS) workshop series (PETS, 2009). While surveillance systems exist which classify between vehicles and loose groups of pedestrians, e.g. (SHAH et al., 2007), there is currently no system available discriminating reliably between pedestrians and cyclists. Recent commercially available pedestrian counting technologies include the modulated light intensity (MLI) (IEE,2010), which does not discriminate between pedestrians and bicyclists.

Automatic bicycle counting technologies are already more established and have similar advantages and disadvantages as pedestrian counting technologies, though desnse groups of bicycles are less likely than pedestrian crowds. As an example, automated bicycle counts have been measured in the city of Vienna, Austria since 2002 with the help of radar technology, and recently with induction loops. Simple rules dicriminate between bicyclists and other objects.

SmartCountPlus builds upon an existing visual indoor people counting technology developed by the AIT Neurinformatics group (SCHRAML et al, 2010a). This highly accurate people counting system has been already installed at a number of indoor locations, including a crowd control systems for a subway station attached to a soccer stadium (SEER et al, 2008). The major objective of SmartCountPlus is to extend this counting technology to be robust against outdoor conditions, where the major contributions are as follows:

- to extend the maximum capturing area of 3.3 meters in order to cope with broader outdoor scenarios,

- to develop embedded clustering and classification algorithms which run on the sensor device and are able to discriminate between pedestrians, pedestrians with umbrellas and bicyclists and calcluate their velocities,

- to perform extensive field tests at various scenarios, and to model classification and counting accuracy as well as dependencies on external data such as weather.

4 SMARTCOUNTPLUS SENSOR DEVICE

The sensor device is based on the principle of stereo vision which aims at duplicating the human visual system by computing a third dimensions (depth) using a pair of vision sensors. With stereo processing, adverse environmental conditions such as rain or cast shadows (which are a major challenge in visual processing systems) can be better met than with a mere 2D visual processing (GRUBB et al., 2004).
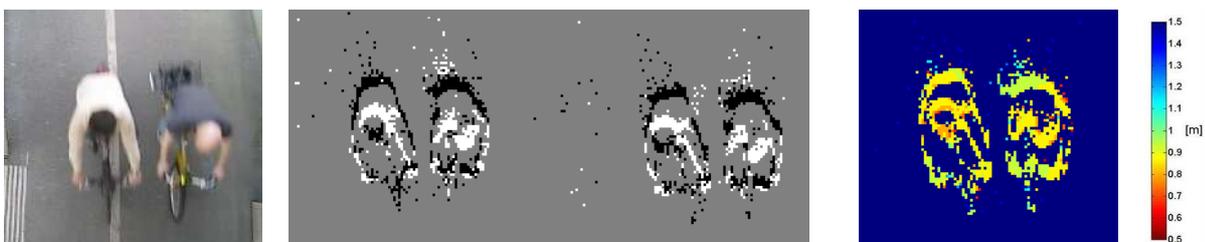

Figure 2 a, b, c, and d

Fig. 2: (a) Still image of two cyclists from a conventional video camera (b) light change events of the two dynamic stereo vision sensors corresponding to the scene in (a), (c) depth map computed by stereo (d), color code indicating range in meters from sensor
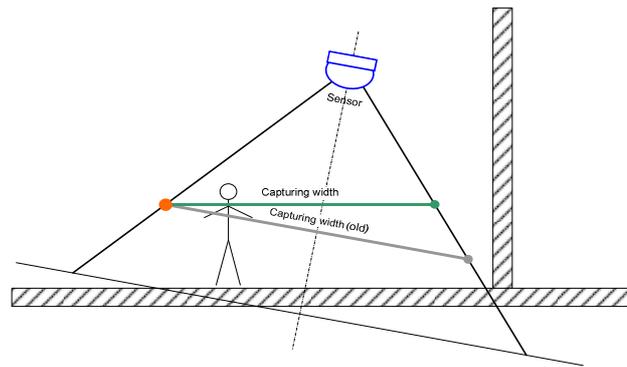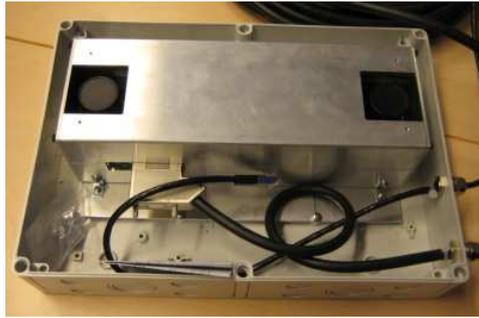
Figure 3 a and b

Fig. 3: (a) SmartCountPlus sensor housing (b) illustration of reduced capturing width when sensor device is mounted in a slanted position.

One vision sensor consists of an array of 128x128 array elements (pixels), where the pixels respond to relative light intensity changes. Note that since only light intensity changes are captured, no classical image in the visual spectrum is ever generated. Fig. 2a) shows a still image of a scene captured with a conventional video camera: The scene contains two riding cyclists, and Fig. 2b) shows the two corresponding two stereo pairs which are generated by the SmartCountPlus sensors: a dark pixel indicates a change from high intensity to low intensity and vice versa. Only the pixel elements which are changing intensity, so called 'address events' are transmitted by the sensor. Fig. 2c) shows the corresponding 'event depth map', where the color indicates the distance from the sensor (see Fig. 2d). Such spatio-temporal depth data are the input for the algorithms discriminating between cyclists and pedestrian.

Note that the image in Fig. 2a) is only for illustration purposes, and the sensor device never captures a conventional image. People can never be recognized in the captured depth data illustrated in Fig. 2b) and c) – such a processing therefore meets privacy concerns which are always raised when capturing visual data.

Fig. 3a) shows the housing of the SmartCountPlus device, including the two lenses of the two stereo vision devices which are separated by 26 cm. When installed in a 'top view' bird's eye position, a cross-section of 4.4 m width can be captured. In contrast to indoor scenarios, top view positions are often hard to obtain in outdoor scenarios, thus requiring mounting the sensor device in a slanted position. A slanted mounting position, however, will reduce the overall width of the captured cross-section, as illustrated in Fig. 3b). Table 1 quantifies the reductions of the capturing width depending for different angles as well as the optimal maximum mounting height. The 'left' and 'right' widths refer to the areas left and right of the dash-dotted line in Fig. 3b).

	optimal	10° slant	20° slant	30° slant
Stereo basis [cm]	26	26	26	26
Slant angle [degr]	0	10	20	30
Mounting height [m]	5.2	5.15	5.0	4.8
Capturing width (left) [m]	2.2	2.77	3.2	3.5
Capturing width (right) [m]	2.2	1.18	0.44	0.0
Capturing width [m]	**4.4**	**3.96**	**3.64**	**3.45**

Table 1: Optimal mounting height and capturing width for cross-sectional counting depending on the mounting angle

5   CLUSTERING AND CLASSIFICATION

The SmartCountPlus stereo vision sensor continuously generates events as a reaction to moving objects crossing the sensor field of view. Fig. 4 provides and overview of the processing steps, which are described in more detail in (BELBACHIR, 2010), (SCHRAML, 2010a) and (SCHRAML, 2010b).

The objetive of clustering is is to group together events belonging to the same moving object (pedestrians, cyclists, umbrellas). The clusters are computed online, meaning that all events are grouped in one step such that individual events are assigned to a cluster at once.
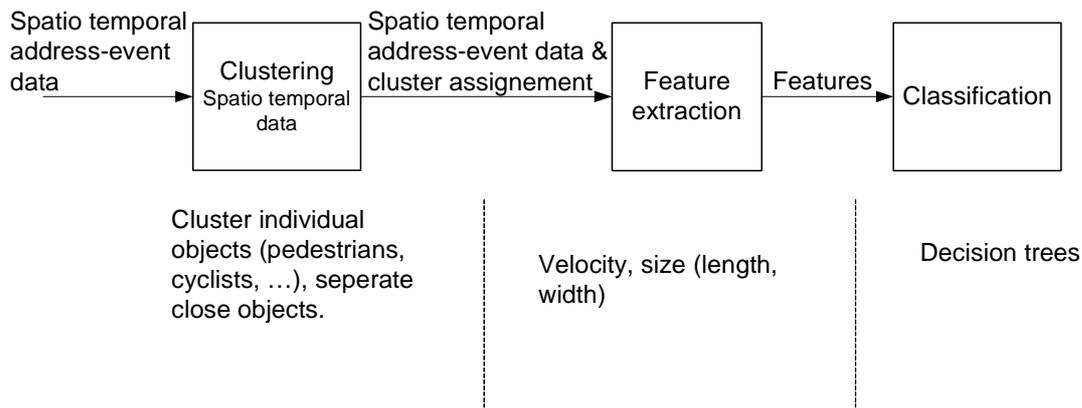
Fig. 4: Steps for processing sensor data as shown in Fig. 2 to classify between different participants

The objective of classification is to recognize the clustered objects' events and separate them into pedestrians and cyclists. After having built clusters from events through moving objects, descriptive cluster features are used to separate between pedestrians and cyclists with the help of a decision tree. We use three features (length, width and passage duration) for the classification as illustrated in Fig. 5. For the decision tree, thresholds on length, width and passage duration are set in order to distinguish between the multiple objects,
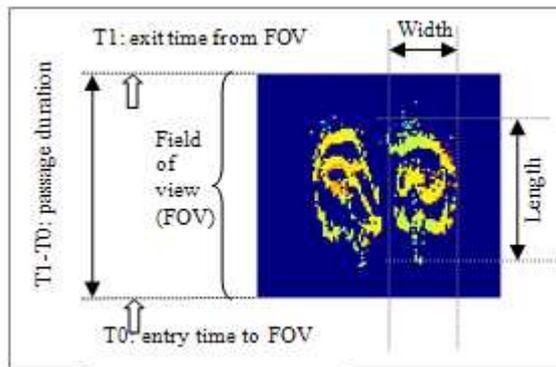

Fig. 5: Illustration of the features used for classification

6   EXPERIMENTAL RESULTS

To evaluate the event-based system and the classification method, we have collected real-world data at the test site shown in Fig. 1b). Test scenarios have been collected with a total of 128 passages (82 riding cyclists; 26 pedestrians, 13 walking cyclists and 7 pedestrians with umbrellas). Fig. 6 shows selected test cases.

Fig. 7a) shows classification results of riding cyclists and pedestrians for multiple scenarios using two dimensions (length to width ratio in the x-axis and passage duration in the y axis). The separating line represents the thresholds used in the decision tree for the classification. The two objects classes are almost linearly, separable. However, running persons can coincide with slowly riding cyclists.

Fig. 7b) and c) present classification results for 2+1 classes (pedestrian and riding cyclist) and 4+1 (pedestrian, riding cyclist, walking cyclist and pedestrian with umbrella), respectively. In these tables only the true positive classification (correctly classified) is represented as a first step. Still a full classification evaluation needs to be performed. It can be noticed that riding cyclists are best distinguishable together with pedestrian and walking cyclist, while pedestrians with umbrella are not efficiently classified. One reason for the bad classification of umbrellas might be the low density of the events and the difficulty to recognize them as one cluster. The other reason is probably the low number of test examples for this classification. This object (umbrella) still needs further investigation with more test data for robust analysis
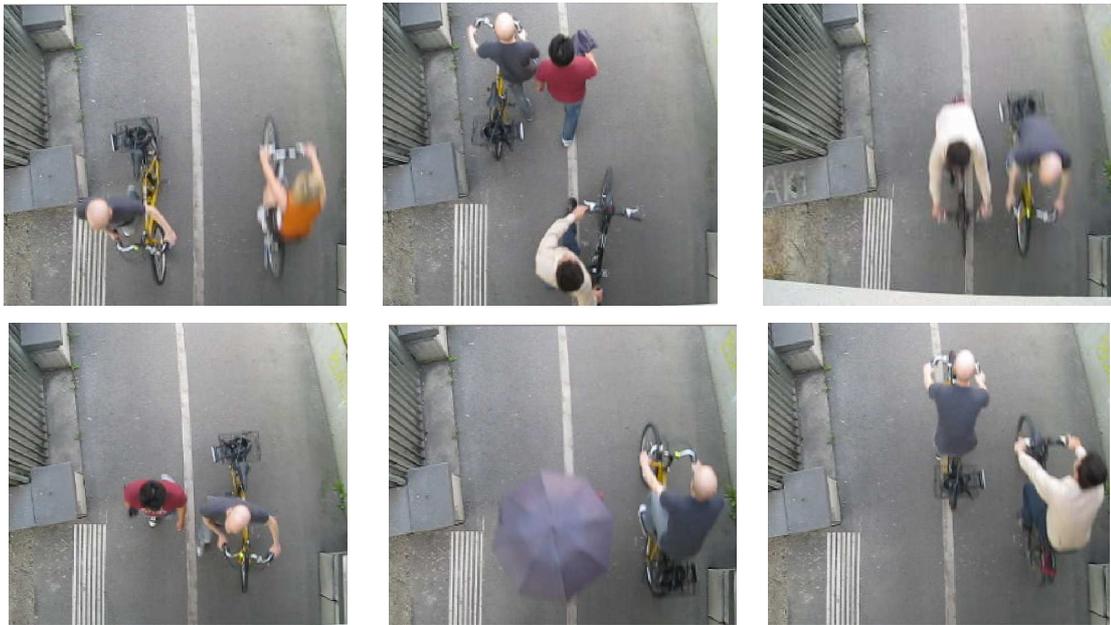
Fig. 6: Selected test scenarios

7   DISCUSSION

While the initial counting results are promising, a sample of 128 passages is clearly too small for representative performance figures. While nearly every commercially available counting technique claims counting accuracies of at least 95%, it remains often unclear for which accumulation interval the counting accuracy has been evaluated: If accuracy figures refer to a time interval of several hours, temporary gross errors could be compensated. Furthermore, the nature of the ground truth data (reference) can help interpretation: Has the ground truth data been directly collected by human observers (with corresponding inaccuracies for high people densities) or with the help of manual video annotation? Future work will therefore include mounting the SmartCountPlus sensor for an extended period of time at different locations. In order to provide a sound basis for evaluation, video footage will be captured for well-defined intervals, in order to obtain a sound model of classification and counting accuracy for different aggregation time intervals.

8   ACKNOWLEDGEMENTS
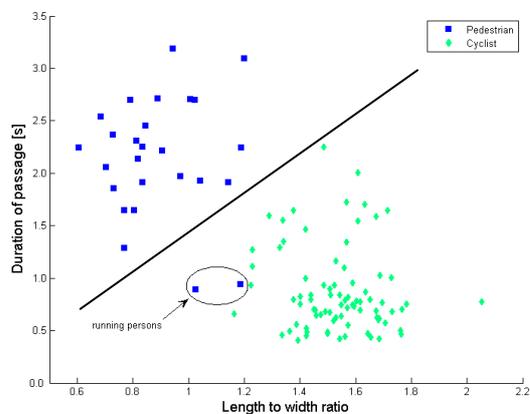
Figure
7a)

Type	Nb. cases	Correctly classified (true positive only)	Classification rate (%)
Riding cyclist	82	82	100
Pedestrian	26	24	92

Figure 7b)

Type	Nb. cases	Correctly classified (true positive only)	Classification rate (%)
Riding cyclist	82	79	96
Pedestrian	26	24	92
Walking cyclist	13	12	92
umbrella	7	3	43

Figure 7c)

Fig. 7: Classification Results (a) classification for riding cyclists and pedestrian using the two features 2D size (length to width ration) and passage duration (b) 2-1 classification, (c) 4-1 classification

9 REFERENCES

BELBACHIR, A.N., SCHRAML, S., BRÄNDLE, N. Real-time Classification of Pedestrians and Cyclists towards Intelligent Crowd Analysis. In: Proc First International Workshop on Socially Intelligent Surveillance and Monitoring (SISM2010), San Francisco, 2010.

CLARK, S.: Walking investment in London: Monitored success and failure. Walk21 Conference Presentation http://www.walk21.com/conferences/newyork.asp (last access April 2010), New York City, 2009.

GREENE-ROSEL, R., DIÓGENES, M.C., RAGLAND, D.R., LINDAU, L.A.: Effectiveness of a Commercially Available Automated Pedestrian Counting Device in Urban Environments: Comparison with Manual Counts. In: Transport Research Board Annual 2008 Meeting, Washington DC, USA, 2008.

GRUBB, G., ZELINSKY, A., NILSSON, L., RILBE, M.: 3D Vision Sensing for Improved Pedestrian Safety. In: Proceedings of the IEEE IVS, pp 19-24, 2004

IEE Homepage, http://www.iee.lu/, (last access April 2010)

NG, W.: NYC Pedestrian Quality Standards. Walk21 Conference Presentation http://www.walk21.com/conferences/newyork.asp (last access April 2010), New York City, 2009.

PETS2009 Homepage, http://www.cvg.rdg.ac.uk/PETS2009/a.html (last access April 2010).

SCHRAML, S., BELBACHIR A.N., MILOSEVIC, N, SCHÖN, P.: Dynamic Stereo Vision System for Real-Time Tracking. In: Proceedings IEEE International Symposium on Circuits and Systems (ISCAS2010). Paris, 2010.

SCHRAML, S., BELBACHIR, A.N.: A Spatio-temporal Clustering Method Using Real-time Motion Analysis on Event-based 3D Vision. In: Proc of the CVPR2010 Workshop on Three Dimensional Information Extraction for Video Analysis and Mining, San Francisco, 2010.

SCHRAML, S., BELBACHIR, A.N., BRÄNDLE, N.: A Real-time Pedestrian Classification Method for Event-based Dynamic Stereo Vision. In: Proc of the CVPR2010 Workshop on Embedded Computer Vision (ECVW2010), San Francisco, 2010.

SCHWARTZ, W., PORTER, C.: Bicycle and Pedestrian Data: Sources, Need, & Gaps. Department of Transportation Buerau of Transportation Statistics BTS00-02. Washington DC, 2000.

SEER, S., BRÄNDLE, N., BAUER, D.: Estimating Pedestrian Movement Characteristics for Crowd Control at Public Transport Facilities. In: Proceedings 11th International IEEE Conference on Intelligent Transportation Systems (ITSC2008), pp 742-747. Bejing, 2008

SHAH, M., JAVED, O., SHAFIQUE, K.: Automated Visual Surveillance in Realistic Scenarios. In: IEEE Multimedia 14, pp30-39. 2007.