STATIONARY VIDEO-BASED PEDESTRIAN RECOGNITION FOR DRIVER ASSISTANCE SYSTEMS

Steffen Schaack
Alois Mauthofer
carhs.communication GmbH
Germany
Ulrich Brunsmann
Labor für Mustererkennung und Computational Intelligence,
Hochschule Aschaffenburg – University of Applied Sciences
Germany
Paper Number 09-0276

ABSTRACT

As statistics have shown, forty-two percent of all injury accidents in Germany's road traffic happen at intersections. Infrastructure-mounted cameras for traffic analysis have been proposed to reduce this number as well as simulation tools, which assist in developing Car-to-Infrastructure (C2I) communication applications in the field of driver assistance, pedestrian, vehicle and traffic safety by a combination of a real application and virtual scenarios.
This paper describes an infrastructure-based vision system for pedestrian and vehicle detection, its integration in the C2X-communication software development framework viilab and the visualisation to display the acquired data in a C2X-vehicle. Two cameras are used to monitor an intersection in the visible spectral range out of different views. With methods of computer vision and machine learning road users are detected and analysed as pedestrians or vehicles for both views. The merged objects' positions are transformed into world coordinates and tracks within the traffic trace are generated. The data can be used in a simulation or can be requested in real time from C2X enabled cars via a roadside unit (RSU) as an environment radar. The performance of the system is discussed.

INTRODUCTION

In the future vehicles will be equipped with technology for car-to-car- and car-to-infrastructure-communication (C2X-communication). Examples for car-to-infrastructure-communication to avoid dangerous situations or an accident might be the broadcast of the traffic light status to cars near a crossing or a warning if a pedestrian crosses the road behind a bend. With forty-two percent of all road traffic accidents intersections are hotspots for accidents involving injury in Germany's road traffic [1]. This project focuses on infrastructure-based image processing at intersections.

Infrastructure-based image processing is already widely used for surveillance applications of road traffic. Examples are traffic control systems that recognise cars moving reverse to the driving direction on motorways or that detect smoke in video images to avoid a fire in a tunnel, frequently assisting security personnel who monitor many video cameras in a coordinating office [2]. Other concepts of infrastructure based image processing are topic of research projects. For instance the project OIS ("Optische Informationssysteme") [3] evaluated a system for adaptive control of traffic lights at a crossing depending on the volume of traffic. The detection mechanism rests upon several stationary mounted cameras monitoring the intersection from different angular fields. Another concept similar to the system introduced in this paper is discussed in the project PUVAME ([4] and [5]), where a detection of pedestrians with a multi camera setup on a park deck is executed. The evaluated system communicates the results of the road based image processing to an omnibus. The aim is to avoid collisions between pedestrian and omnibus in urban road traffic. The focus for this monitoring system is set on intersection and bus stop.

The development of intelligent co-operative systems requires a new generation of development tools which assist the automotive manufacturers and suppliers in developing Car2X communication applications. The software development framework viilab (vehicle infrastructure integration laboratory) ([6] and [7]) is especially designed for Car2X communication application developments. It can be installed as On Board Unit (OBU), Road Side Unit (RSU) or Monitoring Device (MON). Typical vehicle related interfaces (navigation, positioning, Display/HMI or bus systems) as well as infrastructure related interfaces (camera, traffic light) can be included. viilab is used in this project for the demonstration of the integration of infrastructure based data in a C2X environment.

This paper describes the architecture and the integration of a two-camera-based vision system for pedestrian-vehicle detection and tracking and the visualisation to display the acquired data in a C2X-vehicle. The performance of the subsystem is discussed.

REALISATION

For visualisation of pedestrians and vehicles within the simulation tool the following tasks have been solved:

1. Background/foreground separation and segmentation of objects in a region of interest
2. First classification of the segmented objects
3. Transformation of the position of the segmented objects into an ortho picture
4. Datafusion in the ortho picture
5. Tracking of the objects and second classification step
6. Transformation of the objects' positions in the ortho picture into WGS84 coordinates
7. Visualisation inside the simulation tool

Experimental

Figure 1 shows the positions of the two cameras and the corresponding fields of view. The cameras have resolutions of 720x576 pixels and 640x480 pixels, respectively. They are installed on the fourth floor of the university building, the baseline of approximately 20 m corresponds to the width of the building. For the implementation of image processing and machine learning methods the Open Computer Vision Library (OpenCV) and MVTec Halcon 8.0 were used on a 3 GHz dual-core CPU under Windows XP. The video data have been read out serially via Halcon addressing two frame grabbers. The time-lag due to synchronisation was about 52ms, resulting in a maximum positioning error of about 1m, if a maximum speed of 70 km/h is assumed. An overall repetition rate of processed data of 10 Hz was achieved with this system.

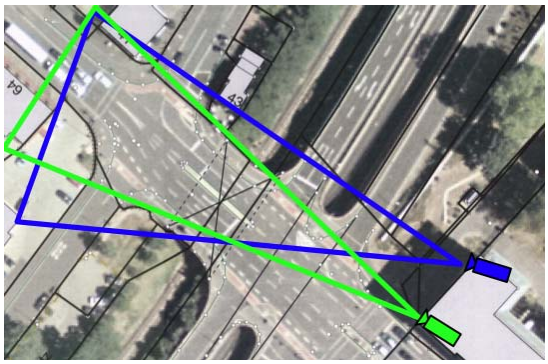

Figure 1. Setup of the cameras to monitor the intersection and field of view.

To implement the communication of the infrastructure based data to a C2X-vehicle the C2X software development framework viilab was used. In the C2X-vehicle both the CAN bus and GPS data can be accessed with the help of viilab to evaluate possible applications of C2X communication. Another part of viilab is the data-visualisation inside the C2X-vehicle, the so called viilab user interface (vui). Based on Mozilla's XUL technology visualisation pages for received data were programmed using JavaScript and scalable vector graphics (SVG) and displayed on the cars' monitor.

Video based pattern recognition

Differencing which means subtracting two images in succession (movement detector) has been implemented for background/foreground separation. In each of the two images a region of interest is used to confine the necessary operations on the observed intersection. The region of interest is visualised in figure 2 and 3 by the outer white polygon surrounding the monitored area. Due to the requirement that the region of interest should cover an identical area of the crossing in both camera pictures, prominent points serve as vertices of the polygons to limit this region.

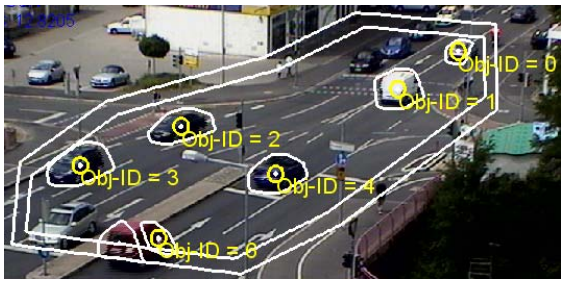

Figure 2. Results of the fusion algorithm for moving objects (Camera picture 1).

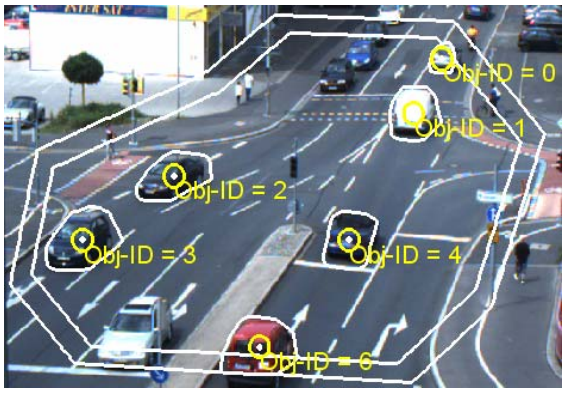

Figure 3. Results of the fusion algorithm for moving objects (Camera picture 2).

To discriminate between pedestrians and vehicles, a two step approach was implemented. In the first classification step features based on the normalised centered moments of second and third order were used. The second classification step is based on tracking statistics and will be described later on. The normalised centered moments of second order (M02, M20, M11) describe the variance of the

object regions from translation and scaling independently. The moments of third order (M30, M03, M21, M12) specify the asymmetry of the region also independently from translation and scaling. A Support Vector Machine (SVM) was trained with these features. For training a feature array was manually generated from a set of objects observed at an image sequence recorded from this crossing. Table 1 shows the confusion matrices of this first classification step for both cameras. The sensitivities for pedestrians and vehicles are in the order of 0.8.

Table 1.
Confusion matrices for the two implemented Support Vector Machines

Confusionmatrix Camera 1		Real Class	
		Pedestrian	Vehicle
Determined Class	Pedestrian'	108	76
	Vehicle'	19	769

Confusionmatrix Camera 2		Real Class	
		Pedestrian	Vehicle
Determined Class	Pedestrian'	67	119
	Vehicle'	26	711

Coordinate transformation and data fusion

The approach for datafusion was to transform the center of mass of all classified objects into an ortho picture of the intersection and to use the pixel coordinates of the ortho picture as reference coordinate system. The ortho picture was a survey map with reference points given in Gauss-Krüger coordinates.
If an identical object is detected by the two sensors within this approach it is assumed that a transformation of its centers of mass should result in two points that are located next to each other in the ortho coordinate system. So the minimum Euclidean distance within the ortho system was used to join objects.

The task of transformation from the two camera coordinate systems into the ortho coordinate system is achieved by artificial neural networks (ANN). For each camera a separate ANN is created to transform an x-/y-coordinate from the camera picture into an x-/y-coordinate in the ortho picture of the crossing. The training vectors were created from pixel coordinate pairs within the ROI of the ortho picture. The backpropagation algorithm was used to train two multilayer perceptrons. Figures 2 to 4 show the results of the implemented fusion algorithm for objects that were detected by the movement detector. Regions that belong to an identical road user obtain an identical ID in the two images as well as in the ortho picture of the scene. Furthermore the outlines of the detected objects are transformed into the ortho picture of the crossing.

With respect to real time processing contour data was approximated by means of the Douglas-Peuker algorithm [8] before this transformation was applied.

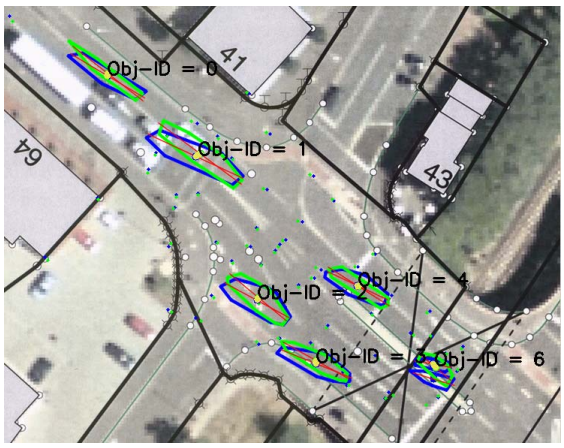

Figure 4. Results of the fusion algorithm for moving objects (ortho picture). The white circles mark the reference points, for which Gauss Krüger coordinates are available.

To use the results of image processing in the C2X-vehicle a final transformation from the ortho coordinate system to the WGS84 (World Geodetic System 1984) coordinate system was applied, because the coordinates which are determined by a GPS receiver are based on WGS84. This second transformation is based on artificial neural networks as described above. An equidistant grid with a point distance of three meters was placed over the ortho picture of the crossing and the correspondent WGS84 coordinates were calculated from the known Gauss-Krüger coordinates. With these pairs of coordinates an additional multilayer perceptron was trained. A mean displacement error and a standard deviation both of about 0.2m were achieved by this method.

Tracking

In order to track the detected objects the following Euclidean distances are calculated:

- The distance of a tracked object's center of mass in the old and in the new picture
- The distance between a coordinate that results from the old center of mass plus its mean measured displacement vector and the object's center of mass in the new picture
- The distance between the new position estimated by a Kalman filter and the position of the center of mass in the new picture

These three distances are summed up to an overall distance giving the last two distances double weight

to adapt their influence. The overall distance is used to assign the newly detected objects to already tracked objects. Newly detected objects that haven't been assigned to an already tracked object yet get added to the list of tracked objects as a new track. To avoid that a quick change of illumination or a reflection can create foreground objects that will get tracked, objects have been detected and assigned to tracked objects several times consecutively before they are valid for communication with the car. Results of the tracking algorithm are classification statistics, object IDs and displacement vectors for the tracked objects. In later steps the displacement vector is used to determine the object heading in the display of the C2X-vehicle. Figure 5 illustrates the tracking result for a pedestrian, figures 6 to 8 show the tracking of vehicles in both camera pictures and in the orthopicture of the crossing.

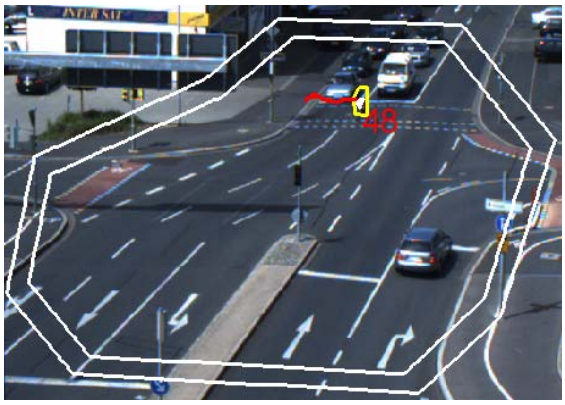

Figure 5. Tracking of a pedestrian visualised in the picture of the second camera.

Objects that are tracked obtain an object ID. Additionally the displacement vector of each tracked road user is displayed in the colour white, the trajectory is displayed in red. When the tracking algorithm is activated the final classification results from the fusion of both acquired images and the classification history collected in the previous tracking steps. In each cycle the tracking algorithm counts how often the road users were assigned to the class pedestrian and to the class vehicle. The class label the object received at most will be the resulting class label of the tracking algorithm. Table 2 shows the confusion matrices of this second classification step for both cameras. The historical tracking data improves the classification results. With enabled tracking sensitivities for pedestrians and vehicles are increased to 0.95. In addition the results of classification are visualised by the colour of the outline of the tracked objects. Pedestrians obtain a yellow (Figure 5), vehicles a green contour (Figure 6 and 7).

Table 2.
Confusion matrices for the two cameras with enabled tracking

Confusionmatrix Camera 1		Real Class	
		Pedestrian	Vehicle
Determined Class	Pedestrian'	55	29
	Vehicle'	0	606

Confusionmatrix Camera 2		Real Class	
		Pedestrian	Vehicle
Determined Class	Pedestrian'	41	16
	Vehicle'	0	663

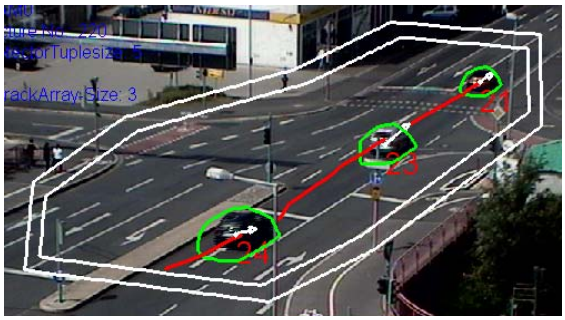

Figure 6. Tracking results displayed in the image of camera 1.

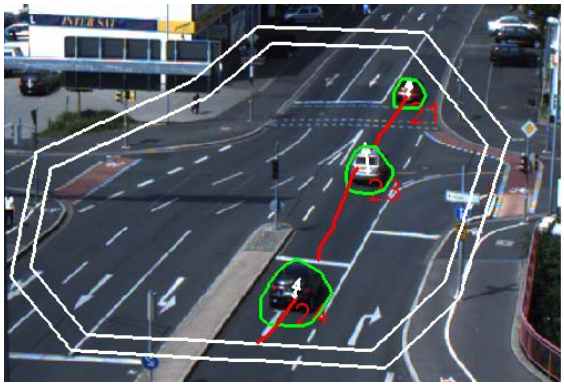

Figure 7. Tracking results displayed in the image of camera 2.

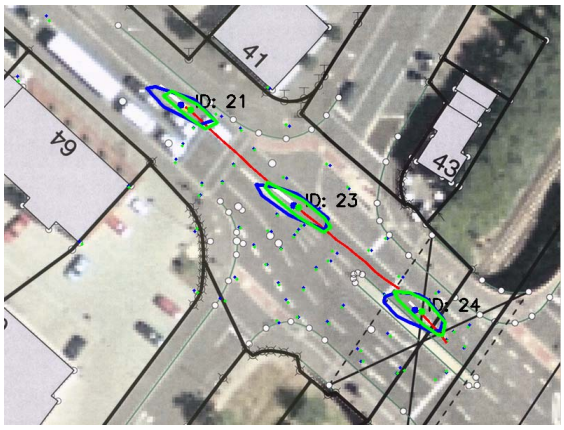

Figure 8. Tracking results displayed in the ortho image.

Communication and visualisation

To transfer the data from the infrastructure to the C2X-vehicle the viilab communication technology based on standard wireless 802.11g technology is used. To exchange the acquired information a SOAP-Server (Simple Object Access Protocol) is implemented in the infrastructure image processing software. From this infrastructure server viilab, which may be used as a simulation environment or an onboard unit, can request information about the objects on the crossing. The SOAP-Server replies to such a request with the WGS84 coordinate and heading, the class and the object ID of all detected objects. If any objects were received from the image processing environment, these objects are visualised in the viilab user interface. To display the received information an environment radar page was created for the vui. Depending on the simulated or real time position and heading of the C2X-vehicle determined by an onboard GPS receiver all detected objects in the surrounding are shown in this environment radar. Figure 10 shows the visualisation page implemented in the viilab user interface.

The communication to viilab was tested with a virtual C2X-vehicle. Figure 9 shows position and heading of such a virtual C2X-vehicle at the observed crossing. Figures 6 to 8 visualise the detection and tracking results in the two camera pictures and the ortho picture of the crossing. Figure 10 finally shows this traffic situation at the crossing in the environment radar of vui with respect to the position of the virtual C2X-vehicle in figure 9.

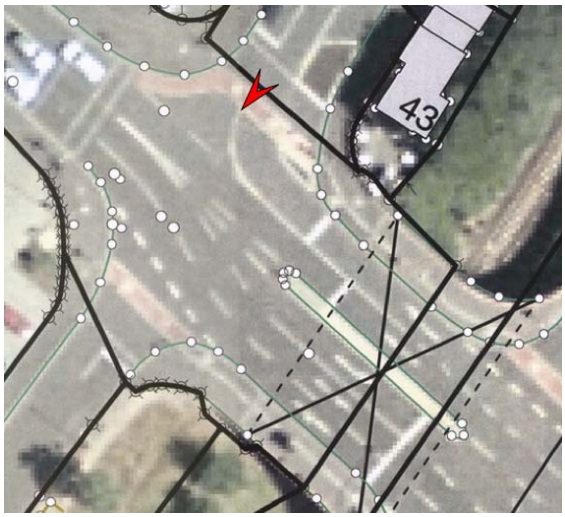

Figure 9. Position and heading of the virtual C2X-vehicle at the crossing.

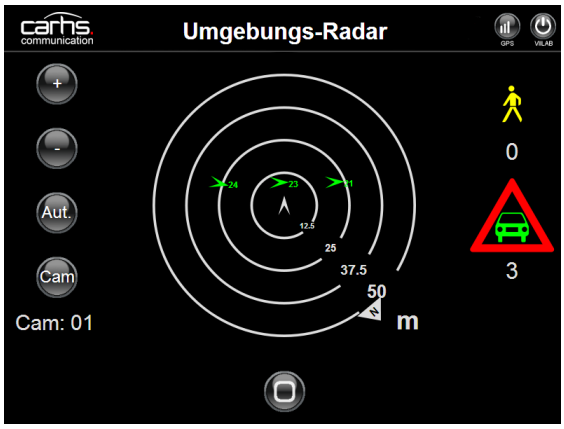

Figure 10. The environment radar visualises the actual traffic at the crossing in front of the virtual car.

CONCLUSIONS

The integration of an infrastructure-based vision system for pedestrian and vehicle detection in viilab, which may be used as simulation tool or as on board unit, and the visualisation to display the acquired data in a C2X-vehicle is described. A two camera approach, a neural net-based camera calibration, a two step classification and a data fusion based on Euclidean metric allow a tracking of road users within a trace of the road in real time even when occlusions occur in a single camera view.

The approach may be adapted to a multi-camera view. To increase the reliability and quality of recognition, classification and positioning the fusion of onboard and offboard sensor data may be considered as a future development target. Due to the fact that cameras for the visible spectral range are used, the current system's time of operation is limited, because good lighting conditions are needed to obtain evaluable images. So in further steps infrared cameras may be added to the system to allow a 24-hour service. Extended training of the first classifier using databases of pedestrians and vehicles as well as integrating additional features, e.g. histograms of oriented gradients, will increase its performance. Another development aim might be the implementation of intelligent assistance into the software of the on board unit. By coupling this infrastructure based technology with C2X-technology benefits for C2X-vehicles in the introduction phase of the new C2X-technology on the market can be expected. Vehicles and pedestrians without communication technology can be detected, localised and tracked on intersections, vehicles with communication technology can receive their positions and use it for intersection assistance.

REFERENCES

[1] Fuerstenberg, K., Hopstock, M., Obojski, A., Rössler, B., Chen, J., Deutschle, S., Benson, Ch.,Weingart, J. and de Lara, A. 2007. INTERSAFE a PReVENT Project, 40.75 Final Report (Project Evaluation and Effectiveness of the Intersection Safety System)

[2] Schaufelberger, W., Gadient, F., Lingwood, S. and Rüthemann, D. 2006. Systeme für die automatische Verkehrsüberwachung (Monitoring) mit digitaler Bildverarbeitung, Forschungsauftrag VSS 1999/265 auf Antrag des Schweizerischen Verbandes der Strassen- und Verkehrsfachleute (VSS)

[3] Dalaff, C., Reulke, R., Ruhé, M., Schischmanow, A. and Schlotzhauer, G. 2003. OIS - An optical information system for road traffic measurement and management, Proc. of the ISPRS Commission IV Joint Workshop "Challenges in Geospatial Analysis, Integration and Visualization II", pp. 209 – 213

[4] Aycard, O., Spalanzani, A., Yguel, M., Burlet, J., Du Lac, N., De La Fortelle, A., Fraichard, T., Ghorayeb, H., Kais, M., Laugier, C., Laurgeau, C., Michel, G., Raulo, D. and Steux, B. 2006. PUVAME - New French Approach for Vulnerable Road Users Safety, Proc. of the IEEE Intelligent Vehicle Symp

[5] Aycard, O., Spalanzani, A., Burlet, J., Fulgenzi, Ch., Dung Vu, T., Raulo, D. and Yguel, M. 2006. Pedestrians Tracking Using Offboard Cameras Proc. of the IEEE-RSJ Int. Conf. on Intelligent Robots and Systems

[6] Mauthofer, A. and Glaab, M. 2008. A new Simulation Method for the Rapid Development of Car-to-X Communication Applications, Internet of Things 2008, 1st International Workshop on Interoperable Vehicles (IOV2008), Zurich, Switzerland

[7] Mauthofer, A., Zang, M. and Glaab, M. 2008. Extended Reality – A new simulation method for the rapid development of car2x communication applications, 15th World Congress on Intelligent Transport Systems ITS, New York, USA

[8] Douglas, D. H. and Peuker, T. K. 1973. Algorithms for the reduction of the number of points required to represent a digitized lineor its caricature, The Canadian Cartographer, Vol 10, No. 2, pp. 112 – 122