A Monocular Image Based Intersection Assistant

Felix Woelk, Stefan Gehrig and Reinhard Koch

***Abstract*— Since cross traffic in inner city intersections is potentially dangerous, its early detection is desirable. In this paper an image-based driver-assistant system is presented. The investigation of optical flow, computed from an image sequence, recorded by a single pan-tilt camera, is used as basis to this system. Its goal is to direct the driver's attention to moving objects in inner city intersections.**

I. Introduction

28% of all accidents in Germany happen in intersection situations, therefore a system assisting the driver in cross traffic situations would be a major achievement in traffic safety. A straight forward and cheap approach to inspect the car environment is to use visual sensors. The observation of complex car environments, e.g. road intersections, with camera sensors, however, results in lots of technical problems. The sensor must be able to focus and track moving objects like pedestrians and cars. Using the human head as an inspiration, pan-tilt cameras are used for this task. Especially while looking into an intersecting road, small and light weight sensors are necessary to enable fast camera movements. Because of this fast movement, heavy stereo rigs are not suitable and the use of a small monocular camera systems is necessary. In the absence of stereo information, alternative algorithms have to be used to investigate the car environment. This work focuses on the optical flow as measured in image sequences recorded by a pan-tilt camera as an appropriate measure for the car environment.

After reviewing some related work, the concept of the optical flow is introduced and the algorithm for the calculation of the optical flow used in this paper is explained. A criterion for detection of independently moving objects by inspecting the optical flow field is derived and degenerative cases are shown. Finally a collision detection procedure is suggested and experiments with synthetic and real data are shown.

System Overview

The demonstrator from the DaimlerChrysler AG, which is used in this work, is called Urban Traffic Assistant (UTA) (fig. 1). The setup of UTA [9] includes a digital camera mounted on a pan-tilt-unit (PTU) (fig. 2), GPS, map data, internal velocity and yawrate sensors, etc. The fusion of GPS and map data will be used to announce the geometry of an approaching intersection to the vision system. The camera then focuses on the intersection. Using the known egomotion of the camera, independently moving objects are detected and the driver's attention can be directed towards them.

Felix Woelk is with the Institute for Computer Science and Applied Mathematic, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany, woelk@mip.informatik.uni-kiel.de

Dr. Stefan Gehrig is with DaimlerChrysler AG, HPC T 728, 70546 Stuttgart, Germany, stefan.gehrig@dcx.com

Prof. Dr. Reinhard Koch is with the Institute for Computer Science and Applied Mathematic, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany, rk@mip.informatik.uni-kiel.de

II. Related Work

Investigating projections of motion fields for the detection of independently moving objects with a moving camera has been subject to many research efforts.

Early work placed restrictions on the allowed camera movement [4], [19], [6], [16], [3], or classified the flow fields into a small number of basic camera movements [14]. Since these restrictions cannot always be fulfilled in practical applications, statistical methods were combined with dimension reduction to reconstruct the egomotion from projected flow fields [7].

Another class of research focuses on the detection of independently moving objects using sparse point correspondences. Some methods require specific camera models [5], others make assumptions about the motion of the object [10].

While this work focuses on monocular image sequences there has also been research using stereo camera systems [1], [17].

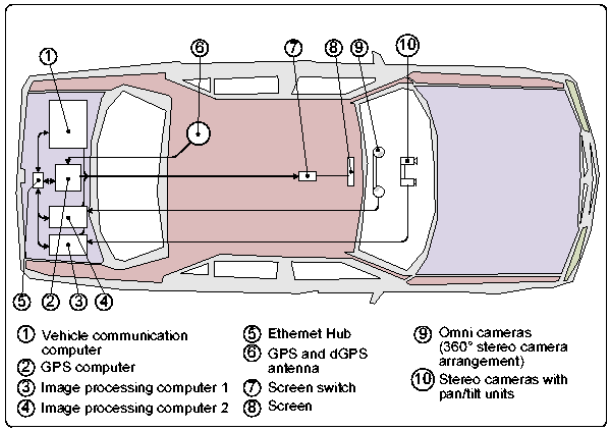

Fig. 1. System architecture of the UTA demonstrator from the Daimler-Chrysler AG.

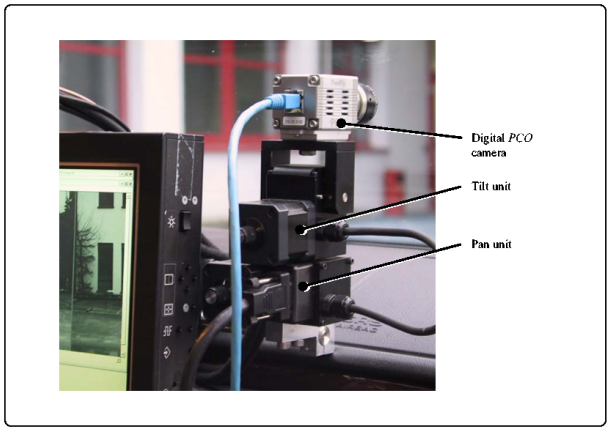

Fig. 2. Digital camera mounted on pan-tilt-unit (PTU) in the UTA demonstrator.

III. Optical Flow

The basic concepts used in the proposed algorithm are introduced in this section. Since visual sensors are used, the concepts relate to capturing and processing of visual data. First the pinhole camera model is described. The homography is introduced as a possibility to correct camera rotation. The fundamental matrix, modeling the relationship between two views of a scene, is described. Finally the notion and calculation of the optical flow field, as used in this work, are given.

Camera Model

A general pinhole camera model is assumed. It can be described by the projection matrix P. P is composed of the calibration matrix K, the 3×3 identity matrix I, the rotation matrix R and the projection center C:

$$x \cong P \cdot X = KR^T[I|-C]\cdot X. \tag{1}$$

A homogenous 3D point X is projected by the 3×4 projection matrix P into the image at coordinates x. $\cong$ denotes projective equality [11]. The rotation matrix R is given by the rotation between the camera and the world coordinate reference frame. The calibration matrix K for digital cameras is given by:

$$K = \begin{pmatrix} f & s & c_x \\ 0 & a\cdot f & c_y \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

Where f is focal length in units of pixel, a is the aspect ration of a single cell on the CCD chip, s is the skew s and $(c_x, c_y)^T$ the principal point of the camera. On modern cameras the skew s can be assumed to be zero and the aspect ratio a can be assumed to be one.

Homography

A general homography H is a plane to plane mapping. The special case of the homography induced by the plane at infinity H_∞ describes the transformation of an image under pure camera rotation:

$$H_\infty = KR_2^T R_1 K^{-1} \tag{3}$$

where K is the calibration matrix of the camera and R_i are the first resp. the second camera orientation. The result of applying H_∞ to an image is a rotation corrected image. It is basically the same image (modulo the borders) as it would have been shot with a camera with different orientation.

Fundamental Matrix

The fundamental matrix F describes the relationship between two views of a static scene. Any point p in the first view is constrained to lie on the epipolar line l in the second view. The fundamental matrix F relates x to l via:

$$l = Fx \tag{4}$$

This constraint can also be expressed for an image point correspondence x_1 and x_2 through :

$$x_2^T F_{12} x_1 = 0. \tag{5}$$

Where F_{12} is the fundamental matrix between the two views 1 and 2.

An important property of all epipolar lines l is, that they intersect in the epipole e itself. The epipole e is the projection of the center of the first camera into the second camera. With pure translational camera movement the epipole coincides with the focus of expansion (FOE) or focus of contraction (FOC), depending on the direction of the camera motion. For a thorough discussion on multiple view geometry see [11].

Optical Flow

Optical flow can be loosely described as the apparent motion of a 2D pattern between two images in a sequence. In a more mathematical sense it is the projection of the 3D motion field into the image plane and thus results from a relative 3D motion between the camera and the recorded scene.

Any arbitrary relative 3D motion between a camera and an object can be described by a translation of the camera and a following rotation of the camera around its projection center (change of the frame of reference or Carlsson-Weinshall duality [11]). Any motion field M can thus be composed from a translational component $M_T(X) = C_2 - C_1$ and a rotational component $M_R(X)$, where C_i is the position of the projection center of the camera at time i. These two parts of the motion field fulfill the constraints $\nabla \times M_T(X) = 0$ and $\nabla \cdot M_R(X) = 0$, i.e. they are source free resp. curl free vector fields. The projection of these two components into the image plane reveals interesting properties. The projection of the rotational motion field $F_R(x)$ is independent from the observed scene and can thus be predicted, given a known camera rotation. The projection of the translational component of the motion field $F_T(x)$ results in a circular flow field, where all flow vectors lie on

a pencil of lines intersecting in the FOE resp. FOC. The length of the translational component of the flow vectors depends on the scene structure. The superposition of $F_R(x)$ and $F_T(x)$ is the observed optical flow $F(x)$:

$$F(x) = F_R(x) + F_T(x) \tag{6}$$

A decomposition of a theoretical optical flow field in its rotational and its translational component is shown in fig. 3.

Computation of the Optical Flow

A large number of algorithms for the computation of optical flow exist [2]. Any of these algorithms calculating the full 2D optical flow can be used for the proposed algorithm. Algorithms only calculating the normal flow (i.e. the flow component component parallel to the image gradient) are, however, inappropriate. The optical flow in this work is calculated using an iterative gradient descend algorithm [12]. It is based on the image brightness constancy equation, stating that the image brightness I is constant over the time t:

$$\frac{dI(x,t)}{dt} = 0 \tag{7}$$

For spatial linear intensity functions I the following holds:

$$(\nabla I)^T F(x) + \frac{\partial I}{\partial t} = 0 \tag{8}$$

Where $I = I(x,t)$ denotes the image intensity, $F(x)$ denotes the flow field at x and t is the time. A linear system for all points in a support window around x is set up:

$$AF(x) + b = 0 \tag{9}$$

With $A = (\nabla I(x_1), \nabla I(x_2), \ldots)^T$ and $b = (I_t(x_1), I_t(x_2), \ldots)^T$. $x_1, x_2, \ldots$ are the image points in the support window and I_t is the partial derivative of the image intensity with respect to the time t. A solution in a least square sense to eq. 9 is given by

$$F(x) = (A^T A)^{-1} \cdot A^T b, \tag{10}$$

if the 2×2 matrix $A^T A$ is invertible. Note that the invertibility of $A^T A$ is correlated with the presence of structure in the support window. Since the solution F(x) is only an approximation for spatial linear image intensities, equation 9 is set up and solved iteratively until the improvement $F(x)$ falls below a given threshold.

For robustness and stability the algorithm is applied to subsequent images of a Gaussian pyramid. The resulting flow of the smaller pyramid image is used as a starting point in the next bigger pyramid image.

In order to deal with single measurement outliers a 5×5 median filter adopted for sparse data is applied to the resulting flow field.

IV. Detecting Moving Objects

The notion of optical flow can be applied independently to any relative motion between a camera and an observed object. Assuming a rigid and static scene and one or more rigid objects moving in the scene, observed by a moving camera, it is therefore feasible to apply the notion of optical flow independently to both the observed scene and the observed objects.

Flow fields induced by pure translation have a simple geometric property. All flow vectors are located on lines intersecting in the FOE, pointing away from the FOE (for an example see fig. 3(d), here the FOE is located in the image center). The length of a single flow vector depends on the observed scene depth and cannot be predicted without knowledge about the recorded scene. In a pure translational camera movement, any flow vector not fulfilling these geometric constraints (i.e. coincidence with the predicted direction) must be located on an independently moving object. In practice a threshold t over the angle between the flow vector and the line given by the image point and the FOE is used as a criterion for the detection of idependently moving objects. The threshold t itself is dependent of the flow length. In real environments the camera movement is by no means restricted to translational movement only. It is therefore required to deal with camera rotation. In our approach, the camera rotation can be derived from known rotation sensor data of the car and can be measured directly from the static scene [15].

Rotation Correction

The optical flow field is composed of a component induced by a known rotation and component induced by the translation with known FOE. If the egomotion of the camera is known (e.g. by internal sensor data), then the rotational component of the flow field F_R, which is independent of the observed scene, can be compensated resulting in the translational component of the flow field F_T. This is possible by simple subtraction (see fig. 4) . Since the rotational component of the flow field is independent of the scene structure, it can be calculated analytically with the use of the homography H,

$$F_R(x) = H \cdot x - x = K R_2^T R_1 K^{-1} \cdot x - x, \tag{11}$$

and the translational component of the flow field can then be recovered via:

$$F_T(x) = F(x) - F_R(x) \tag{12}$$

Degenerate Cases

Since every motion has its own FOE, an additional FOE is present for every motion between the camera and an idependently moving object. Three cases exist where a detection of an independently moving object fails:

- If the FOE of the camera motion and the FOE of the relative motion between the camera and the moving

object coincide, the predicted direction of the translational flow fields are the same and hence the object is not detectable as independently moving. A scenario for this setup is a frontal collision trajectory between the camera and the moving object, another scenario would be any motion parallel to the camera with velocity less than the camera velocity.

- Collinearity: If the object and the two FOEs are collinear and the object is not located in between the two FOEs, the predicted direction of the translational flow fields are the same and hence the object is not detectable as independently moving.
- If there is no relative motion between the camera and the moving object, there is no optical flow and it is hence not detectable as independently moving by means of investigating the optical flow. An obvious scenario is an object moving with same velocity parallel to camera path.

Collision Detection

Given a number of flow vectors belonging to a moving object, a collision detection as known for centuries as the sailor's test for collision (see fig. 5) is possible. If the angle α, under which an object B is seen from an object A, remains constant over time ($\alpha = \alpha'$) and the apparant object size is growing, then a collision will take place. This is equivalent to the fact that the FOE of the relative motion between the camera and the moving object lies within the growing picture of an object in the image.

V. Experimental Results

Experiments were carried out using synthetic images and sensor information as well as images and sensor data gathered with the Urban Traffic Assistant (UTA) demonstrator from the DaimlerChrysler AG [9].

Simulated Data

To test the algorithm a simulated intersection was realized in VRML. Simple block models of houses, textured with real image data, are located on the corners of the intersecting street (fig. 6). A model of a car was used as an independently moving object. Screenshots from a ride through this intersection provided the image data, while the sensor information was calculated from the known camera parameters at the times of the screenshots. Fig. 6 shows some images from the simulated image sequence. Points violating the epipolar constraint are marked with white blobs.

Real Data

The setup of UTA [9] includes a digital camera, mounted on a pan-tilt-unit (PTU) (fig. 2), GPS, map data, internal velocity and yawrate sensors, ... (fig. 1). The fusion of GPS and map data will in future versions be used to predict the geometry of an approaching intersection to the vision system. The camera then focuses on the intersection. The

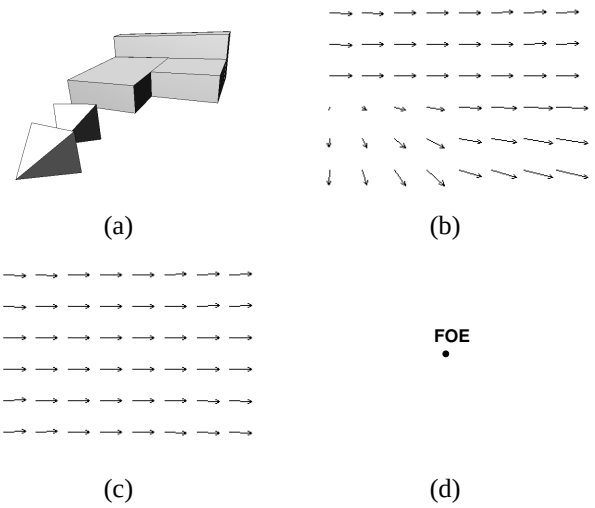

Fig. 3. Theoretical flow fields for a simple scene. The 3D scene is shown at (a). The scene consists of 3 blocks. The camera, displayed as small pyramids, translates towards the blocks while rotating around the y axis. The flow field F as induced by this movement is shown in (b). Its rotational component F_R (c) and translational component F_T (d) with the Focus of Expansion (FOE) are shown in the lower row. Note that no prediction of the flow field, even with known camera egomotion, is possible, if the structure of the observed scene is unknown.

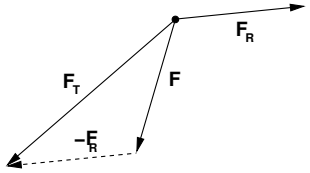

Fig. 4. Rotation compensation of a single flow vector. F is the measured flow, F_R is the rotational component as calculated from the known rotation, and F_T is its translational component.

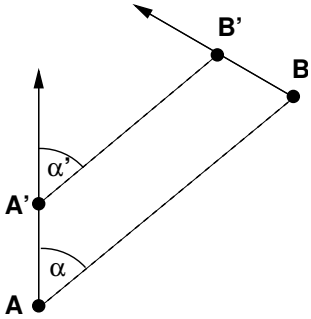

Fig. 5. Sailor's test for collision: If the angle α, under which an object B is seen from an object A, moving itself, remains constant over time ($\alpha = \alpha'$), a collision will take place.

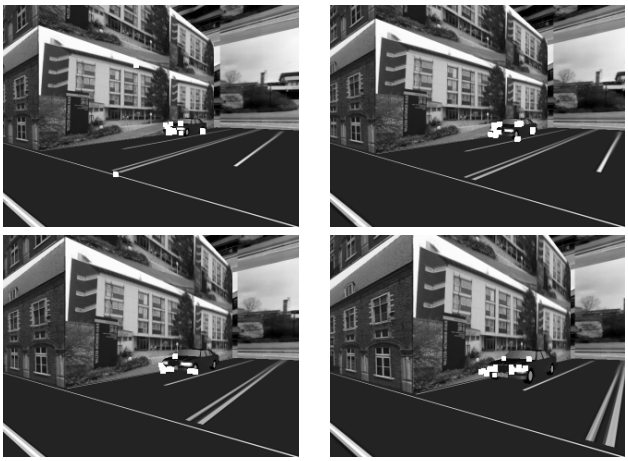

Fig. 6. Some images from the synthetic intersection sequence. The camera is moving on on a straight line, while the car in the image is on a collision course. Moving objects are marked with white points.

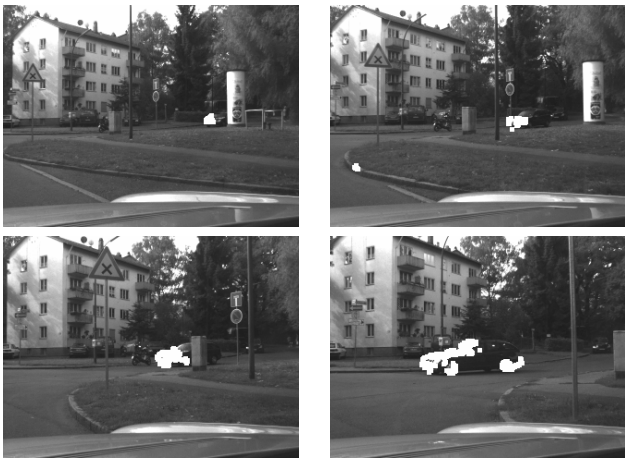

Fig. 7. Some images from a real intersection sequence. Moving objects are marked with white points.

egomotion of the camera is calculated by fusing PTU and inertial car sensors. The result of applying the algorithm to real data is shown in fig. 7.

A single threshold for the violation of the epipolar constraint by an optical flow vector, and thus its location on a moving object exists. Therefore every flow vector with an angle above this threshold results in the detection of an independently moving object, even if this flow measurement is erroneous. For example in the upper right image in fig. 7 there is an outlier located on the curbstone.

Furthermore the quality of the detection of independently moving objects depends on the quality of the FOE position. For the FOE computation the internal speed and yawrate sensors are used to compute the camera position for every frame. The epipole is computed by projecting the center of the first camera into the second. After rotation correction (as explained in chapter IV) the epipole coincides with the FOE. Since speed data is low pass filtered by the sensors and the FOE is calculated using only speed and yawrate, the resulting FOE is not precise when the car is accelerating or rotating. Fig. 8 shows the false positive and false negative rate of the detector on a real world sequence. In the beginning of the sequence (up to frame 50) the car is accelerating while pulling out of the parking lot. This results in imprecise FOE calculation and therefor erroneous detections of independently moving objects. In the end of the sequence (from frame 160) the car is breaking while turning right into the intersecting road. This also results in increased false positive and false negative rates. The image segmentation needed for obtaining the false positive and false negative rates was generated manually. The moving object was hereby approximated by several rectangles.

Timing

TABLE I

MEAN COMPUTATION TIME (*ms*) AND STANDARD DEVIATION OF THE OPTICAL FLOW PER FRAME. THE DEPENDENCY OF THE MEAN OPTICAL FLOW CALCULATION TIME FROM THE WINDOWS SIZE WS (COLUMNS) AND THE EXIT CRITERION ERR (ROWS) IS SHOWN. THEY WERE COMPUTED USING A REAL IMAGE SEQUENCE OF 300 FRAMES WITH IMAGE SIZE 320×240. OPTICAL FLOW WAS CALCULATED FOR 2093.3 ± 333.1 PIXEL PER FRAME.

Err \ WS	0.5	0.1	0.05	0.01
3	15.2±5.4	22.7±5.3	26.8±6.4	35.3±6.5
5	23.1±5.5	32.3±7.4	38.3±8.2	55.0±9.0
7	37.1±8.1	49.7±11.4	60.0±12.5	88.6±14.6
9	55.3±10.8	73.8±16.5	88.7±19.1	132.3±22.2
11	80.8±15.6	107.9±26.4	129.7±30.5	194.2±37.1

Approximately 55 *ms* of the computation time is used for the rotation correction and detection of independently moving objects, while the remaining time is spent estimating the optical flow. The computational impact of the optical

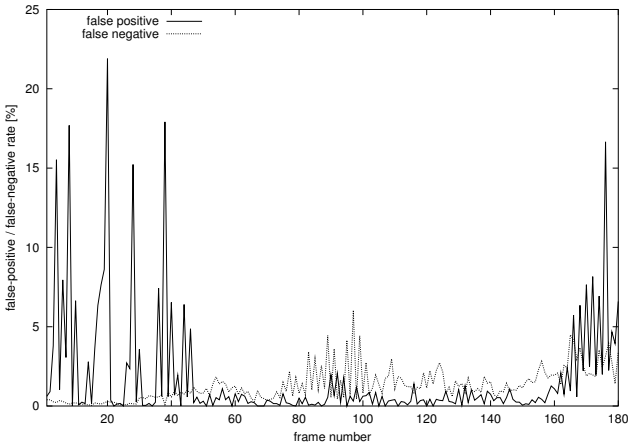

Fig. 8. False positive and false negative rate for a real world sequence. The ground truth image segmentation needed for obtaining these rates was generated by hand. The moving object was hereby approximated by several rectangulars.

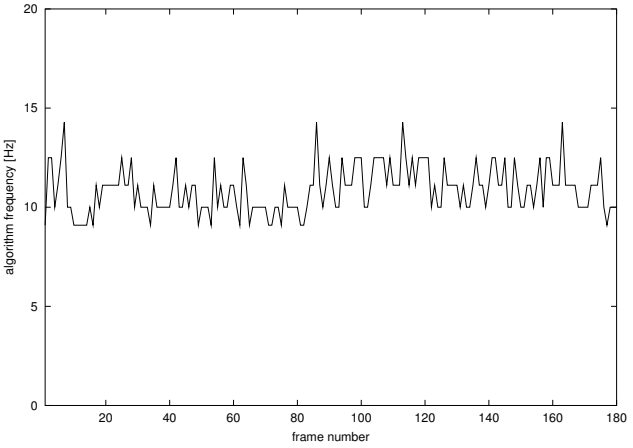

Fig. 9. Calculation frequency for a real world sequence of 180 frames. The overall computation frequency is plotted over the frame number. A support window size of 5×5 and an exit criterion of 0.1 pixel were used. The computation runs at an average of 10.8 ± 0.1 frames per second.

flow algorithm depends heavily on the window size and the exit condition of the iteration (tab. I). The mean overall computation time of the presented algorithm on real image data on a standard 2.4 GHz Pentium IV PC is given in tab. II. The computation frequency in frames per second (fps) is shown in fig. 9 for a real world sequence.

TABLE II

MEAN OVERALL COMPUTATIONAL TIME (*ms*) WITH STANDARD DEVIATION OF THE PROPOSED ALGORITHM. THE DEPENDENCY OF THE OPTICAL FLOW WINDOW SIZE WS (COLUMNS) AND THE EXIT CRITERION ERR (ROWS) IS SHOWN. THE TIMINGS WERE OBTAINED UNDER THE SAME CONDITIONS AS IN TAB. I.

WS \ Err	0.5	0.1	0.05	0.01
3	69.7±7.7	76.6±8.2	79.9±7.8	84.8±8.9
5	79.1±7.8	87.8±9.5	94.3±9.5	110.2±11.2
7	92.9±9.7	106.5±12.9	116.4±13.2	144.3±16.2
9	110.7±11.9	130.0±16.9	144.2±18.9	187.8±22.9
11	137.4±16.8	164.7±27.0	185.8±31.2	249.9±38.1

VI. CONCLUSIONS AND FURTHER WORK

A detection system for moving objects based on the violation of the epipolar constraint by the optical flow was presented. It is usable on any camera motion, given knowledge about the movement. Degenerated cases where no detection is possible were derived. Experiments with real and simulated images and sensor data were made.

Future work includes:

- improvement of egomotion estimation from fusion of car sensor data and visual camera pose tracking from optical flow in static regions [15],
- use of an adaptive threshold for the detection of independently moving objects,
- grouping of flow from moving objects,
- collision warning,
- extraction of the trajectories of the moving objects.

A. Acknowledgments

This work was partly funded by DaimlerChrysler AG and partly by the BMBF INVENT project. See http://www.invent-online.de for details.

REFERENCES

[1] A. Argyros and S. Orphanoudakis “Independent 3D Motion Detection Based on Depth Elimination in Normal Flow Fields.” *Proc. CVPR*, pp. 672-677, 1997.

[2] J.L. Barron, D.J. Fleet, S.S. Beauchemin and T.A. Burkitt, “Performance Of Optical Flow Techniques”, *Proc. CVPR*, Vol. 92, pp. 236-242, 1994.

[3] S. Carlsson and J. Eklundh, “Object Detection Using Model Based Prediction and Motion Parallax.” *Proc. ECCV*, pp. 297-306, 1990.

[4] J. Clarke and A. Zisserman “Detection and Tracking of Independent Motion.” *IVC*, Vol. 14, pp. 565-572, 1996.

[5] J.P. Costeira and T. Kanade “A Multibody Factorization Method for Independently Moving Objects.” *IJCV*, 29(3), pp.159-179, 1998.

[6] W. Enkelmann, “Obstacle Detection by Evaluation of Optical Flow Fields.” *Proc. ECCV*, pp. 134-138, 1990.

[7] S. Fejes and L.S. Davis “What can projections of flow fields tell us about the visual motion.” *Proc. ICCV*, pp. 979-986, Bombay, India, 1998.

[8] C. Fermüller and Y. Aloimonos, “Direct Perception of Three-Dimensional Motion through Patterns of Visual Motion.” *Science*, 270, pp. 1973-1976, Dec. 1995.

[9] S. Gehrig, S.Wagner and U. Franke, “System Architecture for an Intersection Assistant Fusing Image, Map and GPS Information” *Proc. IV*, 2003.

[10] M. Han and T. Kanade, “Multiple Motion Scene Reconstruction with Uncalibrated Cameras.” *TPAMI*, 25(7), pp.884-894, 2003.

[11] R. Hartley and A. Zisserman, *Multiple View Geometry*, Cambridge University Press, 2000.

[12] B. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision” *Proc. DARPA IU Workshop*, pp. 121-130, 1981.

[13] M. Macheline, L. Zelnik-Manor and M. Irani, “Multi-Body Segmentation: Revisiting Motion Consistency.” *Proc. ECCV*, Workshop on Vision and Modeling of Dynamic Scenes, 2002.

[14] R.Nelson “Qualitative Detection of Motion by a Moving Observer” *IJCV*, 7(1):33-46, 1991.

[15] M. Pollefeys, R. Koch and Luc J. Van Gool, “Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters”, *IJCV*, 32(1):7-25, 1999.

[16] D. Sinclair and B. Boufama, “Independent Motion Segmentation and Collision Prediction for Road Vehicles.” *Proc. ECCV*, 1994.

[17] P. Sturm, “Structure an Motion for Dynamic Scenes - The Case of Points Moving in Planes.” *Proc. ECCV*, Vol. 2, pp. 867-882, 2002.

[18] C. Tomasi and T. Kanade, “Shape and Motion from Image Streams: A Factorization Method.” *TR 92-1270, CMU-CS-92-104*, 1992.

[19] P. Torr and D. Murray, “Statistical Detection of Independent Movement from a Moving Camera.” *ICV*, 11(4), pp. 180-187, 1993.