ADOSE – New bio-inspired in-vehicle sensor technology for active safety

J. Kogler, Ch. Sulzbachner, E. Schoitsch, W. Kubinger, M. Litzenberger, AIT Austrian Institute of Technology GmbH

Abstract

Reliable "Advanced Driver Assistance Systems" (ADAS) are intended to assist the driver under various traffic, weather and other environment conditions. The growing traffic requires sensors and systems which handle difficult urban and non-urban scenarios. For such systems new cost-efficient sensor technologies are developed and evaluated in the EU-FP7 project "ADOSE - (reliable **A**pplication-specific **D**etection of road users with vehicle **O**n-board **SE**nsors)", (funded under grant agreement n° 216049) providing the vehicle with a "virtual safety belt" by addressing complementary safety functions.

1 Introduction

The EU-funded project ADOSE is evaluating new sensor technologies and sensor systems. Such sensors are necessary for ADAS like lane departure warning, collision warning, high-beam assist or side impact detection. Fig. 1 illustrates the various sensors from different project members and additionally, the operating distance of each sensor is depicted.

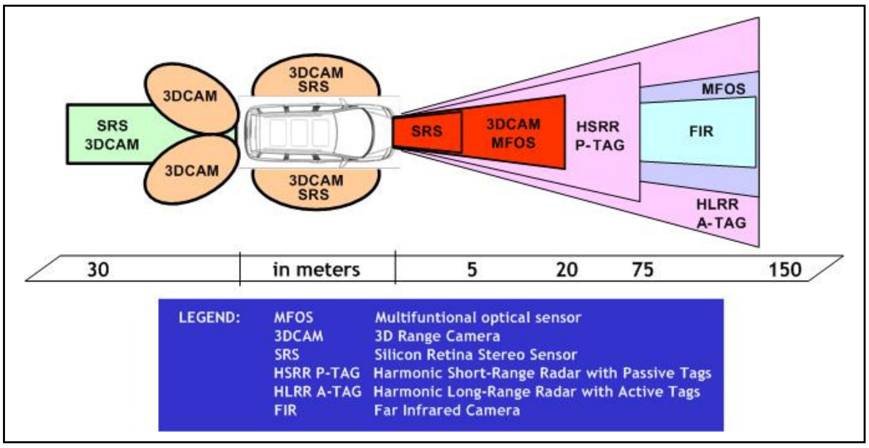

Fig. 1. Overview of all sensor technologies evaluated and considered in ADOSE

The approach of the Austrian Institute of Technology (AIT) towards reduction of the ADAS costs is to use a "Silicon Retina Stereo Sensor" (SRS). The SRS is specifically tailored to serve as a pre-crash warning and preparation sensor for side impacts. Pre-crash applications must reliably react in real time to prepare the vehicle (e.g. activate the pretensioner, preparation of a side airbag) for the imminent impact (which, in case of side impact, cannot be avoided by a reasonable reaction of the impacted vehicle). For the pre-crash sensor, it is necessary to take distance measurements of objects approaching the sensor. Two silicon retinas have therefore been coupled to a stereo vision unit, allowing distance information to be extracted from moving objects in the viewed scenery.

2 Silicon Retina Technology

Derived from the human vision system, the bio-inspired silicon retina sensor is a new type of imager. Conventional optical sensors capture images at a fixed frame-rate. A silicon retina optical sensor provides only timed event-triggered information, which means the sensor delivers information about the illumination changes ('events') in the visual field. The sensor detects intensity changes in positive (ON-event) and negative (OFF-event) direction in an observed scene, with each pixel delivering its address and event data separately and independently. The so-called "address event representation" (AER) was proposed in 1991 by Sivilotti [1] for transferring the state of an array of neurons from one chip to another. An early implementation of an artificial retina has been carried out by Fukushima et. al [2] in 1970 and the first retina imager on silicon basis is described in the work from Mead and Mahowald [3], which have also established the term "Silicon Retina".

This type of sensor is intended to overcome certain obstacles in classical vision systems, which are depicted in Fig. 2.

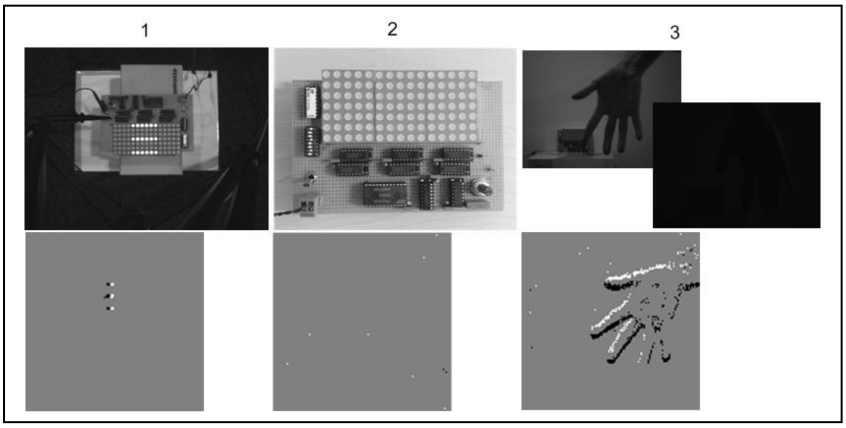

Fig. 2. Advantages of the silicon retina sensor technology, (1) high temporal resolution, (2) data transmission efficiency, (3) wide dynamic range

First, the high temporal resolution allows quick reactions to fast motion in the visual field. Due to the low resolution and the asynchronous transmission of address events (AEs) from pixels where an intensity change has been occurred, a temporal resolution up to 1ms is reached. In Fig. 2 (1) the speed of a silicon retina imager compared to a monochrome camera (Basler A601f@60fps) is shown. The top image in column (1) shows a running LED pattern with a frequency of 450Hz. The silicon retina can capture the LED hopping sequence, but the monochrome camera can not capture the fast moving pattern and therefore, more than one LED column is visible in a single image.

The second advantage is the on-sensor pre-processing because it reduces significantly both memory requirements and processing power. In Fig. 2 (2) the efficiency of the transmission is illustrated. The monochrome camera at top in the column (2) has no new information over time, but the unchanged image must be transferred after an image has been captured. In case of silicon retina imagers underneath no information has to be transferred with exception of a few noise events are visible in the field of view which must be transferred.

The third benefit of the silicon retina is the wide dynamic range up to 120dB, which helps to handle difficult lighting situations, encountered in real-world traffic and is demonstrated in Fig. 2 (3). The top image pair shows a moving hand in an average illuminated room with an illumination of ~1000 lm/m2 and captured with a conventional monochrome camera. The second image of this pair below shows also a moved hand captured with a monochrome camera and an illumination of ~5 lm/m2. In case of the monochrome sensors only the hand in the well illuminated environment is visible, but the silicon retina sensor covers both situations, what is depicted in the image below in Fig.2 (3).

3 Stereo Vision Sensor

The silicon retina (SR) stereo sensor for processing of 3D stereo information uses two silicon retina sensors and Lichtsteiner et. al [4] describes in his work the silicon retina sensor used for the stereo system. Derived from the side impact detection use case defined in the project the stereo vision system must fulfil different requirements. In Fig. 3 the detection area of the camera system is shown. The vehicle (at least 0.5m wide) is approaching the camera system from the side with a maximum speed of 60km/h and the reaction time of the car equipment is assumed to be 350ms. These parameters are responsible for the chosen stereo vision system parameters, which are:

- Detection Range: objects must be detected with high confidence well before activation of countermeasures. Due to the assumption of a system reaction time of 350ms and the maximum speed of 60km/h, a detection range of 6 meters is required. Between 6m and 5m is the main operating distance.

- Field of view (FOV): for the given sensor resolution of 128x128 (pixelpitch 40µm) pixels and a baseline of 0.45m, lenses with a field of view of 30° and a focal length of 8.5mm are chosen. The large baseline is necessary to reach the required depth resolution of three consecutive detections during one meter of movement at a distance of 6 meter.

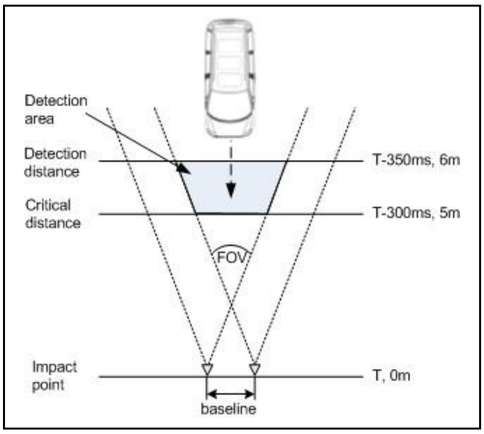

Fig. 3. Overview about the stereo vision system configuration

4 Stereo Vision Algorithm

Conventional area-based stereo algorithms that use monochrome or colour images aim for calculating dense depth images, i.e. providing depth information for every pixel of the input images. The task of stereo vision processing is the matching of corresponding pixels between the left and the right image, the so called "stereo matching". For this problem, a large variety of different approaches can be found in the literature [5]. In Section 4.2 state-of-the-art algorithms for silicon retina cameras are tested. Aside from several pre-processing steps, which are required by some stereo matching approaches, its important for stereo vision algorithms to use calibrated cameras and rectified images which are described in Section 4.1.

4.1 Calibration and Rectification

The acquired data from the cameras are not prepared for line-by-line matching respectively event-by-event matching, because the epipolar lines are not in parallel. Therefore, a rectification of the camera data is carried out which is the work of Schreer [6] described in detail. Before this rectification can be done, the cameras must be calibrated. With conventional cameras the calibration pattern (Fig. 4 on the top) is captured in different views from the right and left camera and the crossing of the pattern are used for the calculation of the camera parameters.

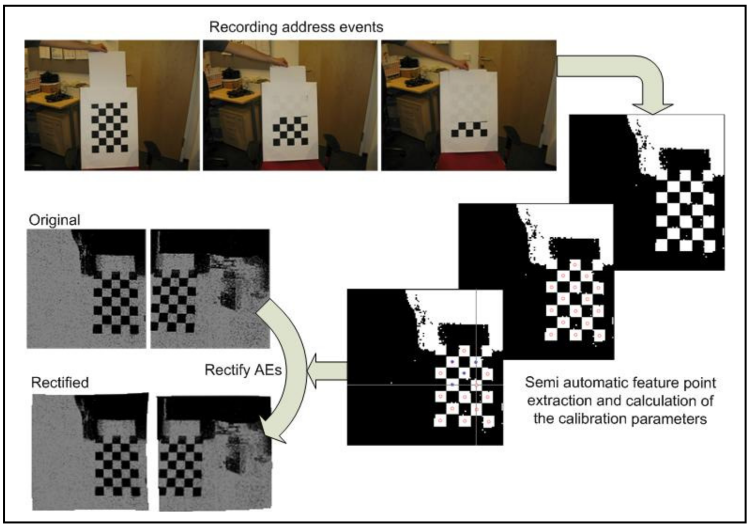

Fig. 4. Calibration and rectification of silicon retina cameras

For silicon retina imagers, it is not possible to capture the calibration pattern if there is no movement more precisely no change in the intensity. In case of silicon retina sensors an alternative approach is necessary. In Fig. 4 on the top the calibration pattern is visible in a stable position and only a white paper is moved up and down in front of the calibration pattern. During this time all address events are collected and written into one output file. The collected address event data are converted into a binary image which is used for the extraction of feature points. Instead of the crossings from the calibration pattern the centres of the squares for extraction of corresponding features are used. The right side in Fig. 4 shows the semi-automatic extraction of the feature points, because not all centres found are supporting the calibration process. For the calibration itself the calibration from Zhang [7] in combination with the calibration toolbox from Caltech for Matlab [8] is used. All data extracted from the binary images are loaded via the external interface into the calibration engine and the results are applied on silicon retina data for the calibration and rectification step. The left side of Fig. 4 shows an example of rectified silicon retina data from the left and right camera.

4.2 Evaluation of state of the art algorithm with silicon retina cameras

In case of silicon retina cameras it is a challenging task to handle the asynchronous incoming address events (AEs) for the stereo matching process. Hess [9] worked with silicon retina data and used a global disparity filter in his work to find a main disparity and combined this information with the outcome of a general disparity filter which evaluated the confidence and possibility of a match.

In our work [13] we evaluate the option to use standard stereo vision algorithms for AE data from silicon retina imagers. There are the area-based approaches, which are described and compared in the work of Scharnstein and Szeliski [10], where we use a "Sum of Absolute Differences" (SAD) algorithm with different window sizes as representative of this category. The second class are feature-based approaches, where in the work from Shi and Tomasi [11] a description of features in general can be found and a more detailed description of an implementation for feature-matching is presented in the work of Tang et. al [12]. For the feature-based approach with silicon retina data a "Centre-of-Gravity" matching is applied. In our work we have compared both state-of-the-art approaches and measured the accuracy of the distance estimation and also the pre-processing step of the data conversion is described in detail. The accuracy of the distance estimation had at least an average relative disparity error of 7% respectively 17%. These results are not satisfying and therefore some improvements took place. In Fig. 5 are new results for both algorithm approaches shown, but with the improvement of calibrated and rectified image data as outlined in Section 4.1. The left side of Fig.5 shows the results of the area-based approach, using 500 image pairs for each result and evaluating three different distances with four different window sizes.

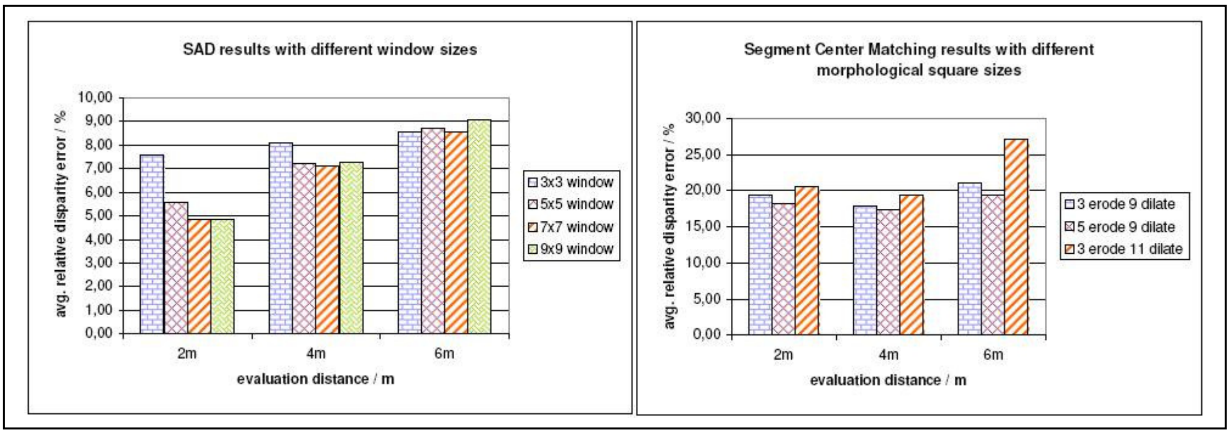

Fig. 5. Results of an area-based (SAD) (left) and of a feature-based (Centre-of-Gravity) (right) algorithm with calibrated and rectified camera data

Due to the used calibrated and rectified data a least an error of 5% can be measured. The right side shows the results of the feature-based algorithm results also in three different distances and with a variety of morphological operations. For each result are 500 image pairs used and the perceivable improvements with the feature-based approach and rectified camera data are minor. The improved results are still not satisfying and therefore a new approach for the stereo matching of silicon retina has to be implemented, where a first outlook of this algorithm is given Section 4.3.

4.3 Stereo algorithm approach specialized for silicon retina data

Conventional block-based and feature-based stereo algorithms have been shown to reduce the advantage of the asynchronous data interface, throttle the performance and do not reach the necessary accuracy for the estimation of the distances. Therefore a novel algorithm approach based on locality and timely correlation of the asynchronous data event streams of both imagers was applied. An overview of the algorithm with all blocks is depicted in Fig. 6.

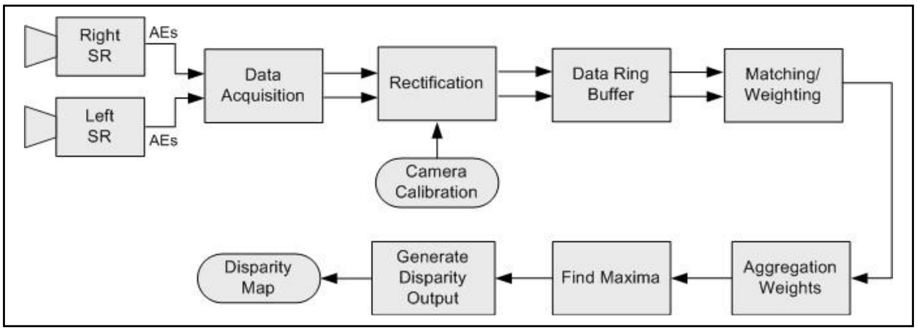

Fig. 6. Blockdiagram of the novel stereo algorithm approach for silicon retina cameras

First, the sensor is calibrated to affect the distortion coefficients and the camera parameters as mentioned in Section 4.1. Then, the events received by the embedded system (described in Section 5) are undistorted and rectified to obtain matchable events lying on parallel and horizontal epipolar lines. The rectified events from the left and right imager are stored in a ring buffer structure to exploit the asynchronous behaviour of the camera. The algorithm reads from the ring buffer and starts processing the stereo data. At first, all events are matched and each matched event gets a weight corresponding to their time difference and is stored into a buffer afterwards. The time correlation and weighting of the matches found is followed by an aggregation, used for the improvement of confidence of the matched events. After the aggregation the search of the maxima starts and the corresponding disparity of the highest weight is written into a result buffer, the so-called disparity map. This approach does not use collecting and image building mechanisms as the approaches described in [13], because the data is directly used from the buffer and matched against the opposite side. Early prototypes showed that by using this algorithm the advantage of the SR technology could be fully exploited at very high processing rates.

5 Embedded System

The embedded system used for the described stereo vision approach to perform data acquisition, and pre-processing is based on a TMS320C6455 single-core fixed-point “Digital Signal Processor” (DSP) from Texas Instruments. Due to the high performance requirements of the stereo vision algorithms, the TMS320C6474 is dedicated to data processing. Both cores are based on a C64x+ DSP core from Texas Instruments. The most significant difference between both DSPs is that the TMS320C6474 consists of three C64x+ DSP cores rather than one. This has a noticeable effect on the Peak MMACS of each DSP. Another difference in terms of data acquisition is that the TMS320C6474 has no adequate parallel interface for connecting the parallel interface of the optical sensor. Fig. 7 depicts the schematics of the embedded system, which compromises two optical sensors that are connected to the adapter-boards to handle the data traffic.

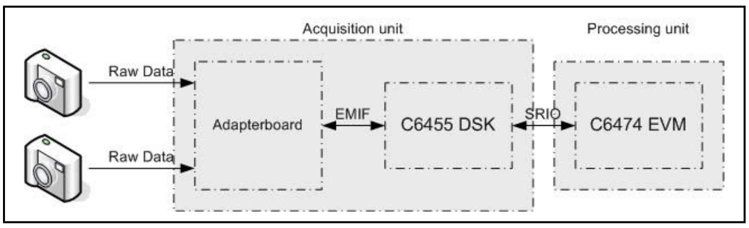

Fig. 7. Overview about the embedded system architecture

The adapter-board is connected to the “External Memory Interface” (EMIF) of the TMS320C6455 DSP starter kit (DSK) for data acquisition. After the acquisition and pre-processing, the data is output over Serial RapidIO™ to the TMS320C6474 evaluation module (EVM), where it is further processed by the stereo algorithm.

6 Conclusion

The paper shows the content of the project ADOSE and especially the task of the AIT in this project. For the chosen application of a side impact detection a stereo vision system with silicon retina cameras has been designed. Based on non-satisfying results from previous implementations a calibration and rectification step took place, but the accuracy of the results with these steps was still not satisfying and therefore new algorithm approaches have been searched and the first new approach has been outlined in this paper. Additional, the embedded hardware platform is shown where the final algorithm is executed.

7 Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n◦ICT-216049 (ADOSE).

References

[1] Sivilotti, M., Wiring consideration in analog vlsi systems with application to field-programmable networks, PhD Thesis, California Institute of Technology, 1991.

[2] Fukushima, K., Yamaguchi, Y., Yasuda, M., Nagata, S., An Electronic Model of Retina, Proceedings of the IEEE (Volume 58 / Issue 12), page 1950-1951, 1970.

[3] Mead, C., Mahowald, M., A silicon model of early visual processing, Neural Networks Journal, Vol 1(1), page 91-97, 1988.

[4] Lichtsteiner, P., Posch, C., Delbrück, T., A 128x128 120dB 30mW Asynchronous Vision Sensor that Responds to Relative Intensity Change, IEEE International Solid-State Circuits Conference (ISSCC'06), San Francisco/USA, 2006.

[5] Ebers, O., Überblick über aktuelle Verfahren zur Tiefenschätzung aus 2D Video Sequenzen, Studienarbeit an der TU Berlin, Institut für Telekomminukationssysteme, FG Nahrichtenübertragung, 2004.

[6] Schreer, O., Stereoanalyse und Bildsynthese, Springer-Verlag, 2005.

[7] Zhang, Z., A Flexible New Technique for Camera Calibration, Technical Report, MSR-TR-98-71, Microsoft Research, 2002.

[8] Bouguet, J.Y., Camera Calibration Toolbox for Matlab, Computer Vision Research Group, Department of Electrical Engineering, California Insitute of Technology, Link: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html, Last Access: 07.12.2008.

[9] Hess, P., Low-Level Stereo Matching using Event-based Silicon Retinas, Semester Thesis, Institute of Neuroinformatics, UNI-ETH Zurich/Switzerland, 2006.

[10] Scharstein, D., Szeliski, R., A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, International Journal of Computer Vision, Vol 47(1-3), page 7-42, 2002.

[11] Shi, J., Tomasi, C., Good Features to Track, Proceedings of the IEEE Computer Vision and Pattern Recognition Conference (CVPR'94), Seattle/USA, 1994.

[13] Kogler, J., Sulzbachner, C., Kubinger, W., Bio-inspired stereo vision system with silicon retina imagers, Proceedings of the 7th International Conference on Computer Vision Systems (ICVS'09), Liege/Belgien, 2009.

Jürgen Kogler → presenting author

Donau-City-Strasse 1
Techgate Tower 5th floor
1220 Vienna
Austria
E-mail: juergen.kogler@ait.ac.at

Christoph Sulzbachner

Donau-City-Strasse 1
Techgate Tower 5th floor
1220 Vienna
Austria
E-mail: christoph.sulzbachner@ait.ac.at

Erwin Schoitsch

Donau-City-Strasse 1
Techgate Tower 5th floor
1220 Vienna
Austria
E-mail: erwin.schoitsch@ait.ac.at

Wilfried Kubinger

Donau-City-Strasse 1
Techgate Tower 5th floor
1220 Vienna
Austria
E-mail: wilfried.kubinger@ait.ac.at

Martin Litzenberger

Donau-City-Strasse 1
Techgate 6th floor
1220 Vienna
Austria
E-mail: martin.litzenberger@ait.ac.at

Keywords: safety, embedded, stereo vision, silicon retina, bio-sensors.