4th International Conference from Scientific Computing to Computational Engineering
4th IC-SCCE
Athens, 7-10 July, 2010

ADDRESS-EVENT MATCHING FOR A SILICON RETINA BASED STEREO VISION SYSTEM

Jürgen Kogler[1], Christoph Sulzbachner[1], Florian Eibensteiner[2], and Martin Humenberger[1]

[1] AIT Austrian Institute of Technology,
Safety and Security Department
Safe and Autonomous Systems
Donau-City-Strasse 1, 1220 Vienna, Austria
e-mail: {juergen.kogler, christoph.sulzbachner, martin.humenberger}@ait.ac.at, web page: http://www.ait.ac.at

[2] Upper Austria University of Applied Sciences
Softwarepark 11, 4232 Hagenberg, Austria
e-mail: florian.eibensteiner@fh-hagenberg.at, web page: http://www.fh-ooe.at

Keywords: Silicon Retina, Address Event, Stereo Vision, Embedded System, Automotive

Abstract. *In this paper we present a stereo vision system based on silicon retina sensors. Derived from the human vision system, these bio-inspired silicon retina sensors are a new type of imagers. Conventional optical sensors capture images at a fixed frame-rate and deliver pixel information of the sensor chip at a certain point in time. In contrast, a silicon retina optical sensor provides only intensity variations (events) in the visual field. The target application of the silicon retina sensor in this work is a stereo vision system for pre-crash/side impact detection. Regarding the asynchronous event-triggered output characteristic of the silicon retina new approaches for data acquisition and processing of the data are necessary. In a first step data from the silicon retina imagers are converted into frames and processed with frame-based stereo vision algorithms. These methods do not benefit from the advantages of the silicon retina technology, thus a new approach exploiting the novelty of the sensor was implemented. This stereo matching approach uses the time information of the delivered address events as a primary matching criterion to find corresponding pixels in the data stream of the left and right sensor. The event-based stereo matching approach is tested with synthetic data, where also an evaluation with ground truth is carried out.*

1 INTRODUCTION

The goal of the EU-funded project ADOSE (reliable **A**pplication-specific **D**etection of road users with vehicle **O**n-board **SE**nsors) is the evaluation of new and improvement of existing in-car sensor technologies, which acquire data for different Advanced Driver Assistance Systems (ADAS) such as lane departure warning, collision warning or high-beam assist. Our aim in this project is the development of a stereo vision system based on silicon retina sensors used for a pre-crash warning/detection of side impacts. Many of the applications used for driver assistance need a sensor which calculates the distance of the objects located in front of the sensor system. The side impact pre-crash warning application needs the distance information to detect fast approaching objects in front of the camera system. For this reason we use a stereo vision system based on two cameras to extract the depth information with a correspondence search in the data of the left and right camera.

This stereo vision system consists of silicon retina sensors, which offers different advantages needed in the field of automotive area. A silicon retina sensor delivers events without a certain frame-rate, which makes the stereo matching to a challenging task. Frame-based stereo matching algorithms uses the full image from the left and right camera at for the correspondence search. In case of silicon retina stereo matching, correspondences must be searched in two data streams delivered independently and asynchronously. Additionally, tasks such as calibration and rectification come along with stereo vision and they differ in comparison to conventional stereo camera systems.

In section 2 the silicon retina technology, the difference to conventional cameras and the address event data format introduction is explained. Then in section 3 the stereo vision system and the requirements of the applications are outlined, followed by section 4 where the stereo vision algorithm is described in detail. In section 5 the test setup is explained and the experimental results are shown. Finally in section 6, we give an outlook to future work and conclude the work.

2 SILICON RETINA SENSOR

The silicon retina sensor is a new type of an optical sensor and is derived from the human eyes. The research goes back to Fukushima et al.[1] in 1970, who implemented the first artificial retina based on electronic standard components. These components emulated the photo receptors and ganglion cells of the human eye and a lamp array visualized the transmitted picture of the artificial retina. In 1988 Mead and Mahowald[2] developed the first retina based on silicon and established the name *Silicon Retina*.

The silicon retina sensor is in comparison to conventional frame-based cameras an event-triggered sensor which delivers intensity changes occurred in the visual field in front of the camera. This kind of sensor is free-running and delivers events continuously and independently in an asynchronous data stream without a certain frame-rate to a receiver. In the case of silicon retina technology the activity at a pixel is called ON-event if the intensity increased and OFF-event if the intensity decreased. These events are packed into an own format, the A*ddress Event* (AE) format, which was proposed by Sivilotti[3] in 1991. After the detection of events, the address (coordinates), the polarity, and a timestamp are written into data packages before the sensor sends the information to a receiver unit.

The silicon retina sensor has three main advantages in comparison to a conventional frame-based optical sensor. These differences are useful in different vision and automotive applications. In Figure 1 the differences of a silicon retina sensor and a conventional monochrome camera are illustrated.

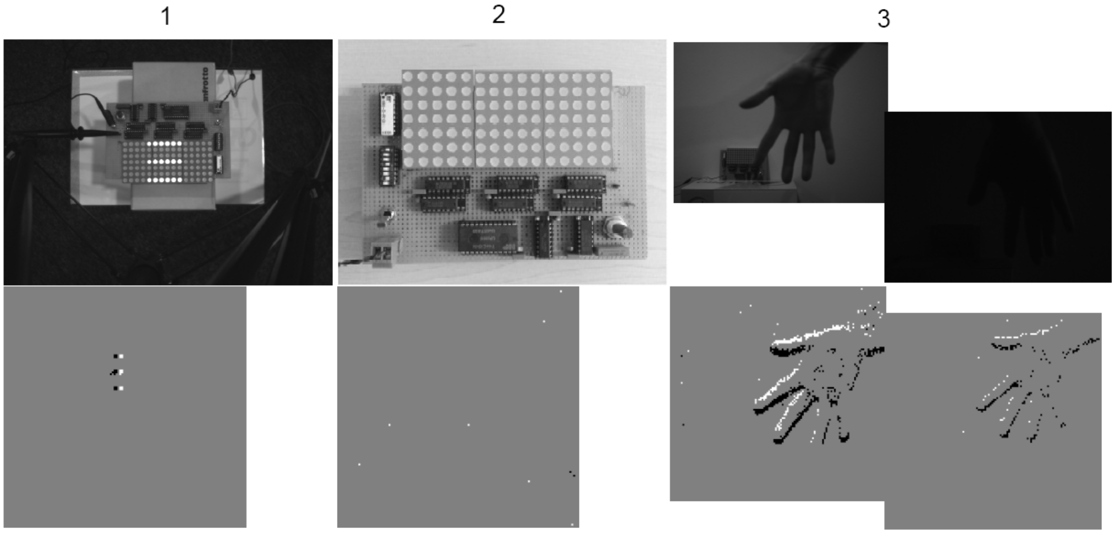

Figure 1. Comparison of a silicon retina sensor (second row) and a conventional camera (first row), (1) high temporal resolution, (2) data transmission efficiency, (3) high dynamic range

In the first column of Figure 1 the temporal resolution is compared. The top image shows a conventional monochrome camera with 60 frames per second which is not able to capture a running light with a speed of 475 Hz. The retina sensor below in column 1 is fast enough making a snapshot of the running light line. The column in the middle demonstrates the data transmission concept. A silicon retina is free-running and sends data only if there are intensity changes in front of the sensor, otherwise the silicon retina sensor sends nothing except noise events. In the top image of the middle column the monochrome camera has to send after a capture request all data even if there is no new information in the visual field. The last column of Figure 1 shows the advantage of the high dynamic range which is very interesting in the automotive area where a lot of different lighting conditions can occur. In the top image pair the result of a monochrome camera in a bright scenario (left) and dark scenario (right) is shown. Here the bright scenario is visible only. The dark environment is, without an adjustment of the camera settings, not possible to capture with enough contrast. In case of silicon retina sensors both lighting conditions results in a clear visible object in front of the sensor.

Within ADOSE we use two versions (generations) of silicon retina sensors which are compared in Table 1.

	Version 1	Version 2
Chip resolution	128x128	304x240
Temporal resolution	1ms	10ns
Pixel pitch	40µm	30µm
Dynamic range	120dB	120dB

Table 1. Two different generations of silicon retina sensors.

Version 1 has a smaller resolution of 128x128 pixels and achieves a temporal resolution up to 1ms. This sensor delivers up to 300k events per second via the address event data interface[4]. In the current sensor generation the chip resolution is 304x240 pixel and has a temporal resolution of up to 10ns which results in a higher data rate. Therefore, a compressed address data format was developed which achieves a performance of 5.125M[5] events per second. Both sensor generations achieve a dynamic range of up to 120dB and the pixel size for the sensor is 40µm in version 1 and 30µm in version 2.

3 STEREO VISION SYSTEM

In ADOSE the goal is to detect fast approaching objects to forecast side impacts. For this reason two silicon retina sensors are placed on a baseline to build up a stereo system. This stereo system is designed for pre-crash warning and consists of the stereo vision system (SVS) and an embedded system for the data acquisition and algorithm execution.

3.1 Stereo vision setup for side impact detection

The SVS must fulfill requirements given by the traffic environment. In Figure 2 a schematic of the side impact scenario including some key parameters is shown.

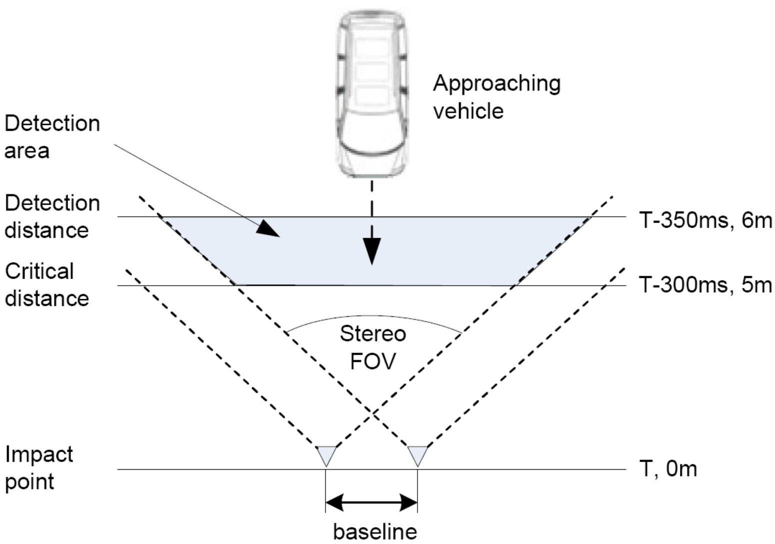

Figure 2. Stereo vision setup for the use in a pre-crash warning side impact detection application.

In the mentioned application the stereo vision system detects a closer coming object and activates the pre-safe mechanism in the car. The speed of the approaching vehicle is defined with 60km/h and the minimal width of 0.5m. For activating all safety mechanisms we assume that the vehicle needs about 300ms which defines the detection range of the camera system. A vehicle with a speed of 60km/h passes a distance of 5m in 300ms, therefore the decision of an impact will occur or not has to be made 5m before the vehicle will have the impact. In Figure 2 the detection distance and the critical distance, where a decision has to be made, are shown. These requirements define the key parameters of the optical system.

3.2 Hardware platform

The data delivered from the stereo vision system has to be processed and a decision about an impact has to be made in real time. For this reason we use a dedicated embedded hardware, which can also handle the increasing amount of data from the next generation of silicon retina sensors. This hardware platform consists of Digital

Signal Processors (DSPs) which handles the data acquisition and the execution of stereo matching and application algorithms. In Figure 3 shows an overview of the used hardware.

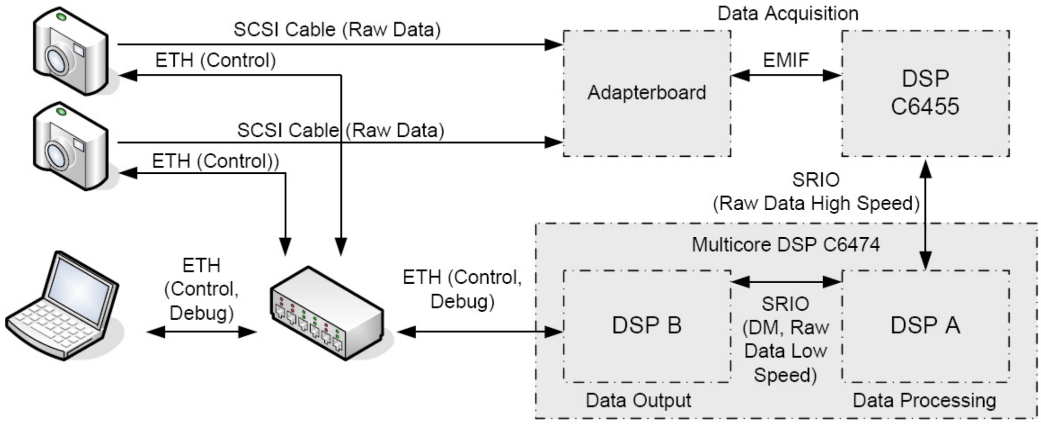

Figure 3. Embedded hardware platform for the data acquisition and execution of the address events from the silicon stereo vision system.

The C6455 DSP is responsible for the data handling from the silicon retina sensors via the adapter board which buffers the data to the C6474 multi-core DSP board which processes the address events data and sends the algorithm results to a visualization unit. A detailed description about the hardware configuration is described in the work of Sulzbachner et al.[6],[7]. The multi-core DSP platform generally enables parallel data processing however, this kind of algorithm can significantly benefit from application depended customizations of the underlying system architecture: hence optimized memory access patterns, special computation units, and massively parallel and pipelined hardware architectures are preferred. Such customized architectures can be implemented on *Field Programmable Gate Arrays* (FPGA), which can exploit the immanent parallelism of stereo-vision algorithms[8].

4 STEREO MATCHING ALGORITHMS

The stereo matching is the main part of the work and is considered in detail. In the case of silicon retina stereo matching two different approaches have to be considered, the frame-based and the event-based matching.

4.1 Frame-based processing

Silicon retina sensors deliver address event streams and no frames like conventional cameras and therefore it is not possible to use frame-based stereo matching algorithms directly without pre-processing of the data. For this reason an address-event to frame converter was implemented, which is described in a previous work[9], and used for processing of area-based[10] or feature-based[11] stereo matching algorithms. In Figure 4 the address-event to frame conversion process is depicted.

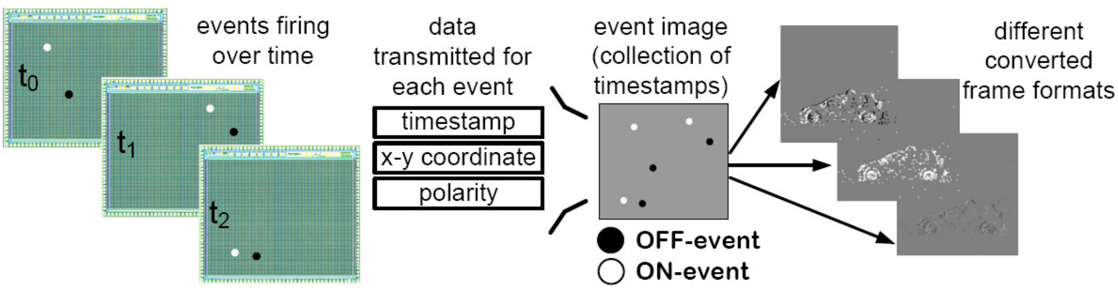

Figure 4. Workflow of the address event to frame converter.

The left side in Figure 4 shows the events over time, which send their coordinates, polarity, and timestamp. This information is collected and after a certain time a full-frame image is written. Depending on the stereo matching approach used, a grayscale or binary image is generated. Experimental results of the frame-based approach showed[12] that the accuracy is not good enough and the advantages such as the high temporal resolution is lost because of the conversions. Due to this fact a new approach was searched which uses the address events directly without additionally converting steps.

4.2 Event-based processing

Due to the asynchronous and frame-less data delivery of the silicon retina sensors, an event-based stereo matching approach, which exploits the characteristics of the silicon retina technology, has to be developed. For this reason a time-correlation algorithm for the correspondence search is used. This algorithm uses the time difference between events for the primary matching costs. In Figure 5 the whole workflow of the event-based algorithm approach is shown.

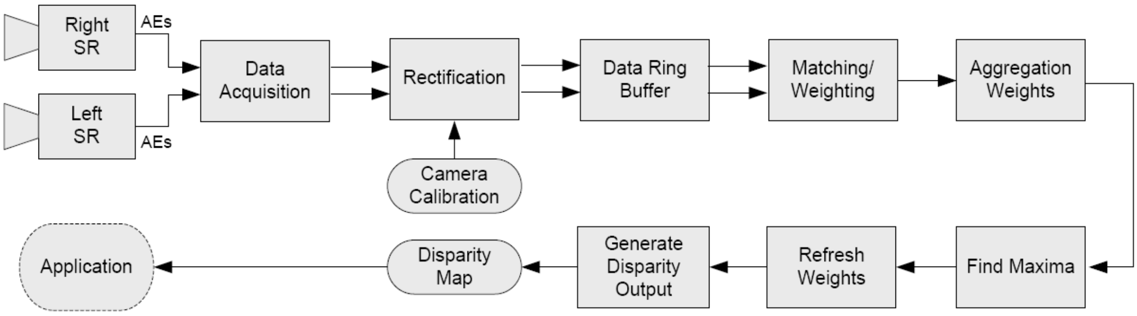

Figure 5. Workflow of the event-based time correlation stereo matching algorithm.

The workflow starts with acquiring the data from the silicon retina sensors, which means the data are read from the adapter board buffers and given to the rectification unit. For the rectification step all needed parameters were calculated in a previous calibration step. The calibration determines the intrinsic, extrinsic, and distortion coefficients plus the rectification matrices for both sensors. For the calibration of silicon retina cameras new methods have been developed, because the calibration pattern used for the calibration of conventional cameras were not suitable for silicon retina sensors. Hence, the event generation is stimulated with movements in front of the calibration pattern and before the execution of the calibration in MATLAB with the Caltech[13] toolbox takes place, a semi-automatic extraction of the feature points is implemented. In the rectification step both sensor planes are transformed in a way that the epipolar lines[14],[15] are in parallel and the correspondence search of the stereo matching algorithm can take place in the one-dimensional space. After the calculation of the calibration parameters with Zhang's[16] method, the parameters for the rectification are written in look up tables.

The rectified events are written into a ring buffer structure which handles the asynchronous and over time amount varying data. In the next step the matching of the events is carried out, where for each oncoming event a corresponding event on the opposite side is searched. For this search all events of the current timestamp as well as events from the past are used. This means also previous events are considered during the correspondence search. This search is carried out within the disparity range in a horizontal line due to rectification. In Figure 6 on the top left side the event buffers are shown which store the current and the historical events.

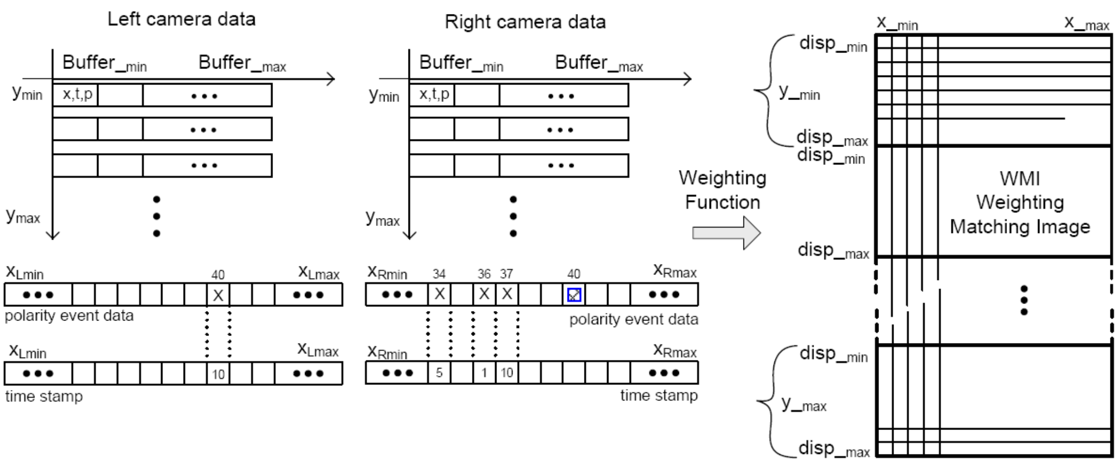

Figure 6. Matching and weighting of corresponding address events and writing of calculated costs into the WMI.

If there are possible matching candidates within the disparity range of actual and historical events, which have the same polarity, the timestamps are used for matching costs calculation. In Figure 6 the matching with an event at the x-coordinate 40 is illustrated. The left event is the reference event and the search takes place on the right side where three candidates with the same polarity and within the considered history are found. Now, the time difference between the timestamp of the left camera and the three found events of the right camera are calculated.

For determination of the costs of a found matched event pair, different weighting functions were used. In Figure 7 all used weighting functions are depicted.

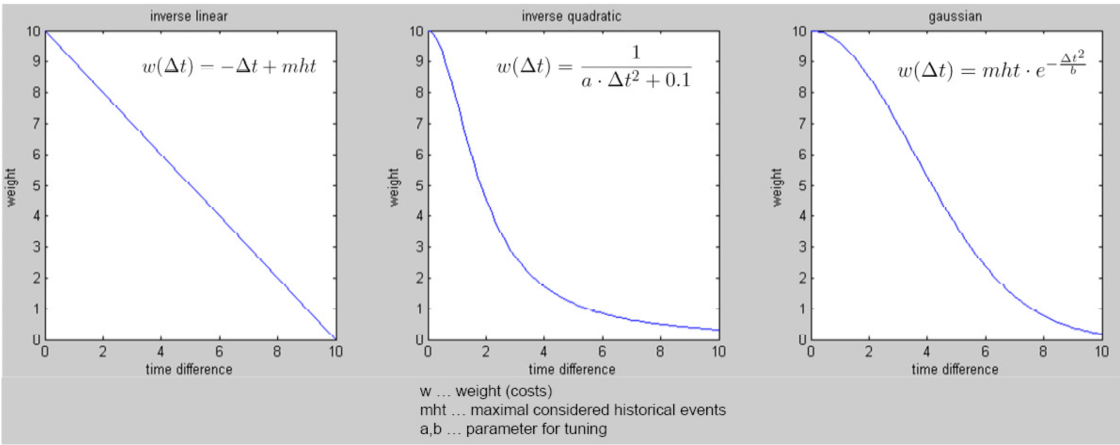

Figure 7. Weighting function for calculating the costs of matched address events.

On the abscissa of the charts the weighting costs which may achieve a maximum of 10 (derived from the maximal considered historical events) and on the ordinate of the charts the time difference is shown. In the example the considered history is 10, which means from the current timestamp of an event 10 timestamps of the past are used for the current calculations. The left function shows a simple inverse linear relation between the time difference Δt and the weight. In the middle chart an inverse quadratic function is depicted which is faster declining and matches more current events and does not consider older events in the same amount. The Gaussian function shown in the right chart of Figure 7 increases, in comparison to the inverse linear function, the weights of current timestamps and decreases the older timestamps. Both functions on the right side can be tuned with a parameter for the adaption to different weighting needs.

All the matched and weighted events are written in the *Weighted Matching Image* (WMI) shown in Figure 6. This data storage is a two dimensional representation of a three dimensional space, where a place for each pixel coordinate and disparity level is reserved. The WMI is a dynamic data storage which is updated each processing cycle and only deleted if a reset takes place. That means all costs entered stay in the WMI for a defined time till they are removed and so the matched costs from previous steps contribute to the results of the current calculations.

The next step of the algorithm is the aggregation of the weights in the WMI. Therefore the WMI structure is transformed logically and a filter kernel works on weights with the same disparity. In the current algorithm a average filter kernel with a variable window size is used. After the aggregation step the maximum costs for each coordinate are searched which represents the best matching disparity. In consideration that the WMI is a refreshing data structure, after the maximum search all weights are checked if they have to be deleted from the WMI and therefore the weight itself is a lifetime counter. Each round the weight is reduced with a defined value till the weight is zero and deleted from the WMI or refreshed by a new match as well as a new weight. The results are written into the disparity map which can be used from the application for further processing.

5 EXPERIMENTAL RESULTS

For the evaluation of the event-based stereo vision algorithm a specific tool was used. It is called *Event-Editor*[17] and gives the opportunity to generate synthetic stereo data which allows a verification of the algorithm because of available ground truth information.

Before the evaluation, the parameters for the synthetic data have to be generated. The detection range is between 6m and 5m meter what gives, considering the system configuration, a disparity range between 15 and 20. In the evaluation we considered a higher range of 35 which evaluates the capability of the algorithm to detect closer objects as well. The simplified model of the silicon retina used for the generation of synthetic silicon retina data is a suitable approximation of the real silicon retina sensor for the algorithm evaluation.

In Figure 8 the evaluation results of the algorithm are shown. Different aggregation window sizes, noise situations and weighting functions are compared. The value in percent of the y-coordinate gives the information how many disparities were calculated correctly in relation to the total amount of events in the current matching step. The evaluation was carried out in three different disparity distances, depicted with the three colors in Figure 8. The three bars at the x-coordinate describe the algorithm evaluation settings. The first part shows the aggregation window size (3x3 or 5x5), the second part shows if there was a noise signal added (aN) or not (wN) and the last part describes the weighting function used.

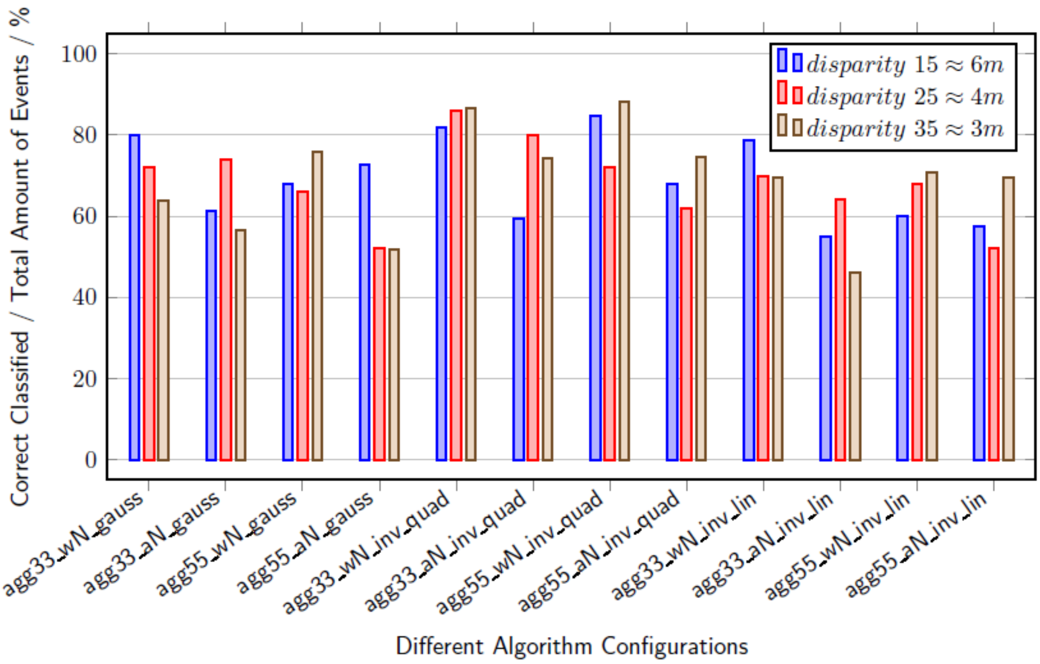

Figure 8. Experimental results of the event-based stereo matching algorithm with different evaluation and algorithm settings.

The results show that the quality of disparity calculation is independent of the evaluated distances. Added noise influences the rate of correct classified disparity values in comparison to noise free data. The aggregation window size has only a small influence but especially when noise is added the amount of correct classified disparities increase if the window size is enlarged. The highest influence has the weighting function (Gaussian function, an inverse quadratic function or an inverse linear function). In case of a linear quadratic function the best results of more than 80% correct classified disparities could be achieved.

6 CONCLUSION AND FUTURE WORK

The address event stereo matching with a silicon retina stereo system is different to conventional stereo vision systems based on frame processing. In previous work we tried to make the silicon retina data with certain conversion steps feasible for frame-based stereo matching algorithms, which results in a loss of the advantages given by the silicon retina technology.

Therefore we implemented an event-based stereo matching approach which uses the asynchronous silicon retina data directly without any conversion steps. In this approach the time was used as the primary matching criterion to find corresponding events from the left and right camera. The results showed that different weighting functions for the time give different matching results and for a better quality and higher robustness the matching costs have to be combined with other matching methods.

The next step is to combine the time-correlation matching costs with other costs to make the algorithm output more robust against different influences such as increased noise or inaccuracies in the rectification step. Additional we want to start with first tests of the application to show that the stereo matching output can be used for the pre-crash/side impact detection application. Another task is the migration from the PC-based solution to the DSP-multicore platform shown in Section 3.2 and continuative the integration of the event-based matching approach into a FPGA, expecting not only a reduction of the latency and an improvement of the throughput of the system but also an improvement of the scalability and the overall energy efficiency of the system.

7 ACKNOWLEDGEMENTS

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] Fukushima, K., Yamaguchi, Y., Yasuda, M. and Nagata, S. (1970), "An Electronic Model of Retina", Proceedings of IEEE, Vol. 58, No. 21, pp. 1950-1951.

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

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

[4] Lichtsteiner, P., Posch, C. and Delbruck, T. (2008), "A 128x128 120dB 15μs Latency Asynchronous Temporal Contrast Vision Sensor", IEEE journal of solid-state circuits, Vol. 43, No. 2, pp. 566-576.

[5] Hofstätter, M., Schön, P. and Posch, C. (2009), "An integrated 20-bit 33/5M events/s AER sensor interface with 10ns time-stamping and hardware-accelerated event pre-processing", *Proceedings of the International IEEE Conference on Biomedical Circuits and Systems, BioCAS, Beijing China, 26-28 November*, pp. 257-260.

[6] Sulzbachner, C., Kogler, J. and Kubinger, W. (2010), "An Embedded High Performance Data Acquisition and Pre-Processing Interface for Event-Based Silicon Retina Data", *Accepted for Proceedings of the International IEEE/ASME Conference on Mechatronic and Embedded Systems and Applications, MESA, Qingdao ShanDong China, 15-17 July*.

[7] Sulzbachner, C. and Kogler, J. (2010), "A Load Balancing Approach for Silicon Retina based Asynchronous Temporal Data Processing", *Accepted for Proceedings of the International Euromicro Conference on Software Engineering and Advanced Applications, SEAA '36 , Lille France, 1-3 September*.

[8] Porter, R. B. and Bergmann, N. W. (1997), "A generic implementation framework for FPGA based stereo matching." *Proceedings of the IEEE Region 10th Annual Conference on Speech and Image Technologies for Computing and Telecommunications, TENCON '10, Brisbane Australia, 2-4 December,* Vol. 2, pp. 461-464

[9] Kogler, J., Sulzbachner, C. and Kubinger, W. (2009), "Bio-inspired stereo vision with silicon retina imagers", *Proceedings of the International Conference on Computer Vision Systems, ICVS '7 , Leuven Belgium, 13-15 October*, Vol. 5815, pp. 174-183.

[10] Scharstein, D. and Szeliski, R. (2002), "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorihtms", International journal of Computer Vision, Vol. 47, No. 1-3, pp. 7-42.

[11] Tang, B., Ait-Boudaoud, D., Matuszewski, B.J. and Shark, L.K. (2006), "An Efficient Feature Based Matching Algorihtm for Stereo Images", *Proceedings of the International IEEE Conference on Geometric Modeling and Imaging, GMAI , London United Kingdom, 5-7 July*, pp. 195-202.

[12] Kogler, J., Sulzbachner, C., Schoitsch, E., Kubinger, W. and Litzenberger, M. (2010), "ADOSE – New bio-inspired in-vehicle sensor technology for active safety", *Proceedings of the International Forum on Advanced Microsystems for Automotive Applications, AMAA '14 , Berlin Germany, 10-11 May,* pp. 155-164

[13] Bouguet, J.Y. (2008), "*Camera Calibration Toolbox for MATLAB*", Computer Vision Research Group, Department of Electrical Engineering, California Institute of Technology (http:// http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

[14] Schreer, O. (2005), *Stereoanalyse und Bildsynthese,* Springer-Verlag, Berlin Heidelberg.

[15] Gonzalez, R.C. and Woods, R.E. (2002), *Digital Image Processing,* Prentice Hall/Pearson Education, Upper Saddle River, New Jersey.

[16] Zhang, Z. (2002), *A Flexible New Technique for Camera Calibration*, Technical Report MSRTR9-71, Microsoft Research.

[17] Sulzbachner, C., Kogler, J. and Eibensteiner, F. (2010), "A Novel Verification Approach for Silicon Retina Stereo Matching Algorithms", *Accepted for Proceedings of the International IEEE Symposium organized from the Croatian Society Electronics in Marine, ELMAR '52 , Zadar Croatia, 15-17 September.*