A Novel Verification Approach for Silicon Retina Stereo Matching Algorithms

Christoph Sulzbachner[1], Jürgen Kogler[1], Florian Eibensteiner[2]
[1]AIT Austrian Institute of Technology, Donau-City-Strasse 1, 1220 Vienna, Austria
[2]Upper Austria University of Applied Sciences, Softwarepark 11, 4232 Hagenberg, Austria
Email: christoph.sulzbachner@ait.ac.at

Abstract—**This paper presents a novel verification approach for silicon retina based stereo matching algorithms. The silicon retina is a new type of bio-inspired sensor technology derived from the human vision system. The sensor only provides event-triggered information and avoids redundant data by delivering intensity changes in a scene. Due to the special characteristics and limitations of this sensor technology, existing verification approaches and performance metrics are not suitable. Thus, a tool for synthetic scene generation, ground truth generation and algorithm verification has been implemented.**

Index Terms—**Silicon Retina, Stereo Matching, Verification, Ground Truth**

I. Introduction

The silicon retina technology is a new type of sensor technology which is completely different to conventional imager sensors. This type of sensor does not deliver images captured at a fixed frame-rate. In fact, the sensor only gives information about brightness variations in a scene at a very high time-resolution. Unaltered parts of a scene those contain no information are not supplied and need not to be processed. Due to the special characteristics of the imager technology, novel stereo algorithms and verification approaches are required to explore the feasible prospects of the sensor.

In the EU-funded project ADOSE[1] we use a silicon retina stereo vision sensor for pre-crash warning/preparation for a side impact collision. Our aim is to detect closer coming objects in a 3D map of a scene. The stereo matching approach is based on time-space correlation of events in a scene captured by two different points of view.

An increasingly large set of image and video data sets exist publicly on the Internet including their ground truth information that is essential for verification. For conventional frame-based stereo vision, the Middlebury Stereo Vision[2] data set exists, that is intended for evaluation of dense frame-based stereo correspondence algorithms. Scharstein and Szeliski [3] provide both a taxonomy of existing algorithms and a test bed for quantitative evaluation of stereo algorithms in their work. For evaluating the performance of stereo matching or the effects of varying some parameters of an algorithm, both use the root-mean-squared (RMS) error measured in disparity units and the percentage of bad matching pixels. Additionally to these statistics, they also focus on three different kinds of regions in their test images: textureless, occluded and depth discontinuity regions.

The Middlebury data sets are intended to be used for conventional stereo matching that is based on processing static frames of a scene without any time information. In contrast, the silicon retina imagers deliver event- and time-based information of a scene. Thus, the Middlebury data set is not adequate and cannot be used for evaluating the performance of silicon retina based stereo matching.

Commonly used approaches are based on hand-labeling or different sensor technologies such as structured light or laser. Hence, for silicon retina a new approach for a verification tool has been developed that affords synthetic generation including the ground truth data and automatically verification with defined performance metrics for silicon retina.

The remainder of this paper is outlined as follows: Section II gives an overview of the silicon retina technology and the data that is delivered by the imager. Section III covers our verification approach. We describe both generating synthetic scenes and ground truth data. Performance statistics of adapted area-based stereo algorithms are presented in Section IV. Finally, we give a conclusion about the work.

II. Silicon Retina Concept

The silicon retina technology goes back to Fukushima et. al. [7] in 1970, who first implemented a model of a retina. Mead and Mahowald [2] in 1988 developed the first silicon based retina.

The silicon retina technology does not deliver sensor data of complete frames or regions of interest of a frame at a fixed frame-rate. This type of sensor uses an event-based concept. Data is only transmitted and thus need to be processed, when variations of brightness are detected in an observed scene. Areas in a scene without any variations are not transmitted.

In ADOSE, we are using a silicon retina imager with an optical resolution of 304×240pixel and a time-resolution of 10ns. For stereo applications, it is very important that both silicon retina imagers have a common time base. Due to the very small time-resolution the complete synchronization is implemented in hardware [8].

A. Address-Event-Representation

Generally, Address-Event-Representation (AER) is a tuple of a time-stamp, the coordinates of the event and a polarity

[1]http://www.adose-eu.org
[2]http://vision.middlebury.edu/stereo

flag, indicating whether the event was an ON- (increasing brightness) or OFF-event (decreasing brightness). Due to the high data rates of the sensor, a compressed data format is used. For minimizing the Address-Event (AE) overhead, the tuple is separated into a time- and an event-message. The time-message delivers a time-stamp with a granularity up to nano-seconds. The event-message consists of the coordinates, where the event has occurred and the polarity.

B. Stereo Approaches

Due to the novelty of the silicon retina technology, existing stereo vision algorithms cannot exploit the potential of the imagers.

Scharstein and Szeliski [3] present an overview and evaluation of many different types of area-based stereo algorithms for conventional optical sensors processing on a frame basis. Kogler et. al. [5] evaluated different stereo matching algorithms working on frame-basis. In this content, a AE frame is a collection of AEs over a defined time-period Δt. They showed methods for building AE frames for a sum of absolute differences (SAD) based stereo matching algorithm.

Generally, a feature-based stereo approach works by matching extracted features from an image. Shi and Tomasi [6] give an overview of features and Tang et. al. [1] cover matching feature points. In [5], we used a *Center-of-Gravity* matching method. Alike the area-based approaches, AE images were computed.

Present research is done in time-space correlation methods for stereo matching to fully exploit the AER concept and the high time-resolution of the silicon retina technology.

III. Verification Approach

Verification of conventional computer vision applications is afford by available data sets. For silicon retina applications no tools or data sets exist. In [5], we used real-world environments for stimulating the algorithms used for verification without ground truth. The most accurate test data including reference data was justified with a measuring tape. This was sufficient for some estimations whether a stereo algorithm is computing principally or not. Predictions of the achieved quality of the stereo matching algorithms were not possible.

For scene and ground truth generation a verification tool was implemented that is shown in figure 1. The tool allows stimulating a stereo matching algorithm with a defined interface with synthetic generated or already existing AER data stereo streams. Only a synthetic generated scene affords generating ground truth data. The tool consists of a fast graphical user interface for visualizing the processing of the stereo matching algorithm and an embedded interpreter for generating a scene and handling the object models and further control methods. The embedded Python[3] interpreter allows accessing the powerful language with all its features and extending the visualization tool with e.g. a debug module for statistical analysis of a scenario. A color legend assists interpreting a pixel's disparity.

[3]http://www.python.org

A. Data Visualization

The verification tool processes the data frameless similar to the interface of a silicon retina imager in AER format. The windows shown in the verification tool are visualized time-slots, defined in equation 1, where $X \in [0..RES_X]$ and $Y \in [0..RES_Y]$ are sets representing coordinates of the imager plane, where RES_X is the horizontal and RES_Y is the vertical resolution of the optical sensor. Due to the polarity of an event, the regarding pixel has an assigned grayscale value. Pixels without an event are handled by a default value.

$$frame_t(x,y) := \begin{cases} 255 & AER_t(x,y) = 0 \\ 0 & AER_t(x,y) = 1 \quad \forall x \in X, y \in Y \\ 128 & else \end{cases} \tag{1}$$

For fast visualization purpose and to minimize the memory usage of the tool the data is ordered by the time-slots and the imager identifier. Thus, for displaying a certain visualized time-slot, the particular time-stamp and identifier have to be found using a search algorithm.

B. Scene Generation

The verification tool supports different types of objects that are intended to be used for scene generation. The root object implements disparity generation at pixel level and thus all higher level objects should be derived from the root object *StereoObject*. Supported types of objects are:

- The root object *StereoObject* implements all basic functionalities including the data exchange with the visualization tool and data containers for handling the AER data including their disparity.
- The higher level objects *StereoPoint*, *StereoRectangle*, *StereoCircle* or *StereoLine* implement the representation of the particular basic geometric objects in the AER data space. The higher level objects can be combined, e.g. for generating a closer coming vehicle.
- It is also possible to use regions of interest (ROI) of events of a recorded visualized time-slot as an object and use it for scene generation.

C. Data Management

Due to the known event information of a synthetic scene, an event stream for both the left and the right silicon retina imager including their ground truth data can be generated. For verification purpose and data exchange between different development environments, the event and disparity information is handled in an extensible mark-up language (XML) based representation.

For advanced visualizing data, functionalities have been implemented for exchanging data with the virtual reality modeling language (VMRL) [11]. Figure 2 shows a processed disparity map of a recorded scene of a pedestrian.

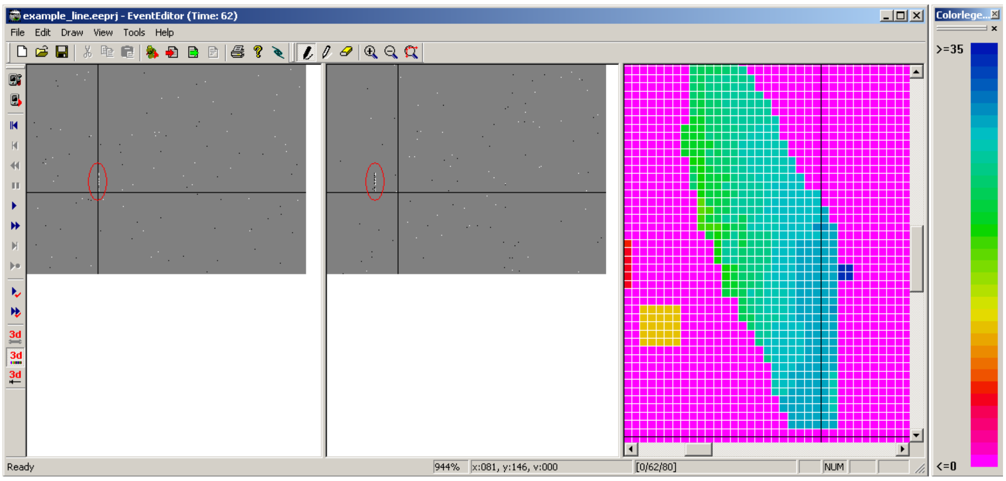

Fig. 1. Verification Tool; scenario shows a moving line from the upper left to the bottom right corner of the visualized time-slot. left window: left imager; middle window: right imager; right window: processed disparity map by the stereo matching algorithm

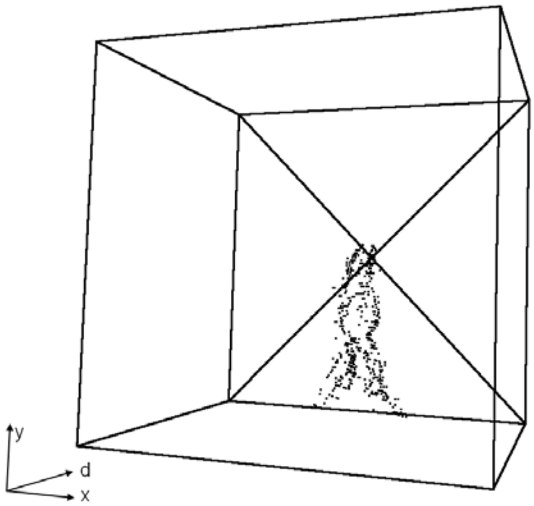

Fig. 2. Disparity map of scene with a pedestrian, processed with an area-based stereo matching algorithm and visualized in VRML.

D. Verification

The verification module *Verify* is not part of the visualization tool and is implemented as an external Python module. Thus, adaption of the module during runtime are possible and debugging is essentially eased.

The current implementation of *Verify* affords comparing the disparity map processed by the stereo matching algorithm at a specific time-stamp with the ground truth information.

A method for visualizing the performance of classifiers are receiver operating characteristics (ROC). Fawcett [9], [10] and Provost and Fawcett [4] give an introduction and practical considerations for ROC analysis in their work. They cover performance metrics and optimization methods such as hit rate, accuracy, or precision.

Within verification of silicon retina stereo matching algorithms, we also address two-class problems. Each instance of the ground truth **GT** is mapped to one element of the set $\{p, n\}$, where p is an existing event and n is a missing ground truth event. The classification model represents the mapping from **GT** to a predictable class, the disparity map **DM** is mapped to the set $\{y, n\}$, where y is an existing and n is a missing disparity event. Based on these combinations a two-by-two confusion matrix can be build.. Metrics evaluation is based on comparing the disparity to the ground truth data set. Due to the functionality of the algorithm used for stereo matching, the verification approach needs to compare actual with previous appeared events. Equation 2 defines both the disparity and the ground truth set for metrics evaluation, where t_h defines the propagation delay of a set.

$$
\begin{aligned}
dm(x,y,t) &:= \int_{t-t_h}^{t} DM(x,y,t)dt \qquad (2)\\
gt(x,y,t) &:= \int_{t-t_h}^{t} GT(x,y,t)dt
\end{aligned}
$$

- A *true positive* is defined by equation 3 for both existing disparity and ground truth data with an error tolerance δ_d.

$$tp(t) := \sum_{x\in X, y\in Y} [|dm(x,y,t) - gt(x,y,t)| < \delta_d] \quad (3)$$

- A *false positive* is defined in equation 4 for the same restrictions as *true positives*, though the error tolerance δ_d is exceed.

$$fp(t) := \sum_{x\in X, y\in Y} [|dm(x,y,t) - gt(x,y,t)| \geq \delta_d] \quad (4)$$

- A *false negative* $fn(t)$ is defined by an existing ground truth and a missing disparity value.
- A *true negative* $tn(t)$ is defined by both a missing ground truth and a missing disparity value.

Based on these performance primitives, further performance metrics can be computed such as true positive rate or false positive rate of a time-slot shown in equation 5.

$$\begin{aligned} tp\ rate(t) &:= \frac{tp(t)}{tp(t)+tn(t)} \\ fp\ rate(t) &:= \frac{fp(t)}{fp(t)+fn(t)} \end{aligned} \tag{5}$$

IV. Evaluation

Figure 3 shows three different synthetic input scenarios (a-c) each evaluated with the shown performance metrics true positive rate (tp rate) and false positive rate (fp rate). Basically, the scenario represents a moving object from the upper left to the bottom right corner, processed with the same algorithm and an increasing level of noise from scenario a to c. With an increasing noise level the true positive (tp) rate decreases and the false positive (fp) rate increases. The abscissa of the plot shows the time-stamps of the computed time-slot.

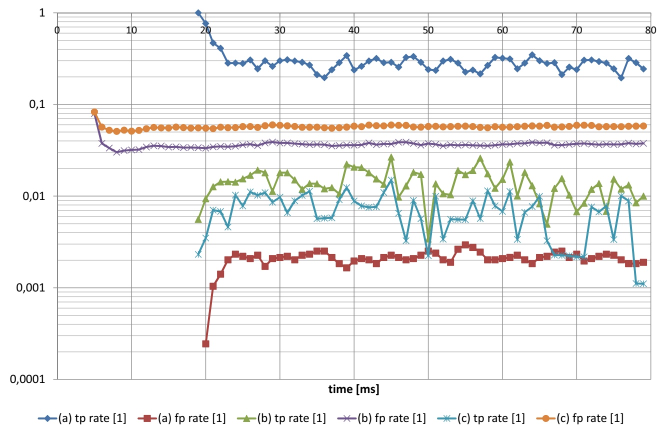

Fig. 3. Different scenes with increasing noise levels from scenario a-c. The evaluation shows a performance analysis of true positive rate and false positive rate of area-based stereo matching algorithms, which is processed visualized time-slots.

Due to the special characteristics of the silicon retina technology the significance of the evaluation results depend on the detected motion in an observed scene, because the majority of a visualized time-slot has no brightness changes and therefore no events occur. Thus using these performance metrics, the *true negative* rate is always high in scenes without motion in comparison to the *true positive* rate.

V. Summary and Outlook

This paper presented a novel verification approach for silicon retina based stereo matching algorithms including synthetic scene and ground truth generation. We showed characteristics and limitations of the silicon retina technology and the definition of the performance metrics. Concluding, we showed an example of three different scenarios and accordingly applying the performance metrics.

The presented approach is suitable for verifying an algorithm but not sufficient for system validation purposes because the tool does not completely reflect the behavior of the silicon retina imager. Hence, a displaying device is in development that allows a replay of the synthetic scenes to act as stimuli for the silicon retina imagers.

Enhancements of the scene generation would be an implementation of a silicon retina model in the base object *StereoObject* to ensure that all generation objects use a more realistic imager model. Up to now, imager specific parameters e.g. noise is implemented individual.

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] B. Tang and D. Ait-Boudaoud and B. J. Matuszewski and L.-K. Shark. An Efficient Feature Based Matching Algorithm for Stereo Images. *Proceedings of the IEEE Geometric Modeling and Imaging Conference (GMAI)*, 2006.

[2] C. Mead and M. Mahowald. A silicon model of early visual processing. *Neural Networks Journal*, 1:91–97, 1988.

[3] Daniel Scharstein and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47:7–42, 2001.

[4] F. Provost and T. Fawcett. Robust classification for imprecise environments. *Machine Learning*, 42(3):203–231, 2001.

[5] J. Kogler and C. Sulzbachner and W. Kubinger. Bio-inspired stereo vision system with silicon retina imagers. *International Conference on Computer Vision Systems (ICVS)*, 2009.

[6] J. Shi and C. Tomasi. Good Features to Track. *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, 1994.

[7] K. Fukushima and Y. Yamaguchi and M. Yasuda and S. Nagata. An Electronic Model of Retina. *Proceesings of IEEE*, 58(12):1950–1951, 1970.

[8] M. Hofstätter and P. Schön and C. Posch. An integrated 20-bit 33/5M events/s AER sensor interface with 10ns time-stamping and hardware-accelerated event pre-processing. *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2009.

[9] T. Fawcett. ROC graphs : Notes and practical considerations for researchers. *Technical Report HP Labratories*, 2004.

[10] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[11] World Wide Web Consortium (W3C). VRML Virtual Reality Modeling Language. http://www.w3.org/MarkUp/VRML, 1995.