# Event-Driven Stereo Vision Algorithm Based on Silicon Retina Sensors

Florian Eibensteiner
Hans Georg Brachtendorf
Upper Austria University of Applied Sciences
Softwarepark 11, 4232-Hagenberg, Austria
florian.eibensteiner@fh-hagenberg.at

Josef Scharinger
Johannes Kepler University Linz
Altenberger Strasse 69, 4040 Linz, Austria
josef.scharinger@jku.at

***Abstract*—In this paper a new stereo matching concept for event-driven silicon retinae is presented. The main contribution of the proposed approach is the correlation of incoming events. As a novelty, not only the spatial information is used, but also the time of occurrence of the events as a part of the similarity measure. Stereo matching is used in depth generating camera systems for solving the correspondence problem and for 3D reconstruction of the sensed environment. In fact, using conventionally frame-based cameras, this is a time consuming and computationally expensive task, especially for high frame rates and spatial resolutions. An event-based silicon retina delivers events only on illumination changes and completely asynchronous in time. The sensor provides no frames, but a time-continuous data stream of intensity differences and thus inherently reduces the visual information to a minimum. This paper focuses on an event-based stereo matching algorithm implemented in hardware on a field programmable gate array (FPGA) that allows a reliable matching of the sparse input event data. Furthermore, the approach is compared to other standard frame-based and event driven stereo methods. The results show that the achieved depth map outperforms other algorithms in terms of accuracy and the calculation performance of the hardware architecture is in the range or still higher than state-of-the-art computing platforms.**

***Index Terms*—Silicon Retina, Stereo Vision, Event-Based, Field Programmable Gate Array, Hardware Architecture**

## I. Introduction

The visual sensing and understanding of the environment in which we live is not only important for human beings or animals, but also for robots in automation applications, autonomous vehicles, or just assistance systems in automotive or aeronautic applications. Particularly vision based systems for 3D reconstruction are ubiquitous in automation applications, where robots entirely independent and autonomously assemble products, detect defects or even sort components based on visual perception. All these applications have roughly two things in common: it is crucial to do the image processing in real-time, and those systems are mostly realized on embedded platforms.

For sensing depth information of the observed environment, different technologies can be deployed which can generally be categorized in two groups, active and passive sensing technologies. As indicated by the name, active sensors emit a signal and evaluate the received echo by measuring, e. g., the round-trip time. For instance, such active technologies are, time-of-flight (TOF) cameras, ultrasonic detectors, laser range finders or laser scanners (light detection and ranging(LIDAR)), radar, light-section, and structured light. In contrast, passive technologies as, e. g., stereo vision, structure from motion, and optical flow, do not emit any kind of signal.

Conventional image sensors as charge coupled device (CCD) or complementary metal-oxide-semiconductor (CMOS) deliver data for each pixel at a fixed frame rate, thus high resolution in space and time lead to huge amounts of data, which makes it difficult to fulfill real-time requirements in particular on embedded platforms with limited resources.

By contrast, the capturing principle of event-based imaging sensors, so-called silicon retinae, is completely different. Here, a single pixel works independently and asynchronously in time and fires events only on changes of illumination. As a matter of fact, event-driven imaging provides no framed data, but a continuous sequence of differences of intensity values over time. Therefore, such a sensor inherently compresses the visual information to a minimum. In many computer vision applications, the usage of sparse event information can be an efficient method for enhancing the performance [1], [2], [3], even on embedded computing platforms.

This sensing concept shows advantages for high dynamic range and high speed applications and is inspired by biological vision systems. By now, the only commercially available event-based silicon retina sensors are shipped by the Austrian Institute of Technology (AIT) and inilabs, a spin-off of the Institute of Neuroinformatics (INI) at the University of Zurich.

Thus, this work introduces an event-driven stereo matching algorithm for sparse silicon retina data, implemented in an FPGA based embedded system to face the demands on computational performance. The matching procedure is directly applied on the event data considering the timing information and the polarity of the events for cost aggregation.

## II. Sensor Technology

As mentioned above, in this work we use two *silicon retina* cameras in a stereo vision system to calculate depth data of a scene. In comparison to conventional CMOS or CCD imagers, every pixel of a silicon retina independently provide events only on changes of the illumination. In general, static parts of the scene, e. g., non-moving objects or stationary

978-1-5090-4591-4/17/$31.00 ©2017 IEEE

background, causes no illumination changes and thus are completely suppressed. This frameless, asynchronous, time-continuous, logarithmic photoreceptor offers three essential advantages [4]:

- The illumination change dependents and an asynchronous generation of event data obviously leads to a significant data reduction because only dynamic parts of the scene are detected and static parts are completely suppressed.
- The design of the analogous pixel array and the event-driven signal generation enables a very high temporal resolution of up to $10\,\mathrm{ns}$.
- The logarithmic measurement of the photo current yields a high dynamic range, therefore the sensor is suitable to be used in applications with fast transient light conditions.

The silicon retina sensor considered in this work (called ATIS, Asynchronous, Time-based Image Sensor) has a spatial resolution of $304\times240$ pixels, a temporal resolution of up to $10\,\mathrm{ns}$, and a dynamic range of $143\,\mathrm{dB}$. Details about the sensor and its characteristics can be found in the work of Posch et al. [5], [6].

For generating events whenever the illumination changes, the sensor uses an illumination change detector circuit. An event is defined as $e(p,t)$ [7], where $p = (x,y)^T$ is the spatial location of the pixel which fires the event, and $t$ is the time of occurrence given in the units of timestamps. One timestamp corresponds to the temporal resolution of the sensor (1 timestamp $\triangleq 10\,\mathrm{ns}$, in the case of the ATIS). Due to the slow motion of the objects in our test cases, we use a temporal resolution of $100\,\mu\mathrm{s}$ for one timestamp. The polarity of an event can either be positive (on-event) or negative (off-event), depending on the direction of the change of the illumination $I$ over a period of time $\Delta t$. The following equation shows the definition of the event's polarity:

$$e(p,t) = \begin{cases} +1 & I(p,t) - I(p,t-\Delta t) > \Delta I \\ -1 & I(p,t) - I(p,t-\Delta t) < -\Delta I \end{cases} \quad , \quad (1)$$

with the adjustable on- and off-threshold $\Delta I$.

Due to the asynchronous behavior of the analog front-end, each pixel work independently, not all pixels of an edge are simultaneously active, and objects' contours become just visible over several timestamps. Therefore a stereo algorithm must consider a certain history in the matching process. Analyses have shown, that the higher the temporal resolution of the sensor is, the larger is the time difference. This difference is mainly a problem between the left and the right sensor, because between corresponding events there can be several timestamps.

In Figure 1 a detailed view of a non-moving and a moving person captured by the ATIS and a conventional monochrome camera system are illustrated. The image details of these sensor systems are not exactly the same, because of the different field of view (FOV), optical characteristics, and various properties. In order to gain the visualization, the event stream is converted into a grayscale image, where off-events are colored to black, and on-events are mapped to white pixels. Areas in the image with no activity, e. g., non-moving objects or stationary background, are represented by gray pixels. converting the events into a grayscale image, all events $e(p,t)$ of one or more timestamps are aggregated and mapped into the grayscale image $I_g(p,t)$ according the following equation:

$$I_g(p,t) = \begin{cases} 255 & e(p,t-h) = +1 \\ 0 & e(p,t-h) = -1 \\ 128 & Else \end{cases} \quad , \quad (2)$$

where $h$ refers the aggregation period, and a grayscale range from $0$ to $255$ with a middle gray value of $128$ is used. Furthermore, if several events occur at the same spatial location $p = (x,y)^T$, older events are overwritten by newer ones.

## III. Related Work

Generally, in the case of stereo matching algorithms for event-based silicon retina sensors a differentiation needs to be drawn between algorithms working directly on event data (matches on- and off-events), and methods using grayscale images generated from event streams.

Stereo matching using neuromorphic sensors got started with the development of the silicon retina sensors, and in 1989 Mahowald and Delbrück [8] presented a stereo matching approach using a cooperative algorithm, applied on event-based data using static and dynamic image features. The algorithm was completely implemented in hardware as an analog CMOS circuit, which is capable to find correspondences between image pairs, delivered by two 1D retina sensor, in real-time.

In the work of Schraml et al. [9] an implementation of an area-based stereo vision algorithm was proposed. Here, different similarity measures were evaluated using grayscale images generated from the event data, whereby the aggregation time is adjustable.

A detailed comparison of area-based matching techniques and a feature-based segment center matching algorithm was evaluated in the work of Kogler et al. [10]. Beside different similarity measures, here several conversion and aggregation methods were additionally analyzed.

One year later, Kogler et al. [11] presented an event-driven algorithm using weighted time differences as major correlation criterion for solving the correspondence problem. The calculation of the matching costs is done by summing up weights, which depend on the time difference of the events in the local neighborhood of the considered pixels. In doing so, different weighting functions such as, linear, quadratic, and Gaussian, were evaluated.

Another strategy for event-based matching was introduced in the work of Rogister et al. [7]. In this approach, events which are in a certain spatial distance to epipolar lines and in a predefined time window are matched, if they fulfill additional constraints, such as polarity equality, uniqueness, ordering, and temporal activity of a pixel.

In the work of Piatkowska et al. [12], an event-driven cooperative approach was presented. The algorithm was analyzed and evaluated in detail in [13] using real world scenarios with moving persons.

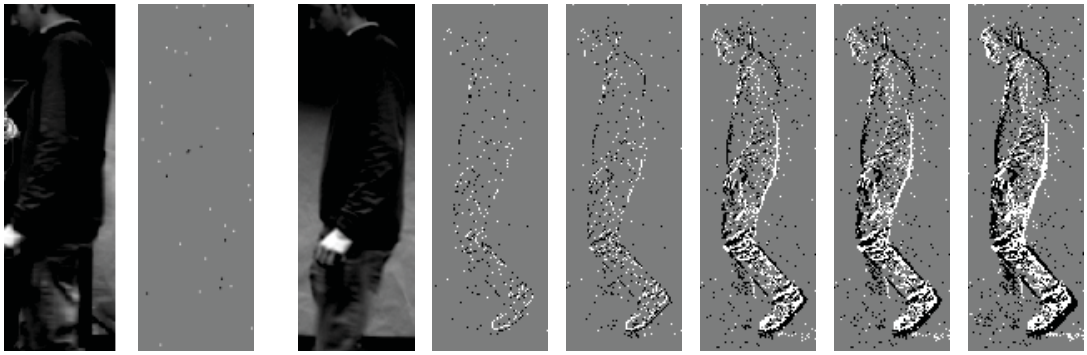



Fig. 1. Comparison of a conventional monochrome sensor and a silicon retina senor. The events are converted into grayscale images, where off-events are colored to black, on-events are mapped to white pixels, and inactive pixels are gray. (a) shows a non-moving person, therefore apart from some noisy pixels there is no output of the silicon retina as depicted in (b), If the person begins moving (a detailed view is shown in (c)), the silicon retina sensor continuously delivers events as illustrated in (d)–(h) Here the events are collected over a time period of 5 ms, 10 ms, 30 ms, 40 ms and 60 ms.

An event-driven panoramic vision sensor was introduced in the work of Belbachir et al. [14]. Here, two silicon retina sensors are mounted on a rotating device, thus the sensor can record a $360^\circ$ view. An event-driven and real-time capable stereo matching system based on this sensor was presented by Schraml et al. [15].

Recently, in the work of Firouzi and Conradt [16] an event-driven stereo matching approach was introduced, using a dynamic cooperative neural network, which processes each incoming event without any buffering.

## IV. Event-driven Stereo Matching

The hardware architecture of the implemented stereo vision core is based on the concept outlined in [17]. As a matter of fact, the pixels stimulated by a moving object which causes illumination changes do not fire simultaneously but asynchronously, and thus, are distributed over several timestamps. An evaluation of the time-delayed event generation behavior of the single pixel is also discussed in the work of Carneiro et al. [18] and in [19]. Hence, the edges and surface textures become complete over a period of several timestamps. Furthermore, this period of time also depends on the ratio between objects' motion speed and the temporal resolution of the silicon retina. Consequently, for the correspondence search, not only events from the current timestamp must be evaluated, but also events that occurred in a certain time window.

In order to exploit the activity based event generation of the silicon retina, the image plane is divided into blocks of a predefined size $m{\times}n$ called segments, e. g., for this work we used segments with a size of $16{\times}16$ pixels. As a consequence, the correspondence search is done only for image fields with a certain event activity and not for each pixel of the whole image. The benefit of this approach is twofold: The correspondence search is only done for regions of interests where events occurred, combined with the advantages of significantly minimizing memory accesses and an efficient parallel memory architecture. Both leads to a faster matching process and in addition to a better utilization of the matching pipeline.

A brief description of the stereo matching approach is shown in Algorithm 1. New incoming events from the left $l$ and the right $r$ sensors are firstly written to an array of image memories $B_{his} \in \mathbb{R}^{c\times b\times m\times n}$ recognizing a certain time window or history $t_{his}$, where $c$ refers to the available channels (left or right), $b$ denotes the amount of segments, and $m$ and $n$ specify the dimensions of a single segment.

$$\begin{array}{l} B_{his}[g][k][i,j] = e(p,t) \\ \forall i \in \{0,\ldots,m-1\} \wedge j \in \{0,\ldots,n-1\} \\ \wedge k \in \{0,\ldots,b-1\} \wedge g \in \{l,r\} \wedge t > t_{cur} - t_{his} \end{array} \quad , \quad (3)$$

where $g$ defines the channel (left or right), $k$ is the segment number, $i$ and $j$ are the coordinates within a segment, $t_{cur}$ is the current time, and $t_{his}$ the recognized history. The spatial position $p$ of an event is mapped onto entries in $B_{his}$ according to the rule

$$\begin{array}{l} p \mapsto k,i,j := \{i = x \bmod m, j = y \bmod n, \\ k = \lfloor \frac{x}{m} \rfloor + \lfloor \frac{y}{n} \rfloor \cdot \lfloor \frac{w}{m} \rfloor \} \end{array} \quad , \quad (4)$$

where $w$ again denotes the width of the sensor array in pixels. Furthermore, Algorithm 1 works on segment level which leads to

$$B_{seghis}(g,k) = B_{his}(g,k,[0:m-1],[0:n-1]) \quad , \quad (5)$$

Events which are older than $t_{his}$ are deleted, and additionally, old events are overwritten by new ones if they occur at the same spatial location. Subsequently, by counting the events for each single segment in the left image memories, a segment queue $q \in \mathbb{R}^b$ is determined. Since the left sensor is used as reference for the depth map, this is not done for the right channel. The amount of events determines not only the order of the segments for the matching process, therefore the queue is sorted descending order, but can also be used for noise filtering. This is done by rejecting all segments from the queue, where the event count is below a certain threshold.

**Algorithm 1** Event-based Stereo Matching Algorithm

**Require:** Two retinas $R_l, R_r$
**Require:** rectified event streams $E_l$ and $E_r$
  **for all** events $e_l(p,t)$ in $E_l$ **do**
    Build history by
    $B_{seghis}(k,l) = \text{merge}(B_{seghis}(k,l), e_l(p,t))$;
    Determine segment queue by
    $q = \text{detEventCount}(B_{seghis}([0:k-1],l))$;
    Prioritize queue by sorting $q_s = \text{sort\_desc}(q)$;
  **end for**
  Build history for right event stream
  $B_{seghis}(k,r) = \text{merge}(B_{seghis}(k,r), e_r(p,t))$;
  **for all** segments $s_i$ in $q_s$ **do**
    Compute disparities according to
    $D = \text{match}(B_{seghis}(s_i,l), B_{seghis}([0:k-1],r)$;
  **end for**
  Do consistency check $D_{cc} = \text{check}(D)$;
  **return** $D_{cc}$

Afterwards, the stereo matching and the consistency check are done, and finally the resulting disparity map $D_{cc}$ is returned.

A more detailed description of the event-driven stereo matching approach and its implementation can be found in [4], [19].

## V. Experimental Results

The event-driven stereo algorithm was evaluated using real world indoor scenarios, where the sensor was static and observes a dynamic scene. To enable a pixel-wise ground truth data generation from arbitrary scenes, a reference stereo camera system is used, which is mounted together with the silicon retina stereo sensor. The depth data from both systems are registered onto each other, enabling the calculation of the average distance error in depth. More information about the reference stereo algorithm and the ground truth evaluation system can be found in work of Humenberger et al. [20] and Kogler et al. [21].

In Figure 2a the test scenario captured by the reference camera system is shown. Here, two persons, one in a distance of $2.5\,\mathrm{m}$, and the second one in a distance of $3.5\,\mathrm{m}$, walking in parallel to the stereo system. The resulting depth data achieved by the reference stereo algorithm is depicted in Figure 2b, where the distance in [m] is encoded by the color – from red (background) to dark blue (nearest distance).

The results of the proposed event-driven stereo algorithm in terms of the average distance error, given in meters [m], is illustrated in Figure 3. In doing so, the size of the correlation window has been changed between 3×3, 9×9, and 15×15, a disparity range of 30 levels is considered, and the history length has been varied between 200, 250, and 300 timestamps. For analyzing the impact of the consistency check on the accuracy of the resulting depth values, it is once disabled (CC Off, red circles) and once enabled (CC On, green triangles), respectively.

For comparison, the results of an area-based stereo matching algorithm using the sum of absolute differences (SAD) as similarity measure are also depicted (see the blue squares in Figure 3). This algorithm uses grayscale images which are generated from the event streams before, and no consistency check is done in this case [22].

As the results shows, if a higher history length is used, the accuracy increases. This is attributable to one reason: the longer the considered history, the more and more events are available which leads to a more confident match. Also clearly visible is the improvement of the results by enlarging the correlation window. As a consequence of the sparse distribution of the events, the probability to match the right event pattern is significant higher if a larger neighborhood is considered. The effect of the consistency check on the quality of the matching results is also remarkable. Here, it is quite interesting that the consistency check leads to the almost same depth error for a 9×9, and 15×15 correlation patch, but it is disabled the bigger search window yields remarkably better results.

The comparison of these results with those achieved by the SAD based stereo matching algorithm clearly shows that the proposed algorithms provides significantly better results in any case. This can be attributed to the correlation approach, using spatial as well as time information of the events, which is particularly robust in the case of sparse input data as provided by silicon retinae.

In addition to the distance error, the density of the depth map and the reliability of the disparity values are also evaluated. These ratios are expressed by $R_D$ and $R_E$. The first metric describes the relation between all input events appeared in the considered history and the amount of calculated disparities and is expressed by

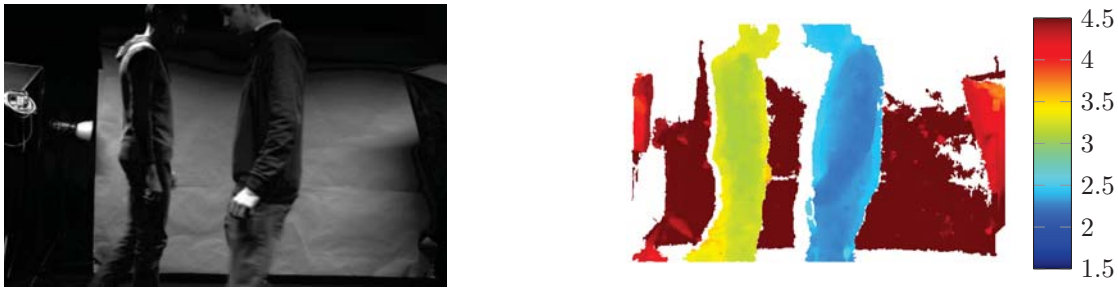
$$R_D = \frac{\text{Calculated Disparites}}{\text{All Input Events}} \quad . \tag{6}$$

The second measure is the ratio of depth results recognized for verification with the ground truth data and calculated disparities, and is calculated by

$$R_E = \frac{\text{Evaluated Disparites}}{\text{Calculated Disparities}} \quad . \tag{7}$$

As shown in Table I, the results are compared with some selected stereo matching algorithms proposed by Kogler [23]. For calculating these results a history length of 300 timestamps has been used for the event-driven matcher and the time correlation approach, the event transform and feature matching algorithms have been tested with a history length of 400 timestamps. All algorithms used a correlation window of 9×9 pixels. Apart from the time correlation approach, which is a line-based stereo matching algorithm.

As a consequence of the frame less and the event-driven mode of operation of the proposed stereo vision algorithm, common performance metrics describing the processing speed of the systems as, e. g., fps or million disparity evaluations per second (Mde/s), both are based on frames per second, are




(a) (b)

Fig. 2. Moving persons, one in a distance of 2.5 m (blue), and the second one in a distance of 3.5 m (green). The grayscale image depicted in (a) was captured by the reference stereo camera system. For the evaluation of the average distance error, the depth map shown in (b) is used as ground truth. Here, the color encodes the distance in [m], while pixels colored in red shades are mapped to the background.

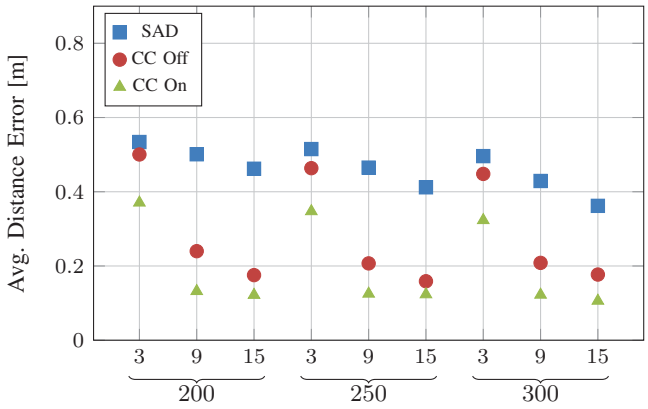



Fig. 3. Average distance error of the event-driven approach compared with an area-based SAD algorithm. A history length of 200, 250, and 300 timestamps has been used, where the correlation window ranges from 3×3, 9×9, to 15×15 pixels. In hardware the consistency check was enabled, whereby a no deviation between the left to right and right to left analysis has been allowed. The results of the SAD based approach have not been optimized by a consistency check.

TABLE I
COMPARISON OF AVERAGE DISTANCE ERROR, $R_D$ AND $R_E$ OF THE PROPOSED EVENT-DRIVEN ALGORITHM WITH SOME SELECTED STEREO MATCHING APPROACHES PROPOSED IN THE WORK OF KOGLER [23].

| Algorithm | Avg. Error [m] | $R_D$ [%] | $R_E$ [%] |
|---|---|---|---|
| Event-driven Matcher (HW) | 0.208 | 97.2 | 79.5 |
| Event Transform | 0.292 | 77.8 | 84.1 |
| Time Correlation | 0.594 | 84.0 | 78.3 |
| Feature Matcher | 0.220 | 3.4 | 97.8 |

difficult to use. However, the latter considers both, image resolution and the disparity range, and thus, it is more convincing. This metric is defined as follows [20]:

$$\mathrm{Mde/s} = 10^{-6} \cdot \mathrm{width} \cdot \mathrm{height} \cdot d_{range} \cdot \mathrm{fps} \quad , \qquad (8)$$

where the spatial image resolution is depicted by width and height, the disparity range by $d_{range}$, and fps is the number of frames per second. The factor $10^{-6}$ is caused by the normalization to million. If a non-iterative and image content unaware stereo algorithm is used, this rate remains constant for different image resolutions. At this point it should be noted that Mde/s refers to the number of finally calculated disparity results, thus the consistency check as an impact on the results.

To get a sense of the performance, the Mde/s was calculated for the case that all segments must be processed – this means the hole image plane must be processed which is in fact the worst case. Again, the correlation window size ranges between 3×3, 9×9, to 15×15 pixels and a disparity range of 30 levels is assumed. The results are listed in Table II. For the sake of completeness, in the case of using the ATIS the fps are between 2383.68 and 484.08. Here, fps is based on the duration calculating a single segment projected to the whole image plane considering the parallelism of the matching pipeline. As evident form real world test scenarios, the calculation of the whole image plane is unlikely even at longer histories. However, side effects regarding image borders and the utilization of the matching pipeline are crucial for the performance.

TABLE II
PERFORMANCE OF THE PROPOSED STEREO VISION ALGORITHM IMPLEMENTED IN HARDWARE MEASURED IN Mde/s. HERE, THE WORST CASE IS ASSUMED WHERE ALL SEGMENTS MUST BE PROCESSED, AND THE DISPARITY RANGE IS ASSUMED WITH 30 LEVELS.

| | CC On | CC Off |
|---|---|---|
| 3×3 | 879.97 | 5217.39 |
| 9×9 | 633.19 | 3699.80 |
| 15×15 | 484.03 | 2834.50 |

Humenberger analyzed in his thesis [20] some frame-based algorithms concerning the Mde/s, e. g., the Mobile Ranger, a SAD based stereo matching algorithm implemented on an FPGA, achieves 996.24 Mde/s even at 30 fps. A census transform based algorithm introduced by Humenberger [20] is evaluated by Jung [24] on different platforms. The achieved processing speed ranges from 1350 to 31.78 Mde/s, depending on the particular hardware platforms, which are a desktop CPU and GPU as well as an embedded CPU and GPU. In comparison with these platforms, the achieved Mde/s rate of the proposed hardware architecture is in the same range or still higher, even though the clock frequency of 100 MHz is relatively low with respect to the clock speed of a CPU or GPU.

Turning the attention to the consumption of hardware resources, synthesizing the proposed stereo algorithm for an Altera Stratix IV FPGA as possible target platform. The amount of required resources, as there are: logic elements, registers and memory bits, mainly depend on the spatial resolution of the used sensor system, the size of the segments, the size of the correlation window, the disparity range, and the considered history length  keeping in mind that in the current solution only on-chip memories are used in order to reduce the latency caused by data storage and handover to a minimum. The synthesis results of the stereo vision core using, e. g., a correlation window of $9\times9$ pixels, considering a history length of 300 timestamps, 30 disparity levels, and a spatial resolution of $304\times240$ pixels are: 43 581 registers, $1.45 \cdot 10^5$ logic elements, and $1.77 \cdot 10^6$ memory bits.

## VI. Conclusion

In this work a hardware implementation of a novel event-driven stereo matching algorithm has been developed. Due to the segmentation of the sensor's image plane, only active areas have to be processes which leads not only to a short latency and high performance, but also exploits the event-driven asynchronous behavior of the silicon retinae. The proposed method was verified with ground truth data and was compared to a standard stereo matching method and state-of-the-art event matching algorithms. The results showed that our approach outperforms these existing algorithms in terms of accuracy, density of the achieved depth map, and reliability of the found matches. Compared to other embedded computing platforms, the massively parallel and pipelined hardware architecture obtain a higher calculation performance, even at a lower clock frequency.

## References


[1] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 510–517.

[2] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Current Opinion in Neurobiology*, vol. 14, no. 4, pp. 481–487, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959438804001035

[3] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, June 2010.

[4] F. Eibensteiner, J. Kogler, and J. Scharinger, "A High-Performance Hardware Architecture for a Frameless Stereo Vision Algorithm Implemented on a FPGA Platform," in *Proceedings of the 10th IEEE Embedded Vision Workshop EVW (held in conjunction with IEEE CVPR)*, Columbus/USA, 2014, pp. 637–644.

[5] C. Posch, D. Matolin, and R. Wohlgenannt, "An asynchronous time-based image sensor," in *Proceedings of the IEEE International Symposium on Circuits and Systems ISCAS*, Seattle/USA, 2008, pp. 2130–2133.

[6] C. Posch, D. Matolin, and R. Wohlgenannt, "High-DR frame-free PWM imaging with asynchronous AER intesity encoding and focal-plane temporal redundancy suppression," in *Proceedings of the IEEE International Symposium on Circuits and Systems ISCAS*, Paris/France, June 2010, pp. 2430–2433.

[7] P. Rogister, R. Benosman, S.-H. Ieng, P. Lichtsteiner, and T. Delbruck, "Asynchronous event-based binocular stereo matching," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 347–353, february 2012.

[8] M. Mahowald and T. Delbrück, "Cooperative Stereo Matching Using Static and Dynamic Image Features," in *Analog VLSI Implementation of Neural Systems*, ser. The Kluwer International Series in Engineering and Computer Science, C. Mead and M. Ismail, Eds. Springer US, 1989, vol. 80, pp. 213–238. [Online]. Available: http://dx.doi.org/10.1007/978-1-4613-1639-8_9

[9] S. Schraml, P. Schön, and N. Milosevic, "Smartcam for Real-Time Stereo Vsion - Adress-Event Based Embedded System," in *Proceedings of the 2nd International Conference on Computer Vision Theory and Applications VISAPP*, vol. 2, Barcelona/Spain, 2007, pp. 466–471.

[10] J. Kogler, C. Sulzbachner, and W. Kubinger, "Bio-Inspired Stereo Vision System with Silicon Retina Imagers," in *Proceedings of the 7th International Conference on Computer Vision Systems ICVS*. Liege/Belgium: Springer-Verlag Berlin Heidelberg, 2009, pp. 174–183.

[11] J. Kogler, C. Sulzbachner, F. Eibensteiner, and M. Humenberger, "Address-Event Matching for a Silicon Retina Based Stereo Vision System," in *Proceedings of the 4th International Conference from Scientific Computing to Computational Engineering IC-SCCE*, Athens/Greece, 2010, pp. 17–24.

[12] E. Piatkowska, A. N. Belbachir, and M. Gelautz, "Asynchronous stereo vision for event-driven dynamic stereo sensor using an adaptive cooperative approach," in *Proceedings of the 3rd Workshop on Consumer Depth Cameras for Computer Vision CDC4CV (held in conjunction with IEEE ICCV)*, Sydney/Australia, 2013, pp. 45–50.

[13] E. Piatkowska, A. N. Belbachir, and M. Gelautz, "Cooperative and asynchronous stereo vision for dynamic vision sensors," *Measurement Science and Technology*, vol. 25, no. 5, pp. 1–8, May 2014. [Online]. Available: http://stacks.iop.org/0957-0233/25/i=5/a=055108?-key=crossref.932c79d8a42320a9f3e178487cd9f29a

[14] A. Belbachir, S. Schraml, M. Mayerhofer, and M. Hofstatter, "A Novel HDR Depth Camera for Real-Time 3D 360° Panoramic Vision," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2014, pp. 425–432.

[15] S. Schraml, A. N. Belbachir, and H. Bischof, "Event-driven stereo matching for real-time 3d panoramic vision," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 466–474.

[16] M. Firouzi and J. Conradt, "Asynchronous Event-based Cooperative Stereo Matching Using Neuromorphic Silicon Retinas," *Neural Processing Letters*, vol. 43, no. 2, pp. 311–326, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11063-015-9434-5

[17] F. Eibensteiner, J. Kogler, C. Sulzbachner, and J. Scharinger, "Stereo-Vision Algorithm Based on Bio-Inspired Silicon Retinas for Implementation in Hardware," in *Proceedings of the 13th International Conference on Computer Aided Systems Theory EUROCAST*, ser. Lecture Notes in Computer Science, Las Palmas/Spain, 2011, pp. 624–631.

[18] J. Carneiro, S.-H. Ieng, C. Posch, and R. Benosman, "Event-based 3D reconstruction from neuromorphic retinas," *Journal of Neural networks*, vol. 45, pp. 27–38, september 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23545156

[19] F. Eibensteiner, "Hardware Architecture of an Event-Driven Stereo Vision Algorithm Based on Silicon Retina Sensors," PhD-Thesis, Johannes Kepler Universität Linz, 2016.

[20] M. Humenberger, C. Zinner, M. Weber, W. Kubinger, and M. Vincze, "A fast stereo matching algorithm suitable for embedded real-time systems," *Journal of Computer Vision and Image Understanding*, vol. 114, no. 11, pp. 1180–1202, 2010.

[21] J. Kogler, F. Eibensteiner, M. Humenberger, M. Gelautz, and J. Scharinger, "Ground Truth Evaluation for Event-Based Silicon Retina Stereo Data," in *Proceedings of the 9th IEEE Embedded Vision Workshop EVW (held in conjunction with IEEE CVPR)*, Portland/USA, 2013, pp. 649–656.

[22] J. Kogler, F. Eibensteiner, M. Humenberger, C. Sulzbachner, M. Gelautz, and J. Scharinger, "Enhancement of Sparse Silicon Retina-Based Stereo Matching Using Belief Propagation and Two-Stage Postfiltering," *Journal of Electronic Imaging*, vol. 23, no. 4, p. 043011, 2014. [Online]. Available: http://dx.doi.org/10.1117/1.JEI.23.4.043011

[23] J. Kogler, "Design and Evaluation of Stereo Matching Techniques for Silicon Retina Cameras," PhD-Thesis, Vienna University of Technology, 2016.

[24] R. Jung, "Echtzeit Stereo-Matching auf einer Embedded GPU Plattform," Masters's Thesis, Upper Austria University of Applied Sciences, Campus Hagenberg, 2015.