Improved Cooperative Stereo Matching for Dynamic Vision Sensors with Ground Truth Evaluation

Ewa Piatkowska[1], Juergen Kogler[2], Ahmed Nabil Belbachir[3], and Margrit Gelautz[2]

[1]AIT Austrian Institute of Technology, Vienna, Austria
[2]Vienna University of Technology, Vienna, Austria
[3]Teknova AS, Grimstad, Norway
ewa.piatkowska@ait.ac.at, juergen.kogler@tuwien.ac.at
anb@teknova.no, margrit.gelautz@tuwien.ac.at

Abstract

Event-based vision, as realized by bio-inspired Dynamic Vision Sensors (DVS), is gaining more and more popularity due to its advantages of high temporal resolution, wide dynamic range and power efficiency at the same time. Potential applications include surveillance, robotics, and autonomous navigation under uncontrolled environment conditions. In this paper, we deal with event-based vision for 3D reconstruction of dynamic scene content by using two stationary DVS in a stereo configuration. We focus on a cooperative stereo approach and suggest an improvement over a previously published algorithm that reduces the measured mean error by over 50 percent. An available ground truth data set for stereo event data is utilized to analyze the algorithm's sensitivity to parameter variation and for comparison with competing techniques.

1. Introduction

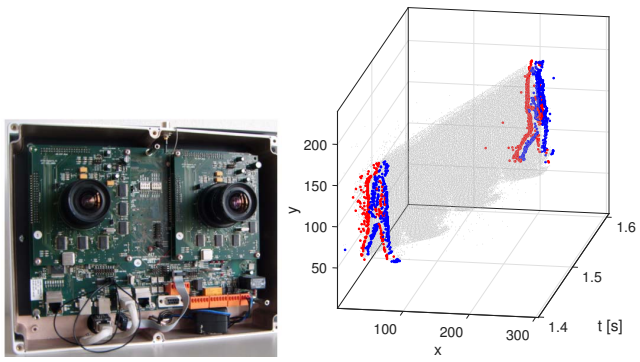
Recently, bio-inspired Dynamic Vision Sensors (DVS) [18, 22] have gained a lot of recognition, mainly due to their advantages over clocked (frame-based) cameras such as high-temporal resolution, wide dynamic range and low power consumption. These features make them perfectly suitable for new trends in robotics, such as high-speed motion analysis and tracking or autonomous navigation in uncontrolled environments. Exciting directions are opening now in event-based vision related to both novel sensor types (e.g. [2] [4]) as well as advanced processing of event data. This includes, for instance, advances in panoramic 3D vision [25, 12], retrieving information from moving handheld cameras for Simultaneous Localisation and Mapping (SLAM) [29] or camera pose estimation [13, 10].

Dynamic Vision Sensors differ from conventional sensors in their construction with respect to pixel circuits and chip architecture. The read-out of the pixel information is initiated by the pixel itself, by firing an event upon a detected relative change in light intensity. An event is defined by the spatial location and the time of its occurrence. Events can either be negative (off-event) or positive (on-event), depending on the polarity of the change of illumination over a period of time. Assuming a fixed DVS pose, the amount of delivered data is significantly reduced because only changes in the scene are detected.

While Dynamic Vision Sensors offer numerous advantages, they also introduce several challenges to the way the data have to be processed afterwards. Most of the well-established computer vision algorithms operate on pixels' values within images, therefore they cannot be explicitly applied to the stream of events. Moreover, a Dynamic Vision Sensor performs feature extraction on the pixel circuit level. Although this proves to be very efficient, it also can produce some amount of erroneous data (noise). Hence, event-based algorithms need to compensate for the limited information from the sensor by more sophisticated processing and reconstruction algorithms. Nonetheless, at the same time it is important to minimise the complexity and processing requirements.

In this paper we are tackling the task of 3D stereo reconstruction from a static stereo DVS setup (Figure 1a). We use a stereo Asynchronous Time-based Image Sensor (ATIS) [22], which has a spatial resolution of 304×240 pixels, a temporal resolution of up to 10ns, and a dynamic range of 143dB. An example of data captured with our stereo sensor setup is shown in Figure 1c. A sequence of frames captured by conventional camera output is compared with the image representation of the event data (events ac-

2160-7516/17 $31.00 © 2017 IEEE
DOI 10.1109/CVPRW.2017.51



(a) (b)

(c)

Figure 1: Event data sequence captured by a stereo DVS built out of two ATIS sensors [22].

cumulated over time with respect to polarity). As we can observe, only edges of the moving objects are captured. In Figure 1b, the same sequence of events is plotted as three-dimensional space-time cloud.

The main contribution of this paper is the development of an improved cooperative stereo matching algorithm which builds upon the original method presented in [21]. As a motivation for our proposed approach, we discuss and justify that single event to event matching based on temporal coincidence is not reliable in more complex scenes. We prove that using a simple window-based event matching can significantly improve the overall stereo algorithm's results and show that the enhanced cooperative algorithm can achieve comparable or better results than some competing techniques.

In the remainder of this paper, we first describe efforts made in the field of event-based stereo matching, focusing on works for static stereo dynamic vision sensors, in Section 2. In Section 3, we discuss the challenges of event to event matching, and present the proposed algorithm. In our experimental evaluation in Section 4, we analyse and evaluate our algorithm using a ground truth [15]. We present our conclusion in Section 5.

2. Related Work

In the literature, we can distinguish two major ways of processing visual event data: image-based (also called frame-based) and event-based algorithms. Members of the first group explore the possibility of adapting events to conventional computer vision algorithms. Events can be converted to images by aggregation over a specific time period, as presented in [26][17], or reconstructed to gray-scale images by more sophisticated algorithms, as shown in [6] or [1]. Once the events are encoded into an image form, conventional stereo algorithms can be applied. Schraml et al. [26] evaluated window-based stereo matching with different matching cost variants, achieving the best performance with Normalized Sum of Absolute Differences (NSAD). Kogler et al. [17] proposed a window-based stereo matching technique applied to a tri-logic image representation that stores the polarity of the last event generated at a particular position. More recently, Kogler et al. [16] suggested an adapted belief propagation and a two-stage filter (2SF) technique and demonstrated that the latter gave better results.

The second way of dealing with event data, the asynchronous processing, is to operate directly on the stream of events. The main challenge in event-based stereo matching is finding correspondences between events. In conventional stereo, two pixels are compared by their values (grayscale, color). Since events do not convey the absolute brightness values, the task becomes more difficult. In most of the approaches, events are compared by their occurrence in time, following the assumption that the same stimulus should trigger corresponding pixels in the left and right view at the same time. In the method proposed by Rogister et al. [24], matching candidates selected from events within a defined temporal window are compared by their Euclidean distance to the epipolar line. Additional constraints have been proposed to reduce matching ambiguity, e.g., matching only events of the same polarity and orientation or eliminating wrong matches by an ordering constraint [24]. Furthermore, in order to smooth the final results, the disparities of events can be averaged over time [17]. Carneiro et al. [5] apply Bayesian filtering to the initial matches projected into 3D space.

Another class of event-based approaches are cooperative stereo algorithms. Cooperative stereo [20] is one of the first algorithms describing the process of stereoscopic vision and its formulation by a computational model. The use of cooperative stereo for processing event data dates back to 1989, when first prototypes of stereo silicon retina were developed by Mahowald and Delbruck [19]. The authors tested the cooperative computation using single line sensors and achieved very good results. Following that, a software version of cooperative stereo for global disparity estimation was implemented by Hess in his semester thesis [11]. More recently, Piatkowska et al. [21] proposed a dynamic co-

operative network that adapts the disparity estimates with each incoming event. Cooperative stereo approaches have also been successfully used in classical computer vision for dense stereo matching, as presented in [30].

Among the most recent methods on event-based 3D reconstruction - even though not directly connected to the work presented in this paper - we can find panorama stereo vision from rotating, stereo line-DVS [26], multiview stereo from a single sensor [23] or simultaneous 3D reconstruction and 6-DoF tracking [13]. The two latter methods apply probabilistic estimation using e.g. Bayesian or Kalman filters to the events with reconstructed gray-scale values.

Some of the above-mentioned algorithms have already been realized in hardware platforms. The work presented in [27] was implemented on a DSP (digital signal processor) and later also on an FPGA (field programmable gate array) [7]. Regarding the event-based matching, the time correlation algorithm from [17] was realized on DSP by Sulzbachner et al. [28] and FPGA by Eibensteiner et al. [9].

3. Algorithm

The main controversy about the frame-based methods of event processing is that accumulation into fixed time frames is limiting the temporal resolution of the dynamic vision sensor. All events at a particular address that occur within one fixed time frame are assigned the same disparity value. On the other hand, frame-based processing is less computationally demanding, therefore eligible for real-time performance, which is highly demanded for many applications. Event-based methods perform matching of single events, based on their coincidence in time, including some other constraints such as polarity, orientation, order, etc. Solving the correspondence problem becomes challenging when dealing with more complex scenes captured by dynamic vision sensors such as objects moving at different speeds, highly textured objects, or cluttered scenes. In such cases, the single event matching based on temporal correlation of events can be insufficient to handle the ambiguities. In the following, we describe several situations when the temporal coincidence assumption could be violated.

3.1. Challenges of event based matching

The assumption that corresponding events should coincide in time between left and right view can be violated due to a number of reasons. As already mentioned in [24], there is no guarantee that corresponding events will have exactly the same timestamps due to hardware limitations. Firstly, the delay between visual stimulus and a pixel's response depends on jitter, which varies across pixels and platforms. Secondly, streaming events over a shared bus can introduce some delays to the event timestamps.

Temporal correspondence is prone to errors in more complex scenes, due to the perspective and relative velocity of the motion. Objects closer to the sensor are naturally projected as bigger, and their relative motion velocity as well as event rate are higher. On the contrary, objects that are farther from the sensor are projected as smaller and generate fewer events. Correspondence between left and right events of background objects may not be found because events from the foreground objects are more likely to be closer in time, thus having a higher weight in temporal matching.

There are many factors that influence the response of a pixel, including (i) external conditions, such as lighting, speed and direction of the moving object, color of the object (contrast); and (ii) internal parameters, called bias settings, that control the pixel's response time, sensitivity, spike rate, etc. It is very likely that these factors will not result in identical responses of the pixels from left and right. First, because bias settings differ across hardware platforms. Second, because the viewpoint makes a difference, i.e. an object's appearance varies while viewed from different angles, hence the generated event sequence would differ in the left and right sensor.

In addition, dealing with moving objects in the scene, the change detected by a pixel depends on the intensity of the object against the background. If the perceived background varies between the left and right view, the magnitude of illumination changes is different. In extreme cases, this can also result in different polarity of corresponding events.

Event-based methods as proposed in the literature allow for asynchronous processing but are still very prone to errors caused by ambiguities in matching. One way of improving the quality of stereo is finding an alternative way of measuring the matching cost. In our approach, we focus on event-based matching using a local neighborhood of events.

3.2. Enhanced cooperative stereo

As mentioned in Section 2, cooperative approaches for event matching have proved effective. In this paper, we are building on the work presented in [21]. In what follows, we briefly describe the cooperative algorithm and highlight and motivate our main changes.

Let $e = (e_x, e_y, e_t, e_p, e_c)$ represent the event by its location (x, y), time t, polarity p and camera c.

$$e = (e_x, e_y, e_t, e_p, e_c) \mid e_x, e_y \in \mathbb{N}, e_t \in \mathbb{R}, \\ e_p, e_c \in \{0, 1\} \tag{1}$$

The set of all events in the input stream is denoted by E. For each event, we search for the set of possible matching candidates M_e among events of the other view within a given disparity range (d_{min}, d_{max}), as defined in Eq. (2). Match-

ing is done symmetrically for the left and right events.

$$\forall_{e\in E} \quad M_e = \{m \mid m \in E, \\ d_{min} < |m_x - e_x| < d_{max}, \\ m_y = e_y, m_t < e_t, m_c = \neg e_c\} \tag{2}$$

In most of the event-based matching algorithms [11, 21] matching candidates are weighted by their similarity to the reference event e as defined in Eq. (3), which incorporates the parameter α. The similarity function ρ for each of the candidates is computed as:

$$\forall_{m\in M_e} \quad \rho(e,m) = \frac{1}{\alpha \cdot |e_t - m_t| + 1} \tag{3}$$

Since single event to event matching is not reliable, we have modified the matching cost function to compare events in their neighborhood (also referred to as matching window). The neighborhood N_e of event e is described by Eq. (4). The radius of matching window $mwin$ is given as an algorithm parameter.

$$N_e = \{n \mid n \in E, \|e_x - n_x\| < mwin, \\ \|e_y - n_y\| < mwin\} \tag{4}$$

For each of the matching candidates m, the window-based matching cost ρ_{win} is defined in Eq. (5).

$$\rho_{win}(N_e, N_m) = \frac{\sum_{\{(e,m)\in N_e \times N_m | e_x = m_x \wedge e_y = m_y\}} \rho(e,m)}{|N_e|} \tag{5}$$

Individual costs are calculated between each of the events in N_e and the corresponding event at the same position in the candidate's neighborhood N_m. Afterwards they are summed and normalized by the amount of events in N_e. As can be seen, we are still using Eq. (3) to compute the similarity between single events within neighborhoods.

For each incoming event, the set of possible matching candidates, weighted as described before, is mapped to the cooperative network as follows:

$$\forall_{m\in M_e} \quad C^*_{e_x,e_y,d} = \rho_{win}(N_e, N_m), \\ \text{where } d = |e_x - m_x| \tag{6}$$

The cooperative network is a three-dimensional structure $C = (X, Y, D)$, consisting of nodes $C_{x,y,d}$ for each disparity $d \in D$ and spatial location $(x, y), x \in X, y \in Y$. Nodes act independently and are connected to two types of neighborhoods: supporting Φ and inhibitory Ψ. The first one (Eq. (7)) implements the smoothness assumption, including support from nodes at the same disparity plane and within a given radius $swin$. Function $\Phi(x, y)$ returns indexes (x', y') of the cooperative network node at position (x, y).

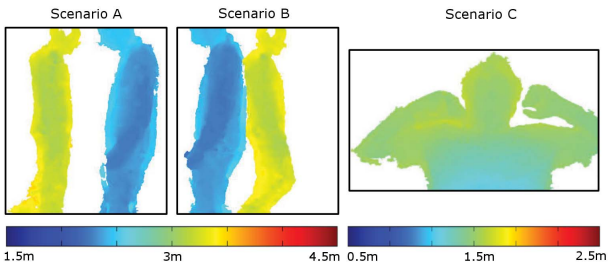


Figure 2: Ground truth test scenarios [15].

$$\Phi(x,y) : (x',y') \mid |x - x'| < swin \\ \wedge \; |y - y'| < swin \tag{7}$$

The second one, defined in Eq. (8), realizes the uniqueness assumption through competition between the candidate nodes along the disparity dimension. Function $\Psi(d)$ returns indexes d' of nodes in an inhibitory neighborhood of the cooperative node at position (x, y, d).

$$\Psi(d) : \; d' \mid d_{min} < |d - d'| < d_{max} \tag{8}$$

The cooperative network is constantly changing as the events are generated. Once candidates of an event e are mapped to the network with initial weights C^* from Eq. (6), the affected nodes $C_{e_x,e_y,d}$ are updated as follows:

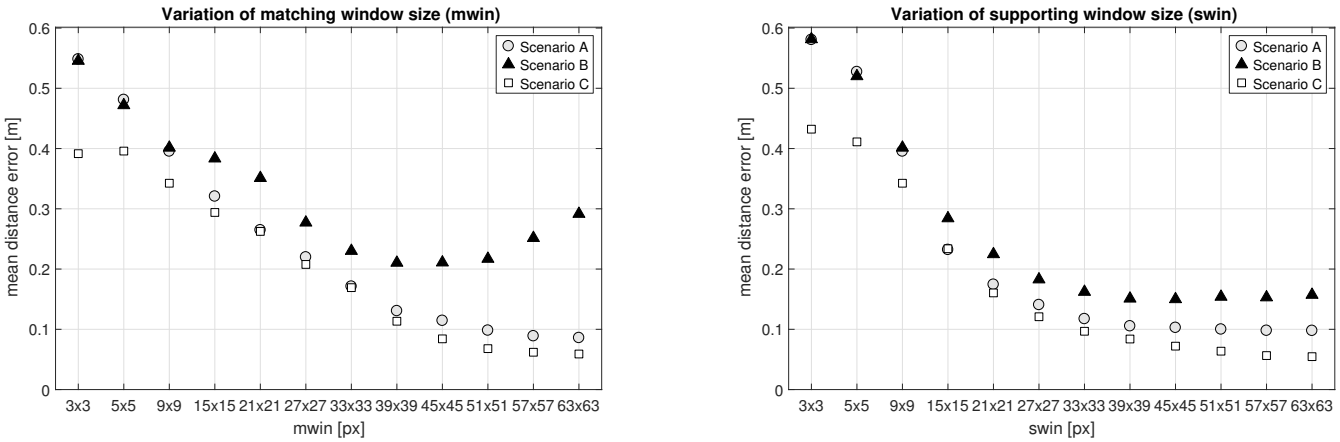
$$C^{n+1}_{e_x,e_y,d} = \left(\frac{\sum_{x',y' \in \Phi(e_x,e_y),d} C^n_{x',y',d} * C^*_{e_x,e_y,d}}{\sum_{x,y,d' \in \Psi(d)} C^n_{x,y,d'}} \right)^{\varepsilon} \tag{9}$$

We have slightly adjusted the cooperative update function, based on conclusions from [30]. Normalization is achieved by division by the sum of nodes from the inhibitory neighborhood. The parameter ε controls the amount of inhibition applied to the cooperative nodes values.

Furthermore, in [21] noise events were assumed to be filtered out by a density threshold in the cooperative network, however, this method was found not to be reliable. In our algorithm, the additional noise removal filter is employed at the stage of mapping an event to the cooperative network, taking into account the initial weights of the matches and density of the neighborhood.

4. Experimental Results

We present an experimental evaluation of the proposed cooperative stereo matching algorithm in terms of parameter selection, accuracy and comparison to competing algorithms. As reference for our quantitative assessment, we



(a) (b)

Figure 3: Results of the proposed stereo algorithm with varying a) matching window and b) supporting window sizes.

swin	15x15			27x27			33x33			39x39		
mwin	15x15	33x33	39x39	15x15	33x33	39x39	15x15	33x33	39x39	15x15	33x33	39x39
A	0.195	0.124	0.110	0.127	0.099	0.096	0.111	0.093	0.092	0.101	0.090	**0.089**
B	0.279	0.172	0.169	0.175	0.150	0.150	0.157	**0.146**	0.152	0.147	**0.146**	0.152
C	0.204	0.137	0.099	0.098	0.089	0.075	0.080	0.084	0.076	0.072	0.075	**0.069**

Table 1: Results given in mean distance error [m] of the cooperative stereo with different sizes of matching window (mwin) and supporting window (swin). Results are presented for all three test scenarios (A, B, C) and the best result per each scenario is shown in bold.

use the ground truth data set for stereo event data which was proposed by [15]. The data set comprises three test scenes - denoted as Scenario A, B, and C - as shown in Figure 2. All test scenarios present indoor scenes captured by a stationary stereo camera set-up. In Scenario A and B, two people are walking in opposite directions at different distances from the camera, with a small amount of occlusions present in Scenario B. Scenario C depicts a person sitting relatively close to the camera and moving the upper body and arms. The ground truth depth for the event data was generated using a conventional stereo system as reference, and its accuracy is estimated to be better than 0.027m for distances up to 2m, with errors increasing up to 0.117m at a distance of 3.5m [15]. For the calibration and rectification of the silicon retina stereo sensor, the calibration toolbox of Bouguet [3] is used, as described in more detail in [8].

4.1. Cooperative stereo analysis

We started our experiments by assessing the influence of the size of the matching window *mwin* and supporting window *swin*, which are the two key parameters within the cooperative network. Throughout our experiments we kept the parameters α, ε, d_{min} and d_{max} at constant values of 0, 0.7, 7, and 70, respectively.

As shown in Figure 3, the window sizes for both parameters were varied individually from 3×3 to 63×63, while the other parameter was kept constant at 9×9. For Scenario A and C, we can observe an improvement of the accuracy with increasing window sizes of over almost the whole tested range. For Scenario B, the plots for *mwin* and *swin* indicate a minimum value for window sizes around 39×39. For values larger than 39×39, further improvements gained for Scenario A and C are also relatively small. The significant gain in accuracy between window sizes of 9×9 and 39×39, with the associated mean errors changing from almost 0.4m to below 0.2m, was accompanied by an increase in runtime of a factor 3.06 (*mwin*) and 1.49 (*swin*) using our unoptimized Matlab code.

The next step of our analysis was to find the best combination of *mwin* and *swin* settings. Table 1 gives the mean error (in meters) achieved by the proposed cooperative stereo algorithm with *mwin*={*15×15, 33×33, 39×39*} and *swin*={*15×15, 27×27, 33×33, 39×39*}. The best results per each of the scenarios are highlighted. We observe

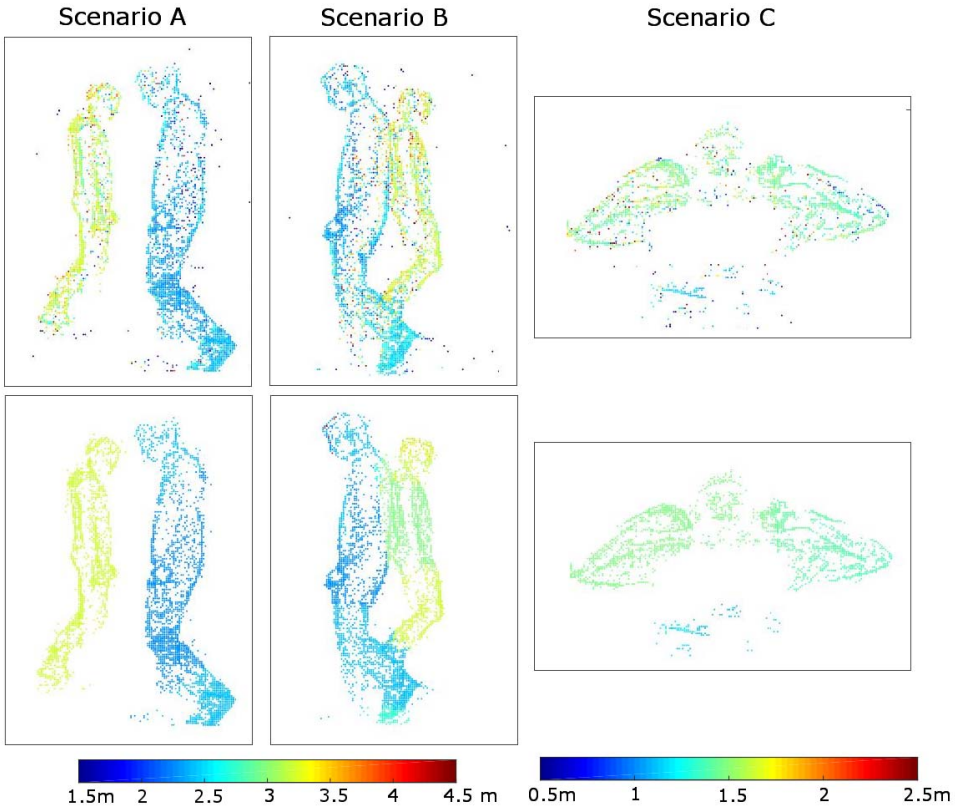


Figure 4: Qualitative results (depth maps) of the cooperative stereo algorithm. The first row presents results of the algorithm from [21], and the second row the results of the proposed improved algorithm.

that the results do not vary much across scenarios. As a consequence, we have selected 39×39 as the size for both *mwin* and *swin* in our further evaluation.

4.2. Comparative evaluation

We compare the results of our improved cooperative stereo algorithm with other frame-based and event-based matching algorithms in Table 2. The first algorithm *T-Corr* is a simple event to event matching method based on time correlation [17]. The second algorithm *SAD+2SF* was introduced in [16]. It is a frame-based matching approach which relies on Sum of Absolute Difference (SAD) as matching function and an additional post-processing technique, the two-stage filter (2SF), that incorporates median filtering. The error rates for *T-Corr* and *SAD+2SF* were taken from the literature [14] and [16], respectively. The third algorithm *CoopSt* [21] is the cooperative stereo using time-based single event matching. Its parameter *swin* is set to the window size of 39×39. Finally, the results of the proposed enhanced cooperative stereo algorithm, with window-based matching used for calculating the initial weights, are listed in the last column of Table 2.

	T-Corr [17]	SAD+2SF [16]	CoopSt [21]	proposed
A	0,581	0,119	0.290	**0.089**
B	0,618	0,222	0.350	**0.152**
C	0,277	0,088	0.207	**0.069**

Table 2: Results comparison using ground truth test scenarios A, B and C.

As expected, the highest mean distance error in Table 2 was found for the basic event to event matching technique *T-Corr*. Comparison of the two right-most columns demonstrates the clear gain in accuracy achieved by the proposed improvement strategy, with the average error (computed over all three scenarios) dropping by over 50 percent to 0,11m in the final result. The proposed cooperative matching also outperforms *SAD+2SF* with error differences much smaller in this case. The improvement achieved with respect to the initial version can also be seen on the depth maps in Figure 4, where noise visible in the initial results is noticeably reduced in the final results.

5. Conclusions

In this paper, we proposed an enhanced cooperative stereo matching technique that calculates similarity over a local neighborhood of each event pair to compute initial matching weights for the cooperative optimisation. We presented an evaluation based on ground truth data to assess the algorithm's reconstruction accuracy and sensitivity to parameter variation. We found that our algorithmic improvements reduced the measured mean error by over 50 percent to 0,11m, with constant parameter settings applied to all three test scenarios. A quantitative comparison with competing stereo algorithms confirmed the usefulness of the proposed approach. As part of future work, further improvements may be incorporated directly into the cooperative network, e.g., by using adaptive local neighborhood operations. Also, additional post-processing techniques could be applied. Furthermore, the algorithm could be evaluated on longer sequences when suitable extensions of the ground truth datasets become available.

References

[1] P. A. Bardow, A. J. Davison, and S. Leutenegger. Simultaneous optical flow and intensity estimation from an event camera. In *Computer Vision and Pattern Recognition (CVPR)*, 2016.

[2] A. N. Belbachir, S. Schraml, M. Mayerhofer, and M. Hofstätter. A novel HDR depth camera for real-time 3D 360° panoramic vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 425–432, 2014.

[3] J. Y. Bouguet. Camera calibration toolbox for MATLAB. Published in the Internet, 2008. Computer Vision Research Group/Department of Electrical Engineering/California Institute of Technology - `www.vision.caltech.edu/bouguetj/calib_doc/index.html` (Accessed 17 May 2017).

[4] C. Brandli, R. Berner, M. Yang, S. C. Liu, and T. Delbruck. A 240×180 130dB 3μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341, 2014.

[5] J. Carneiro, S.-H. Ieng, C. Posch, and R. Benosman. Event-based 3D reconstruction from neuromorphic retinas. *Neural Networks*, 45:27–38, 2013.

[6] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger. Interacting maps for fast visual interpretation. In *International Joint Conference on Neural Networks*, pages 770–776, 2011.

[7] F. Eibensteiner, A. Gschwandtner, and M. Hofstätter. A high-performance system-on-a-chip architecture for silicon-retina-based stereo vision systems. In *International Congress on Computer Application and Computational Science (IRAST)*, pages 976 – 979, 2010.

[8] F. Eibensteiner, J. Kogler, M. Schörghuber, and J. Scharinger. Automated stereo calibration for event-based silicon retina imagers. In *Proceedings of the 6th International Conference from Scientific Computing to Computational Engineering (IC-SCCE)*, Athens/Greece, 2014.

[9] F. Eibensteiner, J. Kogler, C. Sulzbachner, and J. Scharinger. Stereo-vision algorithm based on bio-inspired silicon retinas for implementation in hardware. In *13th International Conference on Computer Aided Systems Theory (EUROCAST)*, Lecture Notes in Computer Science, pages 624–631, 2011.

[10] G. Gallego, J. E. A. Lund, E. Mueggler, H. Rebecq, T. Delbruck, and D. Scaramuzza. Event-based, 6-DOF camera tracking for high-speed applications. *arXiv:1607.03468 [cs]*, 2016.

[11] P. Hess. Low-level stereo matching using event-based silicon retinas, 2006. Semester Thesis, ETH Zurich.

[12] H. Kim, A. Handa, R. Benosman, S.-H. Ieng, and A. J. Davison. Simultaneous mosaicing and tracking with an event camera. In *British Machine Vision Conference (BMVC)*, 2014.

[13] H. Kim, S. Leutenegger, and A. J. Davison. Real-time 3D reconstruction and 6-DoF tracking with an event camera. In *European Conference on Computer Vision (ECCV)*, 2016.

[14] J. Kogler. *Design and Evaluation of Stereo Matching Techniques for Silicon Retina Cameras*. PhD thesis, Vienna University of Technology, 2016.

[15] J. Kogler, F. Eibensteiner, M. Humenberger, M. Gelautz, and J. Scharinger. Ground truth evaluation for event-based silicon retina stereo data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 649–656, 2013.

[16] 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*, 23(4):043011, 2014.

[17] J. Kogler, C. Sulzbachner, F. Eibensteiner, and M. Humenberger. Address-event matching for a silicon retina based stereo vision system. In *4th International Conference from Scientific Computing to Computational Engineering (IC-SCCE)*, pages 17–24, 2010.

[18] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128×128 120dB 30mW asynchronous vision sensor that responds to relative intensity change. In *IEEE International Solid State Circuits Conference - Digest of Technical Papers (ISSC)*, pages 2060–2069, 2006.

[19] M. A. Mahowald and T. Delbruck. Cooperative stereo matching using static and dynamic image features. In C. Mead and M. Ismail, editors, *Analog VLSI Implementation of Neural Systems*, number 80 in The Kluwer International Series in Engineering and Computer Science, pages 213–238. Springer US, 1989.

[20] D. Marr and T. Poggio. Cooperative computation of stereo disparity. In L. Vaina, editor, *From the Retina to the Neocortex*, pages 239–243. Birkhuser Boston, 1976.

[21] E. Piatkowska, A. N. Belbachir, and M. Gelautz. Asynchronous stereo vision for event-driven dynamic stereo sensor using an adaptive cooperative approach. In *IEEE International Conference on Computer Vision (ICCV) Workshops*, 2013.

[22] C. Posch, D. Matolin, and R. Wohlgenannt. A QVGA 143dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. *IEEE Journal of Solid-State Circuits*, 46(1):259–275, 2011.

[23] H. Rebecq, G. Gallego, and D. Scaramuzza. EMVS: Event-based multi-view stereo. In *British Machine Vision Conference (BMVC)*, 2016.

[24] 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*, 23(2), 2012.

[25] S. Schraml, A. N. Belbachir, and H. Bischof. Event-driven stereo matching for real-time 3D panoramic vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 466–474, 2015.

[26] S. Schraml, A. N. Belbachir, N. Milosevic, and P. Schön. Dynamic stereo vision system for real-time tracking. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1409–1412, 2010.

[27] S. Schraml, P. Schön, and N. Milosevic. Smartcam for real-time stereo vision - address-event based embedded system. In *2nd International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 466–471, 2007.

[28] C. Sulzbachner, C. Zinner, and J. Kogler. An optimized silicon retina stereo matching algorithm using time-space correlation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1–7, 2011.

[29] D. Weikersdorfer, D. B. Adrian, D. Cremers, and J. Conradt. Event-based 3D SLAM with a depth-augmented dynamic vision sensor. In *IEEE International Conference on Robotics & Automation (ICRA)*, pages 359–364, 2014.

[30] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):675–684, 2000.