Event-Driven Stereo Matching for Real-Time 3D Panoramic Vision

Stephan Schraml[1], Ahmed Nabil Belbachir[1], Horst Bischof[2]
[1]AIT Austrian Institute of Technology, Digital Safety & Security Department, New Sensor Technologies, Austria. [2]Graz University of Technology, Institute for Computer Graphics and Vision, Austria.

ABSTRACT. This paper presents a novel stereo matching approach for a multi-perspective panoramic stereo vision system, making use of asynchronous and non-simultaneous stereo imaging towards real-time 3D 360° vision. The method is designed for events representing the scenes visual contrast as a sparse visual code allowing the stereo reconstruction of high resolution panoramic views. We propose a novel cost measure for the stereo matching, which makes use of a similarity measure based on event distributions. Thus, the robustness to variations in event occurrences was increased. An evaluation of the proposed stereo method is presented using distance estimation of panoramic stereo views and ground truth data. Furthermore, our approach is compared to standard stereo methods applied on event-data. Results show that we obtain 3D reconstructions of 1024×3600 round views and outperform depth reconstruction accuracy of state-of-the-art methods on event data.

The multi-perspective panoramic stereo vision system consists of a rotating pair of dynamic vision line sensors [1] generating streams of asynchronous events.

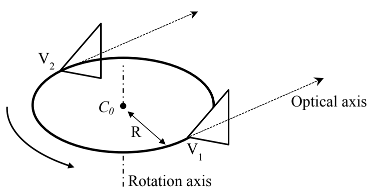

Figure 1: 360° Panoramic stereo vision setup.

In a first step the spatial context from event-streams is restored by transforming the events' timestamp t and address y information into image coordinates, which are used to render a panoramic event map $E(x,y)$. We then define a measure - the *Non-Zero-Distance* (*NZD*) function - that describes the local distribution of events within a segment of E. Segments of E and *NZD* are used to calculate the similarity between two event-sequences:

$$C^{L,R}(x,y,d) = \sum_{w \in W} E^L(x+w,y) * NZD^R(x+w+d,y) \tag{1}$$

where superscripts L,R denote the left and right sensor respectively and W is the patch window size. The principle is illustrated in Figure 2. The partial cost is equivalent to the sum of minimal distances between event positions. In order to maintain symmetry, partial cost is also similarly calculated with E and NZD exchanged:

$$\hat{C}^{L,R}(x,y,d) = \sum_{w \in W} NZD^L(x+w,y) * E^R(x+w+d,y) \tag{2}$$

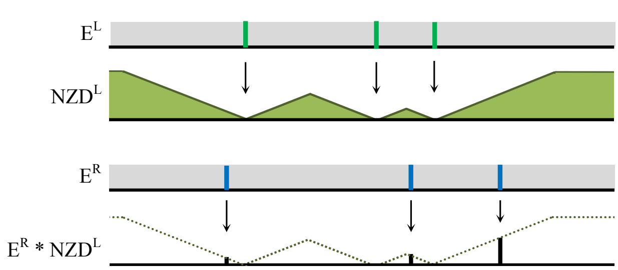

Figure 2: Illustration of the similarity measure. Partial cost calculation is based on minimal distances between event positions. Events are indicated by vertical bars.

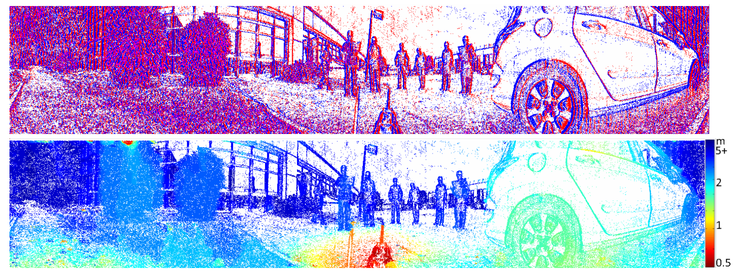

Figure 3: Events from a panoramic view showing the event's polarity (top); ON-events (red) and OFF-events (blue). Event-driven stereo reconstruction result of a real-world scenario recorded at 10 pan/s (bottom).

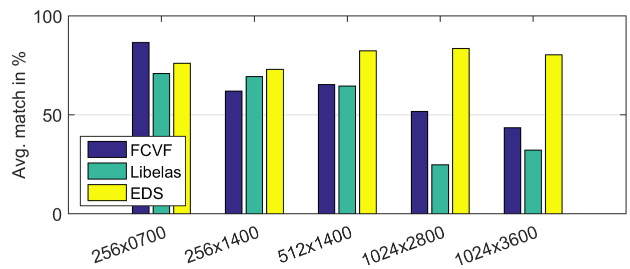

Figure 4: Comparison of average match performance vs. image resolution using two standard stereo matching methods (FCVF [3] and Libelas [2]) applied on event-data and our event-driven stereo approach (EDS).

Finally, the cost of a match is modeled as the sum of these partial costs.

$$C = \begin{cases} C^{L,R} + \hat{C}^{L,R}, & \text{if } (n^L \geq \tau) \wedge (n^R \geq \tau) \\ C_{max}, & \text{else} \end{cases} \tag{3}$$

with $n^j = \sum_{w \in W} E^j(x,y)$ the number of events in a segment and τ is the minimum event count. The matching cost for segments that do not contain enough events is set to a maximum cost C_{max}. In order to reduce ambiguity, a cross-check is performed by exchanging the role of the left (L) (reference) and right (R) sensor.

The stereo matching approach based on non-simultaneous event-driven vision is described in more detail in the paper, as is the evaluation of the proposed method. We found out that, thanks to the novel cost measure, our tailored event-driven stereo approach accurately reconstructs the 3D information of event-data over a wide range of sparsity and parsimony. It thereby outperforms standard state-of-the-art stereo methods on sparse event-data, particularly for real-time 3D 360° vision with high-image resolutions, starting from 0.7 Mpixels. Results on the natural scene show the usability of the method for applications in a natural environment with varying environmental conditions.

[1] A.N. Belbachir, S. Schraml, M. Mayerhofer, and M. Hofstätter. A novel hdr depth camera for real-time 3d 360-degree panoramic vision. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 419–426, June 2014.

[2] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient large-scale stereo matching. In *Asian Conference on Computer Vision (ACCV)*, 2010.

[3] Asmaa Hosni, Christoph Rhemann, Michael Bleyer, Carsten Rother, and Margrit Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 35(2):504 – 511, 2013.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.