REFERENCES

[1] J. Mårtensson and H. Hjalmarsson, "How to make bias and variance errors insensitive to system and model complexity in identification," *IEEE Trans. Autom. Control*, vol. 56, no. 1, pp. 100–112, Jan. 2011.

[2] Y. Fu and T. Y. Chai, "Nonlinear multivariable adaptive control using multiple models and neural networks," *Automatica*, vol. 43, no. 6, pp. 1101–1110, Jun. 2007.

[3] A. S. Kamalabady and K. Salahshoor, "Affine modeling of nonlinear multivariable processes using a new adaptive neural network-based approach," *J. Process Control*, vol. 19, no. 3, pp. 380–393, Mar. 2011.

[4] H. G. Han and J. F. Qiao, "An efficient self-organizing RBF neural network for water quality predicting," *Neural Netw.*, vol. 24, no. 7, pp. 717–725, Sep. 2011.

[5] P. Singla, K. Subbarao, and J. L. Junkins, "Direction-dependent learning approach for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 203–222, Jan. 2007.

[6] P. A. Gutiérrez, C. Hervás-Martínez, and F. J. Martínez-Estudillo, "Logistic regression by means of evolutionary radial basis function neural networks," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 246–263, Feb. 2011.

[7] G. R. Francisco, S. Carlos, and C. Ricardo, "Autonomous mobile robots navigation using RBF neural compensator," *Control Eng. Practice*, vol. 19, no. 3, pp. 215–222, Mar. 2011.

[8] S. Ferrari, F. Bellocchio, V. Piuri, and N. A. Borghese, "A hierarchical RBF online learning algorithm for real-time 3-D scanner," *IEEE Trans. Neural Netw.*, vol. 21, no. 2, pp. 275–285, Feb. 2010.

[9] M. K. Müezzinoglu and J. M. Zurada, "RBF-based neurodynamic nearest neighbor classification in real pattern space," *Pattern Recognit.*, vol. 39, no. 5, pp. 747–760, May 2006.

[10] F. Schwenker, H. A. Kestler, and G. Palm, "Three learning phases for radial-basis-function networks," *Neural Netw.*, vol. 14, nos. 4–5, pp. 439–458, May 2001.

[11] K. Z. Mao and G. B. Huang, "Neuron selection for RBF neural network classifier based on data structure preserving criterion," *IEEE Trans. Neural Netw.*, vol. 16, no. 6, pp. 1531–1540, Nov. 2005.

[12] J. F. Qiao and H. G. Han, "A repair algorithm for RBF neural network and its application to chemical oxygen demand modelling," *Int. J. Neural Syst.*, vol. 20, no. 1, pp. 63–74, Jan. 2010.

[13] H. Peng, J. Wu, G. Inoussa, Q. L. Deng, and K. Nakano, "Nonlinear system modeling and predictive control using the RBF nets-based quasi-linear ARX model," *Control Eng. Practice*, vol. 17, no. 1, pp. 59–66, Jan. 2009.

[14] J. Deng, K. Li, G. W. Irwin, and M. Fei, "Fast forward RBF network construction based on particle swarm optimization," in *Proc. Int. Conf. Life Syst. Model. Simulat. Intell. Comput.*, vol. 6329. 2010, pp. 40–48.

[15] J. Hertz, A. Krough, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley, 1991.

[16] J. Bilski and L. Rutkowski, "A fast training algorithm for neural networks," *IEEE Trans. Circuits Syst. II*, vol. 45, no. 6, pp. 749–753, Jun. 1998.

[17] Z. H. Man, H. R. Wu, S. Liu, and X. H. Yu, "A new adaptive backpropagation algorithm based on lyapunov stability theory for neural networks," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1580–1591, Nov. 2006.

[18] M. S. Al-Batah, N. A. M. Isa, K. Z. Zamli, and K. A. Azizli, "Modified recursive least squares algorithm to train the hybrid multilayered perceptron (HMLP) network," *Appl. Soft Comput.*, vol. 10, no. 1, pp. 236–244, Jan. 2010.

[19] O. Kaynak, K. Erbatur, and M. Ertugrul, "The fusion of computationally intelligent methodologies and sliding-mode control-a survey," *IEEE Trans. Ind. Electron.*, vol. 48, no. 1, pp. 4–17, Feb. 2001.

[20] J. Jiang and Y. M. Zhang, "A revisit to block and recursive least squares for parameter estimation," *Comput. Electr. Eng.*, vol. 30, no. 5, pp. 403–416, Jul. 2004.

[21] K. Li, J. X. Peng, and G. W. Irwin, "A fast nonlinear model identification method," *IEEE Trans. Autom. Control*, vol. 50, no. 8, pp. 1211–1216, Aug. 2005.

[22] J. X. Peng, K. Li, and G. W. Irwin, "A novel continuous forward algorithm for RBF neural modelling," *IEEE Trans. Autom. Control*, vol. 52, no. 1, pp. 117–122, Jan. 2007.

[23] S. Wu and M. J. Er, "Dynamic fuzzy neural networks-a novel approach to function approximation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 30, no. 2, pp. 358–364, Apr. 2000.

[24] S. Wu and M. J. Er, "Dynamic fuzzy neural networks-a novel approach to function approximation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 30, no. 2, pp. 358–364, Apr. 2000.

[25] S. H. Zak, *Systems and Control*. New York: Oxford Univ. Press, 2003.

[26] K. Li, J. X. Peng, and E. W. Bai, "A two-stage algorithm for identification of nonlinear dynamic systems," *Automatica*, vol. 42, no. 7, pp. 1189–1197, Jul. 2006.

[27] R. J. Schilling, J. J. Carroll, and A. F. Al-Ajlouni, "Approximation of nonlinear systems with radial basis function neural network," *IEEE Trans. Neural Netw.*, vol. 12, no. 1, pp. 1–15, Jan. 2001.

Asynchronous Event-Based Binocular Stereo Matching

Paul Rogister, Ryad Benosman, *Member, IEEE*, Sio-Hoi Ieng, *Member, IEEE*, Patrick Lichtsteiner, *Member, IEEE*, and Tobi Delbruck, *Senior Member, IEEE*

***Abstract*— We present a novel event-based stereo matching algorithm that exploits the asynchronous visual events from a pair of silicon retinas. Unlike conventional frame-based cameras, recent artificial retinas transmit their outputs as a continuous stream of asynchronous temporal events, in a manner similar to the output cells of the biological retina. Our algorithm uses the timing information carried by this representation in addressing the stereo-matching problem on moving objects. Using the high temporal resolution of the acquired data stream for the dynamic vision sensor, we show that matching on the timing of the visual events provides a new solution to the real-time computation of 3-D objects when combined with geometric constraints using the distance to the epipolar lines. The proposed algorithm is able to filter out incorrect matches and to accurately reconstruct the depth of moving objects despite the low spatial resolution of the sensor. This brief sets up the principles for further event-based vision processing and demonstrates the importance of dynamic information and spike timing in processing asynchronous streams of visual events.**

***Index Terms*— Asynchronous acquisition, event-based vision, frameless vision, retinas, stereo vision, time impulse encoding.**

I. INTRODUCTION

Current methods to compute high-speed real-time stereo are still too computationally expensive. The generation of 3-D models able to process data in real time beyond the classically used 60 Hz remains a difficult problem. This is mainly due to the high amounts of redundantly acquired data that need to be processed in every incoming frame. Frame-based acquisition of light intensities over regular temporal intervals raises important computational limitations. It lacks temporal dynamics, as it implies a process of integration of light over fixed temporal intervals, which is incompatible with precise timings usually used in neural communication [1]. Biological systems are data-driven, and they encode visual data asynchronously as sparse spiking outputs rather than frames of pixel values [2].

The aim of the event-based neuromorphic field is to emulate biology's use of asynchronous, exceedingly sparse, and

Manuscript received April 27, 2011; revised November 17, 2011; accepted November 19, 2011. Date of publication December 27, 2011; date of current version February 8, 2012.

P. Rogister, P. Lichtsteiner, and T. Delbruck are with the Institute of Neuroinformatics, University/ETH Zurich, Zurich 8057, Switzerland (e-mail: rogister@ini.phys.ethz.ch; patrick.lichtsteiner@espros.ch; tobi@ini.phys.ethz.ch).

R. Benosman and S. Ieng are with Vision Institute, UPMCINSERM-CNRS, Paris 75252, France (e-mail: ryad.benosman@upmc.fr; sio-hoi.ieng@upmc.fr).

Digital Object Identifier 10.1109/TNNLS.2011.2180025

data-driven digital signaling as a core aspect of its computational architecture. Since the pioneering work on the address-event representation (AER) vision system [3], the performance of the latest developed retinas allows one to consider a new paradigm in visual computation [4], [5]. Following biological systems, pixels are independent, they collect and send their own data as a local information from a single pixel independently of the others. The trigger for the pixel readout, or sendout, can be chosen as a threshold on its activity changes. The collected data include the spatial location of active pixels and an accurate time stamping at which a change occurs above a threshold. Additional information such as polarity of events (positive or negative change of light) can be added. Computationally, this representation differs from frames in two respects:

1) it encodes the data in a compressed form in a stream of events;
2) events can be processed locally while encoding the additional temporal dynamics of the scene.

The 3-D reconstruction of scenes using multiple cameras is a fundamental problem in computer vision. In general, few stereo studies have focused on the important link between stereo and motion, and the importance of temporal dynamics. Dynamic information is crucial for stereo matching, as it introduces an additional temporal constraint in the recovery of scenes' structures.

Existing techniques that work on space and time all deal with spatiotemporal volumes of images. They rely heavily on the use of local spatiotemporal representation of luminances to encode orientation [6], or to define an extension of corner-based descriptors to temporal volumes of images [7]. Other techniques rely on heuristics to provide matches from a model-based reasoning [8], or constraints on the temporal derivative of disparity [9]. Some works have concentrated on defining appropriate temporal integration windows as part of the matching process [10]. In other cases, they reinforce disparity estimates from the previous frame using optical flow [11]. Stereo and motion estimation have been combined using partial derivatives [12] as well as Markov random fields [13].

Asynchronous event-based acquisition properties, as we will show, allow a simplification of the 3-D matching process together with a low computational cost. Stereo matching can then be computed in a manner that is not only faster and more efficient, but also more akin to the way in which natural systems process visual information. Artificial vision and neural network theories for depth computation rely mainly on firing rates, but few theories of computation are specifically spike-based. The key idea is to set stereo correspondence on matching temporal occurrences of events, the structure of the binocular stimulus is then reflected by the events' synchrony patterns across the two retinas.

II. Neuromorphic Silicon Retina

Biological retinas, unlike frame-based cameras, transmit less redundant information about a visual scene in an asynchronous manner. The various functionalities of the retina have been incorporated into neuromorphic vision sensors since the late 1980s with the pioneering work of Mahowald [3]. Since then, the most interesting achievement in neuromorphic retinas has been the development of activity-driven sensing. The event-based vision sensors output compressed digital data in the form of events, removing redundancy, reducing latency, and increasing dynamic range compared with conventional imagers. A complete review of the history and existing sensors can be found in [14].

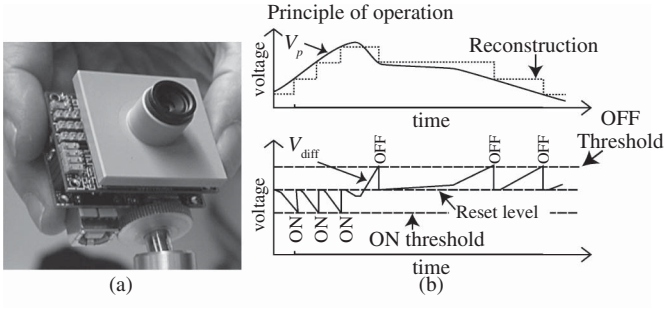

Fig. 1. (a) First-generation DVS sensor with 128 × 128 pixels [15]. (b) Principle of ON and OFF spikes generation of DVS pixels, adapted from [15]. Top: the evolution of pixel's voltage V_p proportional to the log intensity. Bottom: the corresponding generation of ON (voltage increases above change threshold) and OFF (voltage decreases) events, from which the evolution of V_p can be reconstructed.

The dynamic vision sensor (DVS) used in this brief is an AER silicon retina with 128 × 128 pixels [15]. The DVS output consists of asynchronous address events that signal scene reflectance changes at the times they occur. Each pixel is independent and detects changes in log intensity larger than a threshold since the last emitted event (typically 15% contrast). As shown in Fig. 1, when the change in log intensity exceeds a set threshold, an ON or OFF event is generated by the pixel depending on whether the log intensity increased or decreased. Since the DVS is not clocked like conventional cameras, the timing of events can be conveyed with a very accurate temporal resolution of approximately 1 μs. Thus the "effective frame rate" is typically several kilohertz.

The retina pixels also implement the local gain adaptation mechanism, which allows them to work over scene illuminations that range from 2 lux to over 100 klux. When events are transmitted off-ship, they are timestamped using off-chip digital components and then transmitted to a computer using a standard USB connection.

The advantages of the sensor over conventional clocked cameras are that only moving objects produce data, thereby reducing the load of postprocessing.

A. Neuromorphic Stereo Systems

There have been numerous studies on the neural computation of depth. In [16], the Marr–Poggio cooperative stereo algorithm [17] is implemented in a chip. A stereo system is introduced in [3] and [18], which uses the output of two spiking retinas connected to a correlator array. The approach relies on a local inhibition driven along the line of sight and implements the uniqueness constraint (one pixel from the one view is associated with only one in the other, except during occlusions), while the lateral excitatory connectivity

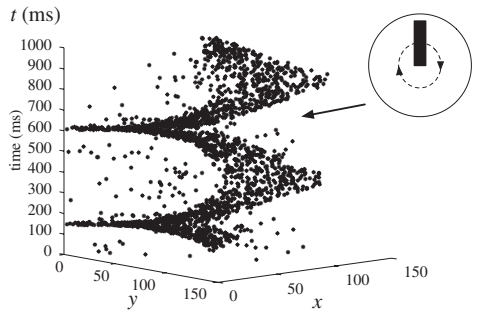

Fig. 2. Space–time representation of events generated in response to a rotating black bar. Each dot represents a DVS event.

gives more weight to coplanar solutions to discriminate false matches from correct ones. Another multichip stereo system uses a combination of AER Gabor-type chips and digital chips to implement disparity-tuned complex neurons constructed according to the binocular energy model [19]. Other neuromorphic stereo approaches use the output of the event-based DVS retinas in a classic frame-based approach to retrieve 3-D information [20]. Another study designed an event-to-frame converter to reconstruct frames and then tested two stereo frame-based vision algorithms: one window-based and one feature-based using center-segment features [21].

B. Visual Input Events

The DVS models the transient responses of the retina [2]. The stream of events from the retina can be defined mathematically as follows. Let $e(\mathbf{p}, t)$ be an event occurring at time t at the spatial location $\mathbf{p} = (x, y)^T$. The values $e(\mathbf{p}, t)$ are set to be -1 or $+1$ when a negative or a positive change of contrast is detected, respectively. We can then define S_{t_1,t_2} the spatiotemporal set of events occurring in the temporal interval $[t_1, t_2]$ as

$$S_{t_1,t_2} = \{e(\mathbf{p}, t) | t \in [t_1, t_2]\}. \tag{1}$$

The absence of events when no change of contrast is detected implies that redundant visual information usually recorded in frames is not carried in the stream of events. Fig. 2 shows an example of the spatiotemporal visualization of a set of DVS events in response to a rotating bar.

III. Event-Based Matching

A. Time Window

Let two retinas $\mathcal{R}_i$ and $\mathcal{R}_j$ observe a common part of a scene. A 3-D point $\mathbf{X}$ moving in space triggers changes of luminance in the field of view of the retinas. Events will be written as $e(\mathbf{p}_k^i, t)$, where superscript i indexes the retina in which the event happens and subscript k indexes the event in the retina $\mathcal{R}_i$. The 3-D point $\mathbf{X}(t)$ generates an event $e(\mathbf{p}_k^i, t)$ when it projects onto the ith retina's focal plane $\mathcal{R}_i$ according to the well-known geometric relationship

$$\begin{pmatrix} \mathbf{p}_k^i \\ 1 \end{pmatrix} = P_i \begin{pmatrix} \mathbf{X}(t) \\ 1 \end{pmatrix}, \quad \text{if } e(\mathbf{p}_k^i, t) \neq 0 \tag{2}$$

where P_i is the projection matrix associated to the retina i [22].

At some time t, a change of luminance caused by a moving edge will affect pixels and eventually generate events in both retinas. The timing of these outputs events will not correspond to the exact timing of the real luminance changes in the physical world. There is a latency in pixels reaching the threshold and generating the events, this latency varies across pixels and retinas. There is also an additional jitter in the timing of the events, due to the on-chip circuits that are needed for transmitting the events off-chip on a shared digital bus. Spatial aliasing also affects events' timing, thus it is impossible to match pixels across retinas based on the exact timing of their output events.

As shown in Fig. 3(a), $e(\mathbf{p}_1^i, t_1)$ and $e(\mathbf{p}_1^j, t_2)$ both correspond to $\mathbf{X}(t)$. Due to the variance in the transmission of the temporal information, it is unlikely that $t_1 = t_2$ [see Fig. 3(b)]. It is also not possible to match events based on the shortest time difference of occurrence. The activity of nonrelated pixels can generate events with timings between two matching events. Events $e(\mathbf{p}_2^j, t_3)$ and $e(\mathbf{p}_3^j, t_4)$ corresponding to other activities in the scene might trigger events that are closer temporally to $e(\mathbf{p}_1^i, t_1)$ than $e(\mathbf{p}_1^j, t_2)$.

Although we cannot use the exact timing or close to the exact timing to discriminate matches, we can define a time window in which true matches are more likely to occur. Thus we define within a time window δ_t [see Fig. 3(b)], so that for an incoming event $e(\mathbf{p}_1^i, t)$ in $\mathcal{R}_i$, there exists a set $S^j(t)$ containing events in $\mathcal{R}_j$ defined by

$$S^j(t) = \left\{ e(\mathbf{p}_k^j, t') \text{ in } \mathcal{R}_j | \forall t',\ |t' - t| < \frac{\delta_t}{2} \right\}. \tag{3}$$

B. Distance to the Epipolar Line

The epipolar geometry is the intrinsic projective geometry between two views [22]. It is independent of scene structure, and depends only on the cameras' internal parameters and relative pose. The fundamental matrix F is computed from a set of point correspondences as introduced in [22]. Let $e(\mathbf{p}_1^i, t)$ and $e(\mathbf{p}_1^j, t')$ be two corresponding events, then the fundamental matrix is defined by the equation

$$(\mathbf{p}_1^j \ 1)^T F \begin{pmatrix} \mathbf{p}_1^i \\ 1 \end{pmatrix} = 0 \tag{4}$$

for any pair of matching events in both retinas' focal planes. F links $\mathcal{R}_i$ to $\mathcal{R}_j$, with $F\begin{pmatrix} \mathbf{p}_1^i \\ 1 \end{pmatrix} = \mathbf{l}_{ij}$. The epipolar line $\mathbf{l}_{ij}$ is in $\mathcal{R}_j$, and it contains all possible matches of event $e(\mathbf{p}_1^i, t)$ in $\mathcal{R}_j$ [see Fig. 3(a)].

We can then use the distances of possible matches—that lie within the temporal window δ_t constraint—to the epipolar lines. The set of possible matches M of an event $e(\mathbf{p}_1^i, t)$ is then given by

$$M(e(\mathbf{p}_1^i, t)) = \left\{ e(\mathbf{p}_k^j, t') \in S^j(t) | \forall k,\ d(\mathbf{p}_k^j, \mathbf{l}_{ij}) < \Delta_p \right\} \tag{5}$$

where $d(\mathbf{p}_k^j, \mathbf{l}_{ij})$ is the Euclidian distance of $\mathbf{p}_k^j$ to $\mathbf{l}_{ij}$, and Δ_p in pixels is a threshold representing the maximum distance allowed. It is generally set to 1 pixel. The main idea of the stereo matching is that, considering a short time interval, it is

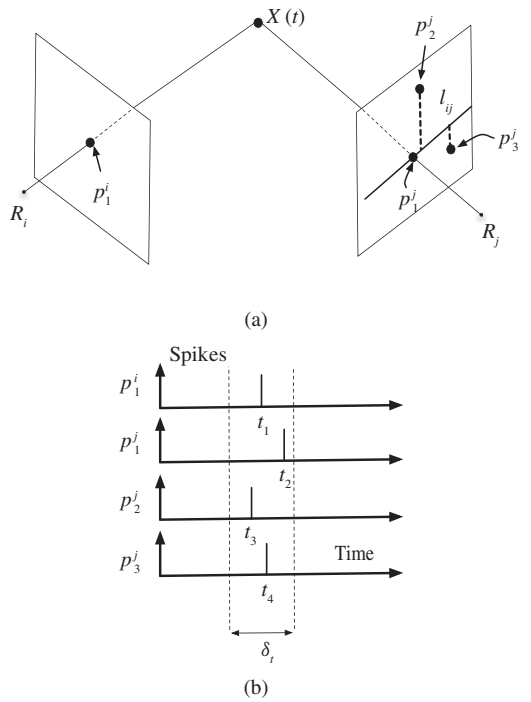

Fig. 3. (a) Two events generated by the real point $\mathbf{X}$ at time t detected by corresponding pixels located at $\mathbf{p}_1^i$ and $\mathbf{p}_1^j$ at slightly different times t_1 and t_2. (b) There are only few possible candidates matching the event $e_1^i(\mathbf{p}_1^i, t_1)$ if the distance to the epipolar line $\mathbf{l}_{ij}$ and a time window around t_1 are used.

Algorithm 1 Stereo algorithm for retina event matching

Require: Two retinas $\mathcal{R}_i$, $\mathcal{R}_j$

Require: F, an estimation of the fundamental matrix for the pose

for all events $e(\mathbf{p}_k^i, t)$ in retina $\mathcal{R}_i$ **do**

Compute the epipolar line $\mathbf{l}_{ij} = F\left(\mathbf{p}_{k_1}^i\right)$

Determine the set of events $S^j(t)$

Extract from $S^j(t)$, the subset $M(e(\mathbf{p}_k^i, t))$ of all possible matches

Select events fulfilling chosen additional constraints such : polarity, uniqueness, orientation,…

end for

return $M(e(\mathbf{p}_k^i, t))$

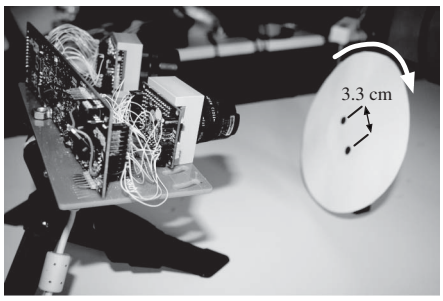

Fig. 4. Two 128 × 128 DVSs mounted in a stereo rig observing a rotating disc on which two points are drawn. The disc rotation frequency is 16 Hz, while the distance between the two dots is set to 3.3 cm.

very unlikely that two uncorrelated events happen at the same time and fall on the same epipolar line.

C. Additional Constraints

Other straightforward constraints enforce the removal of the remaining ambiguities.

1) The polarity of the events, taken into account so that ON events from one retina are matched only to ON events in the other retina.
2) Similarly for OFF events, the uniqueness constraint, applied so that an event cannot be matched with more than 1 event in the other retina.
3) The ordering constraint, which a classical binocular stereo constraint on ray orientation [23].
4) The temporal activity of pixels, where two pixels are matched if their number of total emitted events over a time period is close.

The general matching steps are summarized in Algorithm 1.

IV. Experiments

The stereo matching algorithm has been implemented in the open-source Java software project (jAER). We use two DVSs classically calibrated using the method described in [22]. Fig. 4 shows the two retinas observing a rotating disc on the right. The two retinas are mounted on a synchronization board providing a common timestamp clock for the timing of events in both DVSs. The synchronization board outputs a single stream of events. The synchronization board sends the events to a Pentium 4 laptop through a high-speed USB 2.0 connection.

Fig. 5 presents the results of the event-based stereo matching algorithm on the DVS outputs. Disparity maps are created from the events generated by a pen moved back and forth at three different distances from the two retinas which are separated approximately by 10 cm, as illustrated in Fig. 5. The computed disparity is color-coded from blue to red for disparities results varying between 0 and 127 pixels, respectively. The default value for background unprocessed pixels is 0. The matching used a time window of $\delta_t = 1$ ms, and a maximum distance to epipolar line $\Delta_p = 1$ pixel. Two pixels are matched if their number of emitted events corresponds up to a maximum difference of 3. All additional constraints were applied in all experiments.

As shown in Fig. 5, the disparity is a decreasing function of the depth. Some pixels are not matched, which is due to multiple matching. There are cases where more than one pixel fulfills all the constraints, in this case the match is simply discarded.

The second experiment on disparity uses the DVS outputs in response to two pens simultaneously moving in the field of view of the stereo setup. This configuration is particularly interesting, as the two pens oriented vertically will generate events that will happen at the same time and possibly lie on the same epipolar lines. The ordering constraint [23] helps to discard false matches by setting a forbidden zone along epipolar lines according to the orientation of the line of sight of pixels. The results in Fig. 6 shows that the disparity map

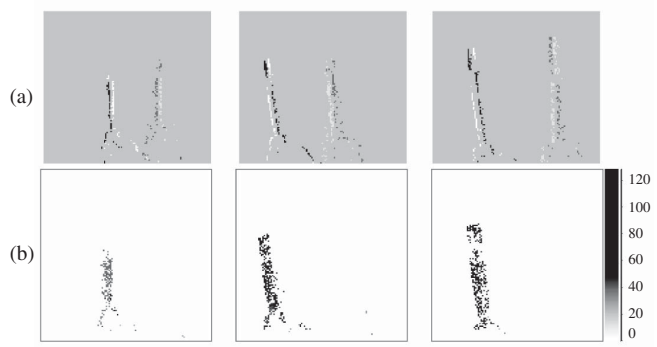

Fig. 5. (a) DVS output in response to a pen being moved back and forth at three different depths: close (left), medium (middle), and far (right) roughly spaced by 10 cm. The left and right retina events are merged into one view using a color code to distinguish between events from the two retinas: green/violet (ON/OFF) for the left retina events and salmon/cyan (ON/OFF) for the right retina events. (b) Corresponding color disparity maps for the pen at the different depths.

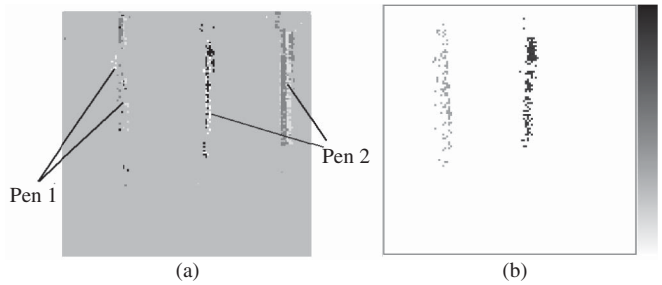

Fig. 6. (a) Raw data and (b) disparity map of two moving pens at different depths.

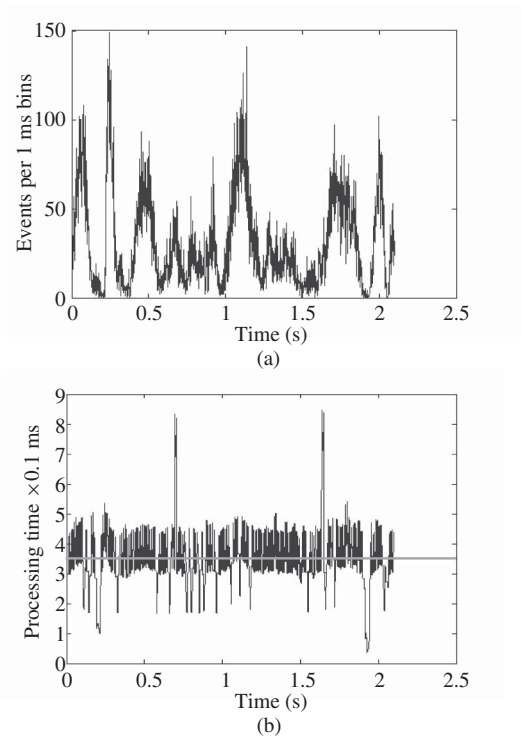

Fig. 7. (a) Number of events per 1 ms. (b) Computation cost corresponding approximately to 30% CPU load.

computed by the two pens is coherent with the depth of the objects with no mix between the events of each pen.

Fig. 7(a) shows the total event rate generated by the two-pen experiment binned over a time interval of 1 ms. Data-driven "bioinspired" DVS sampling removes redundancies at the lowest level, so few operations have to be performed over a short period of time. As shown in Fig. 7(b), the corresponding computational costs are low, giving a mean computation time of 0.3 ms (30% CPU load) measured using a non-optimized Java implementation. The asynchrony of the sensor outputs allows faster acquisition and data transfer, which induce nonsaturation of the CPU, thus allowing real-time processing of binocular data.

The last experiment's aim is to show the accuracy and temporal precision of the matching. We test whether we can reconstruct the distance between two objects independently of their depth from the binocular system. As shown in Fig. 4, the goal of the experiment is to reconstruct the 3-D positions of two rotating dots at a frequency of 16 Hz, each of 1 cm in diameter and spaced by a distance of 3.3 cm.

Due to the speed of the rotation of approximately 100 rad/s, a classical frame-based stereo setup acquiring frames at 60 Hz will be able to see a dot every 2.7 cm. The diameter of a dot being 1 cm, the acquired images will be blurred, thereby inducing imprecision in the 3-D localization of the rotating dots.

The matching is performed directly on incoming events. Independently of the size of matching time interval $\delta_t = 1$ ms, the DVS allows tracking the dot every microsecond. The matched 3-D points are computed at the precision of the delay factor $\delta_t = 1$ ms that is needed to link both retinas' activations. The tracker is then updated at a precision of 1 ms and allows tracking the dot every 0.16 cm in real time giving a speed factor of 17 times faster than a classic acquisition at 60 Hz.

An additional 3-D tracking algorithm is used to define the location of each dot. The process works as follows. When the number of 3-D points reconstructed reaches a given threshold (> 20 points), two spatiotemporal volumes are initialized and positioned around the two sets of computed 3-D points. Then, as used in [24] in a 2-D case, every volume's size and position are updated according to each new incoming reconstructed 3-D point. From the estimated position of each dot, computed for every incoming 3-D point with a precision of 1 μs, it is then possible to estimate the known distance between the dots. Fig. 8(a), (c), and (e) shows a typical example of reconstructed dots over a period of 3 s in response to the rotating stimulus which is placed approximately at three different depths of 30, 40, and 60 cm, respectively, from the stereo retinas' setup. The motion being a pure rotation, the reconstructed trajectories of each dot is a circle.

The estimation of the distance between the two dots is given by Fig. 8(b), (d), (f). The measured results provide an average error of 3.67%, 4.23%, and 5.79% with a standard deviation of, respectively, 2.51%, 3.04%, and 5.67% for 30, 40, and 60 cm, respectively. From these results, we can see that the reconstructed precision is approximately proportional to the

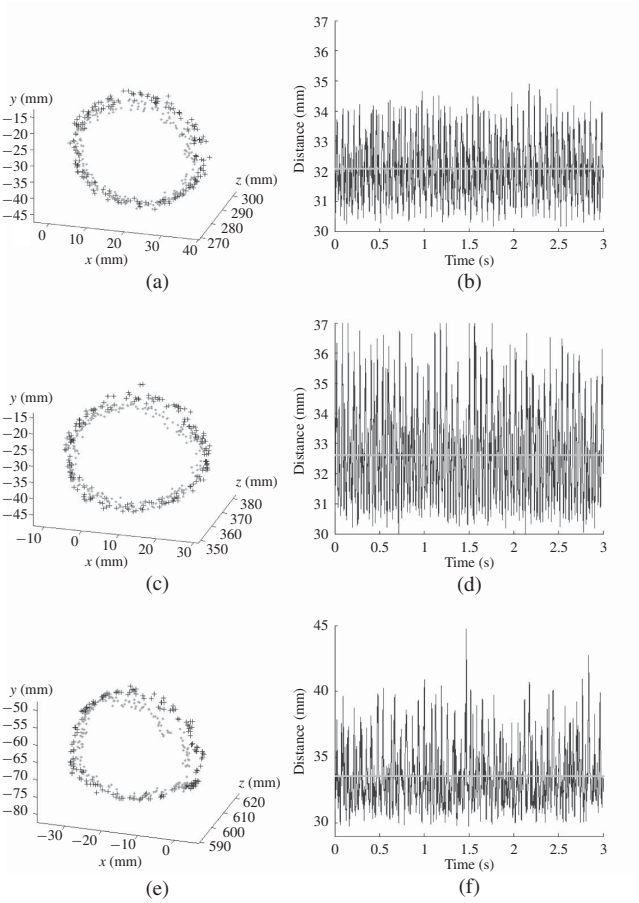

Fig. 8. (a), (c), and (e) Typical example of reconstructed dots over a period of 3 s in response to the rotating stimulus which is placed at three different depths from the stereo setup at 30, 40, and 60 cm, respectively. Since the motion is a pure rotation, the reconstructed trajectories of each dot (given by a different symbol) is a circle. (b), (d), and (f) Estimation of the distance between the two dots. The average percentage of errors for each depth is, respectively, 3.67%, 4.23%, and 5.79%, with a standard deviation of 2.51%, 3.04%, and 5.67%, respectively.

actual depth, albeit with a slight increase of the error due to the noise in the 3-D tracking and reconstruction. The results provide an accurate 3-D localization. The low resolution of the current DVS prototypes (128×128 pixels) restricts the accuracy and limits the use of the setup to relatively close moving objects.

V. Discussion

This brief focuses on synchrony induced by the dynamic content of observed scenes and links coincidence detection properties of neurons to computation [24] rather than by coupling between neurons [25]–[27]. This may provide a partial answer to the controversial question of the role of precise spike timing in neural computation and show how temporal precision and synchrony of spikes are important in computation [28], [29]. Due to the jitter of the used spiking retinas and similarly to biological inter-areal connections [30], precise timing is lost. However, the computational mechanism used in this brief does not critically rely on precise or reproducible spike timing, but on reproducible scene's stimulus-dependent synchrony. Precise spike timing is then not in itself important, and only the relative spike timing carries information about the depth of visual stimuli. This indicates that, in biological systems, the spike timing variations that prevent trial-to-trial reproduction may not affect the exploitation of relative spike timing.

VI. Conclusion

This brief presented an event-based stereo matching algorithm and showed that the asynchronous high temporal resolution properties of the acquisition are particularly efficient in terms of the tradeoff between computational load and accuracy. Most asynchronous event-based sensors do not directly measure light intensities, therefore the standard schemes of stereo computation are no longer appropriate. The sparse nature of the acquired signals allows us to explore new paradigms in 3-D computation. The entirely event-driven algorithm of stereo vision described in this brief takes full advantage of the data-driven neuromorphic signal. The combination of spatial and temporal constraints fully uses the high temporal resolution of neuromorphic retinas. It allows us to produce an optimal stereo algorithm that is able to perform stereo computation at high frequencies using a single standard PC. The use of more retinas should increase the impact of the geometric constraints and introduce more robustness to the system. The asynchronous event-based acquisition is a promising 3-D technology offering yet unexplored potential to overcome current limitations of frame-based 3-D vision. The promising results are much faster than current techniques, as an example, the Microsoft Kinect computes depth at 30 Hz but at a much higher resolution of 640×480 pixels. This technique should be of great use for the robotics and computer vision communities especially in embedded computer vision applications. This technology is promising and will become of higher interest once retinas with larger spatial resolutions are available.

Acknowledgment

The authors would like to thank R. Berner and C. Clercq for sharing their time and help in using the dynamic vision sensor and S.-C. Liu for help with proofreading this brief. They also benefited from both the CapoCaccia Cognitive Neuromorphic Engineering Workshop and the National Science Foundation, Telluride Neuromorphic Cognition workshops, Telluride, Colorado.

References

[1] M. Meister and M. J. B. Ii, "The neural code of the retina," *Neuron*, vol. 22, pp. 435–450, Mar. 1999.

[2] B. Roska and F. Werblin, "Rapid global shifts in natural scenes block spiking in specific ganglion cell types," *Nature Neurosci.*, vol. 6, pp. 600–608, May 2003.

[3] M. Mahowald, "VLSI analogs of neuronal visual processing: A synthesis of form and function," Ph.D. dissertation, Dept. Comput. Sci., California Inst. Technology, Pasadena, CA, 1992.

[4] P. Lichtsteiner, T. Delbruck, and J. Kramer, "Improved ON/OFF temporally differentiating address-event imager," in *Proc. 11th IEEE Int. Conf. Electron., Circuits, Syst.*, Dec. 2004, pp. 211–214.

[5] C. Posch, D. Matolin, and R. Wohlgenannt, "A QVGA 143 dB dynamic range asynchronous address-event PWM dynamic image sensor with lossless pixel-level video compression," in *Proc. IEEE Int. Solid-State Circuits Conf.*, San Francisco, CA, Feb. 2010, pp. 400–401.

[6] G. Granlund and H. Knutsson, *Signal Processing for Computer Vision*. Norwell, MA: Kluwer, 1995.

[7] G. Flitton, T. Breckon, and N. M. Bouallagu, "Object recognition using 3-D SIFT in complex CT volumes," in *Proc. British Mach. Vis. Conf.*, 2010, pp. 11.1–11.12.
[8] M. Jenkin and J. K. Tsotsos, "Applying temporal constraints to the dynamic stereo problem," *Comput. Vis. Graph. Image Process.*, vol. 33, no. 1, pp. 16–32, Jan. 1986.
[9] A. M. Waxman and J. H. Duncan, "Binocular image flows: Steps toward stereo-motion fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 715–729, Nov. 1986.
[10] L. Zhang, B. Curless, and S. Seitz, "Spacetime stereo: Shape recovery for dynamic scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2003, pp. 367–374.
[11] M. Gong, "Enforcing temporal consistency in real-time stereo estimation," in *Proc. Eur. Conf. Comput. Vis.*, May 2006, pp. 564–577.
[12] F. Huguet and F. Devernay, "A variational method for scene flow estimation from stereo sequences," in *Proc. 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–7.
[13] O. Williams, M. Isard, and J. MacCormick, "Estimating disparity and occlusions in stereo video sequences," in *Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 250–257.
[14] T. Delbruck, B. Linares-Barranco, E. Culurciello, and C. Posch, "Activity-driven, event-based vision sensors," in *Proc. Int. Symp. Circuits Syst.*, Paris, France, May–Jun. 2010, pp. 2426–2429.
[15] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor," *IEEE J. Solid State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008.
[16] M. Mahowald and T. Delbrück, "Cooperative stereo matching using static and dynamic image features," in *Analog VLSI Implementation of Neural Systems*, C. M. Ismail, Ed. Boston, MA: Kluwer, 1989, pp. 213–238.
[17] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, no. 4262, pp. 283–287, 1976.
[18] M. Mahowald, *An Analog VLSI System for Stereoscopic Vision*. Boston, MA: Kluwer, 1994.
[19] E. K. C. Tsang and B. E. Shi, "A neuromorphic multi-chip model of a disparity selective complex cell," in *Advances in Neural Information Processing Systems*, vol. 16. Cambridge, MA: MIT Press, 2004.
[20] J. Kogler, C. Sulzbachner, and W. Kubinger, "Bio-inspired stereo vision system with silicon retina imagers," in *Proc. 7th Int. Conf. Comput. Vis. Syst.*, 2009, pp. 174–183.
[21] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, Jun. 1994, pp. 593–600.
[22] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[23] A. Yuille and T. Poggio, "A generalized ordering constraint for stereo correspondence," Artificial Intelligence Laboratory, MIT, Cambridge, Memo Rep. 777, 1984.
[24] R. Benosman, S.-H. Ieng, P. Rogister, and C. Posch, "Asynchronous event-based Hebbian epipolar geometry," *IEEE Trans. Neural Netw.*, vol. 22, no. 11, pp. 1723–1734, Nov. 2011.
[25] J. V. Arthur and K. A. Boahen, "Synchrony in silicon: The gamma rhythm," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1815–1824, Nov. 2007.
[26] G. Indiveri, E. Chicca, and R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 211–221, Jan. 2006.
[27] G. Ermentrout, "Synchronization in a pool of mutually coupled oscillators with random frequencies," *J. Math. Biol.*, vol. 22, no. 1, pp. 1–9, 1985.
[28] M. Abeles, "Role of the cortical neuron: Integrator or coincidence detector?" *Isr. J. Med. Sci.*, vol. 18, no. 1, pp. 83–92, Jan. 1982.
[29] W. Softky and C. Koch, "The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs," *J. Neurosci.*, vol. 13, no. 1, pp. 334–350, Jan. 1993.
[30] J. Bullier, "The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps," *Brain Res. Rev.*, vol. 36, nos. 2–3, pp. 96–107, Jan. 2001.

Frames for Exact Inversion of the Rank Order Coder

Khaled Masmoudi, *Student Member*, *IEEE*,
Marc Antonini, *Member*, *IEEE*, and Pierre Kornprobst

***Abstract*— Our goal is to revisit rank order coding by proposing an original exact decoding procedure for it. Rank order coding was proposed by Thorpe *et al*. who stated that the order in which the retina cells are activated encodes for the visual stimulus. Based on this idea, the authors proposed in [1] a rank order coder/decoder associated to a retinal model. Though, it appeared that the decoding procedure employed yields reconstruction errors that limit the model bit-cost/quality performances when used as an image codec. The attempts made in the literature to overcome this issue are time consuming and alter the coding procedure, or are lacking mathematical support and feasibility for standard size images. Here we solve this problem in an original fashion by using the frames theory, where a frame of a vector space designates an extension for the notion of basis. Our contribution is twofold. First, we prove that the analyzing filter bank considered is a frame, and then we define the corresponding dual frame that is necessary for the exact image reconstruction. Second, to deal with the problem of memory overhead, we design a recursive out-of-core blockwise algorithm for the computation of this dual frame. Our work provides a mathematical formalism for the retinal model under study and defines a simple and exact reverse transform for it with over than 265 dB of increase in the peak signal-to-noise ratio quality compared to [1]. Furthermore, the framework presented here can be extended to several models of the visual cortical areas using redundant representations.**

***Index Terms*— Bio-inspired image coding, frames theory, out-of-core, rank order code, scalability.**

I. Introduction

Neurophysiologists made substantial progress in better understanding the early processing of the visual stimuli. Especially, several efforts proved the ability of the retina to code and transmit a huge amount of data under strong time and bandwidth constraints [2]–[4]. Thus, our aim is to use the computational neuroscience models that mimic the retina behavior to design novel lossy coders for static images. In this brief, we assume that the retina encodes the visual information by the order in which its ganglion cells react to the stimulus - recalling that these cells react through the emission of electrical impulses (the spikes). This choice was motivated by Thorpe *et al*. neurophysiologic results on ultrarapid stimulus categorization [2], [5]. The authors showed that still image classification can be achieved by the visual cortex within very short latencies of about 150 ms. As an explanation, it was stated that: *There is information in the order in which the cells fire*, and thus the temporal ordering can be used as a code.

Manuscript received May 20, 2011; accepted December 3, 2011. Date of publication December 27, 2011; date of current version February 8, 2012.

K. Masmoudi and M. Antonini are with I3S Laboratory-CNRS-UNS, Sophia-Antipolis 06903, France (e-mail: kmasmoud@i3s.unice.fr; am@i3s.unice.fr).

P. Kornprobst is with NeuroMathComp Team Project, INRIA, Sophia-Antipolis 06902, France (e-mail: pierre.kornprobst@sophia.inria.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2011.2179557