Asynchronous Event-Based Hebbian Epipolar Geometry

Ryad Benosman, Sio-Hoï Ieng, Paul Rogister, and Christoph Posch

***Abstract*— Epipolar geometry, the cornerstone of perspective stereo vision, has been studied extensively since the advent of computer vision. Establishing such a geometric constraint is of primary importance, as it allows the recovery of the 3-D structure of scenes. Estimating the epipolar constraints of nonperspective stereo is difficult, they can no longer be defined because of the complexity of the sensor geometry. This paper will show that these limitations are, to some extent, a consequence of the static image frames commonly used in vision. The conventional frame-based approach suffers from a lack of the dynamics present in natural scenes. We introduce the use of neuromorphic event-based—rather than frame-based—vision sensors for perspective stereo vision. This type of sensor uses the dimension of time as the main conveyor of information. In this paper, we present a model for asynchronous event-based vision, which is then used to derive a general new concept of epipolar geometry linked to the temporal activation of pixels. Practical experiments demonstrate the validity of the approach, solving the problem of estimating the fundamental matrix applied, in a first stage, to classic perspective vision and then to more general cameras. Furthermore, this paper shows that the properties of event-based vision sensors allow the exploration of not-yet-defined geometric relationships, finally, we provide a definition of general epipolar geometry deployable to almost any visual sensor.**

***Index Terms*— Asynchronous acquisition, asynchronous sensing, event-based vision, frame-free, frameless vision, neuromorphic electronics, stereovision, time impulse codification, time-based imaging.**

I. Introduction

THE fundamental matrix is a basic tool in the analysis of scenes captured by two uncalibrated cameras. It has been studied only for central cameras where all incoming rays projecting onto the image plane intersect at one point (focal point). In the case of uncalibrated central cameras, it has become customary to refer to this matrix. The fundamental matrix is a key concept: it summarizes the geometric information and enables the recovery of the epipolar geometry for uncalibrated perspective views [1]. It may also be used to reconstruct the scene from two uncalibrated views up to a projective transformation [2], [3].

Epipolar geometry has also been studied in the case of more general cameras, such as omnidirectional sensors [4], mainly central ones. The difference with mainstream cameras is introduced by the sensors' unconventional surfaces of projection. Svoboda and Pajdla [5] extended the classical epipolar geometry of perspective cameras to all central catadioptric cameras, as presented by Baker and Nayar [6]. Smadja *et al*. [7] applied epipolar geometry to the case of a cylindrical surface of projection, while Torii *et al*. [8] and Li [9] applied this geometry to spherical cameras.

In the more general case of noncentral systems, the definition of epipolar geometry is harder to establish. Noncentral cameras are optical sensors that collect light along rays that do not all intersect at one point. Noncentral images can be obtained in several ways. The first one relies on catadioptric sensors using mirrors with noncentral properties. Within this framework, a wide variety of possible mirrors can introduce noncentrality. They are generally computed according to the task to be achieved [10], [11]. Other techniques to produce noncentral images use classical perspective cameras, usually performing constrained motion. Noncentral images are then produced by mosaicing techniques acquired by pushbroom cameras, stereo panoramas, or spherical mosaics [12]–[16].

The definition of epipolar geometry in the most general unconstrained case is still unknown. This is mostly because geometric knowledge about noncentral systems remains scarce. Swaminathan *et al*. [17] and Ieng and Benosman [18] have studied caustic surfaces in catadioptric sensors. A caustic surface corresponds to the envelope of all incoming rays and can be seen as the set of focal points of the visual system. The work done in this area mainly comprises theoretical studies of the framework of special camera motion configurations [19], [20].

Several researchers have attempted to define a general geometric model of an imager. A nice example is described in a paper that changed the standard approach of an image sensor by modeling a camera as a set of raxels [21], defined as the association of a ray of view with a pixel's location. Following this kind of approach, new calibration methods have been presented, notably by Ramalingam *et al*. [22], Sturm [23], and Debaecker *et al*. [24]. This approach also led to more general and original studies, such as those of Kuthirummal and Nayar [25].

In this paper, we propose a complementary contribution to this research stream by providing a framework to estimate the

Manuscript received February 9, 2011; revised June 17, 2011; accepted August 14, 2011. Date of publication September 26, 2011; date of current version November 2, 2011.

R. Benosman and S.-H. Ieng are with the Vision Institute and the Institute of Intelligent Systems and Robotics, University Pierre and Marie Curie, Paris 75252 Cedex 05, France (e-mail: ryad.benosman@upmc.fr; sio-hoi.ieng@upmc.fr).

P. Rogister is with the Institut fur Neuroinformatik, ETHZ, Universitat Zurich, Zurich 8006, Switzerland (e-mail: rogister@ini.phys.ethz.ch).

C. Posch is with the Austrian Institute of Technology, Vienna 1210, Austria (e-mail: Christoph.Posch@ait.ac.at).

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/TNN.2011.2167239

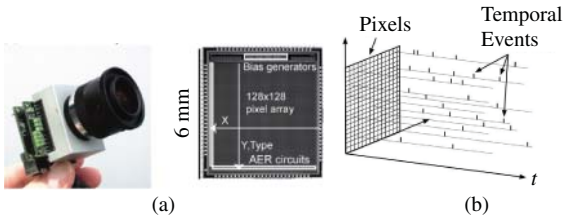

Fig. 1. (a) Protoype of the asynchronous neuromorphic sensor and its integration with a lens and a usb 2.0 interface. (b) Sketch of the codification of light intensities into temporal asynchronous events.

epipolar geometry of a multiple view system using event-based "silicon retina" vision sensors [26], regardless of the geometry, with the sole constraint that the cameras must have common fields of view, i.e., their fields of view must overlap.

We introduce a computational model of asynchronous event-based sensors that will be used to estimate—if exists—the stereo rig's epipolar geometry corresponding to central and more general cases. Using the temporal properties of such sensors, as will be shown, does not require sophisticated minimizations and allows an almost natural determination of the epipolar geometry in even the most general cases.

II. Neuromorphic Event-Based Vision

Modern non-biological information processing systems are known to perform weakly when dealing with complex information from sensory signals that seem to be easily and effortlessly processed by animal or human brains. The superior performance of biological systems can still only partly be accounted for, but it seems obvious that the processing architecture and style of computation in nervous systems are fundamentally different from what is currently deployed in artificial synchronous information processing [27]–[31].

In the late 1980s, Mead showed [32] that standard semiconductor technology can be used to implement circuits that mimic neural functions and to realize building blocks that emulate the operation of their biological role models, i.e., neurons, axons, ganglions, etc., thus enabling the construction of biomimetic devices that combine the benefits of silicon very large scale integration technology with the processing paradigms of biology. The notion underlying this paper on such neuromorphic systems is to emulate biology's use of asynchronous exceedingly sparse data-driven digital signaling as a core aspect of its computational architecture. Neuromorphic systems, like the biological systems they model, process information using energy-efficient, often massively parallel, event-based methods.

A major line of research in biomimetic systems has been in the context of sensory transduction, most notably in vision. Since the seminal implementation of a "silicon retina" by Mahowald and Mead [33] in 1989, a variety of biomimetic vision devices has been proposed and implemented [34]. In the context of imaging and vision, it is crucial to observe that biology has no notion of an image frame as a useful representation of visual information. On the other hand, the frame-based approach is the unchallenged standard in machine vision, and many techniques for information extraction (e.g., cosine

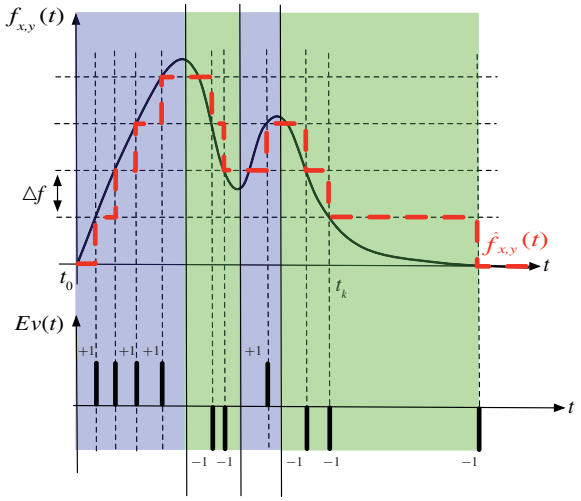

Fig. 2. Codification of pixel gray-level variations into temporal contrast events following the T codification process (5).

transform, perceptual linear prediction, SIFT, mel cepstrum, etc.) have been developed for this type of scene representation. Neuromorphic vision sensors feature massively parallel preprocessing of the visual information at the pixel level, combined with highly efficient asynchronous event-based information encoding and data communication. The temporal contrast address-event representation dynamic vision sensor (DVS) [35] used in this paper is built around an array of autonomous pixels asynchronously generating continuous-time "spike" events that encode relative changes in pixel illumination.

The device mimics the magnocellular transient pathway of the human retina and senses the dynamic information present in natural scenes at temporal resolutions much higher than most conventional image sensors can achieve. In addition, the device largely suppresses the temporal redundancy common to all frame-based image acquisition techniques. The temporal contrast events are communicated off-sensor in the form of continuous-time digital data packets conveying the responding pixel's x, y array address. Data transfer is optimized for real-time operation, trying to maximally preserve the event's crucial timing information. Finally, the sensor principle discards dc information, the DVS does not deliver classical image data in the form of absolute intensity information. 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 event it emitted (typically 15% contrast). 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 the order of microseconds [35].

Fig. 1 shows typical DVS pixel signal waveforms. In response to positive and negative photocurrent gradients, temporal events of two polarities, referred to as ON and OFF events, are generated by the pixel circuit. The gradient is encoded in inter-event intervals (see Fig. 2).

The extremely high temporal resolution of this vision sensor allows the design of new methods usually not even imaginable with the widespread standard frame-based devices that capture images at frame rates rarely higher than a few hundred frames per second. Because each pixel can be individually refreshed every 1/10 000th of a second, the event-based sensors can capture very fast dynamic scenes at "effective frame rates" of tens of kilohertz. Many popular algorithms can be drastically improved.

1) Highly accurate object segmentation in the context of fast motion which can provide improved accuracy for tasks such as face or gesture recognition.
2) Ability to handle and generalize complex geometric constraints (geometric calibration, epipolar geometry, etc.) in nonconventional (e.g., catadioptric) imaging systems.

Many limitations encountered in conventional frame-based vision are expected to be solved or reduced to some extent. Our paper focuses only on the problem of noncalibrated stereo vision, which is a prerequisite for many advanced event-based vision applications.

III. From Images to Frameless Spatiotemporal Luminance Events

This section presents an event-based formalism that is sufficiently generic to encompass DVS processing and to be extendable to other variations of event-based vision sensors.

A gray-scale image is a regular $M \times N$ array of positive numbers that measures the intensity of light incident to the pixels of a charge coupled device or complementary metal-oxide semiconductor sensor located at an image plane. A sequence of images I_{seq} is taken periodically at temporal steps Δt, between instants t_0 and $t_0 + n\Delta t$, such that

$$I_{seq} = \{I(t_0), I(t_0 + \Delta t), \ldots, I(t_0 + n\Delta t)\} \tag{1}$$

with $n \in \mathbb{N}$.

Given a pixel at $[x, y]^t$, if we express its intensity as a unidimensional function of time t

$$\begin{aligned} f_{x,y} : [0, T] &\mapsto \mathbb{R}^+ \\ t &\mapsto f_{x,y}(t). \end{aligned} \tag{2}$$

A digital image taken at time t is the set of unidimensional functions $f_{x,y}$ such that

$$I(t) = \{f_{x,y}(t)\}_{\substack{x \in [0, M-1] \\ y \in [0, N-1].}} \tag{3}$$

Using (2), an image sequence (1) can also be written as a set of 1-D functions $f_{x,y}(t)$ expressing the varying values of pixels in time. We can then write

$$I_{seq} = \{f_{0,0}(t), f_{0,1}(t), \ldots, f_{M-1,N-1}(t)\} \tag{4}$$

for $t \in [t_0, t_0 + n\Delta t]$ and $n \in \mathbb{N}$.

I_{seq} is then composed of a set of discrete functions $f_{x,y}(t)$, the values of which are provided using a global constant sampling process. All values are provided at the same time, the variations of functions $f_{x,y}(t)$ can only be detected at fixed times. Using a fixed global frequency has the effect of not only reducing the dynamics of the change detection process in $f_{x,y}(t)$, but it also produces a large amount of redundant data. This type of acquisition is incompatible with a compact representation of the data, it consumes resources because it acquires, transmits, and stores unnecessary redundant values. There are several ways to quantize $f_{x,y}$ according to its values.

Let us define $\{t_k\}$, the set of times of the signal sampling, assuming that $\forall k \in \mathbb{N}, t_{k+1} > t_k$, and t_0 is the initial time. Some standard sets are the following:

$$T = \left\{t_k \mid |\mathcal{F}(f_{x,y}(t_k)) - \mathcal{F}(f_{x,y}(t_{k-1}))| = \Delta_f\right\}. \tag{5}$$

The signal is sampled each time the variation of the magnitude of $\mathcal{F}$ is equal to Δ_f. In principle, there is no forward method to choose $\mathcal{F}$, it must be selected according to the task to be performed. In what follows, we will set $\mathcal{F}$ as the identity function in order to study the general properties of codification based on relative changes. In practice, there is no elegant formulation of T, as it is difficult to make an assumption about f to have f^{-1}, since this contradicts the random nature of light changes in scenes. Once T is set, $Ev(x, y, t)$ can then be defined as

$$Ev(x, y, t) = \delta(t, t_k) \cdot \text{sign}\left(f'_{x,y}(t)\right). \tag{6}$$

$\delta()$ is the Kronecker delta function and sign() is the sign function of a real number taking value in $\{-1, 1\}$.

$Ev(x, y, t)$ gives a more compact representation of $f_{x,y}$ using Δf, its values are in the set $\{-1, 0, 1\}$. The value 0 corresponds to an absence of change of $f_{x,y}$ which remains in an amplitude less than Δf, while $+1$ and -1 indicate a change, the direction of which is given by the sign.

Following the architecture of the DVS, $\mathcal{F}$ is set as the *log* function. This choice was motivated by two main considerations. The first comes from the goal to provide for a wide dynamic range of pixel intensities. The second is more connected to biomimetic issues, namely to design a logarithmically encoding pixel that is sensitive to (temporal) contrast as in vertebrate retinae [35].

IV. Computation of the Epipolar Configurations

A. Classic Framework (Frame-Based)

The fundamental matrix is generally determined from a set of point correspondences [36]–[38]. The most widely used method employs the eight-point algorithm introduced by Longuet-Higgins for computing the essential matrix [39]. The eight-point algorithm has the great advantage of being linear, which allows it to be easily computed. If eight-point matches are known, then the fundamental is the solution of a set of linear equations. Otherwise, it can be found by solving a linear least squares minimization problem. Other more or less complex algorithms have been proposed. Several can be found in [36], [40], or [41]. It has long been thought that the eight-point algorithm is usually excessively sensitive to noise in the matched points. Hartley [38] showed that the eight-point algorithm implementations do not sufficiently take numerical considerations into account. He then introduced a normalization process of image points, using simple transformations

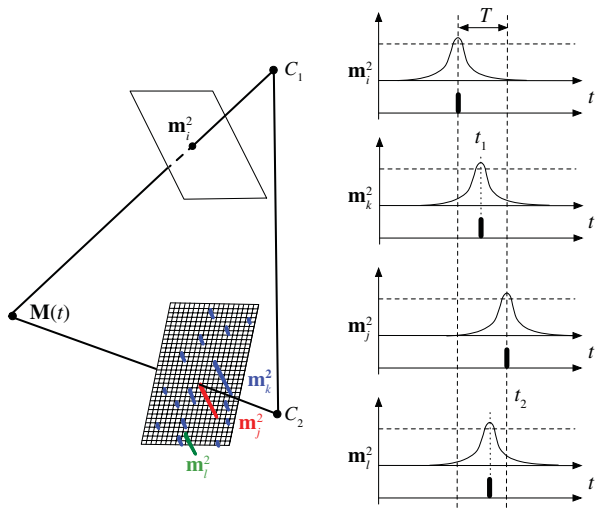

Fig. 3. $\mathbf{M}(t)$, a 3-D point appearing in the scene performing an unknown motion at time t, evokes response from pixel $\mathbf{m}_i^1$ at time t_1 and $\mathbf{m}_j^2$ at time t_2. Due to the nonzero difference $t_2 - t_1$ (of mean latencies) of each given pixel pair, the autonomous operating principle of the pixel circuits, and the asynchronous array readout, other pixels (e.g., $\mathbf{m}_k^2$ or $\mathbf{m}_l^2$) that have no *a priori* relationship with $\mathbf{M}(t)$ can generate events at the sensor output in close temporal proximity to t_1, even closer than the actual response from $\mathbf{m}_j^2$.

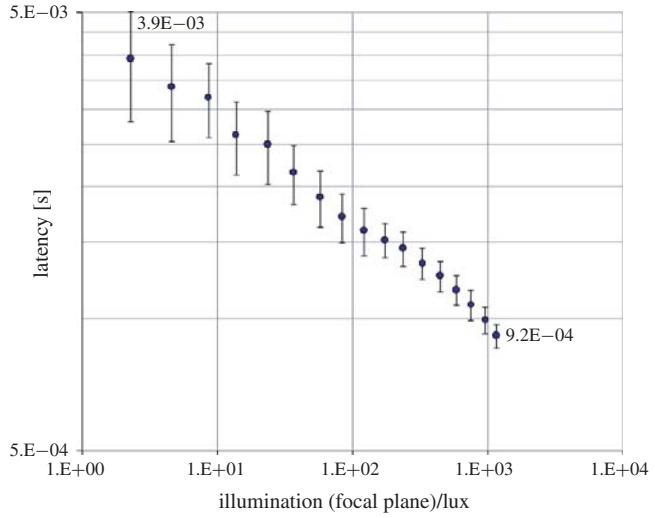

Fig. 4. Latency as a function of chip illumination. Error bars represent the jitter.

(translation and scaling) to introduce stability into the process of estimating the fundamental matrix.

The fundamental matrix is defined by the equation

$$\mathbf{p}'^T F \mathbf{p} = 0 \tag{7}$$

for any pair of matching points $\mathbf{p}' \leftrightarrow \mathbf{p}$ in two images. With sufficient number of matched pairs $\mathbf{p}'_i \leftrightarrow \mathbf{p}_i$, (7) can be used to compute the unknown matrix F. In particular, writing $\mathbf{p} = [u \ v \ 1]^T$ and $\mathbf{p}' = [u' \ v' \ 1]^T$, each point match gives rise to one linear equation in the unknown entries of F. The coefficients of this equation are easily written in terms of the known coordinates $\mathbf{p}$ and $\mathbf{p}'$. Specifically, the equation corresponding to a pair of points $\mathbf{p}$ and $\mathbf{p}'$ will be

$$\begin{aligned} uu' f_{11} + uv' f_{21} + uf_{31} + vu' f_{12} + vv' f_{22} \\ + vf_{32} + u' f_{31} + v' f_{23} + f_{33} = 0 \end{aligned} \tag{8}$$

with f_{ij} being the ith row, jth column entry of F. The row of the equation matrix may be represented as a row vector $[uu' \ uv' \ u \ vu' \ vv' \ v \ u' \ v' \ 1]$. From all the point matches, we obtain a set of linear equations of the form

$$A\mathbf{f} = 0 \tag{9}$$

where $\mathbf{f}$ is a 9-vector containing the entries of the matrix F, and A is the equation matrix. Nowadays, computer text books use this method, and the reader can find all the details of the computation of the fundamental matrix in [1].

B. Event-Based Framework

Let two asynchronous event-based DVS sensors $\mathcal{C}_1$ and $\mathcal{C}_2$ observe a common part of a scene. A 3-D point $\mathbf{M}$ moving in space triggers changes of luminance in the field of view of the sensors (Fig. 3). Events will be written as $Ev(\mathbf{m}_k^i, t)$, where $\mathbf{m}$ is the spatial location $\mathbf{m} = (x, y)^T$, superscript i indexes the sensor in which the event happens, and subscript k indexes the event in the retina $\mathcal{C}_i$. The 3-D point $\mathbf{M}(t)$ generates an event $Ev(\mathbf{m}_k^i, t)$ when it projects onto the ith sensor's focal plane $\mathcal{C}_i$ according to the classic known geometric relationship

$$\begin{pmatrix} \mathbf{m}_k^i \\ 1 \end{pmatrix} = P_i \begin{pmatrix} \mathbf{M}(t) \\ 1 \end{pmatrix} \tag{10}$$

where P_i is the projection matrix associated to $\mathcal{C}_i$ [1].

Following the temporal properties of the asynchronous event-based sensors, in an ideal case, the set of corresponding $Ev(\mathbf{m}_k^i, t)$ should be time-stamped with equal values, as they are the consequence of the same event that happens at a given time. We can then expect to be able to perform a direct match on their times of arrival.

Unfortunately, due to the constraints of electronic architecture, events rarely arrive at the same time. The DVS pixel is nonideal in many respects. It has limited contrast sensitivity, nonzero and variable latency from visual stimulus to event generation, random jitter in the event timing, and non-stimulus-related background event activity due to noise and leakage phenomena. At the system level, additional impacts on the event timing precision due to limited communication channel bandwidth, data collisions, and asynchronous bus arbitration are observed.

In a typical application environment, most of these effects can be neglected, and only latency and jitter are of concern. The latency of the pixel response is the time from the occurrence of an illumination change at the photodiode to the corresponding event output. This variable delay time is, in typical illumination conditions, limited by the photoreceptor bandwidth, which is proportional to the photocurrent and hence to pixel illumination (see Fig. 4). For a given scene, the event latency can be assumed to be similar and constant for a pair of pixels being stimulated at comparable illuminations and by equal temporal contrast.

The jitter, defined as the standard deviation of the latencies of one pixel stimulated by identical temporal contrasts, is

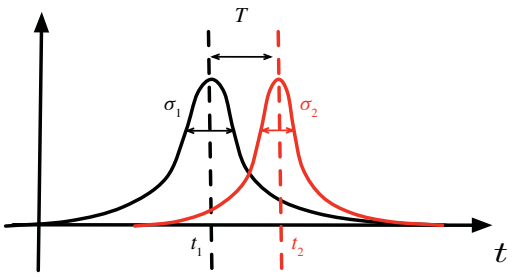

Fig. 5. Distributions of event timings of a pair of pixels responding to a number of identical temporal contrast stimuli at time $t = 0$. Pixel 1 has a mean latency of t_1 and a standard deviation of latencies of σ_1, also known as jitter (pixel 2 likewise).

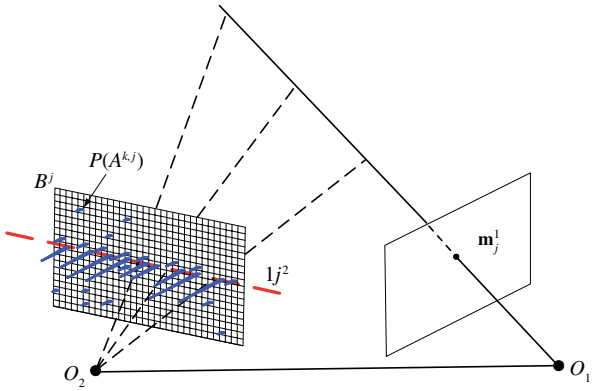

Fig. 6. Epipolar geometry structure from the coactivity probabilistic field and pixel's activities.

caused by photocurrent and dark current shot noise in the photoreceptor as well as readout noise in the pixel circuitry. It can be modeled as Gaussian distributed around a scene background illumination-dependent mean latency which is constant but different from pixel to pixel and is due to process parameter variations across the array (fixed-pattern noise) and from chip to chip. This jitter, apart from effects on latency due to the different illuminance of pixels, is the dominant source of pixel-event timing variations.

Measurement results for the latency and jitter of a single DVS pixel as a function of pixel illumination are shown in Fig. 4. For a typical indoor application illumination range, as present in the experiments in Section V, of 10–300 lux, the jitter of one pixel is of the order of 100–200 μs.

Taking into account the experimentally established pixel circuit characteristics, the distributions of event timings of a pair of pixels responding to a number of identical temporal contrast stimuli at time $t = 0$ can be depicted as shown in Fig. 5. Pixel 2 with mean latency of t_2 will, in most cases, deliver its event after pixel 1, but not always. The time difference $T = t_2 - t_1$ is not determined and can only be described in probabilistic terms. These uncertainties in pixel response times prevent the aimed-for stereo matching of single pixel events solely based on the exact timing of event occurrence. It is hence necessary to establish a search window in time, centered at t_1, during which the occurrence of t_2 is expected and consequently only an event falling into this time window is accepted for stereo matching.

The mean difference $T = t_2 - t_1$ (of mean latencies) of each given pair of pixels across two large arrays of pixels is zero. However, a mean upper bound to the difference of mean latencies for any given pair of pixels has to be taken into account for establishing the size of a search window. Experimentally establishing the upper bound to the difference of mean latencies of all possible pairs of pixels from two sensor arrays is very costly. An optimum search window size can be established theoretically only if this parameter is known. At this point, we can only estimate its value based on our knowledge of the semiconductor fabrication process's parameter variations, as published by the foundry, to be on the order of a few hundred microseconds. Following these observations, our experiments used time search window sizes between 500 μs and 2 ms.

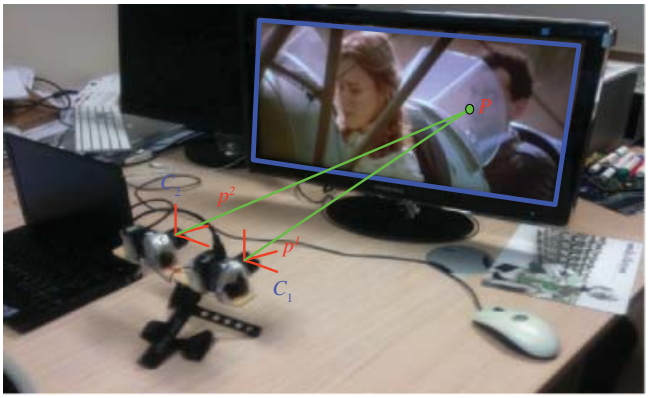

Fig. 7. Experimental setup the two DVS retinas are placed in front of an light-emitting diode (LED)-lit liquid crystal display (LCD) screen. The screen's position was changed to provide several levels of depth. For each position of the screen, the probability coactivation field was computed for 300 points. For each depth, to each pixel $p^1 \in \mathcal{C}_1$ corresponds a unique pixel $p^2 \in \mathcal{C}_2$.

The nonzero $t_2 - t_1$, together with the autonomous operating principle of the pixel circuits and the asynchronous array readout strategy, leads to a further complication of the situation. Assume a moving scene point $\mathbf{M}(t)$ that will generate corresponding events from pixels $\mathbf{m}_i^1$ at time t_1 and $\mathbf{m}_j^2$ at time t_2. As shown in Fig. 3, when pixel $\mathbf{m}_i^1$ generates an event at time t_1, other pixels (e.g., $\mathbf{m}_k^2$ or $\mathbf{m}_l^2$) that have no *a priori* relationship with $\mathbf{M}(t)$ can also generate events in close temporal proximity to t_1, even closer as the actual response from $\mathbf{m}_j^2$. This can be due to other scene-related stimulation or also to background noise activity. Two fundamental strategies for resolving such event collision conflicts are conceivable. Either all events that occur during the search window are accepted as equiprobable matching partners, or the whole event is discarded if more than one possible partner appears in the search window.

C. Coactivation Sets

If $\mathbf{M}(t)$ produces two events $Ev(\mathbf{m}_j^1, t_j)$ and $Ev(\mathbf{m}_k^2, t_k)$, we then define $A_{k,j}$ as the set of

$$A_{k,j} = \left\{ t_k \in [t_j - T, t_j + T] \,\middle|\, Ev\left(\mathbf{m}_k^2, t_k\right) \neq 0 \text{ if } Ev\left(\mathbf{m}_j^1, t_j\right) \neq 0 \right\}. \quad (11)$$

Algorithm 1 Algorithm to estimate coactivation probability field B^j for every $m_j^1 \in \mathcal{C}_1$

1: **for all** pixels $\mathbf{m}_j^1 \in \mathcal{C}_1$ **do**
2: **for all** position of the screen **do**
3: Compute B_{pos}^j at the current position
4: $Posmax^j \leftarrow$ find the position of the pixel with the highest activity
5: update B^j with the new computed B_{pos}^j
6: **end for**
7: **end for**
8: Fit a line $\mathbf{l}^j$ on positions $Posmax^j$
9: return B^j and $\mathbf{l}^j$ for every $\mathbf{m}_j^1 \in \mathcal{C}_1$.

$A_{k,j}$ is the set of all the temporal activations of pixel $\mathbf{m}_k^2 \in \mathcal{C}_2$ happening within a time window T from those of pixel $\mathbf{m}_j^1 \in \mathcal{C}_1$. Using the definition of $A_{k,j}$ given by (11), we can write

$$P\left(A_{k,j}\right) = P\left(Ev\left(\mathbf{m}_k^2, t_k\right) \middle| Ev\left(\mathbf{m}_j^1, t_j\right)\right). \quad (12)$$

This is obviously as stated previously for $t_k \in [t_j - T, t_j + T]$. We can then write

$$P\left(A_{k,j}\right) = \frac{P\left(Ev\left(\mathbf{m}_k^2, t_k\right) \cap Ev\left(\mathbf{m}_j^1, t_j\right)\right)}{P\left(Ev\left(\mathbf{m}_j^1, t_j\right)\right)}. \quad (13)$$

If B^j is the matrix of elements $P(A_{k,j})$, then it contains the probability of all temporal appearance of pixels in $\mathcal{C}_2$ for the pixel $\mathbf{m}_j^1 \in \mathcal{C}_1$

$$B^j = \begin{bmatrix} P(A_{1,j}) & P(A_{2,j}) & \cdots & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ \cdots & \cdots & \cdots & P(A_{k,N}) \end{bmatrix} \quad (14)$$

with N being the sensor's total number of pixels. According to noise parameters and the period of observation T, most of $A_{k,j} = 0$, i.e., B^j is a sparse matrix which can then be seen as a coarse expression of the corresponding epipolar line of pixel $\mathbf{m}_j^1 \in \mathcal{C}_2$.

The probability of pixels' activation is directly linked to their possibility of match. If B^j is observed over large periods of time, it is then expected that the values of the probabilities $P(A_{k,j})$ of corresponding pixels $\in \mathcal{C}_2$ to $\mathbf{m}_j^1$ are high, while noncorresponding pixels are low, following an appearance frequency directly linked to the sensor's noise and events appearing in scenes. The probability field B^j contains a geometrical structure equivalent to an epipolar line in the focal plane (see Fig. 6). In the present case, the epipolar structure is a line that can be estimated using robust fitting techniques on most on active pixels. This process, as shown in Fig. 6, associates a line $\mathbf{l}_j^2 \in \mathcal{C}_2$ corresponding to the epipolar line of point $\mathbf{m}_j^1 \in \mathcal{C}_1$. In a second stage, when more than eight lines are estimated—in the case of a central perspective planar sensor—the fundamental matrix can then be estimated following a simple robust DLT alogorithm using the relation $\mathbf{l}_j^2 = F\mathbf{m}_j^1$ for every j.

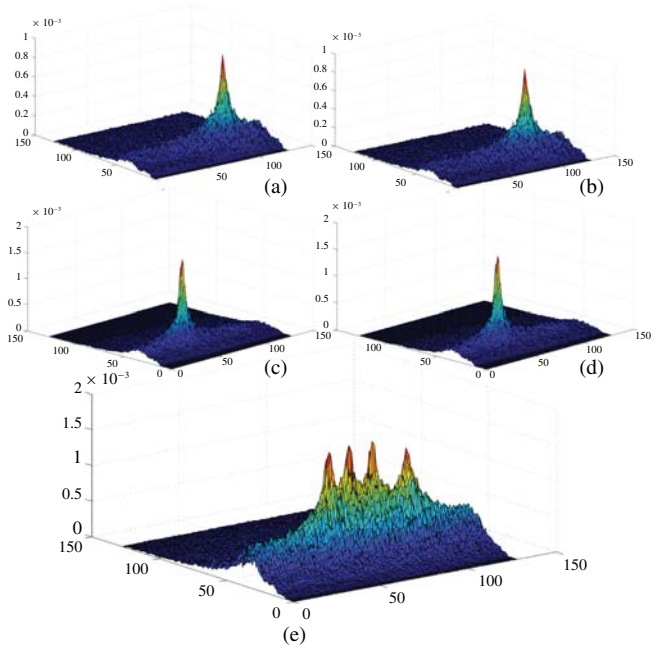

Fig. 8. Activation probability field B^j of a pixel (e) using four conditions (a)–(d) of the screen.

D. Remarks

Notice that this estimated fundamental matrix is computed in a very unusual way. There was no need to define or match any pattern in scenes. The fundamental matrix appeared implicitly from the coactivation of pixels and allowed epipolar lines to appear spontaneously. It then becomes possible, in a second stage, to provide an estimation of the fundamental matrix. This unusual estimation process relies directly on the observation of Hebb which states *"If the inputs to a system cause the same pattern of activity to occur repeatedly, the set of active elements constituting that pattern will become increasingly strongly interassociated. That is, each element will tend to turn on every other element and (with negative weights) to turn off the elements that do not form part of the pattern. To put it another way, the pattern as a whole will become auto-associated"* [42].

The estimated fundamental matrix can then be the result of temporal coactivations of pixels' activities and can therefore be a considered a Hebbian fundamental matrix.

V. Experimental Evaluation

For the experimental evaluation of the epipolar geometry, two DVS sensors were used [35]. We placed the two cameras in front of an LED-lit LCD screen and, as discussed in Section IV-B, identified the optimal search window experimentally. The optimum window size was found to be 1 ms (see Fig. 7 for a photograph of the experimental setup).

The screen's position was changed, providing several levels of depth. For every position of the screen, the probability coactivation field was computed for 300 points of $\mathcal{C}_1$. The screen continuously displayed a movie. We used many different movies, the choice of a specific movie did not significantly influence the result.

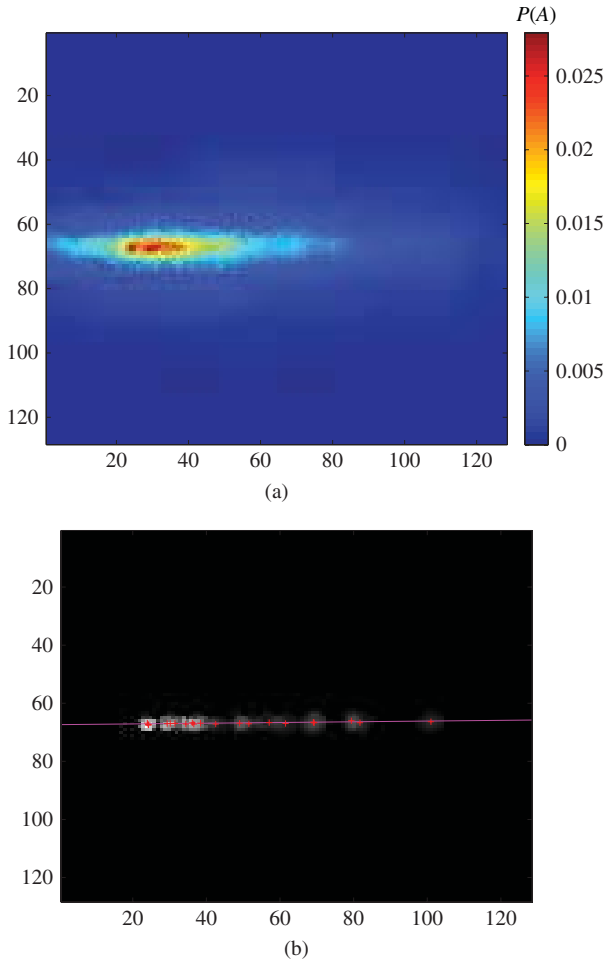

Fig. 9. Geometric structure of the activation peaks corresponding to the epipolar line of the monitored pixel. (a) Complete probability field activity B^j. (b) Thresholding the activation peaks of B^j reveals the geometric line structure of the epipolar geometry of the corresponding central setup.

A total of 24 random screen positions were used to provide stimuli at different depths. For each position, each pixel of $\mathcal{C}_1$ was associated to its coactivity probability field B^j. Fig. 8 shows the experimental results for four different screen positions for a single pixel. A peak in the activation map, clearly visible for each position, corresponds, as shown in Fig. 7, to a matched point of the monitored pixel at the screen's depth. When all experimental positions were combined, the geometric alignment of the peaks corresponded to the epipolar line. The general procedure to obtain B^j for every pixel of $\mathcal{C}_1$ that provides data to estimate the epipolar geometry is given by algorithm (1). Fig. 9 shows the result of this operation for a pixel of $\mathcal{C}_1$.

A. Algorithms

Following are descriptions of the main algorithms used. They mainly differ in how the best pixel activity is computed and which method and data were used to compute $\mathbf{l}^j$.

1) "Max-Fitting": Raw Position of the Maximally Active Pixels: In this algorithm, the epipolar line is robustly fitted to the raw position of the maximally active pixels for every screen's position.

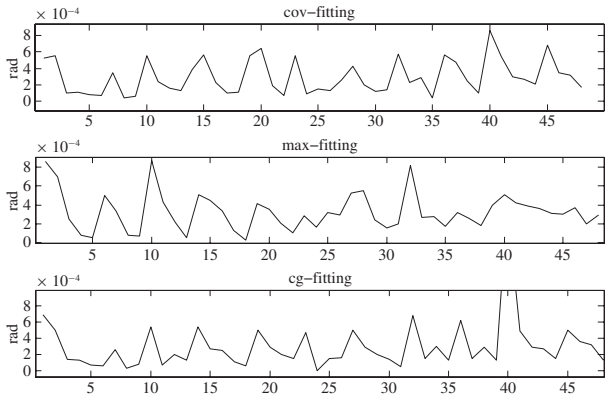

Fig. 10. Angle (y-axis) between max points (x-axis), uncertain fitting, and center of gravity, epipolar lines were estimated using the standard fundamental matrix computation.

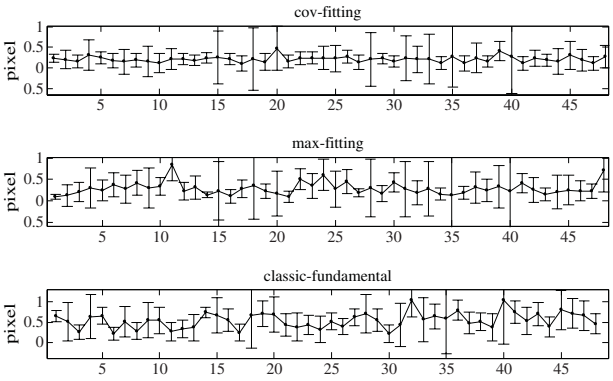

Fig. 11. Pixellic distance (y-axis) between max and gravity center points (x-axis) with estimated epipolar lines.

2) "Cg-Fitting": Gravity Center of B^j_{pos} Threshold $\in [0, 1]$: This algorithm differs by thresholding every B^j_{pos} in order to obtain the best active pixels. The computed position of the maximally active pixels is the mean of the position of remaining pixels. This algorithm allows us to get a subpixel position resolution of the max.

3) "Cov-Fitting": Covariance Fitting of l^j Using Max $B^j_{pos} >$ Threshold $\in [0, 1]$: This algorithm is very similar to the preceding one. It differs by the fitting of $\mathbf{l}^j$, which introduces an uncertainty measure into the location of the max position—highest coactivation location—using a covariance matrix computed from the position of the remaining thresholded points.

B. Results

We tried the various algorithms with five different videos displayed on the LCD screen. The exact choice of the video movie is of no relevance as long as it contains movements. Each sample was presented to the cameras for a duration of 20 min at every screen position. We made no specific choice for the sequences used, they were chosen randomly. In order to provide a "ground truth" and to allow comparison, we estimated the fundamental matrix of the system using the classic computation scheme with the toolbox provided in [43]. We used a blinking LED to produce a set of scene points, using

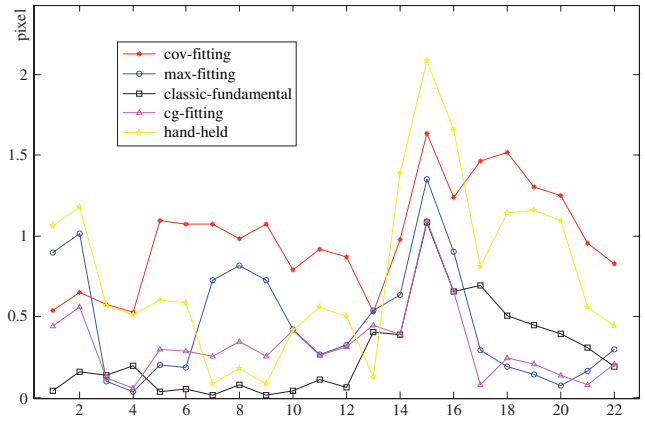

Fig. 12. Pixellic distance (y-axis) between manually matched points (x-axis) used to compute the "ground-truth" classic fundamental matrix and the epipolar lines obtained by the estimation methods. Point 15 is clearly an outlier for all the estimation methods.

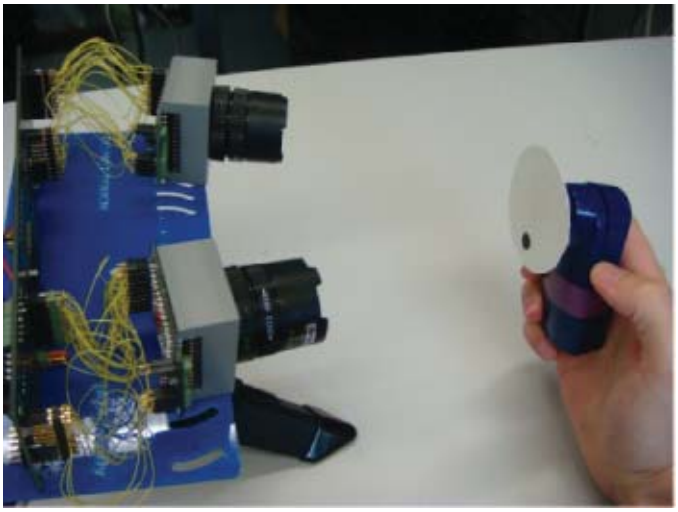

Fig. 13. Experimental validation protocol 3-D reconstruction of the locations of the center of gravity of a rotating disc, tracked and reconstructed in three dimensions using the estimated fundamental matrix (disc device hand-held).

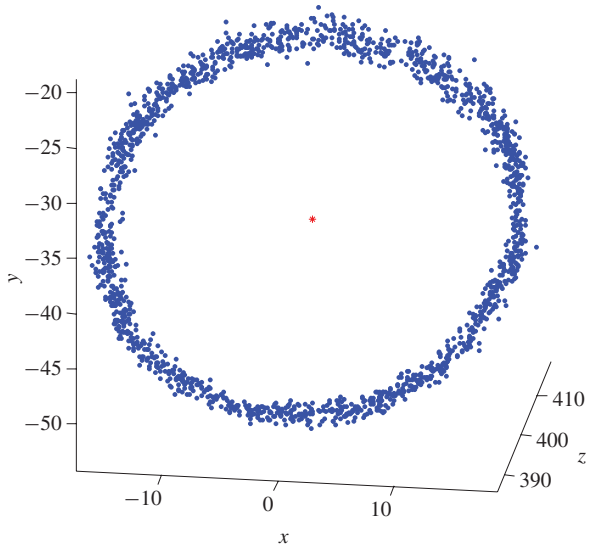

Fig. 14. 3-D reconstruction (in cm) of centers of gravity of the disc's locations in space. The mean estimated radius is 1.09 cm while the ground truth is 1 cm.

the center of gravity as an input for the robust estimation process. A total of 22 points were used. In the following experiments, the estimated fundamental matrix will be referred to as the classic fundamental.

Several criteria were analyzed to test for the accuracy of each presented algorithm. The fundamental matrix produced by the classic method was used as the ground truth.

1) Lines produced by the fundamental matrices estimated with each algorithm were compared with the classically computed ones. Fig. 10 shows the slope differences of epipolar lines with respect to the lines produced by the classic method.
2) Fitting errors are shown in Fig. 11 and compared in Fig. 12. One can see that the mean errors are less than a pixel, with standard deviations of the same order of magnitude. An examination showed that, if the classic method usually provides fewer fitting errors, the event-based methods using the centers of gravity or the local maxima detection can perform as efficiently. Most of the techniques provide usable results. Fig. 12 shows that both the "max-fitting" and the "cg-fitting" algorithms provide a precision better than one pixel. Notice that, in itself, choosing the maximum pixel as an input to the fitting of the epipolar line is sufficient. The "cov-fitting" algorithm provided the least accurate results because the covariance matrix depends on the orientation of the screen, thereby introducing unnecessary inaccuracies.
3) "Hand-held" estimation (Fig. 12) corresponds to experiments in real scenes. The stereo rig has been moved freely in a normal indoor environment. The overall results show that the method works in real applications and is not restricted to an experimental setup with a static vision system and the use of a screen.

The hand-held fundamental matrix was used in a 3-D reconstruction process using a DVS stereo rig. As shown in Fig. 13, the stereo system observed a disc rotating at 200 rpm. Every 2 ms, the center of gravity of all incoming events was extracted from both images and reconstructed in three dimensions. The intrinsic parameters of each camera were previously estimated using classic techniques as described in [1]. Fig. 14 shows the reconstructed locations of the center of the rotating disc. The estimated radius of rotation of the disc from all the 3-D locations was 1.09 cm, while the real value was 1 cm. The dispersion around the circular rotation plane, 1.1 mm, was mainly due to the limited resolution of the sensors (128×128) and the fact that the rotating device was hand-held.

Further experimental results were acquired using a set of frames built from the event-based sensor signals. This approach is not meant to be applied to frames, but is proposed for the purpose of illustration and to allow comparison with classic fundamental estimations. Frames were built by integrating events over short time intervals. The illumination dynamic was produced by means of a flashing light. Fig. 15 shows the produced stereoscopic images with manually selected pixels in

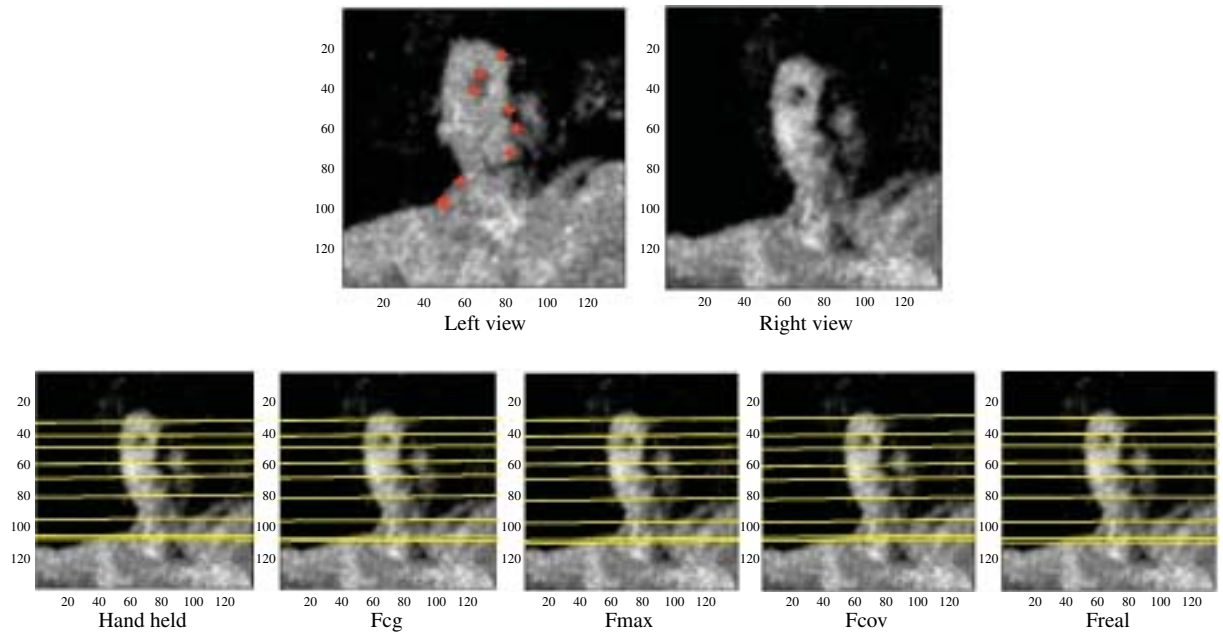

Fig. 15. Top: Pair of stereoscopic images with selected points in the left view. Bottom: The associated epipolar lines obtained with the estimated fundamental matrix: Fcg using "cg-fitting," Fmax using "max-fitting," Fcov using "cov-fitting," Freal corresponding to the "classic-fundamental" ground-truth estimation, and finally the "Hand-held" where the stereo rig was freely moved in an indoor environment.

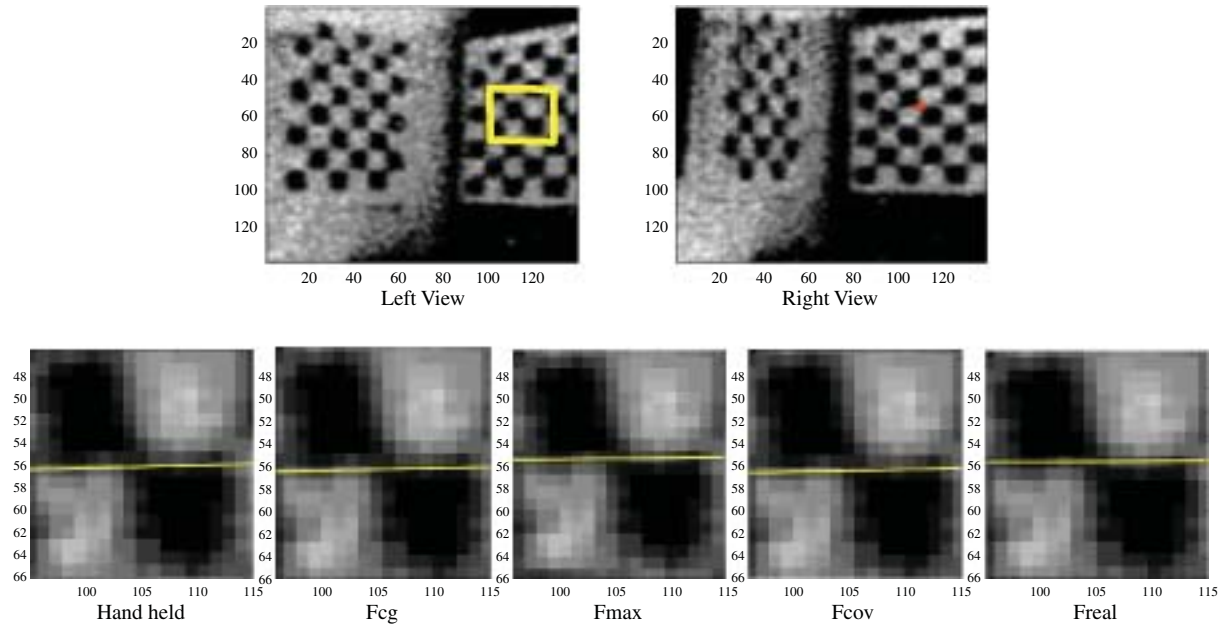

Fig. 16. Top: Chessboard stereo pair with a selected corner marked in the left image. Bottom: Results for the aforementioned versions of fundamentals. The epipolar lines are all going through the associated corner in the right image up to some estimation errors.

the left view, the epipolar lines were computed according to the fundamental matrices estimated with the aforementioned algorithms. The same test was carried out with a chessboard (see Fig. 16). The selection of a corner in the left image produced the associated epipolar line in the right one. The zoomed image clearly shows the accuracy of each computed lines going through the corresponding corner in the right image.

C. Remarks

Using an LCD screen as a stimulus generator allowed us to speed up the experiment by generating dynamic events at different depths. Since the screen produced frames, it is legitimate to ask whether the proposed method is applicable to a standard frame-based camera. To change the intensity of a pixel, the liquid crystal of the screen must reorient itself. The time needed to realize the change depends on the initial and final intensity values. If the camera can capture the intensity change dynamics, then the method described above will work by distinguishing pixels according to the intensity dynamisms. For an LCD screen with a mean response time of 16 ms, the camera should be at least five to ten time faster: i.e., the camera should be able to capture at 300–600 f/s. These frame rates are

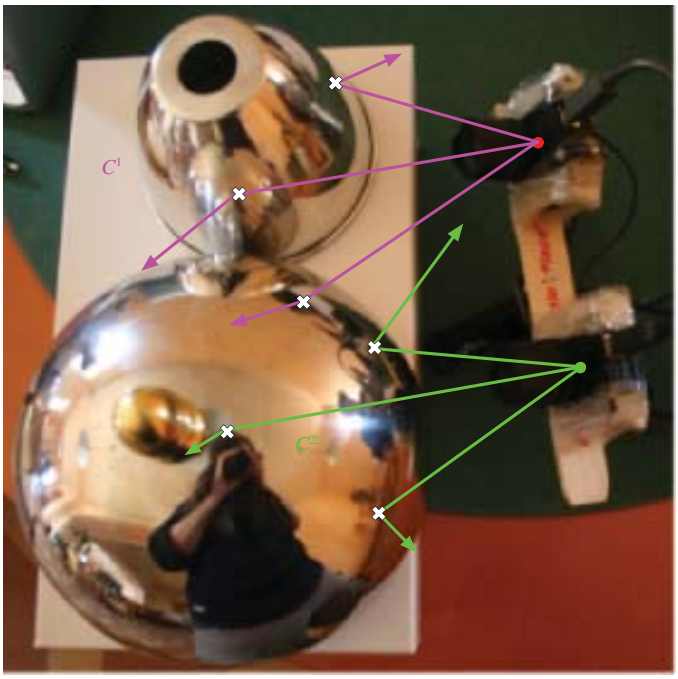

Fig. 17. Noncentral catadioptric setup with event-based cameras. Each sensor, combined with its reflective surfaces, is a noncentral vision system. The mounting is designed to probe the stereoscopic geometry of the device from coactivation sets computed with the probability map.

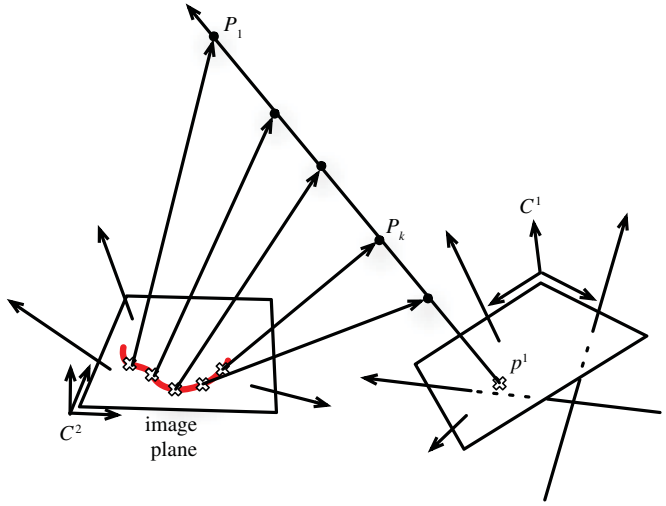

Fig. 18. Example of a nonperspective stereo system the equivalent of the epipolar line is built with coactivation selection. The schematic serves only for the purpose of illustration and represents a particular case for which a pixel on the right sensor is mapped to a curve on the left sensor. Nothing can guarantee the existence of such a curve.

only possible with high-speed cameras dedicated to a specific use. On the other hand, this dynamic is perfectly within the range of temporal resolution of a DVS. The number of events generated is a function of the change of value encoding the dynamic properties of the scene.

VI. Probing New Geometries

The epipolar constraint condenses the geometric essence of a central stereoscopic vision system: 3-D scene points and camera centers are related by geometric loci called epipolar planes, which are projected onto epipolar lines in the image planes. Extensive work has investigated central stereovision with frame-based cameras: the cameras' geometry can be recovered by estimating the fundamental matrix, even for uncalibrated systems. Both operations are time and resource consuming, as all the frame's pixels must be processed exhaustively.

Nonconventional sensors are expected to produce a nonconventional geometric relationship, which is finally projected into the image space. One representative example is the central omnidirectional catadioptric sensor for which the epipolar lines are replaced by conic curves [5], [44]. In an even more general context, if an epipolar curve exists, it can only be computed numerically, due to the noncontinuity of the light sampling process or some singular configuration. Using standard image processing techniques, estimating geometric constraints can be difficult or even impossible because of nonlinearities that limit the detection and matching algorithms. Event-based sensing is adapted to dynamic contexts, and the possibility of directly segmenting and matching events through coactivation maps obviously opens new perspectives. Pixels matched over time with coactivation maps will produce the epipolar curves that can be used to analyze and characterize the geometric configuration.

To illustrate the concept, we used a noncentral stereo system that relied on two DVS cameras observing a spherical reflector and another reflector of unknown shape (see Fig. 17). This vision system is clearly noncentral. The process of estimating an activation map was performed as explained in the previous section. In the present case, shown in Fig. 18, a pixel $p^1 \in \mathcal{C}^1$ intersects a set of lines corresponding to the pixels' line of view of $\mathcal{C}^2$. The intersections' results are only restricted by the sensors' geometry—more likely a curve in this example. The results of the coactivation of a set of connex pixels are presented in Fig. 19. In the present case, the activation map of each pixel $p \in \mathcal{C}^1$ is close to a curve and corresponds to its common field of view with the rays of $\mathcal{C}^2$.

It seems obvious that the method could deal with more complex geometries that do not necessarily follow a compact parametric formulation. In that case, how could such information be compactly represented?

In such cases, we would expect each pixel to correspond to an activation map. The solution lies in the storage capacities, which current technologies can encompass easily. The current sensor prototype has a resolution of 128×128, let us then imagine that higher resolutions will be obtained in the near future, such as 1024×768. At this resolution, if every pixel's activation map must be stored, and assuming a binary coding of the information on the pixels' intersections, the total cost per pixel would be $(1024 \times 768)^2$ bits. Thus, the total cost for all maps is 72 Gbytes. However, because the activation maps are sparse structures, the total number of pixels of an epipolar line compared to the total number of pixels of the sensor represents at most 0.17%.

This is the upper bound of the ratio, as it corresponds to the length of the diagonal divided by the total number of pixels. If a sparse representation is used, then the total

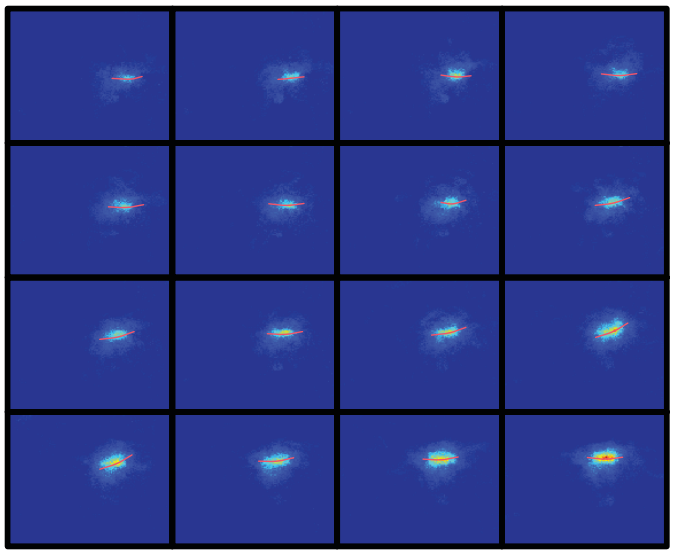

Fig. 19. Probability maps for a set of columnwise consecutive pixels in the left sensor. For each pixel, the coactivated set is shown with the maximum value underlying the best fit in the right sensor.

memory needed to store such a system would be 120 Mbytes $\backsimeq$ 0.17×72 Gbytes. This drastically reduces the needed storage space. The additional use of compression techniques and fast memory access, such as lookup tables, are expected to speed up the process further.

VII. Discussion

The geometric foundation of stereovision has been considered well established for a long time, but this paper shows a totally original way to tackle the problem with the help of mechanisms inspired by biological vision. In this new approach, time, which is usually considered a secondary variable of the visual signal, prominently contributes to the characterization of the vision system.

Because biological vision is naturally adapted to time-based processing in a highly parallelized manner, bio-inspired mechanisms open new paths to solving vision problems in elegant ways and, more importantly, allow us to consider new perspectives to solve usually nonsolvable problems using standard tools.

The use of precise timings appears to be a key element of this computational principle. Information is not only carried by the nature of events (e.g., which type of ganglion cells fire) but also in the precise temporal and spatial locations at which they occur. The adjunction of time in perception seems to open promising perspectives for which principles and properties are yet to be discovered.

The example of epipolar geometry shown in this paper gives an idea of the powerful potential of an event-based formalism in conjunction with the adequate hardware: we have built a tool which is similar to a contact probe. It measures directly the geometric locus of the epipolar constraint without the need for any prior knowledge of a binocular system. As it is, it can, in principle, be used to explore even more complex multioccular (or equivalently noncentral) systems.

VIII. Conclusion

This paper provided an event-based formulation of the epipolar constraint for a stereovision system formed by two asynchronous event-based cameras. Instead of exclusively using geometric constraints to estimate the fundamental matrix, the presented method showed that the use of time allows the estimation of the epipolar constraint that appears as a temporal consequence of spatial "coactivations" of pixels. A direct consequence of the presented approach was its ability to retrieve the epipolar geometry even in cases where it cannot be written in a closed form. The fundamental link between two or more views appears then naturally as a probabilistic wiring of pixels. The presented event-based framework provides a general tool to establish and estimate the geometric constraints of any stereovision system regardless to its geometrical structure.

One of the most important properties of this approach was its intrinsic ability to automatically define and extract the needed features to compute the fundamental matrix. It was not required to predefine them. The experimental results show the applicability of the method to perspective binocular systems but also to more complex configurations such as noncentral systems for which we have currently no real knowledge. The use of high temporal resolutions in computation should impact other fields of computer vision in the near future by allowing the reformulation of ill-posed techniques in a temporal manner similar to biological system computations.

References

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[2] O. D. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?" in *Proc. Eur. Conf. Comput. Vis.*, 1992, pp. 563–578.

[3] R. Hartley, R. Gupta, and T. Chang, "Stereo from uncalibrated cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Champaign, IL, Jun. 1992, pp. 761–764.

[4] R. Benosman and S. Kang, *Panoramic Vision: Sensors Theory and Applications*. New York: Springer-Verlag, 2002.

[5] T. Svoboda and T. Pajdla, "Epipolar geometry for central catadioptric cameras," *Int. J. Comput. Vis.*, vol. 49, no. 1, pp. 23–37, Aug. 2002.

[6] S. Baker and S. Nayar, "A theory of single-viewpoint catadioptric image formation," *Int. J. Comput. Vis.*, vol. 35, no. 2, pp. 1–22, 1999.

[7] L. Smadja, R. Benosman, and J. Devars, "Hybrid stereo configurations through a cylindrical sensor calibration," *J. Mach. Vis. Appl.*, vol. 17, no. 4, pp. 251–264, 2006.

[8] A. Torii, A. Imiya, and N. Ohnishi, "Two and three view geometry for spherical cameras," in *Proc. Workshop Omnidirect. Vis.*, 2005, pp. 1–8.

[9] S. Li, "Full-view spherical image camera," in *Proc. 18th Int. Conf. Pattern Recognit.*, Hong Kong, 2006, pp. 386–390.

[10] R. Swaminathan, S. Nayar, and M. Grossberg, "Designing mirrors for catadioptric systems that minimize image errors," in *Proc. Workshop Omnidirect. Vis.*, 2004, pp. 1–12.

[11] R. Hicks, "Designing a mirror to realize a given projection," *J. Opt. Soc. Amer. A*, vol. 22, no. 2, pp. 323–330, 2005.

[12] P. Rademacher and G. Bishop, "Multiple-center-of-projection images," in *Proc. SIGGRAPH*, 1998, pp. 1–8.

[13] R. Gupta and R. I. Hartley, "Linear pushbroom cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 9, pp. 963–975, Sep. 1997.

[14] H. Shum, A. Kalai, and S. Seitz, "Omnivergent stereo," in *Proc. Int. Conf. Comput. Vis.*, 1999, pp. 1–8.

[15] S. Seitz, "The space of all stereo images," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, vol. 1. Vancouver, BC, Canada, Jul. 2001, pp. 26–33.

[16] S. Nayar and A. Karmarkar, "360 Ã–360 mosaics," in *Proc. Comput. Vis. Pattern Recognit.*, 2002, pp. 388–395.

[17] R. Swaminathan, M. D. Grossberg, and S. K. Nayar, "Caustics of catadioptric cameras," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, vol. 2. Vancouver, BC, Canada, Jul. 2001, pp. 2–9.

[18] S. Ieng and R. Benosman, *Geometric Construction of the Caustic Surface of Catadioptric Non-Central Sensors*. New York: Springer-Verlag, 2006, ch. 1, pp. 39–54.
[19] T. Pajdla, "Epipolar geometry of some non-classical cameras," in *Proc. Comput. Vis. Winter Workshop*, 2001, pp. 159–180.
[20] T. Pajdla, "Stereo with oblique cameras," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 161–170, 2002.
[21] M. D. Grossberg and S. K. Nayar, "A general imaging model and a method for finding its parameters," in *Proc. Int. Conf. Comput. Vis.*, 2001, pp. 108–115.
[22] S. Ramalingam, P. Sturm, and S. Lodha. (2005, Jun.). *Toward Complete Generic Camera Calibration*. pp. 1093–1098 [Online]. Available: http://perception.inrialpes.fr/Publications/2005/RSL05a
[23] P. Sturm. (2005, Jun.). *Multi-View Geometry for General Camera Models*. pp. 206–212 [Online]. Available: http://perception.inrialpes.fr/Publications/2005/Stu05
[24] T. Debaecker, R. Benosman, and S. Ieng, "Cone of view camera model using conformal geometric algebra for classic and panoramic image sensors," in *Proc. 8th Workshop Omnidirect. Vis.*, Marseilles, France, 2008, pp. 1–11.
[25] S. Kuthirummal and S. K. Nayar, "Flexible mirror imaging," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[26] K. Boahen, *Communicating Neural Ensembel Between Neuromorphic Chips*. Norwell, MA: Kluwer, 1998, pp. 229–259.
[27] O. Braddick, *Neural Basis of Visual Perception*. Amsterdam, The Netherlands: Elsevier, 2001, pp. 16269–16274.
[28] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gomez-Rodriguez, L. Camunas-Mesa, R. Berner, M. Rivas, T. Delbruck, S. Liu, R. Douglas, P. Hafliger, G. Moreno, and A. Civit, "CAVIAR: A 45 k neuron, 5 M synapse, 12 G connects/s AER hardware sensory-processing-learning-actuating system for high-speed visual object recognition and tracking," *IEEE Trans. Neural Netw.*, vol. 20, no. 9, pp. 1417–1438, Sep. 2009.
[29] R. Serrano-Gotarredona, T. Serrano-Gotarredona, A. Acosta-Jimenez, and B. Linares-Barranco, "A neuromorphic cortical-layer microchip for spike-based event processing vision systems," *IEEE Trans. Circuits Syst.*, vol. 53, no. 12, pp. 2548–2566, Dec. 2006.
[30] E. Chicca, P. Lichtsteiner, T. Delbruck, G. Indiveri, and R. J. Douglas, "Modeling orientation selectivity using a neuromorphic multi-chip system," in *Proc. IEEE Int. Symp. Circuits Syst.*, Island of Kos, Greece, Sep. 2006, pp. 1235–1238.
[31] T. Y. W. Choi, P. Merolla, J. Arthur, K. Boahen, and B. E. Shi, "Neuromorphic implementation of orientation hypercolumns," *IEEE Trans. Circuits Syst.*, vol. 52, no. 6, pp. 1049–1060, Jun. 2005.
[32] C. Mead, "Adaptive retina," in *Analog VLSI Implementation of Neural Systems*, C. Mead and M. Ismail, Eds. Boston, MA: Kluwer, 1989, pp. 239–246.
[33] M. A. Mahowald and C. Mead, "The silicon retina," *Sci. Amer.*, vol. 264, no. 5, pp. 76–82, May 1991.
[34] K. Boahen, "Neuromorphic microchips," *Sci. Amer.*, vol. 292, no. 5, pp. 55–63, May 2005.
[35] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 $\times$ 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.
[36] Q.-T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo, "On determining the fundamental matrix: Analysis of different methods and experimental results," INRIA, Le Chesnay, France, Tech. Rep. RR-1894, 1993.
[37] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 161–195, Mar. 1998.
[38] R. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 6, pp. 580–593, Jun. 1997.
[39] H. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, Sep. 1981.
[40] P. A. Beardsley, A. Zisserman, and D. W. Murray, "Navigation using affine structure from motion," in *Proc. Eur. Conf. Comput. Vis.*, 1994, pp. 85–96.
[41] R. Hartley, "Euclidean reconstruction from uncalibrated views," in *Proc. 2nd Eur. - U.S. Workshop Invar.*, 1993, pp. 1–18.
[42] D. Hebb, *The Organization of Behavior*. New York: Wiley, 1949.
[43] *MATLAB Functions for Multiple View Geometry* [Online]. Available: http://www.robots.ox.ac.uk/~vgg/hzbook/code/
[44] S. B. Kang, "Catadioptric self-calibration," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Hilton Head Island, SC, Jun. 2000, pp. 201–207.

Ryad Benosman is an Associate Professor with University Pierre and Marie Curie, Paris, France, leading the Natural Computation and Retina Models Laboratory, Vision Institute, Paris. His work covers neuromorphic visual computation and sensing. He is currently involved in the French retina prosthetics project and in the development of retina implants. He is an expert in complex perception systems, which cover the conception, design, and use of different vision sensors covering omnidirectional 360° wide-field view cameras, variant scale sensors, and non-central sensors. He is among the pioneers of the domain of omnidirectional vision and unusual cameras and still very active in this domain. He has been involved in several national and European robotics projects, mainly in the design of artificial visual loops and sensors. His current research interests include the understanding of the computation operated along the visual system's areas and establishing a link between computational and biological vision.

Sio-Hoï Ieng is an Assistant Professor with University Pierre and Marie Curie, Paris, France, and a member of Vision Institute, Paris. He worked on the geometric modeling of catadioptric sensors assumed to be noncentral and their link to the caustic surface. His current research interests include computer vision, with special reference to the understanding of general visual sensors, exploring cameras networks, and studying the connection between biologic and artificial vision.

Paul Rogister received the Ph.D. degree from the Institute of NeuroInformatik, Eidgenössische Technische Hochschule, Zurich, Switzerland.

His current research interests include the development of automated behavioral neuromorphic stereo-vision monitoring systems that enhance the precision and reliability of behavioral assessment in experiments investigating the impact of nervous system lesions and rehabilitation, in particular for spinal cord injury and repair of the forelimb function.

Christoph Posch received the M.Sc. and Ph.D. degrees in electrical and electronics engineering and experimental physics from the Vienna University of Technology, Vienna, Austria, in 1995 and 1999, respectively.

He was with the European Organization for Nuclear Research, European Laboratory for Particle Physics, Geneva, Switzerland, from 1996 to 1999, on analog compelementary metal-oxide-semiconductors (CMOS) and BiCMOS integrated circuit design for particle detector readout and control. From 1999 onward, he was with Boston University, Boston, MA, where he was engaged in applied research and analog/mixed-signal integrated circuit design for high-energy physics instrumentation. In 2004, he joined the newly founded Smart Sensors Group, Austrian Institute of Technology (formerly Austrian Research Centers), Vienna, Austria, where he has been a Principal Scientist since 2007. He has authored more than 70 scientific publications and holds several patents in the area of vision and image sensing. His current research interests include neuromorphic analog very large scale integration, CMOS image and vision sensors, and biology-inspired signal processing.

Dr. Posch is a recipient and co-recipient of several scientific awards including the Jan van Vessem Award for the Outstanding European Paper from the IEEE International Solid-State Circuits Conference in 2006 and the Best Live Demonstration Award from the IEEE International Symposium on Circuits and Systems in 2010. He is a member of the Sensory Systems and the Neural Systems and Applications Technical Committees of the IEEE Circuits and Systems Society.