

Estimation of Vehicle Speed Based on Asynchronous Data from a Silicon Retina Optical Sensor

M. Litzenberger, A.N. Belbachir, *Member, IEEE*, N. Donath, G. Gritsch,
H. Garn *Senior Member, IEEE*, B. Kohn, C. Posch and S. Schraml

Abstract — This work presents an embedded optical sensory system for traffic monitoring and vehicles speed estimation based on a neuromorphic “Silicon-Retina” image sensor, and the algorithm developed for processing the asynchronous output data delivered by this sensor. The main purpose of these efforts is to provide a flexible, compact, low-power and low-cost traffic monitoring system which is capable of determining the velocity of passing vehicles simultaneously on multiple lanes. The system and algorithm proposed exploit the unique characteristics of the image sensor with focal-plane analog preprocessing. These features include sparse asynchronous data output with high temporal resolution and low latency, high dynamic range and low power consumption. The system is able to measure velocities of vehicles in the range 20 to 300 km/h on up to four lanes simultaneously, day and night and under variable atmospheric conditions, with a resolution of 1 km/h. Results of vehicle speed measurements taken from a test installation of the system on a four-lane highway are presented and discussed. The accuracy of the speed estimate has been evaluated on the basis of calibrated light-barrier speed measurements. The speed estimation error has a standard deviation of 2.3 km/h and near zero mean.

I. INTRODUCTION

THE problem of traffic speed monitoring has been tackled since long time, using non-optical sensors, like microwave RADAR (*Radio Detection And Ranging*) [1], audio systems [2] or induction loops, optical sensors like LIDAR (*Laser Imaging Detection And Ranging*) [3], [4] and, rather recently, using video-based systems.

The past decade has seen several attempts to provide image processing tools for traffic surveillance. In [5], a tutorial on the existing video processing techniques for traffic surveillance is presented. In [6], [7] and [8] the authors provide processing methods for estimating the vehicle speed from a sequence of video frames. The proposed methods make use of object recognition and background maintenance/subtraction on the successive frames to improve the estimation accuracy.

Manuscript received March 20, 2006; revised June 30, 2006.

All Authors are with the Austrian Research Centers, ARC Seibersdorf research GmbH, Tech Gate Vienna, Donau-City Str. 1, 1220 Vienna, Austria.

Corresponding author: martin.litzenberger@arcs.ac.at, Tel +43-50550-4111, Fax +43-50550-4125.

A patent application covering the system concept and the address-event processing algorithms has been filed under Austrian patent application number A1011/2005.

One significant problem with high resolution video systems is the high data volume delivered by the image sensor. Readout and processing of the largely redundant data ultimately face limitations due to computational effort and power consumption. In this paper, an embedded traffic data sensor based on a novel, bio-inspired “Silicon-Retina” (SR) imager [9], [10], [11] with focal-plane analog preprocessing is presented. This sensor combines sparse asynchronous data output with high temporal resolution and low latency, high dynamic range and low power consumption. In contrast to many other non-video based traffic data acquisition systems, the sensor has a high degree of freedom regarding its mounting position and is able to service several lanes simultaneously. This paper describes an embedded system built around this sensor, and demonstrates its performance for vehicle speed estimation in a real test environment.

The paper is structured as follows. In section II, the embedded sensory system is described. Section III describes the test site and shows data real world data. The algorithm for vehicles speed estimation is described in Section IV. The experimental results after the application of the presented algorithm on the real data are discussed in Section V. Section VI contains conclusions.

II. DESCRIPTION OF THE SILICON-RETINA EMBEDDED SENSORY SYSTEM

In this section, the temporal derivative SR sensor and the embedded traffic monitoring system [12] are briefly described.

In contrast to traditional CCD or CMOS imagers that encode image irradiance and produce constant data volume at a fixed frame rate, irrespective of scene activity, the SR sensor contains an array of autonomous, self-signaling pixels which individually respond in real-time to relative changes in light intensity by placing their address on an asynchronous arbitrated bus. Pixels that are not stimulated by a change in illumination are not triggered, hence static scenes produce no output. Because there is no pixel readout clock, no time quantization takes place at this point.

The sensor operates largely independent of scene illumination, directly encodes object reflectance, and greatly reduces redundancy while preserving precise timing information. Because output bandwidth is automatically dedicated to dynamic parts of the scene, a robust detection of fast moving vehicles at variable lighting conditions is

achieved. The scene information is transmitted event-by-event to a DSP via an asynchronous bus. The pixel location in the imager array are encoded in the event data that are reflected as i,j coordinates in the resulting image space in the form of *Address-Events* (AE) [13]. An effective way of processing AE data takes advantage of the efficient coding of the visual information by directly processing the spatial and temporal information contained in the data stream.

The high dynamic range of the photosensitive element ($>120\text{dB}$ or 6 decades) makes the SR imager ideal for applications with uncontrolled light conditions.

Fig. 1 depicts the general architecture of the concerned embedded sensory system, which comprises a SR imager, a First-In, First-Out (FIFO) buffer memory and a Digital Signal Processor (DSP). The SR imager and the DSP consume in total roughly 2.5 W of electrical power.

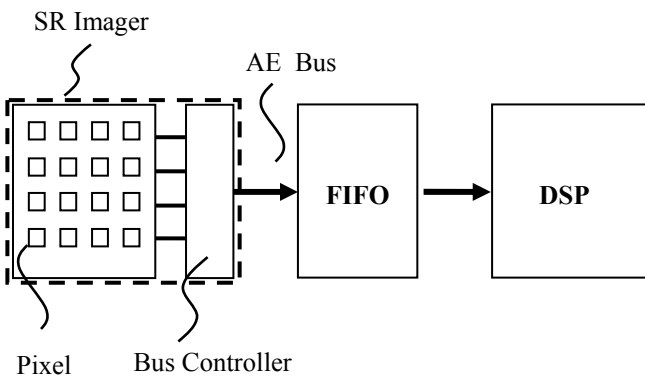


Fig. 1. Schematics of the general layout of the embedded traffic data sensor with a silicon-retina imager.

The results and data presented in this paper are produced by a system using a 64×64 pixel resolution version of the SR imager [11]. The location (address) of the event generating pixel within the array is transmitted to a FIFO on a 16-bit parallel bus, implementing a simple 4-phase handshake protocol. The FIFO is placed between the SR sensor and the DSP to cope with peaks of AE activity and is capable of handling up to 40 MHz memory access frequency. In the processing stage, every AE received by the DSP is labeled by attaching the processor clock ticks with 1ms precision as a time stamp. These data are the basis for the vehicle speed estimation.

The DSP also runs the AE processing algorithms for the vehicle speed estimation presented in the following sections. The full processing comprises the AE acquisition and time stamping, the vehicle recognition, counting, classification and speed estimation. The following sections of this paper focus on the speed estimating algorithms.

The images in Fig. 2 show a comparison of a still video picture and 64×64 pixel SR imager AE data of a highway traffic scene. In order to visualize the AE data, events have been collected for a 20 millisecond interval and rendered like a video frame. The different gray shadings encode pixel activity per unit time. Note that the white and black vehicles

both have very similar representation in the AE data stream illustrating the sensitivity of the imager even to small contrast changes.

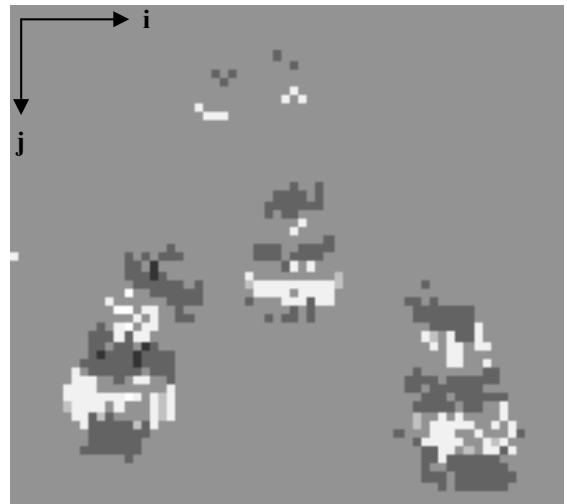
Detailed technical specifications of the embedded traffic data system can be found in [14]

III. EXAMPLE DATA AND TEST SITE

Fig. 3 depicts an example illustration of 4 seconds of an AER data stream, produced by vehicles moving along the highway section depicted in Fig. 2a. The pixel event activity is plotted for the imager i and j axis versus time. The events rate is encoded in grey levels from 0 (white) to 10 kevents/s (dark). Each vehicle produces compact trace of AE's in (t,i) and (t,j) space when moving towards the imager.



(a)



(b)

Fig. 2. Still image from (a) a conventional video camera (b) and representation of the AE data stream from the 64×64 SR imager. The axis of the imager coordinate system i,j and world coordinate system x,y are displayed in the images.

In total one vehicle trace typically consist of not more than 1000 AE's. In Fig 2a the traces of four vehicles, enumerated from 1 to 4, can be distinguished. Note that vehicles No. 1, 2 and 4 are driving on the right hand lane and No. 3 on the center lane.

A four-lane highway section was used as test site for the vehicle speed estimation. The embedded system containing the SR imager is mounted overhead in the middle of the highway at 11 m above the road surface. The sensor faces the oncoming traffic. The AE data are recorded and processed simultaneously in the embedded system. Single vehicle velocity results are transferred via TCP/UDP in XML format to a server in a traffic control center where they are stored in a SQL data base. The system is designed to send its data to any arbitrary IP client.

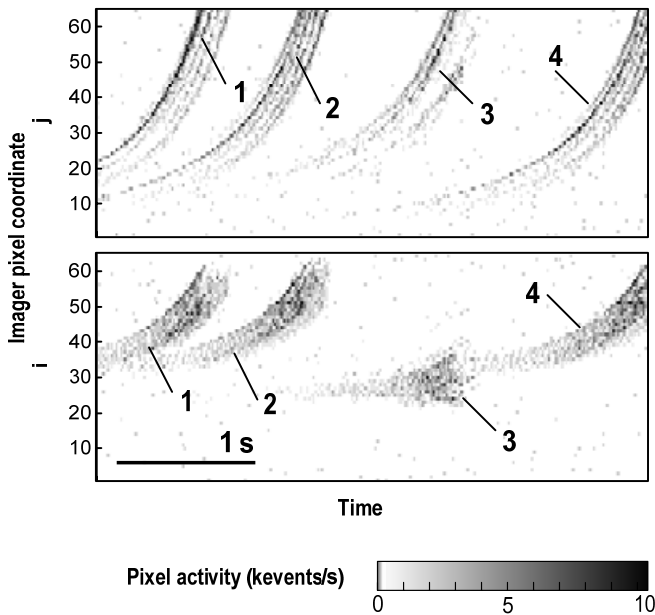


Fig. 3. Address-event data recorded with the temporal derivative SR imager in image space i, j vs. time showing the traces of four vehicles in motion.

IV. ALGORITHM FOR VEHICLE SPEED ESTIMATION

This section describes an efficient and robust method for vehicle speed estimation from AE data streams. The method is based on the estimation of the slope of the AE point cloud representing the leading or trailing edge of an approaching or departing vehicle, respectively.

In a first step vehicles are detected in the unstructured AE data stream by a vehicle detection algorithm detailed in reference [12]. The next processing step is to isolate the trace of the leading or trailing edge of a vehicle. For this purpose the cumulative activity of a pixel row j over time is monitored with a temporal resolution of 10ms. Note that this temporal resolution of the AE processing corresponds to about 100 frames-per-second traditional video processing.

The trigger-time at which the activity exceeds (or falls below) a certain threshold is stored for leading (or trailing) edges for each row j . For a coarse speed estimate, the

observation of two trigger-times might be sufficient. However, due to sensor imperfections and resulting noisy edges, it is advantageous to derive trigger-times for a number of rows N within a predefined region of interest (ROI) corresponding to a certain lane. By this measure a new point cloud in (j, t) is obtained and the amount of data is further reduced to N points (j, t) . This point cloud represents the track of the edge of the vehicle. For the vehicle speed estimation, the trace of the edge is required in world coordinates space and time (x, t) .

The geometric projection from imager coordinate j into world coordinate x is based on known parameters such as mounting height, viewing angle and optical aperture and yields a point cloud in (x, t) . For the transformation $j \rightarrow x$ a look-up-table for the considered imager rows in each ROI is used. Fig 4a shows an example for the resulting point cloud for a vehicle extracted from real-live data recorded at a highway. The negative slope of the point cloud corresponding to a vehicle moving towards the sensor system in the $-x$ direction (corresponding to $v < 0$) is clearly visible. The higher density for points with smaller distance (i.e. small x) is explained by the denser geometric projection of the imager rows onto the ground near the sensor system.

For the (x, t) - point cloud of the leading edge consisting of N points, the velocities from every combination among these points is computed. More precisely, we calculate $N \cdot (N-1)/2$ velocities according to

$$v_{kl} = \frac{x_k - x_l}{t_k - t_l} \quad (1)$$

with $l, k = 1..N$. Note that, as $v_{kl} = v_{lk}$ holds, this value is considered only once. These velocities are entered in a histogram with range 20 to 300 km/h. In Fig. 4b the velocity histogram for the leading edge shown in Fig. 4a is depicted. The empty histogram bins for velocities >220 km/h are not shown for a better visibility of the histogram peak.

The remaining part of the algorithm is to find the position of the maximum in the histogram, which is the vehicles velocity estimate. For the example given in Fig. 4b the result of the velocity estimation is 40 km/h.

The advantages of the histogram method are that it is simple, robust and easy to implement on the embedded platform. The method is able to produce good velocity estimates even in case of noisy or spurious AE data by robustly finding the slope of the most dominant velocity in a point cloud from the highest peak in the velocity histogram. Furthermore the width of the distribution around the maximum value is used as a confidence measure for the quality of the velocity estimate. It allows for rejecting results from broad distributions or distributions without a pronounced maximum originating from measurement artifacts.

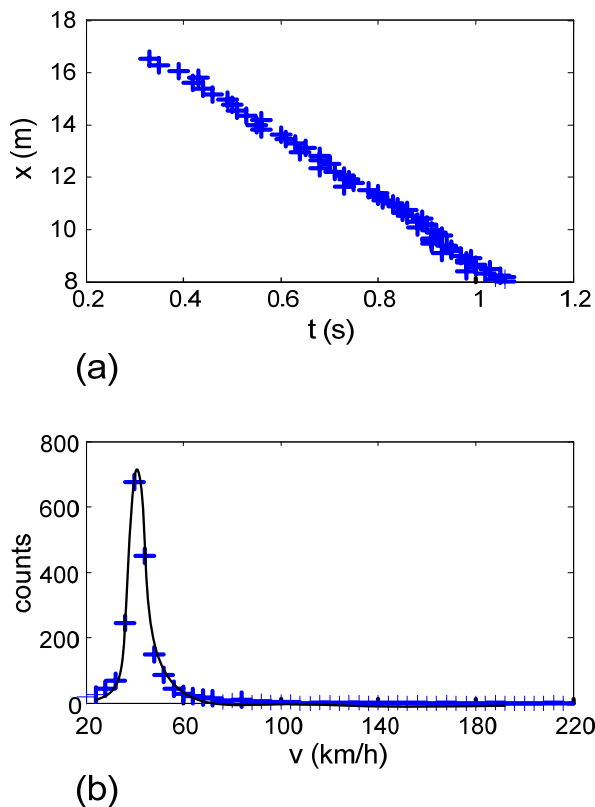


Fig. 4. The leading edge of a vehicle extracted from AE data in x, t space (a) and the corresponding speed histogram (b). The x -coordinate in (a) corresponds to the ground-distance from the sensor system.

V. RESULTS

A. Evaluation of the accuracy

To evaluate the accuracy of the speed estimation algorithm the embedded system using a 128×128 pixel version of the retina sensor [10] was mounted seven meters above a two lane test track and AE data of passing vehicles were recorded under daylight conditions. Two calibrated light-barrier speed measurement units with a resolution of 0.1 km/h and a precision $< 0.1 \text{ km/h}$ were used to measure reference speeds at distances 7 m and 16 m from the sensor base point. These speed data were used for the verification of the speed-estimation algorithm. Cases where the measured reference speeds of the two light-barrier systems differ by more than 1 km/h have been rejected. A total number of 273 test cases have been evaluated for the approaching and departing traffic flow direction.

Assuming that the system has been calibrated to near zero mean error, the quality of the speed estimation for a set of measurements is given by the width of the standard deviation of all errors.

Figure 5a shows the speed estimation result versus reference speed for the approaching ($v < 0$) and departing ($v > 0$) traffic directions. The dashed line in the figure indicates zero error ($y = x$). Figure 5b shows the relation between standard deviation and speed. The standard

deviation of the error is around 3% . The mean errors over all measurements are -0.83 km/h and 0.61 km/h , and the standard deviations are 2.3 km/h and 1.8 km/h for the approaching and departing direction, respectively.

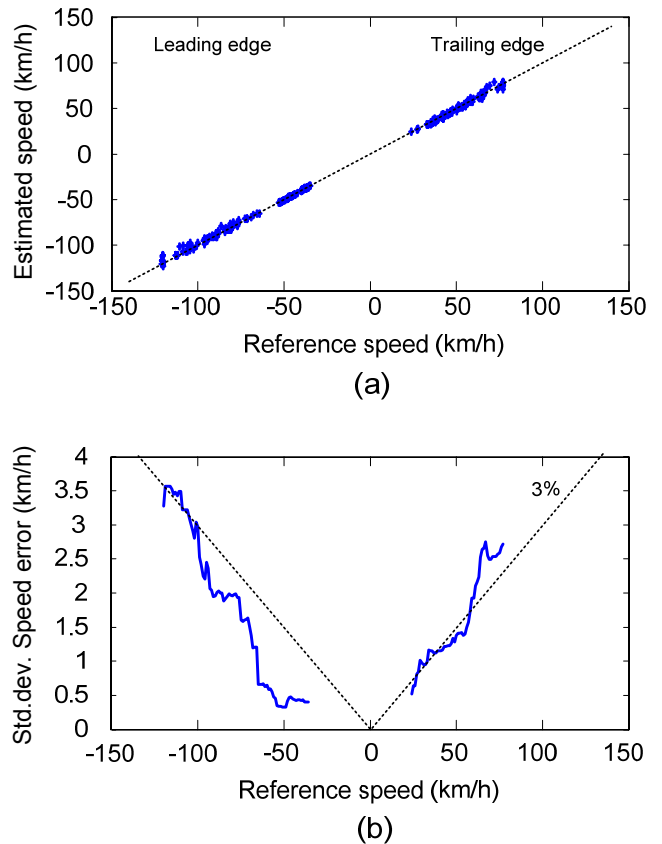


Fig. 5. (a) Speed estimation result versus reference vehicle speed and (b) speed estimation error for departing ($v > 0$) and approaching traffic ($v < 0$). The dashed lines indicate 3% standard deviation.

B. Data from test site

Fig. 6 shows a typical result of the evolution of the average vehicle speeds acquired at the four-lane test site over the course of a full day using a 64×64 pixel sensor. The average speed is calculated from the single vehicles data detected in 15 min intervals and shown for each lane, with lane 1 (4) being the slow (fast) lane.

The evolution of the average speed during night hour's show increased roughness due to the small number of vehicles in the $0:00$ to $5:00$ hours period. On lane 4 very few vehicles have been driving from $0:00$ until $4:00$ hours which explains missing velocity data during this interval. The average velocities on all lanes show two distinct minima caused by a morning ($7:00$) and evening ($16:00$) traffic congestion. During daytime and late evening hours the average velocities for the separate lanes are distributed as expected with the slowest and the fastest averages speeds on the slow (right) and fast (left) lanes, respectively.

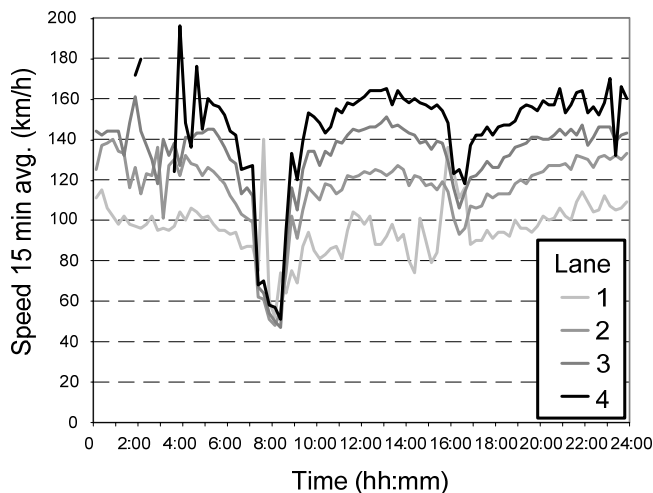


Fig. 6. Typical results of average velocities in 15 min intervals for a four lane highway over the course of a full day. Lane No.1 and 4 denote the slow and fast lanes, respectively.

An important aspect for good velocity estimation is the ability to find the leading (trailing) edge of the vehicle which is typically formed by the shadow of the approaching (departing) vehicle cast on the ground. The good sensitivity of the used SR imager to relative intensity changes enables to robustly detect this shadow, which is always present even under diffuse lighting conditions. Only for this edge the assumption of the geometric projection based on the system mounting height holds, leading to the best velocity estimate. During night operation the imager is sensitive to the vehicles headlights and the parameters for the coordinate projection are adapted accordingly to assure a high quality velocity estimate.

The use of more generalized ROI's of the imager instead of pixel rows and an adapted geometric transformation from imager to world coordinates allows to use the system also in a side mount position.

VI. CONCLUSIONS

Vehicle speed estimation concept and algorithms for an embedded traffic data sensor have been presented. The applied data processing works with a temporal resolution of 10 ms on an AE data stream delivered by a SR temporal derivative image sensor. The data processing algorithm benefits from the capability of the SR imager to detect relative intensity changes and on its efficient asynchronous communication. The presented algorithm estimates vehicle speed in a 20 to 300 km/h range and has been implemented on a compact embedded system. Real live average vehicle velocity data acquired on a test site at a four lane highway has been presented for a 24 hour period.

The efficient data pre-processing of the SR imager highly reduces the computational effort as compared to traditional video-based traffic surveillance systems, enabling a low-cost, embedded DSP implementation. The combined power consumption of the imager and the signal processor is about 2.5 W.

ACKNOWLEDGMENT

The authors would like to thank Tobi Delbruck and Patrick Lichtsteiner from the Institute of Neuroinformatics at ETH Zürich for their work and support on the temporal derivative imager and the Austrian highway authority ASFiNAG for realizing and supporting the test site.

REFERENCES

- [1] J. Langheim, J.F. Henrio and B. Liabeuf, "ACC Radar System – Autocruise – with 77 GHz MMIC Radar," *International Conference of ATA Florence*, 1999
- [2] R. López-Valcarce, C. Mosquera and F. Pérez-González, "Estimation of Road Vehicle Speed Using Two Omnidirectional Microphones: a Maximum Likelihood Approach," *EURASIP Journal of Applied Signal Processing*, vol. 8 pp.1059-1077, August 2004
- [3] A. Rakusz, T. Lovas and A. Barsi, "LIDAR-Based Vehicle Segmentation," in *International Conference on Photogrammetry and Remote Sensing*, Istanbul, Turkey, 2004.
- [4] A. Kirchner, U. Lages and K. Timm, "Speed Estimation with a Laser Rangefinder," *Proceedings of ICARCV'96 4th International Conference on Control, Automation, Robotics and Vision*, pp. 1875 – 1879, Singapore, 1996.
- [5] V. Kastinaki, M. Zervakis and K. Kalaitzakis, "A Survey of Video Processing Techniques for Traffic Applications," *Image and Vision Computing*, vol. 21, pp. 359-381, 2003.
- [6] L. Grammatikopoulos, G. Karras and E. Petsa, "Automatic Estimation of Vehicle Speed from Uncalibrated Video Sequences," *Proc. Of the International Symposium on Modern Technologies, Education and Professional Practice in Geodesy and Related Fields*, pp. 332-338, Bulgaria, November 2005.
- [7] D.J. Dailey, F.W. Cathey and S. Pumrin, "An Algorithm to Estimate Mean Traffic Speed Using Uncalibrated Cameras," *IEEE Trans. On Intelligent Transportation Systems*, Vol.1, No2, pp. 98-107, June 2000.
- [8] D.J. Dailey and T.N. Schoepflin, "A Correlation Technique for Estimating Traffic Speed from Cameras," *Transportation Research Board Annual Meeting, TRB no. 03-3414, Washington*, 2003.
- [9] P. Lichtsteiner, T. Delbruck, and J. Kramer, "Improved ON/OFF Temporally Differentiating Address-Event Imager," in *11th IEEE International Conference on Electronics, Circuits and Systems (ICECS2004)*, Tel Aviv, Israel, pp. 211-214, 2004.
- [10] P. Lichtsteiner, C. Posch and T. Delbruck, "A 128x128 120dB 30mW Asynchronous Vision Sensor that Responds to Relative Intensity Change," in *IEEE International Solid-State Circuits Conference (ISSCC2006)*, San Francisco, USA, February 2006.
- [11] P. Lichtsteiner and T. Delbruck, "64x64 Event-Driven Logarithmic Temporal Derivative Silicon Retina," in *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Nagano, Japan, 2005.
- [12] D. Bauer, P. Bühler, N. Donath, H. Garn, B. Kohn, M. Korber, M. Litzenberger, J. Meser, C. Posch and P. Schön "Embedded Vehicle Counting System with 'Silicon Retina' Optical Sensor" *Workshop on information Optics 2006, Toledo, Spain*, June 5-7, 2006.
- [13] K. Boahen, "Point-to-Point Connectivity Between Neuromorphic Chips Using Address Events," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47 pp. 416-433, 2000.
- [14] http://www.smart-systems.at/products/products_video_tds_en.html