ROBOT SOCCER USING OPTICAL ANALOG VLSI SENSORS

Dr. Vlatko Bečanović

Fraunhofer Institute for Autonomous Intelligent Systems, Schloss Birlinghoven,

D-53754 Sankt Augustin, Germany

becanovic@ais.fraunhofer.de

Tel.: + 4941 14 2854
Fax :+ 4941 14 2342

Dipl. Ing. Ramin Hosseiny

Fraunhofer Institute for Autonomous Intelligent Systems, Schloss Birlinghoven,

D-53754 Sankt Augustin, Germany

hosseiny@ais.fraunhofer.de

Tel.: + 4941 14 2104
Fax.:+ 4941 14 2342

Dr. Giacomo Indiveri

Institute of Neuroinformatics, Winterhurerstrasse 190, CH-8057 Zürich, Switzerland

giacomo@ini.phys.ethz.ch

Tel:. +41 1 635 3024
Fax.: +41 1/ 635 3053

ROBOT SOCCER USING OPTICAL ANALOG VLSI SENSORS

V. Bečanović.[1], R. Hosseiny, [1] and G. Indiveri[2]

[1]Fraunhofer Institute for Autonomous Intelligent Systems, Schloss Birlinghoven,
D-53754 Sankt Augustin, Germany. E-mail:
{becanovic,hosseiny}@ais.fraunhofer.de

[2]Institute of Neuroinformatics, Winterhurerstrasse 190, CH-8057 Zürich, Switzerland. E-mail:
giacomo@ini.phys.ethz.ch

Abstract: In this paper we show how a combination of low dimensional vision sensors can be used to aid the higher level visual processing task of colour blob tracking, carried out by a conventional vision system. The processing elements are neuromorphic analog VLSI (aVLSI) vision sensors capable of computing motion and estimating the spatial position of high-contrast moving targets. In these devices both sensing and processing is done on the chip's focal plane. The neuromorphic sensors can calculate optical flow, position of sharpest edge, and motion of sharpest edge, in real-time. The processing capability of the system is investigated in a mobile robotics application. Firstly, for visualization and evaluation purposes, a correlation analysis is performed between the data collected from the neuromorphic vision sensors and the standard vision system of the autonomous robot, then, we process the multiple neuromorphic sensory signals with a standard auto-regression method in order to achieve a higher level vision processing task at a much higher update rate. At the end we argue why this result is of great relevance for the application domain of reactive and lightweight mobile robotics, at the hands of a soccer robot, where the fastest sensory-motor feedback loop is imperative for a successful participation in a RoboCup soccer competition.

Key Words: Neuromorphic vision sensors, analog VLSI, reactive robot control, sensor fusion, RoboCup.

1. Introduction

In recent years the robotics community has formulated a challenge to fellow researchers in academia and industry, to design and build a fully autonomous humanoid robot that can out-compete the fittest human athletes in the year of 2050. The sport is soccer and the aim is that a team of fully autonomous soccer playing robots shall win against the human world champion. This is the RoboCup challenge formulated by the Japanese researcher Hiroaki Kitano and the official RoboCup federation [1]. International competitions are held annually, where scientists exchange knowledge and ideas at the RoboCup symposium that is open to all participants. The focus is interdisciplinary and research areas range from electronics, mechatronics, robotics and the physics of fuel cell batteries, to computer science and artificial intelligence.

In our lab aVLSI technology is exploited in fast moving mobile robotics, especially RoboCup, where soccer-playing robots perform at high speed. Our robots are competing in the mid-sized league that allows wheeled robots of roughly 50 by 50 cm with the maximum weight of 80 kg. They are equipped with infra-red distance sensors in order to have fast and reliable obstacle avoidance, odometry together with an augmenting gyroscope in order to reduce the error in the odometry measurements, and contact sensitive bumper sensors. The robots use a differential drive for movement, a pneumatic kicker for shooting and two small movable helper arms to prevent the ball from rolling away. The most important sensory inputs are streamed in via FireWire bus [2] from a digital colour camera. The conventional part of vision processing is software based and consumes most of the calculation resources on-board the robot [3].

One of the most difficult tasks in the RoboCup environment is to pass the ball from one player to another. This requires first of all that the robot can control the ball, that is, be in possession of the ball so that it can be kicked in any direction and this while the robot is in motion. The ball needs to be close to the robot in order to be successfully mastered. This can be achieved by carefully controlling the velocity and position of the robot relative to the ball. The closer the ball the lower the relative velocity must be in order for it not to bounce off due to its lower momentum. In order to solve this very demanding problem the

robot has to know where the ball is located at each instant, which requires a fast read-out and processing of the sensory information.

The update rate of the optical sensory system can be increased and the calculation load by the digital vision system reduced by the use of dedicated aVLSI vision sensors. Other factors such as the low weight, low power consumption and small size of the aVLSI sensors also contribute to a more streamlined design and increase robot performance. In this paper we combine four types of neuromorphic vision sensors with the aim to replace, or aid a higher level standard vision system that operates at a lower update rate. Due to their smaller field of view, the sensors can only replace the standard vision system in a confined area in the immediate vicinity of the robot, and the output signals from the neuromorphic vision sensors gives no information of what hypothetical objects are analyzed in the scene. Any moving object can trigger an output from multiple aVLSI sensors if their respective fields of view coincide with the moving object. In principle reading out analog signals from our dedicated sensors is equivalent to the calculation of certain motion properties, such as position and/or velocity for a particular object. This information is usually computationally expensive to calculate in traditional CPU based digital architectures. Selected fields of view can be mapped onto the active focal regions of the aVLSI sensors by proper alignment and the use of appropriate lenses. In this way, multiple data streams are collected from a larger scene and a robust calculation can be performed with sensor fusion.

This paper is structured as follows: in sec. 2 a description of our robot platform is given. The vision sensors used in the experiments are presented in sec. 3. In sec. 4 we investigated how the vision system can be complemented with a set of neuromorphic vision sensors. Here, we present data collected during three experimental runs with one of our robots. We show that this data is suitable for further higher level processing. We close with an indication of a direction for future research, namely the design and evaluation of selective attention systems, and, the use of neural networks for improved dynamical prediction of higher level visual data.

2. Our robot platform

Our soccer playing robot has actuators in the form of motors to drive the robot and to turn a panning camera. A valve is used to kick the ball pneumatically and small robot arms attached to the left and right side of the robot keeps the ball in front of the kicker plate. Besides the optical sensors; camera and neuromorphic vision sensors, it has four infrared distance sensors, a contact sensitive bumper strip with rubber shield, and odometry at the two actuated wheels of the robot. This is augmented by a gyroscope for fast turning movements. All of these peripheral devices are controlled by three 16 bit micro controllers [4]. They are interconnected with a bus interface (CAN), which is standard that is used in German automobile industry. Motor current control is performed with one of the aforementioned micro controller modules [5]. A second micro-controller reads out the encoder of the panning camera unit and controls its motor. The third and last controller takes care of all analog signals on the soccer robot, like distance readings of infrared sensors and also the A/D conversion of the neuromorphic sensor signals. Micro controller modules communicate with the small notebook PC via the CAN-bus interface. The notebook PC operates the main behaviour program and the operating system can be either Windows or LINUX. The cyclic update rate is 30 Hz (~33 ms) which is governed by the frame rate of the digital camera.

For this experiment we increased the observation rate for the neuromorphic sensors to the maximum effective sampling rate of the micro-controller module that is used which is ~2 kHz (0.5 ms). In the various experiments the signal is down-sampled to 153 Hz in the first experiments and up to 520 Hz in the more complex experiment done at the end.

The robot vision system does colour blob tracking of multiple objects and delivers information from tracked objects such as position of geometrical centre, bounding box and pixel area. In our experiments only the position of the geometrical centre of the tracked object will be used to train the system. Other parameters like pixel area are only used indirectly, in order to prepare data for the training phase of the system by removing noisy information from distant objects and other artefacts. The vision software used

for the experiments is a free software developed at the Carnegie Mellon University and used be many robot teams in RoboCup tournaments [3].

3. Neuromorphic vision sensors

Neuromorphic technology mimics processing principles in biology, which for our sensors means that processing power is highly distributed and that circuits are analog. Images are processed directly at the focal plane level where photoreceptors, memory elements and computational nodes share the same physical space on the silicon surface. Typically each pixel in a neuromorphic sensor contains local circuitry that performs, in real time, different types of spatio-temporal computations on the continuous analog brightness signal. These sensors perform data reduction, as they transmit only the result of the vision processing off-chip, without having to transmit the raw visual data to further processing stages. In CCD cameras, or conventional CMOS imagers, the brightness at the pixel level is measured and merely the gain level is calculated according to the average brightness level of the whole scene.

The specific computational function of a neuromorphic sensor is determined by the structure of its architecture and by the way its pixels are interconnected. Since each pixel processes information based on local sensed signals and on data arriving from its neighbours, the type of computation being performed is fully parallel and distributed. An apparent drawback of this design methodology is given by the fact that the resolutions that can be achieved are typically lower than the ones achieved by CCD cameras (*i.e.* these chips have a low *fill-factor*). The quality and resolution of the pixel output of some of these vision chips seem poor at first. However, we should keep in mind that these sensors have been designed to perform data compression. For instance, in a one-dimensional target-tracking task (*e.g.* along the horizontal plane), the output of a neuromorphic sensor would be only one single analog continuous value, encoding the coordinate of the target's position. In tasks such as vehicle guidance or autonomous navigation, low resolution is not a limiting factor. Insects have fewer pixels than even the cheapest hand-held CCD camera, yet they can avoid obstacles much more efficiently than any existing machine-vision system.

The aVLSI sensors used in our experiments are made using standard 1.6 and 0.8 micron CMOS technologies. Specifically we use three 1D vision chips and a single 2D chip. The one-dimensional sensors are: a tracking chip [6], a correlation-based velocity sensor [7], and, a single chip comprising both tracking and correlation-based velocity measurements (cf. Fig. 1). The 2D sensor is a gradient based optical flow chip [8]. The optical flow chip is the most complex and computes the mean optical flow on its focal plane providing two analog output voltages. The correlation-based velocity sensor delivers the mean right or left velocity computed throughout it's whole 1D array, and the 1D tracker sensor provides an analog output voltage that indicates the position of the highest contrast moving target present in its field of view.

4. Experiment

The purpose of the experiment is to investigate the plausibility of neuromorphic vision sensors to aid higher level vision processing tasks, in particular colour blob tracking, which is a standard real-time vision processing application that is commonly used on mobile robots. The test consists of two stages; firstly to investigate if the sensors can be made sensitive to a moving primary coloured object, and secondly, to validate this against a somewhat cluttered background. The first stage is performed to investigate the precision of the prediction made from the fused sensory readings. The second stage is performed to investigate if there is enough discrimination against background patterns, that is, to investigate the robustness of the object tracking task when the robot is moving. If both stages are successful, this would imply that a set of neuromorphic vision sensors, sensitive to different types of motion, could complement a standard camera based digital vision system in a local domain of the scene.

4.1 The experimental set-up

The experiment is performed at the RoboCup playfield at our institute [9]. The field measures 6 x 10 meters and there are the natural static obstacles present, e.g. two goals, corner markers, and a protective fence. A RoboCup soccer-playing robot is set up with four neuromorphic vision sensors, (cf. Fig. 2), with their respective fields of view partially overlapping (cf. Fig. 3). All sensors are calibrated for an indoor light

luminosity regime. The calibration is performed by adjusting on-chip bias currents to maximize the contrast sensitivity. The red ball gives a relatively low contrast to the green carpet used in RoboCup tournaments; so photographic colour filters are used to enhance the contrast. The filters used are gelatine-based and discriminate against complementary red colours, thus the contrast to the official red ball is enhanced. The same filter is used for all four sensors in the experiment.

The normally panning camera head is fixated in order for the vision system to be a reliable reference signal. The sensors are mounted on the robot so that their fields of view do not overlap significantly. It is also made certain that their fields of view are within the field of view of the camera based (standard) vision system (cf. Figs. 2 and 3). The RoboCup soccer playing robot is fully autonomous and is operated by a behaviour based program that was used by our team at the last world championships in Padua Italy [10],[11]. The test field is prepared with white lines that are located in a dense non-uniform grid and with an average spacing of about one meter. On the field there is a red soccer football. The robot will constantly try to approach the red ball in different manoeuvres, and this will ensure a high event rate for a correlation analysis in the second stage of experiments. For the first stage the robot is stationary and the ball is moved according to certain patterns that ensure an even distribution of events when projected onto the focal plane of the digital vision system.

4.2   Stationary experiments

The experiment consists of data collection from the neuromorphic vision sensors and the digital vision system of a stationary robot. Each sensor will provide two output signals (except the 1D tracker chip, where only a single output is used). All in all this gives 7 neuromorphic sensor signals. For visualization purposes each neuromorphic sensor signal is discretized by shifting a moving window average in time and applying a simple threshold in order to get trig-points. This makes it possible to correlate the activity of neuromorphic vision sensors with the standard vision reference signal which will be used for the purpose of visualization. The active regions can be spotted from the respective standard vision system reference correlated with the trig-points obtained from the sensor signals. In our case the vision reference is the

standard digital vision system, that delivers colour blob data consisting of centroid positions of the biggest red object in the field of view of the digital camera. All signals are time-coded on the CAN-bus interface, also the visual cues from the standard vision system, which gives very accurate timing in the measurements (cf. Fig. 4).

4.3   Moving robot experiment

The experiment consists of data collection from the neuromorphic vision sensors and the digital vision system of a fully performing robot. The sensors are mounted as in the stationary robot experiment and data is here continuously collected for 7 minutes and 10 seconds at a sampling rate of 2 kHz (down-sampled to 520, 260 and 130 Hz). During this time the robot tries to continuously approach the ball and will during its manoeuvres frequently pass lines on the floor, which will influence the tracking task of the red ball. Optimally the system should recognize what sensory input belongs to white lines and what input belongs to the red ball. The experiment is validated by tracking red and white objects with the standard vision system of our robot. The red object corresponds to the red ball and white objects correspond to the white lines present in the playfield. The reference information of the red object is used for the model fitting and the white in order to investigate how robust the system is against disturbances from non-moving, ground fix objects. A model is calculated that has two extra inputs from the robot velocity measurements, that is a total of 9 inputs is used for this experiments. The idea is that the two extra inputs from the robot velocity measurements will help the system to recognize the moving background better. The number of outputs and the vision reference is in accordance with the previous two experiments (cf. Fig. 4)

4.4   Experimental results

Three experiments were performed in order to investigate, firstly, the precision in the prediction from the fused sensory information, and secondly, the robustness of the object tracking task itself. The first step here consists of two stationary robot experiments treated in sections 4.4.1 and 4.4.2, and the second step is a moving robot experiment that is treated in sec. 4.4.3. The data is evaluated by comparing the results from a

standard dynamical prediction model that is fitted on training data and evaluated on test data. A root mean square error is calculated relative to the reference signal from the standard vision system.

The prediction model used for the two stationary robot experiments is a multivariable *ARX* model of 4'th order. The model, which is part of the Matlab™ system identification toolbox is performing parametric auto-regression that is based on a polynomial least squares fit [12]. For the dynamic experiments the best overall model was chosen in the range of up to a 15 ARX coefficients (15'th order ARX model). A lower update frequency requires polynomials that are less smooth and thus will be of higher order. For a higher update rate the change in the signal for each sample is lower, thus a lower order model is then usually enough. Other effects, such as the signal to noise ratio of the sensors and the fact that the two prediction channels are coupled adds some complexity to problem and the above reasoning is only to be seen as a rule of thumb and does not strictly apply for all cases. This can also be seen in Table 2, where an update rate of 260 Hz will have a most optimal ARX model of lower order (3'rd order ARX) than with an update rate of 520 Hz (6'th order ARX).

4.4.1 Stationary robot experiment I

In the first experiment the robot is not moving and the camera and neuromorphic vision sensors detect a single moving red RoboCup soccer football. The ball was moved so that it passed the robot along horizontal paths. The fields of view of the neuromorphic vision sensors were divided into four zones that were partially overlapping, and, within the field of view of the standard vision system. During the experiment the ball was thrown 25 times back and forth in each zone, but in random order, so that the data set would be easily split into a training and testing set of equal size. By this procedure the distribution would be close to uniform in the spatial domain and normal in the temporal domain. The distribution in the spatial domain can be seen in the resulting scatter-plots, where the vision reference signal in this experiment is correlated against trig-points from active sensors (cf. Fig. 5). The prediction efficiency is given in Table 1, where the horizontal (x-channel) over-all RMS error is about 13 %, which for the

horizontal camera resolution of 320 pixels would mean an error of about 40 pixels. This corresponds well to the fact that the resolution of the neuromorphic sensors is between 10 and 24 pixels.

4.4.2 Stationary robot experiment II

In the second experiment, that is performed with a non moving robot and the same boundary conditions as the first experiment, the ball was moved so that it passed straight towards the robot hitting it and bouncing off in an elastic collision, where the ball with its significantly lower momentum got deflected. During the experiment the ball was thrown 25 times back and forth in different zones, but in random order and at the same point of impact, so that the data set would be easily split into a training and testing set of equal size. The point of impact was in the middle of the kicker-plate located in front of the robot. The results here indicate similar efficiency as for the first stationary robot experiment for estimating the horizontal trajectories of the red ball, but with a better efficiency in the estimation of the vertical component (cf. Table 1). This is most probably due to the trajectories of the ball that were more frequently in the active region of the 1D tracker (TR) sensor that was vertically aligned so that it gave an optimal measurement of the vertical component (cf. Figs. 2, 3). An example from the stationary robot data set used in this experiment is given in Figs. 6 and 7, where the predicted result for the horizontal and vertical blob position is plotted with a solid line and the "ground truth" reference signal is plotted with a dotted line. The activity of all the sensors is indicated as trig-points on top of the reference signal. In Fig. 6 it can be seen that for a rich activation of the various sensors the correspondence to the vision reference is good, but, when there is not enough of sensory information there will be some discrepancies. This is clearly seen in Fig. 7, where the plot describes the vertical component of the ball approaching the robot and being bounced off 3 times in a row. In particular, between 302 to 303 seconds in Fig. 7, one can see that the lack of sensory information will reduce the downward prediction of the change in blob position, but this is not as crucial when the system lacks information in the upward prediction (between 308 and 309 seconds in Fig. 7). In this case the system learns by hart that the ball will bounce off the robot, which is actually desirable in our case.

4.4.3   Moving robot experiment

The experiment is validated by tracking red and white objects with the standard vision system of our robot. The red object corresponds to the red ball and white objects correspond to lines present in the playfield. The number of outputs is in accordance with the previous two experiments. The reference information of the red object is used for the model fitting and the reference of the white objects (corresponding to white lines) is only used to indicate trig-points. This information is in turn used for visual inspection and the calculation of the efficiency of discrimination against white lines. The system was trained with 75% of the full data set and tested with the remaining 25%. The results are presented in Table 2, where the over-all RMS error is calculated for the test data for sampling frequencies of 130, 260 and 520 Hz. There are also RMS errors calculated in trig-points for the case when only the ball was visible (red object only) and when the red ball was visible with occluded background (red object and white line). It can be seen from Table 2 that the efficiency seems to be slightly improved at higher update rates and that the ball can be recognized in occluded scenes with an efficiency better than the over-all efficiency for the horizontal x-channel, and, for an efficiency close to the over-all efficiency for the vertical y-channel. This might be the result of the better coverage of the sensors for the horizontal motion, where the 1D velocity tracker (V), the integrating tracker (I-TR) and the horizontal optical flow (OF-X) contribute to the measurement for horizontal motion, as contrary to the vertical channel, where only the tracker sensor (TR) and the optical flow vertical channel (OF-Y) contribute to the measurement. Further, the sensors that contribute to the vertical channel are all directed toward the ground and with their fields of view overlapping (cf. Fig. 2), which makes their combined active region smaller than for sensors with their optical axis parallel to the ground and their fields of view in different directions (as is the case for the sensors that contribute to the horizontal channel).

4.4.4   Discussion

To track moving objects in the spatial and chromatic domains, as well as discriminate against a somewhat cluttered background during robot motion, is of fundamental importance in mobile robotics. The spatial discrimination is here achieved by the placement of the sensors, so that they map the scene in a way that

make optimal use of their functionality. For example, two of the sensors used on our robot are placed as found on flying insects in nature, with their optical axis almost parallel to the ground and looking towards both the right and the left side of the robot with no overlap in their respective fields of view. Another strategic placement is the tracker (TR) sensor that is looking towards the ground and thus will give an output signal that is proportional to the distance to the highest contrast edge that is in its field of view (on the focal plane). The shadow from objects on the ground might here of course produce a false estimate in this "by eye" interpretation of the sensory signal, but the underlying principle is clear. The chromatic discrimination is achieved by reducing the intensities of non-wanted colours by applying a colour filter, and this raises the question of how to discriminate between different moving objects with similar colour intensity profiles. This question remains open. An idea could be to analyze the temporal dynamics, e.g. the characteristics of the motion pattern, in order to extract additional information about each hypothetical object, and in this way improve the efficiency of the system.

5. Future directions

Fast intelligent sensor devises like the neuromorphic vision sensors are advantageous for reactive behaviour based robotics [20], where sensors are influencing actuators in a direct way. The high update rate and the higher level of abstraction given from neuromorphic sensory data, makes it possible to react almost instantly to a given sensory reading, thus faster decisions can be taken. This fact creates possibilities to distribute a decision closer to the sensor which implies that it could be suitable to de-centralize decision making mechanisms in the over-all robot control architecture. This is the way biology does it and there is interesting work done with the implementation of central pattern generators (CPGs) [13], where lower level locomotion of biped robots are modulated with the local and self sustained CPG modules. In biology, the level of de-centralized decision making is high, and a question that here arises is: how much information is it possible to process directly form early visual sensory cues? Some related work done here shows that

recurrent neural networks can be trained to achieve higher cognitive tasks such as self localization [14]. This type of approaches give very reactive systems.

As a first step in the visual process it is common to isolate regions of interest, or objects, that then are tracked, within the field of view of the visual system in question. For a multi sensor system this could impose some difficulties, but there are different approaches to solve this. One approach is to integrate each sensor individually, with a corresponding control loop, that directly actuates the driving motors of the robot. Several such sensor-actuator modules can then be integrated in a hybrid control architecture. In an experiment performed with our goalie robot, it is shown that it is possible to integrate a single neuromorphic vision sensor in such a way. The physical limit of the robot is also reached regarding its reaction time. From an initial visual stimuli to a measured current in the actuating robot motor, it takes roughly about 40 ms. It will take the robot about 300 ms until it moves a distance of less the 30 cm [15], [16]. The time needed to perform the action is in this case an order of a magnitude greater than the delay in the sensory to motor signal path. This is a clear demonstration of the importance to move calculation power closer to the sensory elements (cf. Fig. 8).

An alternative approach would be to use selective attention mechanisms, where the total activity in the scene is represented as a *saliency* map. In a saliency map the characteristic features of the scene, such as contrast, colour opponency, and/or optical flow are extracted and transformed into a heuristic topographical representation. The saliency map could be used to more easily select the regions in the visual scene that should be analyzed, thus reducing the amount of data that needs to be transmitted to the vision processing modules and giving high processing priority only to regions of interest. Neuromorphic systems that combine vision sensors and processing chips to implement selective attention mechanisms have already been proposed [17].

In this paper we circumvent the design of a signal processing hierarchy, and instead use a conceptually more simple method. A sensor-fusion approach where all the sensory data are fed into a dynamic prediction model together with signals form the vision system as training reference. This gives a local prediction

system, whose trained output corresponds to higher level visual information, in our case the tracked blob data in that particular sub-domain of the visual scene. Experiments in this direction were made in an office environment [18] where local sensory data was used to successfully train a dynamical prediction network [19]. Our neuromorphic devices are similar when considering the locality of their range and the low dimensionality of their output data, thus implying that this method might be well suited for performing the sensor fusion task as well.

6. Conclusions

In our work we investigate if the output signals from a small number of neuromorphic vision sensors can perform the elementary vision processing task of colour blob tracking. We also try to connect this approach to other ideas of how to integrate neuromorphic vision sensors in architectures suitable for small mobile robot platforms. For our experiments we use a soccer playing robot as a test-platform, but are looking for a general application domain that can be used for all types of mobile robots, especially smaller robots with limited on-board resources. Those robots can benefit from a fully neuromorphic vision system, which provides high speed performance together with low power consumption and small size. The robot experiments that were performed indicate that optical analog VLSI sensors with low-dimensional outputs give a robust enough signal, and, that the visual processing tasks of object tracking and motion prediction could be solved with only a few neuromorphic vision sensors in a local region of the visual scene. In general it can be concluded that neuromorphic vision sensors could be attractive for solving control problems in mobile robotics, especially for reactive behaviour based robotics, where the information path from sensor to actuator is short.

References

[1] http://www.robocup.org/

[2] http://www.1394ta.org/

[3] J. Bruce, T. Balch, M. Veloso, Fast and inexpensive color image segmentation for interactive robots, *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2000, Vol. 3, 2061–2066.

[4] http://www.infineon.com/

[5] S. Kubina, Konzeption, Entwicklung und Realisierung – Micro-Controller basierter Schnittstellen für mobile Roboter" *Diploma thesis at GMD Schloss Birlinghoven*, 2001. (in German).

[6] G. Indiveri, Neuromorphic Analog VLSI Sensor for Visual Tracking: Circuits and Application Examples", IEEE Transactions on Circuits and Systems II, Analog and Digital Signal Processing, 46:(11) 1337-1347, 1999

[7] J. Kramer, R. Sarpeshkar, C. Koch, Pulse-based analog VLSI velocity sensors", IEEE Transactions on Circuits and Systems II, Analog and Digital Signal Processing, 44:(2) 86-101, 1997

[8] A. Stocker, R. J. Douglas, Computation of Smooth Optical Flow in a Feedback Connected Analog Network", Advances in Neural Information Processing Systems, 11:, Kearns, M.S. and Solla, S.A. and Cohn, D.A (Eds.), MIT Press, 1999

[9] R. Hosseiny, Fusion of Neuromorphic Vision Sensors for a mobile robot*, Master thesis RWTH Aachen*, 2003. (in German).

[10]   A. Bredenfeld, G. Indiveri, Robot Behavior Engineering using DD-Designer, *Proc. IEEE/RAS International Conference on Robotics and Automation (ICRA),* Seoul, Korea, 21-26 May, 2001.

[11]   A. Bredenfeld, H.-U. Kobialka, Team Cooperation Using Dual Dynamics, *Balancing reactivity and social deliberation in multi-agent systems* (Hannebauer, Markus[Hrsg.]: Lecture notes in computer science, 2001), 111 – 124.

[12]   L. Ljung, System Identification: Theory for the User, Prentice Hall, 1999.

[13]   M. A. Lewis, Perception Driven Robot Locomotion, *Journal Robot Society of Japan*, *20*(3), 2002, 51-56.

[14]   R. Mudra, R. Hahnloser, R. J. Douglas, Neuromorphic Active Vision Used in Simple Navigation Behavior for a Robot, *Proc. 7'th Int. Conf. On Microelectronics for Neural Networks*, Grenada, Spain April 7-9, 1999, 32-36.

[15]   V. Bečanović, A. Bredenfeld, P. G. Plöger, Reactive Robot Control using Silicon Retina Sensors, *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, Washington DC, USA, May, 2002, 1223-1228.

[16]   V. Bečanović, G. Indiveri, H.-U. Kobialka, P. G Plöger,. A. Stocker, Silicon Retina Sensing guided by Omni-directional Vision, *Proc. ninth IEEE conf. on Mechatronics and Machine Vision in Practice (M2VIP)*, Chiang Mai, Thailand, ISBN 962-442-228-1, September 10-12, 2002, 112-118.

[17]   G. Indiveri, Neuromorphic VLSI device for implementing 2-D selective attention systems, *IEEE Transactions on Neural Networks*, *12*(6), 2001, 1455-1463.

[18]   J. Hertzberg, H. Jäger, F. Schönherr, Learning to Ground Fact Symbols in Behavior-Based Robots, *Proc. 15th European Conference on Artificial Intelligence.* (F. van Harmelen ed.), IOS Press, Amsterdam, ISBN 1-58603-257-7, Lyon, France, 2002, 708-712.

[19]   H. Jäger, Adaptive Nonlinear System Identification with Echo State Networks, *Proc. NIPS*, Vancouver, B.C., Canada, 14-16 Dec, 2002.

[20]   R. Brooks, A robust layered control system for a mobile robot, *IEEE Journal of Robotics and Automation*, Vol, RA-2, No. 1, 1986, 14-23.

Tables:

Table 1. First and second stationary robot experiment – test data: The overall RMS error for the x-value and y-value of the centroid of the pixel blob delivered by the standard vision system (SVS). RMS errors of sensors are calculated only in their trig-points, thus the lower and irregular sample size. The RMS error is calculated as the difference between the object position given by the vision reference and the one predicted with the 4'th order ARX model.

Stationary robot Data Set I: (153 Hz, 4'th order ARX)	X Channel RMS Error	Y Channel RMS Error	Sample size
Over all SVS test data:	**0.1295**	**0.1920**	38044
SR Opt. Flow:	0.1101	0.2069	569
SR Tracker:	0.06250	0.1449	4647
SR Velocity:	0.2405	0.2505	126
SR Int. Tracker:	0.1089	0.2304	112
Stationary robot Data Set II: (153 Hz, 4'th order ARX)			
Over all SVS test data:	**0.1386**	**0.1245**	37974
SR Opt. Flow:	0.1586	0.1236	236
SR Tracker:	0.1416	0.1172	1004
SR Velocity:	0.1803	0.1210	387
SR Int. Tracker:	0.1316	0.1396	161

Table 2. Moving robot experiment – test data: The overall RMS error for the x-value and y-value of the centroid of the pixel blob delivered by the standard vision system (SVS). RMS errors of the standard vision system are calculated for: (i) all test data, (ii) when a red object is present within the range of the sensors and (iii) when a red object and white line/s are present. The RMS error is calculated as the difference between the object position given by the vision reference and the one predicted with the corresponding ARX model.

| Moving robot Data Set: (130 Hz, 12'th order ARX)	X Channel RMS Error	Y Channel RMS Error	Sample size
Over all SVS test data:	**0.2574**	**0.2808**	13967
SVS Red object only:	0.2293	0.2331	758
SVS Red obj. & White line:	0.2195	0.2714	320
(260 Hz, 3'rd order ARX)			
Over all SVS test data:	**0.2471**	**0.2679**	27936
SVS Red object only:	0.2241	0.2328	829
SVS Red obj. & White line:	0.2113	0.2983	363
(520 Hz, 6'th order ARX)			
Over all SVS test data:	**0.2485**	**0.2568**	55872
SVS Red object only:	0.2247	0.2163	829
SVS Red obj. & White line:	0.2116	0.2571	361

Figures:
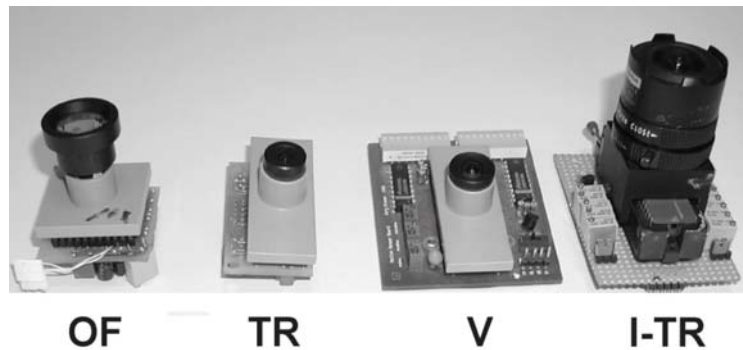
Figure 1: The four aVLSI sensors used in the experiments. From the left: 2D Optical Flow sensor (OF), 1D sharpest edge tracker chip (TR), 1D sharpest edge velocity chip (V), and the 1D integrating edge tracker chip (I-TR).
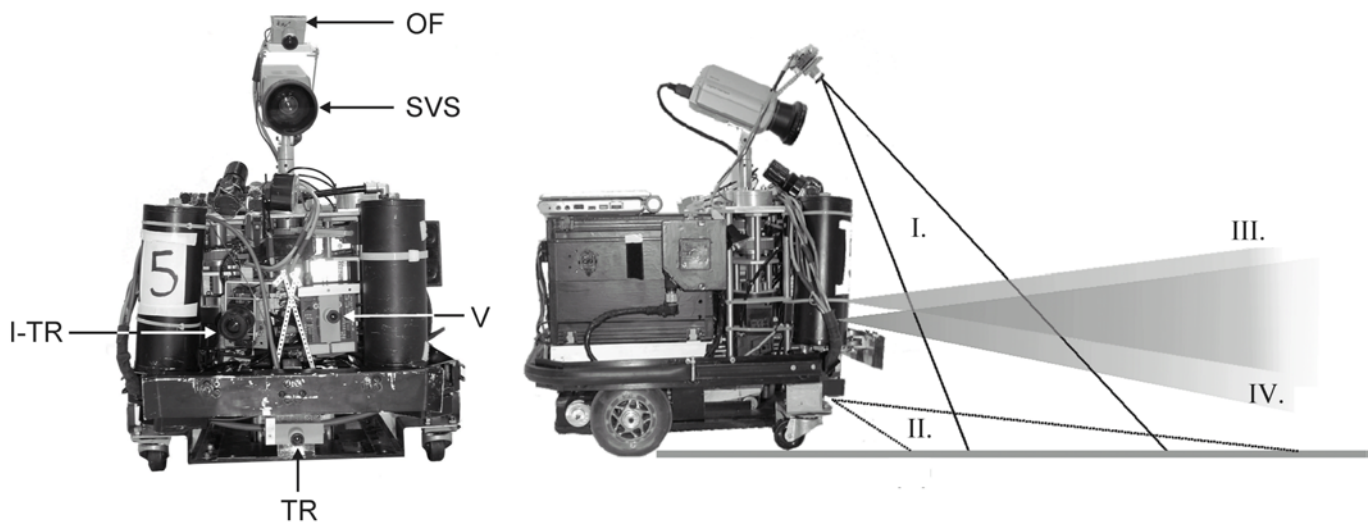

Figure 2: The four aVLSI sensors mounted on our robot (left) with their respective fields of view (right). The 2D optical flow sensor (OF) is pointing straight towards the ground (cf. I). The absolute edge tracker (TR) is also pointing towards the ground (cf. II). It is mounted at a somewhat lower angle and with its pixel array vertically aligned. The 1D velocity tracker (V) and the 1D integrating tracker (I-TR) are directed as a divergent stereo pair and with their resp. pixel arrays horizontally aligned (cf. III and IV).
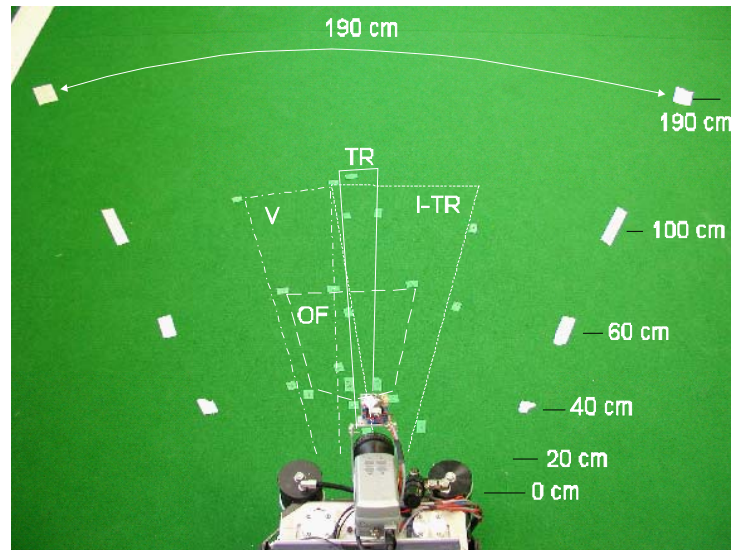
Figure 3: Photo of the experimental set-up for the stationary robot experiments as seen in a birds eye perspective. The imagined active regions of the various neuromorphic vision sensors are indicated with thin white lines. The regions correspond to the: 2D optical flow sensor (OF), the 1D absolute edge tracker (TR), the 1D velocity tracker (V) and the 1D integrating edge tracker (I-TR). The field of view of the standard vision system is marked with thick white markers on the ground.
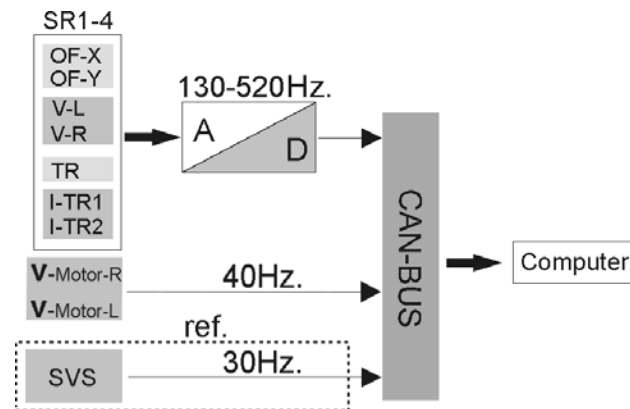

Figure 4: The data acquisition architecture used in the robot experiments. The neuromorphic vision sensors (SR1-4) are sampled at 130, 153, 260 and 520 Hz. In the moving robot experiment extra velocity information is used from both actuated robot wheels at 40 Hz. The standard vision system (SVS) is used as reference and is sending blob data at 30 Hz. All signals are time coded on the CAN-bus and stored on the robot computer.
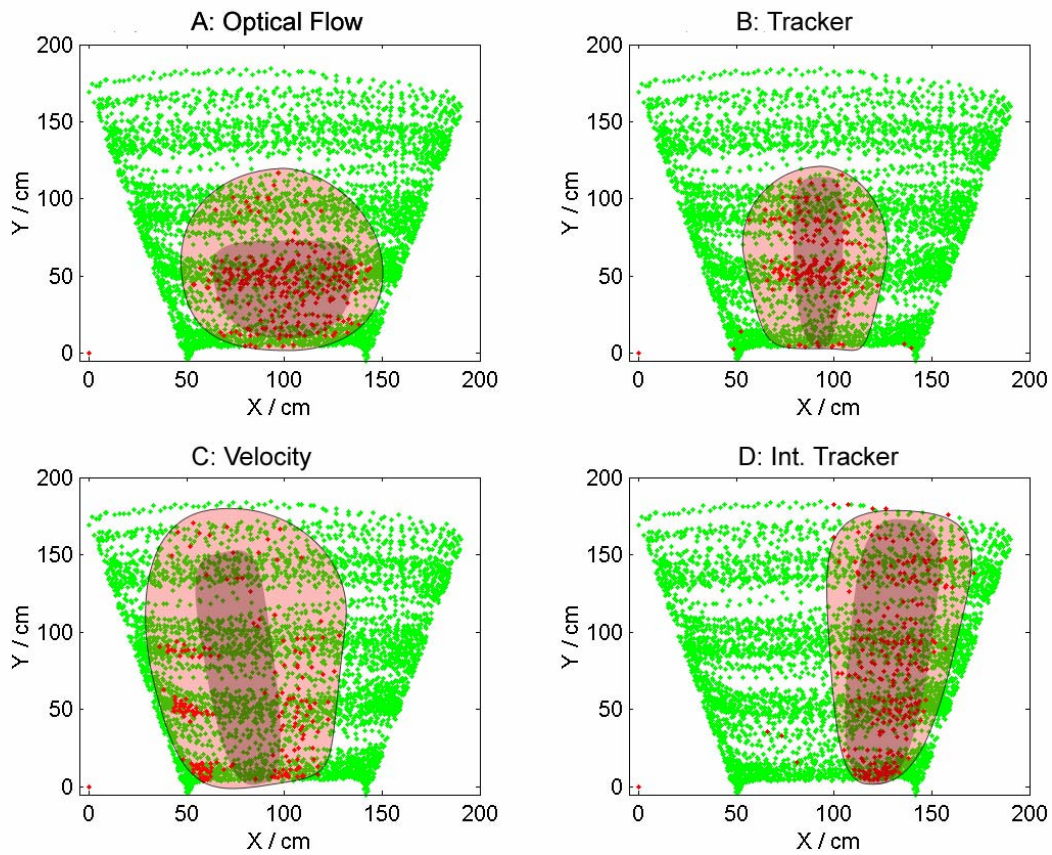
Figure 5: Experimental data from the first stationary robot experiment where the red vision reference, from the standard vision system, is correlated against the activities of the sensors. The dark grey regions indicate the fields of view of the various sensors used in the experiments and the light grey regions indicate the distribution of the blob centroid positions for the triggered vision blob data (cf. Fig. 3).
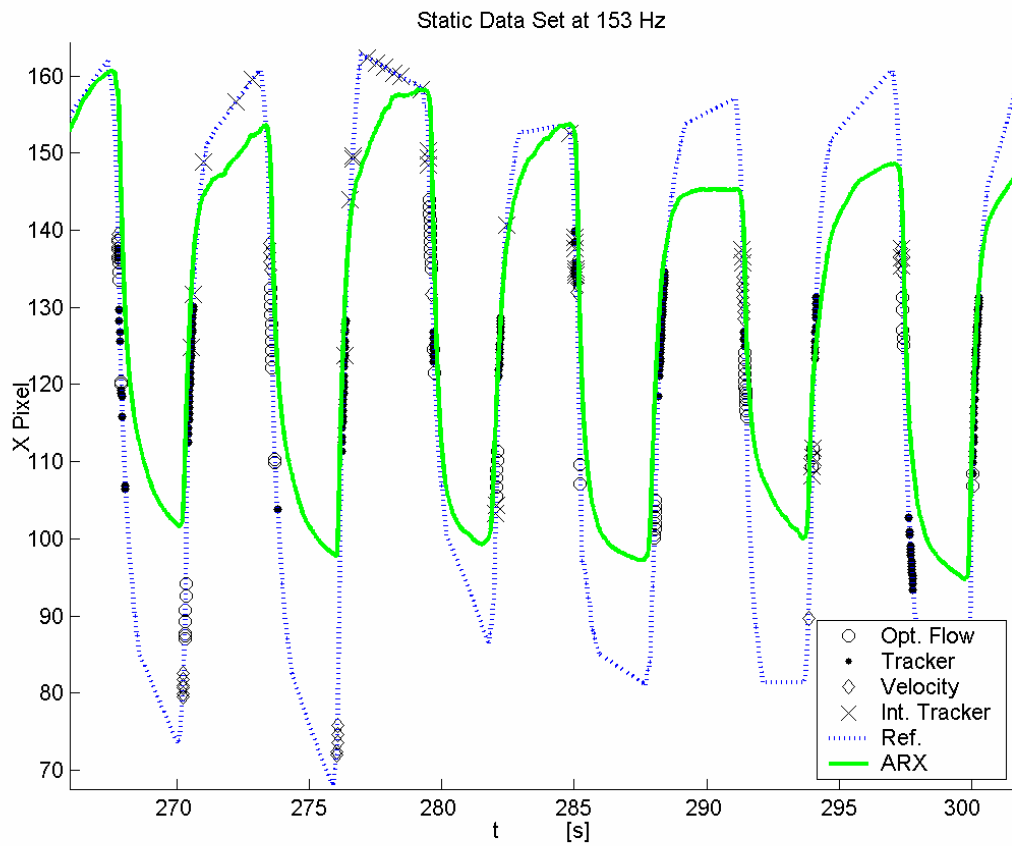
Figure 6:   An example from the stationary robot experiment for the red channel of the standard vision system. The predicted result for the *horizontal* blob position is plotted with a solid line and the "ground truth" reference signal is plotted with a dotted line. The activities of all the sensors are indicated as trig-points on top of the reference signal.
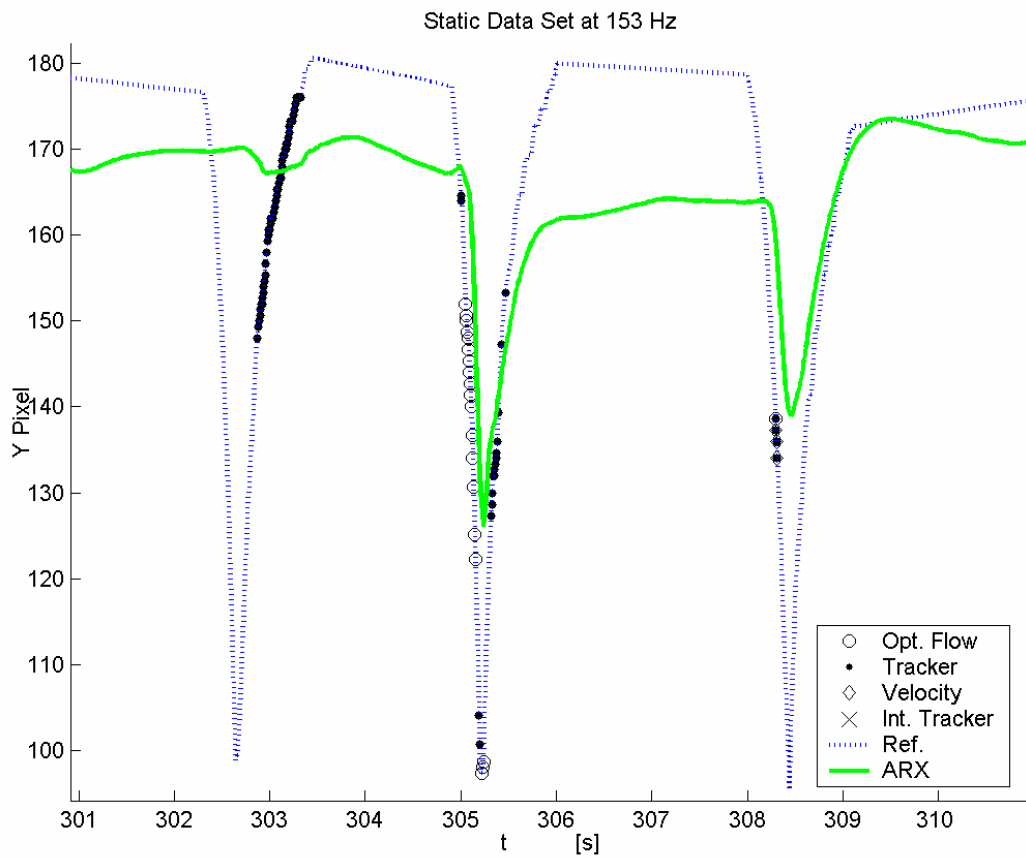
Figure 7:   An example from the stationary robot experiment for the red channel of the standard vision system. The predicted result for the *vertical* blob position is plotted with a solid line and the "ground truth" reference signal is plotted with a dotted line. The activities of all the sensors are indicated as trig-points on top of the reference signal.
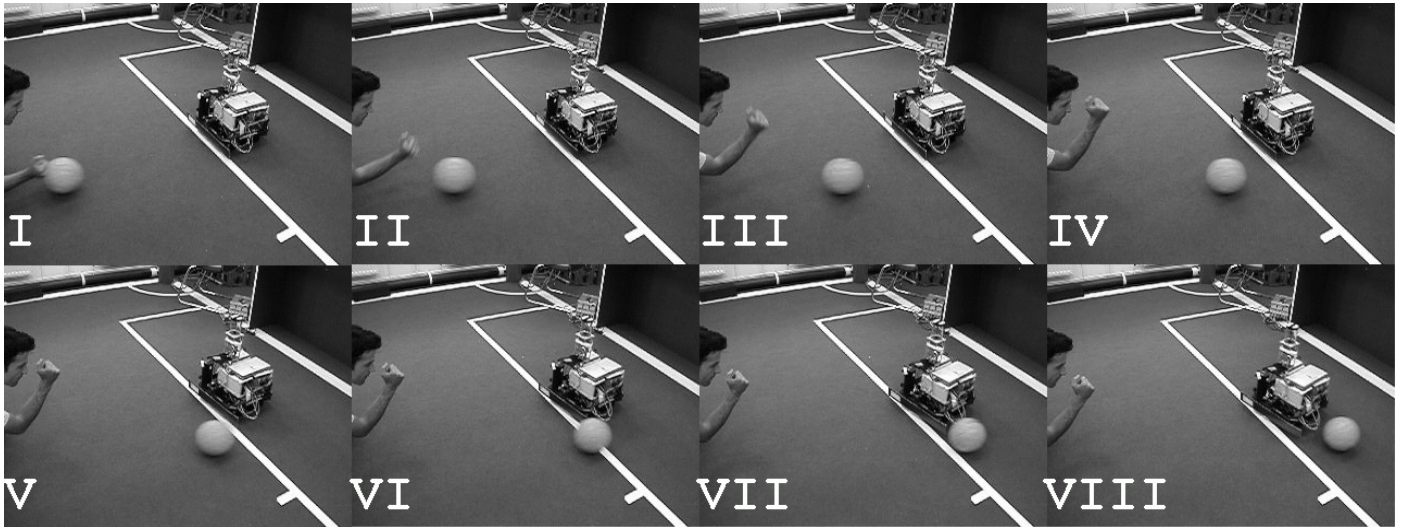
Figure 8: This picture sequence ('I' to 'VIII', 15 frames/sec.) depicts a penalty situation. The depth of the goal area is 1 m, the width of the goal is 2 m. The ball of diameter 22 cm is placed 1 m in front of the goalie, already in a sensitive area of a neuromorphic vision sensor. In picture 'I' the ball is accelerated to about 3.5 m/s. In 'IV' the robot has started to move (starting position pinpointed) thus the observed delay till movement is less then 200 ms. This can be attributed to at least 15 ms motor "run-up" time, 20 ms CAN bus delay and at least 100 ms acceleration time against the moment of inertia before the snap shot. The robot accelerates with 2.5 m/s^2. Although it can go the required 10 cm in 280 ms, in principle, it still misses the deadline of 300 ms which the ball needs to pass by, in picture 'VII', because of the aforementioned delays.

Vlatko Bečanović is a post-doctoral researcher at the Fraunhofer Institute for Autonomous Intelligent Systems in Germany. He graduated at the Stockholm University (Sweden) with a B. Sc. in Physics in 1997, and in 2000 he received a PhD degree in Experimental Particle Physics from the Royal Institute of Technology (KTH) in Sweden. His research interests are sensor fusion, recurrent neural networks, biologically inspired machine vision systems and their application to fast autonomous mobile robots. His research is at the moment focused on the implementation of neuromorphic technology in the domain of autonomous mobile robotics. Mr. Bečanović has been involved in RoboCup activities since 2001. He has been a regular member of the middle sized league teams *GMD-Robots* and *GMD-Musashi*. He was also in the humanoid league organizing committee for the 7'th international RoboCup competitions held in Padua Italy in 2003.

Ramin Hosseiny is a doctoral researcher at the Fraunhofer Institute for Autonomous Intelligent Systems in Germany. He received his Diploma (M.Sc.) in Electrical Engineering from the University of Technology Aachen (Germany) in 2003. His research interests are biologically inspired vision systems, neuromorphic technology, embedded system design and neural networks in robotics. In particular Mr. Hosseiny's research direction is the practical application of a novel analog spiky recurrent neural network in the domain of autonomous mobile robotics. He is also involved in the development of a neuromorphic system for visual odometry through ego motion estimation with optical flow.

Giacomo Indiveri is a Research Assistant at the Institute of Neuroinformatics of the Swiss Federal Institute and the University of Zurich. He graduated in electrical engineering from the University of Genoa, Italy in 1992 and won a post-graduate fellowship within the "National Research Program on Bioelectronic technologies" from which he graduated (cum laude) in 1995. His current research interests include the design and implementation of neuromorphic systems for modelling selective attention neural mechanisms, and for exploring the computational properties of networks of silicon integrate and fire neurons. Indiveri is co-teacher of two classes on the analysis and design of analog VLSI Neuromorphic Systems at the Swiss Federal Institute of Zurich, co-organiser of the Workshop on Neuromorphic Engineering, held annually in Telluride, and co- author of the book "Analog VLSI, Circuits and principles", by Liu et al. from MIT Press 2002.