# Object Tracking Using Multiple Neuromorphic Vision Sensors

Vlatko Bečanović, Ramin Hosseiny, and Giacomo Indiveri[1]

Fraunhofer Institute of Autonomous Intelligent Systems,
Schloss Birlinghoven,
53754 Sankt Augustin, Germany
{becanovic, hosseiny}@ais.fraunhofer.de

**Abstract.** In this paper we show how a combination of multiple neuromorphic vision sensors can achieve the same higher level visual processing tasks as carried out by a conventional vision system. We process the multiple neuromorphic sensory signals with a standard auto-regression method in order to fuse the sensory signals and to achieve higher level vision processing tasks at a very high update rate. We also 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.

**Keywords:** Neuromorphic vision sensors, analog VLSI, reactive robot control, sensor fusion, RoboCup.

## 1 Introduction

In our lab aVLSI technology is exploited in fast moving mobile robotics, e.g. RoboCup, where soccer-playing robots perform at high speed. The robot that is used in our experiments is a mid-sized league robot of roughly 45 by 45 cm with the weight of 17 kg. It is 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 robot uses 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 [1] from a digital color camera. The conventional part of vision processing is software based and consumes the most of the calculation resources on-board the robot [2].

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

[1] Institute of Neuroinformatics, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.E-mail: giacomo@ini.phys.ethz.ch

D. Nardi et al. (Eds.): RoboCup 2004, LNAI 3276, pp. 426–433, 2005.

while the robot is in motion. The ball needs to be close to the robot in order to be successfully controlled. 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.

This paper is structured as follows: in section 2 a description of our robot platform is given. The neuromorphic vision sensors used in the experiments are presented in section 3. In sections 4 and 5 we investigate how the vision system can be aided with a set of neuromorphic vision sensors. Here, we present data collected during experimental runs with one of our robots. We show that this data is suitable for further higher level processing. In the conclusions we point out the importance of the results that were achieved.

## 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 [3]. They are interconnected with a bus interface (CAN), which is a standard in German automobile industry. A notebook PC operates the main behavior 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 the experiments 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 color blob tracking of multiple objects and delivers information from tracked objects such as position of geometrical center, bounding box and pixel area. In our experiments only the position of the geometrical center 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 artifacts. 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 [2].

## 3 Neuromorphic Vision Sensors

Neuromorphic vision chips process images directly at the focal plane level. 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. Data reduction is thus performed, as they transmit only the result of the vision processing off-chip, without having to transmit the raw visual data to further processing stages. Standard CCD cameras, or conventional CMOS imagers merely measure the brightness at the pixel level, eventually adjusting their gain to the average brightness level of the whole scene. The analog VLSI sensors used in our experiments are made using standard 1.6 and 0.8 micron CMOS technologies. They are small 2x2 mm devices that dissipate approximately 100mW each. Specifically, they a 1D tracking chip [5], a 1D correlation-based velocity sensor [6], a single 1D chip comprising both tracking and correlation-based velocity measurements, and, a gradient based 2D optical flow chip [7] (cf. Fig. 1). The 2D optical flow chip is the most complex and computes the 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 the whole 1D array in two separate output channels, 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.

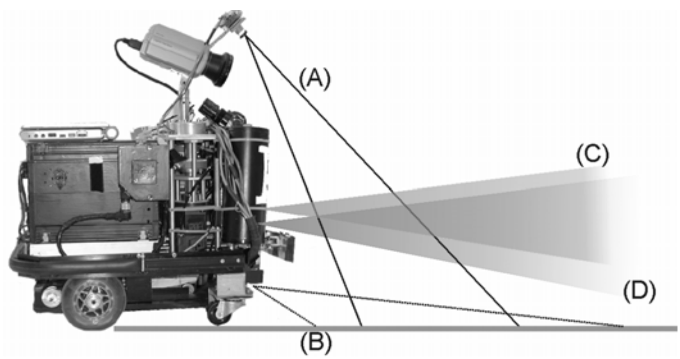


**Fig. 1.** Four aVLSI sensors mounted on the robot with their respective fields of view: The 2D optical flow sensor (A) is pointing straight towards the ground and also the absolute tracker (B) is pointing towards the ground. The absolute tracker (B) is mounted at a somewhat lower angle and with its pixel array vertically aligned. The 1D velocity tracker (C) and the 1D integrating tracker (D) are directed as a divergent stereo pair and with their respective pixel arrays horizontally aligned

## 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 color 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 colored object, and secondly, to validate this against a somewhat cluttered background. The first stage is performed to investigate the precision of the prediction 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 aid a standard camera based digital vision system in a local domain of the scene.

The experiment consists of data collection from the neuromorphic vision sensors and the digital vision system of our soccer robot. The RoboCup soccer playing robot is fully autonomous and is operated by a behavior based program that was used by our team at the last world championships in Padua Italy [8],[9]. 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.

Three experiments were performed, two stationary experiments followd by a moving robot experiment at the end [10]. In the stationary epxeriments 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. In the moving robot experiment the robot will constantly try to approach the red ball in different maneuvers. During this time the robot will 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.

## 5 Experimental Results

The first step here consists of two stationary robot experiments treated in section 5.1, and the second step, which is a moving robot experiment is treated in sec. 5.2. The data is evaluated by comparing the results from a standard dynamical prediction model. 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 [11]. 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).

### 5.1 Stationary Robot Experiment

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 uniformly distributed in the spatial domain and normally distributed in the temporal domain. The prediction efficiency is given in Table 1. For example, 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 40 pixels, which corresponds well to the fact that the resolution of the neuromorphic sensors is between 10 and 24 pixels.

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, where the ball with its significantly lower momentum got deflected in an elastic collision. During the experiment the ball was thrown 25 times back and forth in different zones, but in rando`m 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 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). An example from the stationary robot data set used in this experiment is given in Figs. 2 and 3, 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.

**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 |

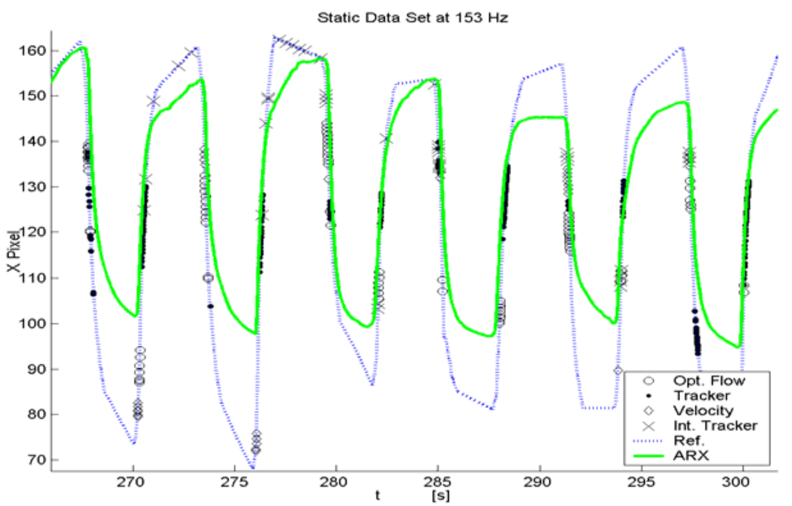


**Fig. 2.** An example from the stationary robot experiment for the red channel of the standard vision system. The predicted result for the *horizontal* blob position is plotted with a solid line and the "ground truth" reference signal is plotted with a dotted line. The activity of all the sensors is indicated as trig-points on top of the reference signal

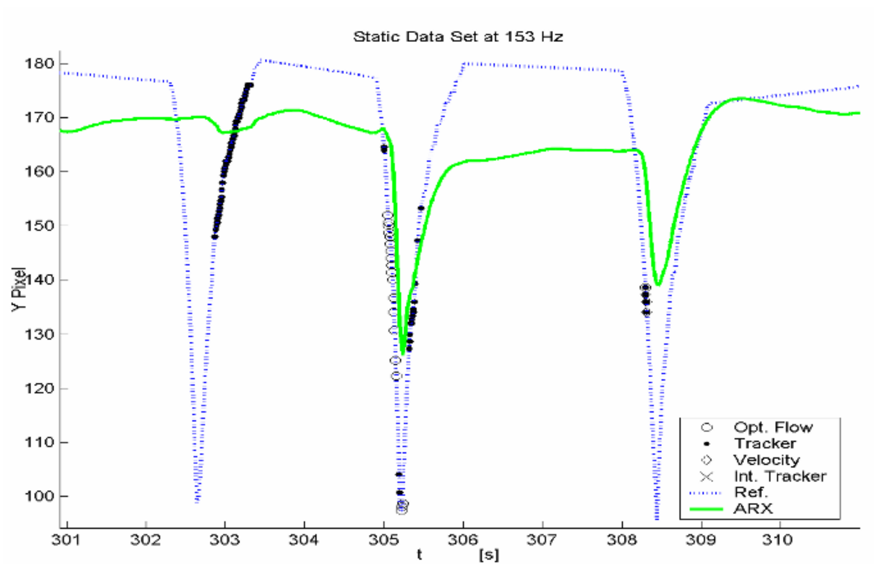


**Fig. 3.** An example from the stationary robot experiment for the red channel of the standard vision system. The predicted result for the *vertical* blob position is plotted with a solid line and the "ground truth" reference signal is plotted with a dotted line. The activity of all the sensors is indicated as trig-points on top of the reference signal

### 5.2 Moving Robot Experiment

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) on a fully performing robot, where the robot during this time tries to approach the ball in different maneuvers. The experiment is validated by tracking red and white objects with the standard vision system of the robot, where the red object corresponds to the red ball and white objects correspond to lines present in the playfield. The reference information of the red object is as before used for the model fitting and the reference of the white objects (corresponding to white lines) is only used to indicate trig-points to be 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 close to over-all efficiency).

**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 |

## 6 Summary and 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 object tracking. 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 neuromorphic vision systems, which provide high speed performance together with low power consumption and small size which is advantageous for reactive behavior based robotics [12], where sensors are influencing actuators in a direct way. In general it can be concluded that the results of the robot experiments presented 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 can be solved with only a few neuromorphic vision sensors analyzing a local region of the visual scene.

## Acknowledgments

The authors would like to thank Dr. Alan Stocker, from the Center for Neural Science, New York University, for providing the 2D optical flow sensor. Special thanks are due to Stefan Kubina, Adriana Arghir and Dr. Horst Günther of the Fraunhofer AIS for help regarding the set-up of the robot experiments. This work is funded by the Deutsche Forschungsgemeinschaft (DFG) in the context of the research program SPP-1125 "RoboCup" under grant number CH 74/8-2. This support and cooperation is gratefully acknowledged.

## References

[1] http://www.1394ta.org/
[2] 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.
[3] http://www.infineon.com/
[4] S. Kubina,"Konzeption, Entwicklung und Realisierung – Micro-Controller basierter Schnittstellen für mobile Roboter" *Diploma thesis at GMD Schloss Birlinghoven*, 2001. (in German).
[5] 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
[6] 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
[7] 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
[8] 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.
[9] A. Bredenfeld, G. Indiveri, "Robot Behavior Engineering using DD-Designer", *Proc. IEEE/RAS International Conference on Robotics and Automation (ICRA)*, 2001.
[10] R. Hosseiny, "Fusion of Neuromorphic Vision Sensors for a mobile robot", *Master thesis RWTH Aachen*, 2003.
[11] L. Ljung, System Identification: Theory for the User, Prentice Hall, 1999.
[12] R. Brooks, "A robust layered control system for a mobile robot", *IEEE Journal of Robotics and Automation*, Vol, RA-2, No. 1, 1986.