PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 50, NO. 3–4, PP. 281–294 (2006)

COGNITIVE MODEL FOR IMAGE SENSOR MOTION CONTROL

Barna RESKÓ[1], Péter KORONDI[1,2] Péter BARANYI[1] and Hideki HASHIMOTO[2]

[1]Computer and Automation Research Institute of the Hungarian Academy of Sciences
Budapest University of Technology and Economics
[2]Integrated Intelligent Systems Japanese–Hungarian Laboratory
University of Tokyo, Tokyo, Japan
e-mail: rbarna@datatrans.hu
e-mail: korondi@get.bme.hu
e-mail: baranyi@tmit.bme.hu

Received: April 28, 2006

Abstract

This paper presents a neurobiology and cognitive psychology inspired model, implemented using a neural-network-like parallel computational strategy. The goal of the model is to show how a visual cortex inspired system can control camera alignment using a given camera hardware setup, in a similar way to the brain's controlling eye movements. The system (computational model and camera hardware) are integrated into an experimental environment, the Intelligent Space, a room of ubiquitous computing. The intelligent space is based on several Distributed Intelligent Network Devices (DIND). A DIND has a sensor input, integrated intelligence and a communication interface. The model presented in this paper serves as the integrated intelligence component of a particular DIND. The proposed model implemented in a parallel hardware performs real time operation.

Keywords: Cognitive vision, stereo alignment, visual cortex.

1. Introduction

The main goal of this paper is to present a neurobiological and cognitive psychological analogy based cognitive model. The model is based on the biological architecture and cognitive functionalities of the mammalian visual cortex, with a special respect on the binocular cells. The model can solve the problem of stereo alignment, analogically to the brain fixating the eyes on the same spot of the visual field, allowing 3D vision. As a consequence of the model's architecture, it can also be applied in object (or face) tracking, as described later in this paper. Besides the possibilities of practical applications of the model, it also aims to extend the limits of classical computation.

In order to show why cognitive models can give the necessary boost, consider the example where a test person has to determine whether there is a cat or something else in the shown image, and press a button according to the decision. Such a task is impossible for a computer to perform today, yet a human can do it reliably in half a second or less. This result becomes more shocking if we know that the "processing time" of the basic processing unit of the brain (a typical neuron) is in the range of milliseconds, while the basic processing unit (a logic gate) of a modern silicon-based computer is 5 million times faster. The answer for how the "slow" brain can solve this task lies in its special architecture and particular information representation and processing. It is thus our belief that in order to step beyond the borders of today's computer systems' architectures the basic way of information representation and processing has to be changed. For new ideas we turn to existing cognitive systems in biological architectures to study them, because they already bear the solutions that we are seeking for. A cognitive system is implemented in a biological neural network, where simple units of computation are connected in a very complex structure. Our research goal is to turn the cognitive information processing system into engineering models which can later be organized into a cognitive psychology inspired model running on a biology related computational architecture.

Our work has received inspiration from research about biological visual systems, [1, 2, 3]. These works reveal that the mammalian visual cortex is populated with neurons responsible for the perception of different visual cues. Among these cues we can find the binocular cues, which allow the binocular perception of space. In this paper a similar model is proposed for camera alignment. This is not to say that the model presented is necessarily identical with biological visual systems. The ultimate criterion of our work is performance from a technical point of view.

The proposed model is embedded in a given camera hardware. Similarly to the human eyes, the camera hardware also has two cameras, which can be rotated together in the vertical direction, and separately in the horizontal direction, as shown on *Fig. 1*. Due to these mechanical constraints, the model deals only with horizontal camera motions.

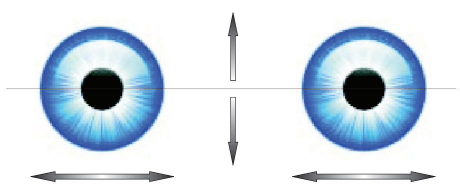

Fig. 1. The human eyeballs can move together in the vertical direction, and separately in the horizontal direction

The camera system with the control module of cognitive intelligence is integrated into an experimental environment, the 'Intelligent Space', proposed by Hashimoto Laboratory at The University of Tokyo. Intelligent Space is a space, which has distributed sensory intelligence and it is equipped with actuators. Actuators are used to provide information and physical support to the inhabitants. The various devices of sensory intelligence cooperate with each other autonomously, and thus the whole space gains high intelligence. The Space is populated with agents, and each agent obtains the sensing information from multiple input attributes of the Intelligent Space, such as cameras or microphones. Agents also obtain augmented information from other agents. Intelligent Space recomposes the whole space from each agent's sensing information, and returns intuitive and intelligible reactions to man. In this way, Intelligent Space is the space where man and agents can act mutually.

A space becomes intelligent, when its fundamental components, the Distributed Intelligent Network Devices (DINDs) are installed in it. A DIND consists of three basic modules that provide three main functions:

- **sensor module** monitors the dynamic environment within the Intelligent Space
- **computational module** processes the sensed data, extracts information, and makes decisions
- **communication module** communicates the extracted information and decisions towards other DINDs or inhabitants.

The intelligence of the DIND mainly comes from the intelligence and computational performance of the integrated computational module. The integrated classical mathematical and soft-computing based computational methods usually perform well in Intelligent Space, however in some aspects it is easily possible to face their limits. In some cases they are not robust enough, or simply due to their architecture they cannot deal with some aspects of the problem.

The organization of this paper is as follows. In Section 2 the cognitive model for eye alignment will be presented, along with its modification for object tracking. Section 3 includes a short description of the Intelligent Space and the DIND, in which the proposed model is integrated. In Section 4 the test results are presented, along with the possible hardware implementations of the model in Section 5. Finally in section 6 conclusions with a discussion of the results is made.

2. Cognitive Model for Camera Alignment

2.1. Background

The model described in this section was inspired by the functionalities of binocular neurons in the mammalian visual cortex. Binocular perception of space is based on binocular neurons that receive input from both eyes. These neurons can be classified by their functionalities described by two parameters: the horizontal positions on the two retinas of the image patches they receive as input. If there is a shift between the two horizontal positions, the neuron is responsive to a certain disparity or depth. On a geometrical basis, the neurons can be re-classified to be described by two different parameters: the actual position on one retina and the depth (depending on the position on the other retina). If a neuron receives a similar stimulation from the two eyes, it is because the image projected to those particular retinal areas are similar. Such a stimulation will cause an intense firing of the neuron. The actual neurons firing at a time describe the binocular depth of a scene at different locations of the visual field. By separately moving the eyes along the horizontal direction, neuron classes of different functionalities will perform intense firing.

Suppose that the eyes are aligned to the same visual feature of the scene. Such a state has two important consequences:

- The neuron receiving both inputs from the center of the two fovea will perform intense firing.
- This neuron belongs to the class which is responsive to the center of the fovea on one retina, and zero depth.

Now suppose that the eyes are not aligned to the same visual feature (this happens when one looks with crossed eyes). Such a state also has two consequences:

- The neuron receiving both inputs from the center of the two fovea will not perform intense firing.
- The neuron representing the center of the fovea and non-zero depth will perform intense firing.

The goal is to control the eyes in a way that both of them fixate on the same visual feature. Humans do this by intuitively choosing the feature (with one eye) and aligning the other eye to envision the same feature. In other terms, the goul is to activate a neuron that is associated with the center of the fovea and with a certain depth. According to the depth a control signal to move the eyes is produced until the neuron associated with the center of the fovea and zero depth is activated, as a result of correct eye alignment. Further details about binocular perception and eye alignment is available at [1].

In the following sections we present a computational model based on the binocular cells of the visual cortex. The goal is to control two cameras so that they are aligned on the same visual feature. A small technical modification of the described model can be used for object tracking.

2.2. Pixel Pair Matching

A given camera hardware was used with two cameras that can be moved together vertically and separately along the horizontal direction (*Fig. 1*). The proposed model is embedded in this camera system.

The goal is to align the optical axis of the cameras to envision the same object of the visual field. The input is two sub-windows of the projected image of the two cameras. The following assumptions were made in the choice of the input sub-windows.

If we consider the biological analogy of the camera alignment, a master-slave relationship between the two eyes or cameras can be supposed, providing a master and a slave input image sub-window for the model. The master image serves as the target, to which the slave image has to be matched during the alignment process. The slave image is provided by the slave camera. The model has to adjust the optical center of the slave camera so that a similar feature P is projected on the

Fig. 2. Image parts fed into the ANN. Master image (left) and Slave image (right)

optical center of both the master and the slave camera. A pixel and its surrounding as a 9×9 square is considered as the master sub-window. The task of the model is to find a 9×9 sub-window on the slave image that best matches the sub-window of the master image.

Not all the pixels on the slave image are taken into consideration. Due to mechanical constraints, the camera can only be rotated horizontally. As a result all the 9×9-pixel-sized sub-windows on the horizontal line of the slave image are compared to the single sub-window in the center of the master image, thus the interesting image parts are the 9×9 square on the master image and an $X_{max} \times 9$ band on the slave image, where X_{max} is the width of the slave image in pixels (*Fig. 2*).

A colour image can be decomposed to three images according to the RGB colour components. In our system 8-bit gray scale images are used. Because the intensity value of each pixel is sensitive to changes in contrast and illumination, comparing only the intensities of the pixels and their surroundings to acquire a matching degree does not yield acceptable results. The gradient of each pixel is also taken into consideration during comparison of two sub-windows. We suppose that three measures for each pixel are known, which are intensity, contour strength and contour orientation. Denote I_M^p and I_S^p the intensity of the master and slave pixels being matched, D_M^p and D_S^p the contour strength of the master and slave pixels being matched and O_M^p and O_S^p the contour orientation of the master and slave pixels being matched. The matching degree of a pixel pair $p_M(x_M, y_M)$ and $p_S(x_S, y_S)$ is defined as the average of the absolute difference between the corresponding feature values of the pixels. Denote $d(p_M, p_S)$ as the matching degree, which can be written in the following form:

$$d(p_M, p_S) = \frac{|I_M^p - I_S^p| + |D_M^p - D_S^p| + |O_M^p - O_S^p|}{3} \tag{1}$$

This choice has been made because it gives a true degree of similarity between pixels, it can be computed by a neural network-like parallel computational tool using parameters that are easy to preadjust.

In the next section the proposed cognitive model and its neural-network-like parallel implementation will be presented.

2.3. Model implementation

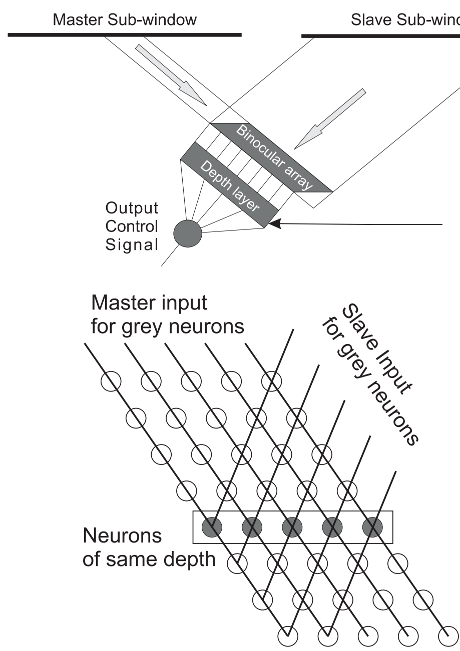

Fig. 3. The overview of the proposed model (top) and a section of the 3D array of binocular neurons (bottom). Each horizontal group of neurons represents a different depth for the same image part (envisioned in the master image).

The structure of the proposed implementation of the model is shown on *Fig. 3*. A neural-network-like computational strategy was used for two main reasons:

- It imitates the inspirator of the model, the brain, which is a vast neural network on the physiological level.
- To allow an extensive parallelization of the functionalities for a potential parallel hardware implementation.

The structure of the proposed implementation is shown on *Fig. 3* (top). The system is composed of 5 layers. Each layer is responsible for a different task (input, matching degree, decision, output), which yields a clear structure that is easy to understand and further develop. In the first layer we suppose to have the input data from the interesting image regions, i.e. the 9×9 square on the center of the master image and the $X_{max} \times 9$ band from the slave image. These data are fed into the input layer of the network.

The outputs of the input neurons form two buses, a master and a slave bus, as an analogy of the optic nerves behind the human eyes. The master bus M for each pixel p of the 9x9 window on the master image carries the I_M^p,D_M^p and O_M^p values, thus groups $3 \times 9 \times 9 = 243$ outputs of the input layer. Let $M_{i,j}$ denote the I_M^p,D_M^p and O_M^p values that belong to the pixel $p(i, j)$ in the master window. The slave bus S carries the I_S^p,D_S^p and O_S^p values of each pixel on the $X_{max} \times 9$ band on the slave image, thus groups $3 \times 9 \times X_{max}$ are outputs of the input layer. Let $S_{k,j}$ denote the I_S^p,D_S^p and O_S^p values that belong to the pixel $p(k, j)$ in the band on the slave image. In the next layer the two buses are crossed in a 3-dimensional array $\mathbf{A}$ of $9 \times 9 \times (X_{max} - 8)$ neuron triplets. This builds up the second layer of the proposed network. A two dimensional section where j is set to a constant value is shown in *Fig. 3* (bottom).

In array $\mathbf{A}$ each neuron triplet calculates the absolute difference of the I_M^p; I_S^p, D_M^p; $D_S^p and O_M^p$; O_S^p values with respect to the pixels from which their inputs are arriving from.

$$\begin{cases} \mathbf{A}_{i,k,j}^{(I)} = |M_{i,j}^{(I)} - S_{i+k,j}^{(I)}|, \\ \mathbf{A}_{i,k,j}^{(D)} = |M_{i,j}^{(D)} - S_{i+k,j}^{(D)}|, \\ \mathbf{A}_{i,k,j}^{(O)} = |M_{i,j}^{(O)} - S_{i+k,j}^{(O)}|, \\ i, j \in [1..9], k \in [1..X_{max} - 8], \end{cases} \tag{2}$$

where for example $M_{i,j}^{(I)}$ denotes the intensity value arriving from the $p(i, j)$ pixel of the master image.

The third layer neurons of the network are situated in a similar 3D array as those in the second layer. The difference is that in the third layer there is only one neuron in each cell of the array. The outputs of the three corresponding neurons in layer 2 are connected to the inputs of the neurons in layer 3. Denote the array containing the third-layer neurons $\mathbf{B}$, then the operation performed by a third layer neuron is described as

$$\mathbf{B}_{i,k,j} = \frac{\mathbf{A}_{i,k,j}^{(I)} + \mathbf{A}_{i,k,j}^{(D)} + \mathbf{A}_{i,k,j}^{(O)}}{3} \tag{3}$$

This means that a third layer neuron simply calculates the mean value of the outputs coming from the three second-layer neurons. The output is the same as it was written in *Eq.* (1). In other words, $\mathbf{B}_{i,k,j}$ equals the mean difference between the pixels $p(i, j)$ of the master image window and $p(k+i, j)$ of the slave image band. It is to note that if k is increased by one, then $\mathbf{B}$ will contain the mean difference between the same pixel on the master image and the next pixel to the right on the slave image.

To compute the matching degree between two pixels, their surroundings also have to be taken into consideration. The array $\mathbf{B}_{i,c,j}$ where c is a constant, contains the average difference between the 81 pixels in the 9×9 master window and the c^{th} slave window. To get the matching degree of a pixel pair, simply the mean value of the 81 difference values is calculated, and subtracted from 255, the maximum value of the difference. This is done by the fourth layer of the network, which contains $X_{max} - 8$ neurons with 81 inputs for each of them. The output of the k^{th} neuron equals to the matching degree of the master window and the k^{th} window on the slave image.

The original goal was to adjust the optical center of the slave camera on the point P that is seen in the center of the master sub-window. The goal can be considered to be achieved when the highest matching degree is given by the $X_{max} - 8/2^{th}$ neuron. In other words the highest matching has to occur in the central neuron of the fourth layer.

It is the task of the fifth layer to decide if the highest matching value is in the middle of the fourth layer, and if not, then in which direction from the center it is situated. In the fifth layer there is only one special neuron, which outputs -1, 0 or 1 according to the input values. The matching degree values from the previous layer are fed into this special neuron. If the maximum of the inputs is situated to the right from the central input of the neuron, it outputs -1, while if the maximal value is to the left, it outputs 1. If the maximal input is the central input itself, the output will be 0. Finally a simple analog hardware tells the servo-motor mechanism to drive the camera to the appropriate direction in order to achieve an optimal matching.

2.4. Modifications for Object Tracking

With a simple modification the above described model can be used for object tracking. The master camera is replaced by an adaptive image memory. A special feature of this memory is that it can adapt its contents to the changing of the environment. After reset the first image is stored in the image memory. Further on the previously described neural model takes the contents of the image memory as master input, and calculates the necessary camera motions. If the output of the model suggests that no camera motion is necessary, meaning that camera is aligned on the desired object, the content of the image memory is updated by overwriting the old content with a combination of the old image and the new image. Denote $\mathcal{F}_{old}$ the image stored in the image memory, and $\mathcal{F}_{new}$ the image provided by the camera. The next image that will be stored in the image memory will be

$$(1 - \lambda)\mathcal{F}_{old} + \lambda\mathcal{F}_{new}, \tag{4}$$

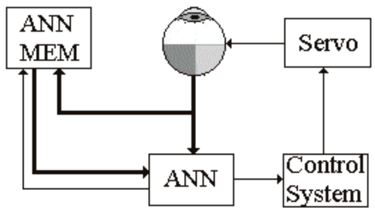

Fig. 4. The system setup using the modified model for object tracking

where λ is a constant decay parameter.

The image memory updates itself according to *Eq.* (4) if the control signal from the model is *neutral*, otherwise the camera image is neglected by the image memory. The system setup for object tracking is shown on *Fig. 4*.

3. Experimental Environment: Intelligent Space

Intelligent space is a limited space (room or building, street or area, or even a whole country), which has ubiquitous computing type computing and sensory intelligence. The sensors might be various types of equipment, such as cameras, microphones, haptic devices, weight sensors, or any other devices that collect information on the state of the space [6]. A conceptual figure of the Intelligent Space is shown in *Fig. 5*.

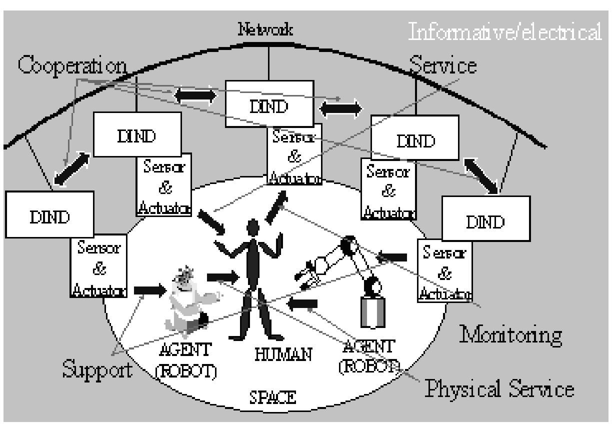

Fig. 5. Intelligent Space concept.

Conventionally, there is a trend to increase the intelligence of a computer to perform more and more complicated tasks or of a robot (agent) operating in a limited area. The I-space concept is the opposite of this trend. The surrounding space has ubiquitous computing and sensory intelligence instead of highly intelligent agents and a concentrated high performance computer. An agent without any sensor or own intelligence can operate in the I-space [7]. Similarly an intelligent supervisory center can take conclusions and initiate actions based on information collected in a building, a city neighbourhood or a large geographical area over the telecommunication network covering it. The difference of the conventional and I-space concept is explained in *Fig. 6*.

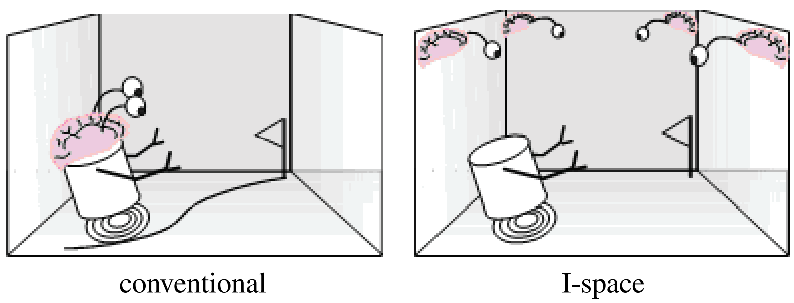

Fig. 6. The conventional concept vs. I-space concept

The various devices of ubiquitous sensory intelligence cooperate with each other autonomously and the whole space has ubiquitous computing background. This is true even if there is a supervision system involved, which is acting as an autonomous agent itself. Each agent in the space has sensory intelligence. (Or has intelligent inputs coming from other agents.) An intelligent agent has to operate even if the outside environment changes, so it needs to switch its roles autonomously and knowing its role it can still help and support humans within the space. I-space having ubiquitous computing recomposes the whole space from each agent's sensory information and returns intuitive and intelligible reactions to human beings. In this way, I-space is the space where human beings and intelligent ubiquitous computing agents can interact mutually.

3.1. Basic Concept of DIND

The key concept called the Distributed Intelligent Network Device (DIND) consists of three basic elements. These are the sensor, the processor (computer, neural network or even a brain) and the communication device. Thus, a DIND is a unit based on three functions that the dynamic environment, which contains people, vehicles and robots, etc., is monitored by the sensor, the information is processed into a form easily captured by the clients in the processor and the DIND communicates with other DINDs through a network or a supervision system, which is itself an autonomous agent. *Fig. 7* (top) shows the basic structure of a DIND. By installing DINDs into the whole I-space, they will perform information exchange and information share by communicating mutually through the network/supervision. The space thus equipped will be altered into an intelligent ubiquitous computing environment [9].

Let's think when people move in space. How they observe, judge and react. A human mainly looks at the space using the eyes, he chooses the required information (recognition), analyses this information, gives a suitable determination (decision), moves legs based on this determination, and do interaction to space (action). As shown in *Fig. 7* (middle), our interactions to space are performed by the repetition of recognition, decision, and action. When a robot performs the same work, decision and action can be said comparatively easier than recognition. If space is too sophisticated, it is very difficult to acquire data by the sensors and to extract only the required information from it.

However, the vision system of a human can solve this problem relatively easily. The system of the eye and the brain can perform similar tasks as a DIND - observe, recognize and react. The speed and precision of this process is amazing. The actual similarity between the DIND concept, the human interaction to space and the vision system can be seen in *Fig. 7*.

Inspired by this similarity between the processes of a DIND and the mental processes of a human, especially in vision, the model for camera alignment and object tracking was integrated into the Intelligent Space for testing and to provide new services within the space.

4. Testing and Evaluation

To evaluate the proposed system, a mechanism was built with cameras mounted on the top. The mechanism was similar to the human eyes: the two cameras could move independently in the horizontal direction, but only together in the vertical one. The proposed model was implemented in a C++ program and was controlling the servo mechanisms that moved the cameras.

An interesting result is the matching degree between the target image and the various image windows situated on the slave image band.

Below an image pair is presented with the matching degree graph and the direction suggested by the system.

On the presented pair (*Fig. 8*) of images the center of the master image (top) corresponds to a pixel that is situated to the left from the center of the slave image (bottom). The matching degrees are shown on *Fig. 9.* In this case the peak is to the left from the center, meaning that the slave camera has to be rotated to the left.

Using a real-time simulation of the system, the camera was able to follow objects like a human face. To start the test, the human had to be still, while the system learnt the face. Afterwards the face could be moved, and the camera was following it. When the camera reached the desired orientation, the image memory

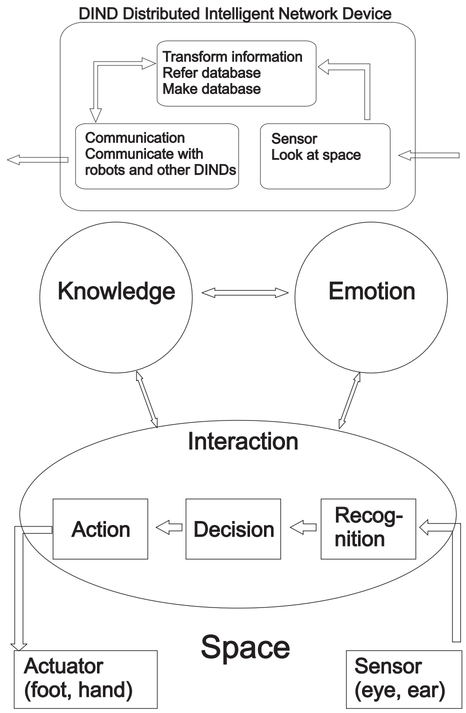

Fig. 7. The similarity between a DIND, the human interaction to space and the human vision system that follows an object moving in space.

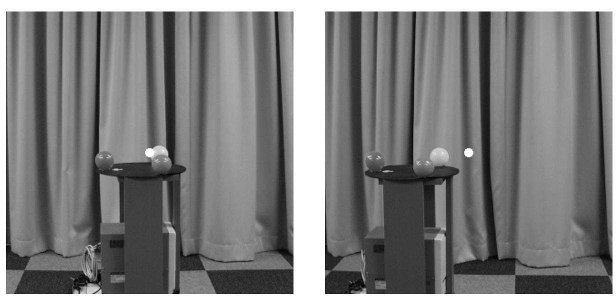

Fig. 8. Master image (left) and slave image (right). The image centers are indicated by the white dots on both images. The matching pixel on the slave image is located to the left from the center of the slave image.

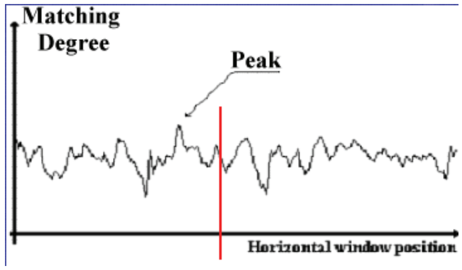

Fig. 9. The matching degree of the image pair. The peak representing the best match is to the left from the center.

was updated. This allows slow changes in the object being followed, and makes the system robust against noise and other perturbations.

5. Hardware Implementation

One of the most advantageous properties of the proposed model's architecture is its native parallelism. A software implementation running on the fastest *von Neumann*-based processor cannot provide fast, $O(1)$ time responses even if it is extended with Single Instruction Multiple Data (SIMD) instruction sets like Intel's Streaming SIMD Extensions (SSE). Parallel processing with multiple simple computational elements, on the other hand, can provide tremendous speedups. The Field-Programmable Gate Array (FPGA) is such a microelectronic device the programming logic of which can be set up according to the users' needs, and some models even allow to be reconfigured during operation time. Thus, the proposed architecture can be implemented in an FPGA, and then can be used as a co-processor or accelerator card in a PC environment to solve dedicated tasks. Moreover, it can be a stand-alone image processing device that solves the task without the execution of any conventional algorithm.

The proposed architecture requires about 300 000 computational elements to perform on a 320-pixel-wide image. The state of the art FPGA has about 6 000 000 logic cells that is sufficient for about 100–200 000 computational elements. This number copes with what is necessary, while the processed image is still relatively small. However, the famous Moore's Law also applies to FPGAs saying that in about every two years the number of transistors on a silicon chip doubles, thus the number of logical cells is expected to double, too. In addition, some of the modern FPGAs also have the capability of being re-programmed in runtime. Applying this feature allows the use of only one chip for the processing that can be done by re-programming the architecture for each task, sacrificing extra processing time. In conclusion, a primary visual cortex based image contour detector chip can be realized in near future by some compromises.

A simple (low resolution) version of the model is being implemented in an FPGA. A serious bottleneck in this solution is the small number of parallel input/output data that can be transmitted to and from the FPGA. Apart from this problem, the implementation will give ground to test and evaluate the model operating on a dedicated hardware.

6. Conclusion

This paper proposes a neurobiology and cognitive psychology inspired model for image capture device alignment and object tracking. Implemented on a parallel computational architecture, the model shows the potential of cognitive vision systems. Furthermore, the model embedded in a camera system and integrated into Intelligent Space shows that it is also applicable for scientific and industrial purposes.

The modification of the system with the introduction of a forgetting image memory is an example of hybrid cognitive-engineering solutions. In such a solution modules of different types are integrated into a system to aggregate the advantages of each of them.

Until today, only the processor-based implementation of the model is available, which does not allow to measure its true performance. However, a real parallel implementation in an FPGA is on its way.

References

[1] http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/say-cheese/marc ia/mfinal.html
[2] http://www.ubiq.com/hypertext/weiser/UbiCACM.html
[3] http://www.theintelligentoffice.com
[4] http://www.ai.mit.edu/projects/iroom/projects/
[5] http://www.ada-exhibition.ch/
[6] LEE, J.-H. – H. HASHIMOTO, H., Intelligent Space - Its concept and contents -, *Advanced Robotics Journal*, Vol. 16, No. 4, 2002.
[7] HASHIMOTO, H., Intelligent Interactive Space based for Robots, *Proceedings of the 2000 International Symposium on Mechatronics and Intelligent Mechanical System for 21 Century*, p. 26, 2000.
[8] PAN, H. – MANIC, M.– LI, X.– WILAMOWSKI, B., Multilevel Logic Multiplier Using VLSI Neural Network, *ICIT*, Maribor, Slovenia, 2004.
[9] HASHIMOTO, H., Intelligent Space - How to Make Spaces Intelligent by using DIND?, *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC'02)*, 2002.10
[10] APPENZELLER, G.– LEE, J. H. – HASHIMOTO, H., Building Topological Maps by Looking at People: An Example of Cooperation between Intelligent Space and Robots, *Proceedings of IROS97*, 00, pp1326-1333, September, 1997.
[11] HASHIMOTO, H.– LEE, J. H. – ANDO, N., Self-Identification of Distributed Intelligent Networked Device in Intgelligent Space, *Proceedings of the 2003 IEEE International Conference on Robotics and Automation (ICRA '03)*, pp. 4172-4177, 2003.09, Taipei, Taiwan, ISBN 0-7803-7737-0