Perspectives of use of the Cognitive Model of Visual Attention in Real-world Applications

Rudolf Andoga[1], Ján Kováč[2], Ladislav Madarász[1]

[1]Department of cybernetics and artificial intelligence, FEI TU Košice

[2]Department of management and economics, SjF, TU Košice

Abstract: *By visual perception of a scene including different objects and for the need of interaction with certain target object located in the scene, it is necessary for the system to aim it's attention at certain target object. This mechanism is one of the principles of vision and likewise many biologically motivated systems it can be profitably used in many real-world applications. Proposed model is an implementation of the mechanism of visual attention in computer simulated environment.*

Keywords: *Cognitive model, neural network, object classification, assembly process and systems, supply systems.*

1 Introduction

Fields of research in artificial intelligence are today traditionally inspired by functions of real biological systems. Although it is often easier to find a specific engineering solution for specific problem, biologically motivated A.I. systems generally show higher flexibility by adaptation to enviromental conditions that are not known in advance. In this article we propose a model of object oriented visual attention, that comes from neurophysical knowledge and psychophysical experiments with the emphasis on the use of the model in real-world engineering applications.

A system or biological organism that interacts with an environment where are situated many different objects may need to manipulate or visually observe one of the objects. To accomplish that task the object has to be selected as a target of operation between other objects - distractors. Important aspect of this selection is also the reduction of input information by visual perception of the scene, thus making it easier to process that information for the system. In this article we will describe the mechanism of visual attention and it's biological motivation together with the perspective of use of attributes of visual attention and proposed model in real-world applications.

2 Biological motivation

2.1 Biological model

In the basic principle the visual information in visual cortex is processed in two basic streams (fig. 2.1). At first, it is ventral stream, or so called „what stream“ that is responsible for identification of objects situated in the visual field. This identification occurs in hierarchically highest layer in visual cortex, AIT (anterior inferotemporal cortex). In this layer one group of active neurons belongs to one class of objects. The second stream responsible for processing of visual information is the dorsal or so called „where stream“. This stream tells where is positioned object that is identified by the ventral „what“ stream. The ventral stream runs through layers V1, V2, V4, PIT and AIT and the dorsal stream runs through layers V1, V2, V4 and goes then to PG (parietal cortex) region, which is interconnected with other somato-motoric functions of the brain (fig. 2.1). The figure is showing specific region of parietal cortex, the LIP region that is responsible for eye movement. The basic principle of visual attention is to get the position of target object among other distractor objects. This process is not fully explored and understood yet, but it is supposed that information about the position of target object is gained by recurrent propagation of signal in ventral stream of visual cortex from hierarchically highest layer, AIT, to lower, retinotopicaly organized layers. Information about the position of the object is then processed by the dorsal stream and passed to other areas of neural system. The whole process of visual attention can be schematically described as follows. Visual information represented on layer V1 is processed in forward manner through hierarchically organized layers V2, V4, PIT and AIT. In AIT layer the objects in visual field are identified and also the selection of target object occurs in this layer at this point. Consequently the signal is then processed in backward manner from the layer AIT to lower layers and by some mechanism (which will be later described), the position of target object is obtained. This process is called selection of position based on identity of the object. Information about position of the target object is then processed in the dorsal stream through layers V2, V4 and PG and consequently the system will move it’s eyes toward the target object. Proposed system models exactly this mechanism and this structure.

TE – anterior inferotemporal cortex

TEO – posterior inferotemporal cortex

PG – parietal cortex

LIP - lateral intraparietal area

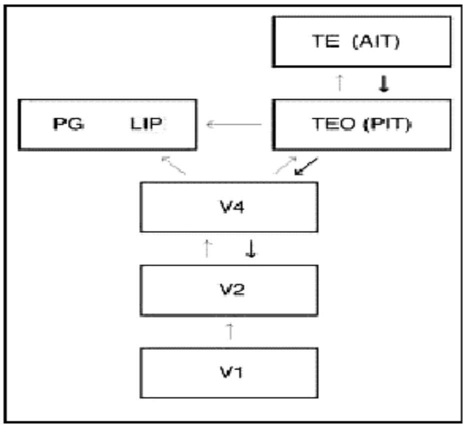

Fig. 2.1: Scheme of the biological model of visual cortex

3 Proposed model

3.1 Structure

Our aim is to model the mechanism of visual object oriented attention and to simulate the psychophysical experiments described in [2]. Consecutive features of the model emerge then from those basic experiments. Basic structure of the proposed model is based on real biological model described in [1] and showed on (fig 2.1.).

The organization of layers in the neural network of proposed model is analogical to organization of layers in visual cortex (fig 2.1). The input information from visual field is represented in layer V1. Layers V1, V2, V4 and PIT have retinotopical organization of neurons in them, that means that they maintain information about the shape and position of objects as they are presented on retina. Next figure shows the basic structure of the proposed model (fig 3.1)

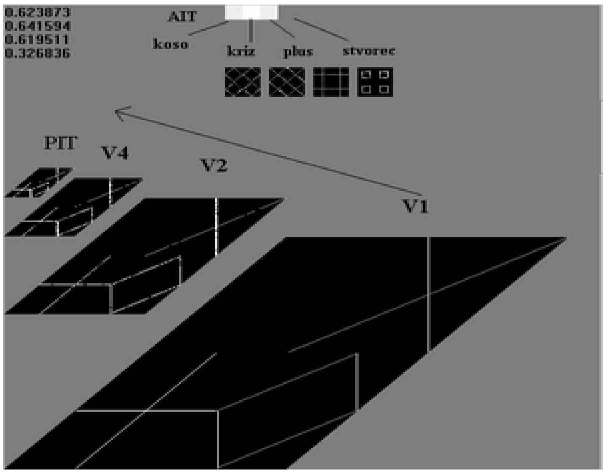

Fig. 3.1: Visualization of the proposed model

Figure 3.1 is showing the single layers V1, V2, V4 and PIT in isometric view. That means that for visualization of the layers is used 3 - dimensional projection of 2-dimensional layers. Single layers can be seen as 2-dimensional matrixes of neurons. Those layers are then ordered hierarchically, that means that they can be seen as layers one above each other in 3-dimensional space. That illustrates the next figure.

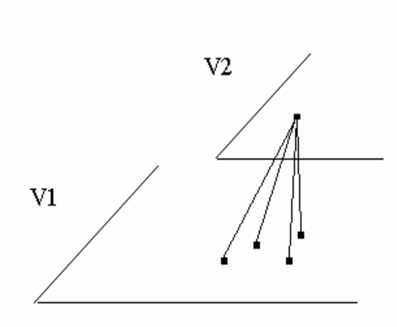

Fig. 3.2: Schematic view of the layers

The input layer V1 is a 2-dimensional matrix of neurons with dimensions (256x256), that means 65536 retinotopicaly organized neurons. This dimension allows the system to process picture, or visual scene with resolution of 256x256 points. For desired change of resolution only the change of dimension of input layer is needed. Bright point in the visual field is represented with neuronal activity $x_{in\{i,j\}} = 1$. Other levels of brightness can be represented by corresponding activity from interval <0;1>. Activation function of neurons in layer V1 is as follows

$$f_{in}(x) = x_{in} \tag{3.1}$$

Layer V2 is similarly as layer V1 a 2-dimensional matrix with dimensions (128x128) and every neuron of this layer is connected with four neurons of previous layer V1. That means that receptive field of every neuron in V2 covers 2-dimensional submatrix (2x2) of neurons in V1. This fact means that every neuron in V2 population is sensitive only to one of the possible line orientations by learning (vertical line, horizontal line, diagonals). Layer V2 is responsible for the first approximation of objects presented on input. It is necessary to mention that single submatrixes of V1 belonging to neurons in V2 do not overlap each other. Next approximation of input objects is done by V4 layer. Analogically every neuron from layer V4 covers a (2x2) submatrix of neurons from layer V2. V4 has dimensions (64x64) neurons. The last approximation layer is the PIT layer with dimension (32x32) with neurons covering again (2x2) submatrixes of V4. The highest layer AIT is responsible for identification of input objects. For the simulation of real biological experiment as described in [2] we used AIT layer consisting of 4 neurons, belonging to four input objects. Decreasing of neuron populations in single layers relates to rising of receptive fields of neurons in visual cortex. Although in real visual cortex the counts of single neurons in single layers and also the interconnection between single layers aren't that deterministic, the approximation we use in the proposed model is sufficient for simulation of biological experiments and technical solutions. Activation function of single neurons in layers V2, V4 and PIT is as follows:

$$f_{in}(x) = \frac{2}{1+e^{-\beta x}} - 1 \tag{3.2}$$

This activation function maps zero input to zero output of single neuron. Neurons in AIT layer have activation function equal to identity. For network weight adaptation in feedforward manner Hebbian learning rule is used:

$$\Delta w_{ij}(t) = \nu x_i(t) x_j(t) \tag{3.3}$$

The use of Hebbian learning rule for network weights adaptation has background in biological neural system. This results not just in biological realism but it also gives emergent behaviour to the model in the way of quickly adapting to new objects and ability to identify deformed or displaced input objects. The model is able to adapt within 10 cycles of learning of previously unknown objects and can identify objects in parallel way invariant to their position.

Propagation of signal in backward manner needed for the extraction of „where" information is carried out by reciprocal synapsies. Those are projected in analogy with real neural system from hierarchically highest layer AIT into lower layers. That means that this feedbackward stream runs through layers AIT->PIT->V4->V2 with layer V2 being the output layer.

3.2 Operation of the model

Operation of the proposed model emerges from it's structural layout and is trying to model visual attention with acceptable structural simplifications. Learning, or weight adaptation ongoes only in feedforward manner, where single objects are presented to the system together with weight adaptation using Hebbian learning rule. Gradually the network is trained to identify all objects invariant to their position. In experiments, after training, the network is able to recognize one to four objects presented on the input V1 layer invariant to their changing position. In realized experiments we suppose the objects to appear on four different positions. If all four objects are present in layer AIT four corresponding neurons are active.

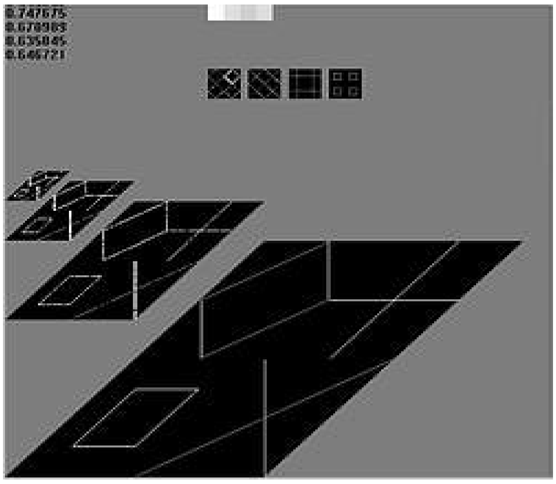

Fig. 3.3: Four objects in layer V1

Model simulates then biological experiment done on primates described in [2]. Process of the experiment is as follows. At first stage the model is trained to identify chosen objects (square, vertical cross, diagonal cross, diamond) that are presented in four segments of the input field. One example of setup and identification of input object is shown in fig 3.3. After this training the experiment consists of three basic steps.

1. First step of the experiment represents training of object that will later become the target of visual attention.

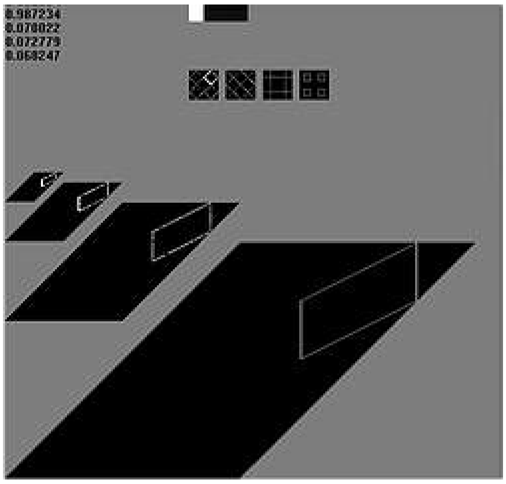

Fig. 3.4: Training of attention for object diamond

Approximately 10 adaptation cycles are needed to train the network to have an object as a target of it's attention. The adaptation goes on only in feedforward manner and synaptic weights are adapted using again Hebbian learning rule. This as all steps are analogical to real psychophysical experiments.

2. During second step of experiment are presented distractor objects and target object on input layer V1. The processing of this input information is then realized in feedforward manner from layer V1 to the AIT layer. All neurons in AIT layer are active, corresponding to the presented objects. Highest neural activity can be observed by neuron corresponding to object diamond.

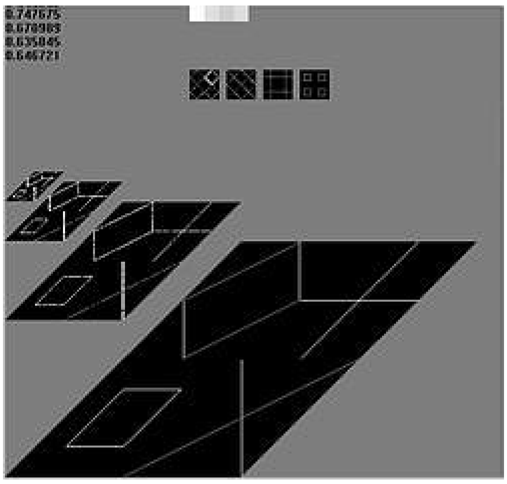

Fig. 3.5: Input field with distractor objects

3. Feedback activity in the network and position selection is the last step of experiment. This is the key process concerning visual attention problem. This whole process can be described as follows. At first a competition in layer AIT occurs. This competition is realized by strategy winner takes all (WTA – strategy). Neuron selected by this strategy is the winner, and should correspond to input target object. Consequently the feedback activity is run from this neuron to hierarchically lower layers. This means that the signal is propagated through synaptical connections running from the AIT layer to lower layers. In this case it is neural path AIT->PIT->V4->V2, so that the V2 layer is the output layer, where the retinotopical information about the position of target object is obtained. The selection of this information is done by the resonance interaction of feedforward (V1->AIT) and feedback (AIT->V2) signal. This means that the activity of neurons that where activated by feedforward signal and in next step also by feedback signal is strengthened and activity of other neurons is inhibited or just falls off. .

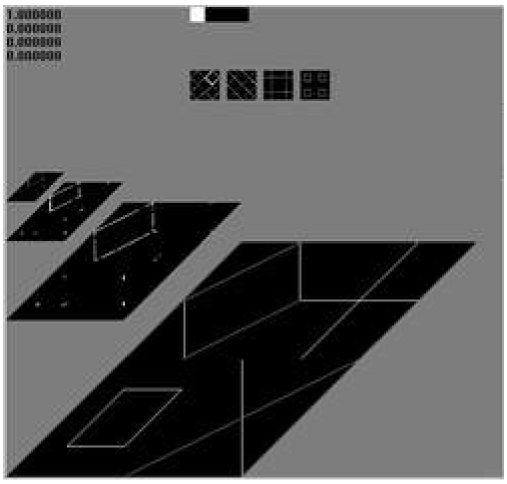

Fig. 3.6: In the layer V2 is selected position of target object

The figure shows as the position of target object (in our case it is diamond) is selected in the V2 layer. Layer V1 is visualized only to see the input patterns and that position of diamond object is correctly selected in layer V2..

4 Features of the model

The basic experiment realized by the proposed model that demonstrates an ability of the model to simulate real psychophysical experiment as described in [2]. After training the model is able to keep track of target object in changing environment by subsequent position changes of target and distractor objects. This feature can be used with great advance by application use of the model. Next very important feature of the model is it's ability to quickly adapt to different shapes of presented objects and is able to identify them. That brings another strong feature which is the ability of the model to deal with shape deformation of objects. This approximation of more complex or deformed shape of an object is done by approximation layers V2, V4 and PIT. In conclusion we can say that the model is capable of parallel shape and position invariant classification of input objects. In the final phase it can also keep track of target object. Those features can be used with great advance in technical assembly process.

Previous experiment shows tracking of only one object in four distinct quadrants. For real-world applications such four quadrants are not sufficient. This is showing another feature of the model and that is that only by correct selection of training patterns input layer can be generally divided into *n* distinct segments. This input layer can also be expanded and also other layers can be expanded in the same way. For better approximation of input objects layers can be also added. The elimination of segments in input layer can be achieved by implementation of modified learning rules based on short term memory of synaptic weights. The last but important feature of the model is that it can work very in real time.

5 Implementation of the model in automated assembly of products

A problem in automated manipulation and assembly is so called disorder of assembled parts, which expresses complexity by positional orientation. The degree of disorder is dependant on the number of possible positions the part can take in the manipulation or assembly process. Disorder in orthogonal coordinate grid is expressed by the number of values of coordinates and angles needed to set the position of certain axis of the part. The degree of disorder can be expressed by the next relation>

$$\mathbf{U = w * s * k} \tag{5.1}$$

Where: w – the count of rotation of a single angle needed for orientation (sphere w = 0, cylinder w = 2, common parts w = 3)

s – the count of unequal sides of the part lying in opposite direction

k – constant taking in account layout of surfaces (k = l – decomposition of surfaces with reference to line, k = 2 – surfaces decomposed in plane, k = 3 – surfaces decomposed in space)

The degree of disorder conforms to needed proceedings for orientation of the part to settle it into desired position. The most important is the position of the part that it steadily takes based upon it's geometry, centre of gravity and constant action force. Probability of natural orientation is defined as a ratio of favorable and all possible positions of the part.

For practical solutions of mentioned problems, classification systems are used. For example by flat parts we suspect that most of the non-rotational parts can be inscribed into tetragonal or triangular shapes with one to three planes of symmetry.

Complex systems for description of parts from the view of automated manipulation and orientation possibilities, describe their basic shape, substantial features and symetry. Special simplifications are used by this process: shell, degenerated shell, similarity and symetry. Those systems are very helpful by solving problems of automated manipulation and assembly. They allow the parts to be classified into groups that show up similar problems by automated feed and orientation in assembly. For similar manipulation and assembly problems are suitable flexible feeding and orientation systems that are able to manipulate with whole groups of parts. An advantage here is direct connection to automated manipulation and orientation systems (shell of the part – handling parts of industrial robot – elements of symetry – feeding vibrating and non-vibrating systems, etc.) [11]

Supply systems and subsystems in structures of assembly systems have significant rank, because they affect their technical, technological and economical characteristics. It comes of the fact that about 80% of time in assembly processes is used for manipulation and transport operations. Supply systems and subsystems provide many functions [11]:

- create operational supply of assembled part,
- orient assembled parts in space and time according to demands of assembly operations,
- in case of need supply separation of assembled parts from the stream, or their binding into a stream,
- semi-automatically, or automatically supply assembly units with assembled parts,
- realize supervisory and supervisory – blocking functions,
- affect the reliability of assembly systems.

Supply systems and subsystems that supply mentioned functions can have variable structure.

Technical complexity of classical supply systems and subsystems can be eliminated by flexible programmable automated systems. Accuracy and reliability of programmed orientation of parts is highly dependant on correct identification of shape and position that target part, or object takes. This comes from the need of provision of stable certainity of recognized positions of oriented object of one kind in one system without it's complicated converting. Programmable supply systems and subsystems have sensoric modules in their configuration, efficient mechanical units, control system and susceptible software equipment.

For designation of position that an object takes in space are important sensoric modules. They do not control just the correctness of position (correct, non-correct), but also orientation of the part in space, dimensions or other characteristics. The basic demand is high flexibility of recognition ability for different kind of objects.

Information about recognized object are processed in control system of the unit or at higher level of control of assembly system. Processed information are distributed as control information to executive units that execute corresponding functions. Important fact is that those units should be programmable. Their construction solution depends on the approach to realized activity (individual, continual). Functional activity is derived from rotational, direct, reversible or combined movements of realization units and elements

Control systems of programmable supply units and subsystems execute many functions:

- processing of information from sensoric units and modules,
- correct evaluation of position of a part and designation of sequence for executive units
- distribution of executive instructions to driving units,
- control and blocking of non-regular activity,
- synchronization and optimalization of activity according to the demands of assembly process.

Software equipment based on the use of cognitive model of visual attention characterizes a new approach to mentioned problems.

References

[1] Van der Velde, F., De Kamps, M.: From Knowing What to Knowing Where, JOCN Paper, 2001.

[2] Chellazi, L., Miller E. K., Duncan J.: A neural basis for visual search in inferior temporal cortex, Nature, 363, 345-347, 1993.

[3] Kandel., E. R., Schwartz J. H., Jessell., T. M.: Principles of neural science (third edition), Appleton and Lange press., 1991.

[4] De Kamps, M. : Using a recurent network to bind form, color and position into unified percept, Neurocomputing journal, 38-40., 2001.

[5] Suzuki, W., Saleem, K.S., Tanaka K.: Divergent backward projections from anterior inferotemporal cortex..., Neural computation 422: 206-208, 2000.

[6] Brunel, N.: Hebbian learning of context in recurrent neural networks, Neural computation 8: 1677 – 1710, 1999.

[7] Andoga, R. : Diplomová práca – Modelovanie objektovo založenej pozornosti s využitím umelých neurónových sietí, 2003.

[8] Andoga, R., Dobeš, M.: Objektovo orientovaná pozornosť a jej modelovanie s využitím neurónových sietí, Slovensko – České rozpravy o umelej inteligencii, 2003.

[9] Madarász, L., Líška, O., Kováč, J.: Computer Supported Modeling of Assembly Systems. In: International Regional DAAAM – CEEPUS Workshop on Intelligent Machines and Technologies for the 21st Century. 27-29, May 1999, Miskolc, Hungary, pp. 277 – 282.

[10] Madarász, L., Líška, O., Kováč, J.: Topology Generating of Cell Structure of Production Systems. INES 2001, Helsinki, Stockholm, September 16-18, 2001, Finland and Sweden, ISBN 952– 5–0689–x, pp. 363-366.

[11] Kováč, J., Svoboda, M., Líška, O.: Automatizovaná a pružná montáž. , © Vienala Košice, 2000, 200s , ISBN80-7099-504-1.