Implementation of positional object recognition in automated assembly

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

[1]Department of Cybernetics and Artificial Intelligence, Technical University of Košice
[2] Department of Management and Economics, Technical University of Košice
Letná 9, 042 00 Košice, Slovakia

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. One of such applications we discuss in the proposed article are automated assembly systems and subsystems. Those systems demand today high degree of adaptability and flexibility. Biological systems and their models already posses those features. Proposed model is an implementation of the mechanism of visual attention in computer simulated environment.

Keywords: *automated assembly, neural network, object recognition, robots, vision, cognitive modeling*

1 Introduction

Nowadays a great interest from manufacturers of machinery, electrotechnics and electronics is being put into assembly processes. Steep development of electrotechnics and electronics industry influences also development of automated assembly in a very favorable way. The development of electronics helps to increase the level of machine intelligence and to solve problems in automated assembly on this basis. On the other hand production and assembly of electrotechnical and electronic products demands better automation equipment and more efficient realization systems.

The interest of manufacturers, but also users leads to demands to lower the expenses needed for assembly and the demands also lead to increasing the productivity of realization works. As an analytical tool for evaluation of effects of these efforts, the time needed to produce the products is used. Modern automated assembly systems allow to considerably shorten the average production time, they allow to increase the production quality and provide other effects needed to increase the competitiveness.

Serial productions need to react dynamically to demands of users and customers. These demand acceleration of innovation cycles. Flexibility is becoming an important heading of development and modernization of assembly systems. Flexibility, which as a category of behavior impacts the thinking and activity of people and is transmitted to assembly technics and assembly systems. It's most important aspect is that flexibility allows the new technologies to efficiently adapt to demands for innovation mainly at production surfaces.

Development of assembly processes is also considerably affected by the development of robotics and artificial intelligence. According to the current prognosis the assembly can become a field with highest count of robots involved. These robotic systems have expelled manual work from assembly lines mainly in automobile industry.

High quality demands are employed in functions of assembly systems, one of the main ones is the failure-free operation of technical devices. It is essential to incorporate necessary systems of failure-free production at every workstation or in the whole production system. It is essential to improve the quality management of production and to eliminate defective products in nearly every operation. In this case there is no need of final control for the final product (nor the statistical).

Assembly technologies, that are characterized by high degree of automation, flexibility and reliability and that insure the high quality were developed and reworked on the basis of advanced research taking in account systematic approach, integration demands and high degree of knowledge from other interdisciplinary branches.

There is only one possibility how to fully supply and develop the processes of production and projecting of assembly systems and it is only with supplement of advanced methods and techniques having mostly simulation characteristics. Only with the use of simulation it is possible to examine the behavior of assembly systems, outputs, costs and other project attributes.

The essential step for increasing of the productivity in automated assembly is an analysis of information flows and their optimalization and modern technical interpretation. New principles and newly developed types of devices with higher degree of artificial intelligence are being applied. Also new methods of production lines leveling are being applied and developed. In flexible assembly the use of assembly robots is very essential together with an equipment of these. Modern assembly systems demand also new kinds of auxiliary devices that secure supplement for assembly units and devices. Attention is also greatly focused on application of modern systems of control, programming, inspection and identification of production environment..

2 Flexible Assembly Facilities

Except the basic assembly units, the programmable supply devices and units are important for flexible assembly facilities and subsystems. Important aspect, upon which the accuracy and reliability of programmed orientation depends is the designation of position that the oriented assembled unit takes in space. This comes from the need of securing stable certainty of position of oriented objects of different kind in one assembly unit without the need of its complex adjustment or reprogramming.

Programmable supply units and subsystems in include in their configuration:

- sensory modules,
- efficient mechanical units and components,
- control system and software.

For the designation of position that an object takes in space the sensory modules are important. Visual systems and tactile sensors are most commonly used. Compared to classical control units, the sensory modules execute broader functions. They do not control only the correct position (correct/incorrect), but also the space position of an assembled part, its dimensions and other characteristics. The main demand to this technology is high flexibility of the recognition ability to identify different kinds of objects.

Information about the object that is being identified is processed in the control system of the unit or on the higher levels of control of the assembly unit. Processed information is distributed as control information to realization units that execute appropriate functions. Important fact is that the units should be programmable or even adaptive. Their construction depends upon the method of realized activity (individual, uninterrupted). Functional activity is derived from rotational, direct, reversible or combined movements of realization units.

Control systems of programmable supply systems and subsystems carry out more functions:

- processing of information coming from sensory units and modules,
- correct evaluation of the position of assembled part and designation of sequence for realization units and components,
- distribution of realization instructions to motor units,
- supervision and blocking by non-regular operation,
- synchronization and optimalization of operation according to demands of the assembly process.

Assembly robots equipped with object recognition system are nowadays used in practical solutions. Solutions that are characterized with original approaches are also known today. Real-world systems employ today the progress achieved in hardware and software field of research. Flexible automated assembly demands low-cost programmable building elements.

Assembly robots make up only a part of all flexible assembly systems. Problems in implementation are not connected to the robot alone, but with its interconnection and reprogramming with other build-up elements. Problems that need to be solved are identified in the following areas:

- Selection of suitable product basis for realization in flexible automated assembly
- Recognition, presentation and assembly of non-oriented parts moving upon conveyor belt
- Proposal of suitable realization units of flexible assembly system

Situation when a robot is surrounded by containers of parts and equipped with special handling and assembly heads and controlled by complex hierarchical control systems causes that investments into the robot are often lower than investments into all peripheral devices.

A problem in setting up of the assembly systems is also the selection of suitable product basis. Certain products aren't always suitable for assembly in such systems. Parts of such products are often different in weight, dimensions and other characteristics. It is very difficult to achieve specified target with only one assembly device without its following enormous complication of construction and thus increasing its price.

Ideal case occurs when a robot is able to place and get randomly oriented object. Typical system uses for example a camera with high resolution with field of view set according to axial system of the robot, so it could reach the object and designate its position. In this process there is a need to perform calculations where complex transformation is needed for the movement of the robot from its standard coordinates position into position of randomly positioned object. System that perform functions of separation of unrecognized object, or systems that are able to identify shapes of mixed object of diverse kind that can even touch or cover each other. Principle of object recognition upon a conveyor belt is illustrated on fig. 2.1

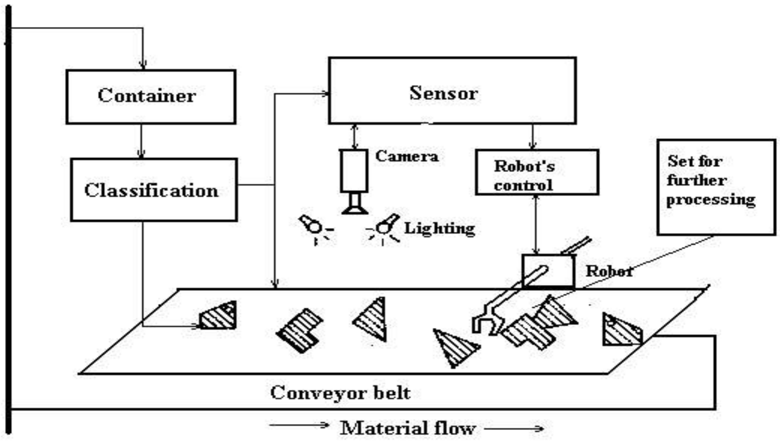

Figure 2.1: Principle of recognition of parts upon a conveyor belt

3 Visual object oriented attention

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.

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.

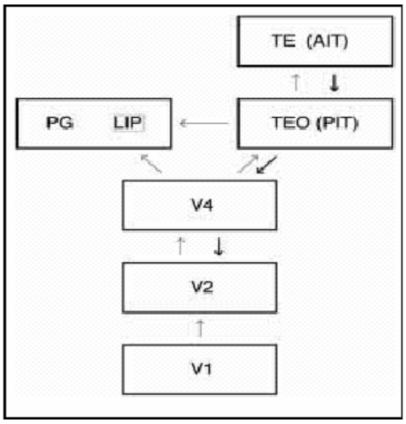

Figure 3.1: Scheme of the biological model of visual cortex

4 Proposed Model

4.1 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 3.1.)

The organization of layers in the neural network of proposed model is analogical to organization of layers in visual cortex (fig 3.1). 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 4.1)

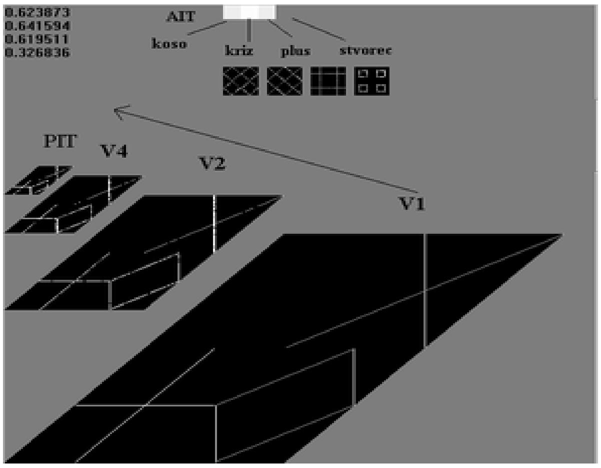

Figure 4.1: Visualization of the proposed model

Figure 4.1 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.

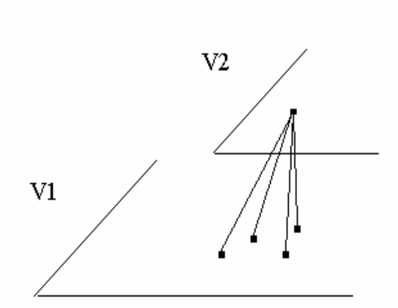

Figure 4.2: Schematic view of the layers

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>.

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.

For network weight adaptation in feedforward manner Hebbian learning rule is used. 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 behavior 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 synapses. 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.

4.2 Operation

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.

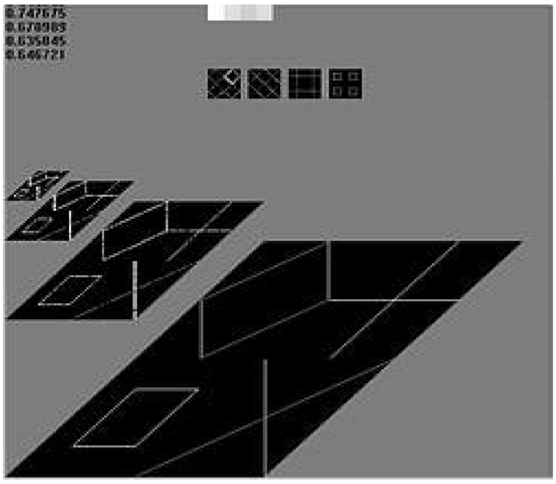

Figure 4.3: Four objects in layer V1

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 4.3. 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.

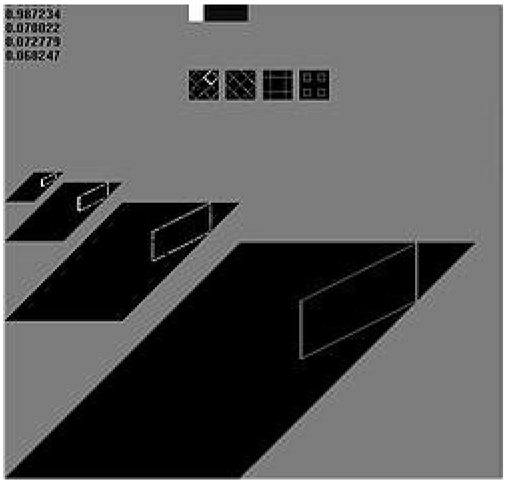

Figure 4.4: Training of attention for object diamond

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.

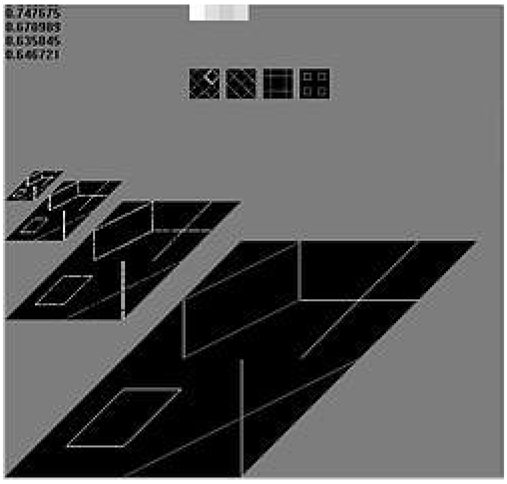

Figure 4.5: Input field with distractors

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 synaptic 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.

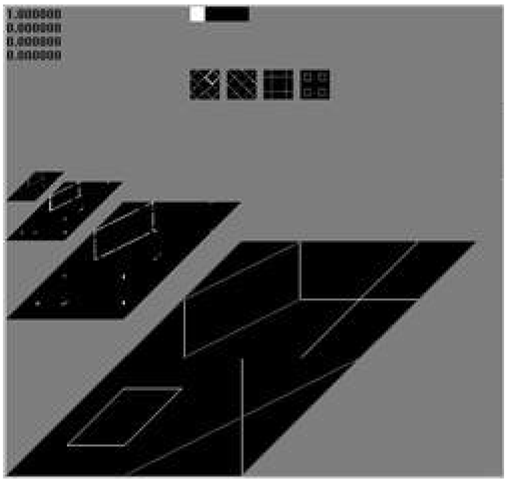

Figure 4.6: In the layer V2 is selected position of target object

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.

5 Proposed Model and automated assembly

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.

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.