2006 International Joint Conference on Neural Networks
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

A Visual Perceiving and Eyeball-Motion Controlling Neural Network for Object Searching and Locating

Jun Miao, Xilin Chen, Wen Gao, and Yiqiang Chen

Abstract—**This paper proposes a visual cognitive neural network for automatic object searching and locating. The model consists of two sub-networks. One is a visual perceiving network, which simulates human eyes to input image signals and recognize an object's direction and distance in terms of a high-level perceiving neuron's maximum response. The other one is an eyeball-motion controlling network, which simulates that human brain's high-level perceiving neurons transfer their responses to eyeball-motion controlling muscle cells to change eye's gaze to the position of the object that the perceiving system is attentive to or interested in. The system is applied to human face features searching and experiments show a promising result.**

I. INTRODUCTION

Human and animals' perception and cognition have been rather difficult problems, among which visual perception and cognition are focuses and fundamental fields. One of the main questions is that people don't know how all the visual neurons work together though a lot of knowledge have been obtained on a single neuron or a small group of neurons through physiological and anatomical experiments. Scientist from psychology, neural physiology, computer science, information science, mathematics and philosophy are experiencing their excitation and exhaustion in exploration of the natural principles of human and animals' vision.

During the past half-century, many theories and hypotheses are suggested for visual perception. One of the most well known models is Marr's vision computational theory[1], whose framework consists of three representations: primal sketch, 2.5D sketch and 3D model, which understands vision as an information processing task that converts a numerical image representation into a symbolic shape-oriented representation[2]. However, this strict hierarchical architecture makes learning difficult and "only toy problems (block worlds, Lambertian surfaces, smooth contours, controlled illumination, and the like) can be successfully solved"[3]. Different from Marr, Fukushima built the Cognitron and Neocognitron models for visual pattern recognition[4-5], which simulate receptive field characteristics of animal's visual simple and complex cells and achieved significant success.

From middle 1980s to early 1990s, Carpenter and Grossberg founded the Adaptive Resonance Theories(ART)[6-8]-a group of new cognition theories and related visual perception architectures[9-10]. ART theories and models are incorporated with quite a few elements from neural biology and psychology, which has produced a strong impaction on the cognitive and recognition researches. In order to avoiding the "curse of interconnecting wires"[17] leaded by full connectivity of Hopfield networks, Chua and Yang proposed Cellular Neural Network (CNN)[18-19] to simulates cellular structure in local excitatory connections between network units, which can be used for early visual processing.

In last decade, there came some new models, theories and relative architectures on perception and cognition. Different from CNN, used for midlevel visual perception, LEGION[20-21] is a neural network of oscillators with local excitatory connections between oscillators and global inhibition via a global inhibitor, which was proposed by Terman and Wang to address the binding problem (how the responses of local feature-responding neurons in different areas of the brain are bound together to form global percepts). As for high-level visual perception, Riesenhuber and Poggio proposed a quantitative model to explain aspects of higher-level visual processing such as object recognition[14]. At the aspect of cognitive theories and architectures, one of the representatives is J. Anderson's "Atomic Components of Thought"(ACT) and ACT-R theories[11-12], which simulate human cognition and can be used to understand how people organize knowledge and produce intelligent behavior. The other one is EPIC[13], a cognitive architecture proposed by Kieras and Meyer, which include quite comprehensive simulations for human's auditory, visual, tactile perception and ocular, vocal, manual motor controlling.

Except for ART, ACT-R and EPIC, most of the above models, when used in visual perception, seldom concern active vision functions, which seem indispensable for human being's visual cognition. For example, due to lacking context-reasoning ability on objects' positions and scales, when used for object searching and locating, perceiving systems cannot make decision where to search and has to traverse a series of shrunk images in a blind and time-consuming mode. Quite different from it, human can

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, No.6047304311 and No.60303018, "100 Talents Program" of CAS, Shanghai Municipal Science and Technology Committee (No.03DZ15013), the program for New Century Excellent Talents in University (NCET-04-0320), Beijing Science & Technology New Star Plan, Open Funding of Multimedia and intelligent software lab of Beijing University of Technology and ISVISION Technologies Co., Ltd.

J. Miao is with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China (phone: +86-10-58858300-313; fax: +86-10-58858300-399; e-mail: jmiao@ict.ac.cn).

X. Chen, W. Gao and Y. Chen are with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China (e-mail: {xlchen, wgao, yqchen}@ict.ac.cn).

0-7803-9490-9/06/$20.00/©2006 IEEE

repeatedly reason and control his eyeballs to rotate for changing gaze quickly to acquire richer and richer visual information to approach and locate interested objects. Its neural mechanism[22] is that in the brain superior colliculus and relative oculomotor nucleus control eye movement due to low-level perception of salient features or high-level cognition of objects spatial relationship. Some work has been done to simulate oculomotor mechanism for "where-what" recognition [23] and autonomous robot vision [24]. Different from their functional simulation approaches, according to the cognitive physiological principles and corresponding neuron inter-connecting structure, this paper proposes a visual cognitive neural network model consisting of visual perceiving and eyeball-motion controlling sub-networks for automatic and active visual object searching and locating.

In the following paragraphs, firstly the whole visual cognitive neural network is suggested in section II. Then the two sub-networks are introduced respectively in section IIA and IIB. Section III gives the detailed cognizing and learning mechanism. In section IV, experiments on human facial features searching and results are discussed. Finally, conclusion and future directions are given in last section.

II. A VISUAL COGNITIVE NEURAL NETWORK

Fig. 1 and 2 illustrate a cognitive neural network architecture and its system implementation for visual object searching and recognition.

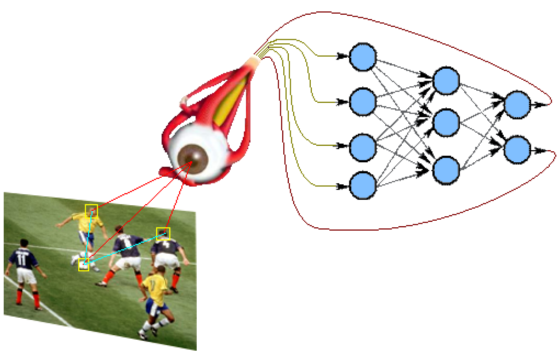

Fig. 1. Visual cognitive neural network architecture

The neural network consists of two sub-networks. One is a visual perceiving neural network, which simulates human eyes to input original image signals and recognize an object's direction and distance in terms of a cognizing neuron's maximum response. The other one is a motion controlling neural network, which simulates that human brain's cognizing neurons transfer their responses to eyeball-motion controlling muscle cells to change eye's gaze to the position of the object that the perceiving system is attentive to or interested in.

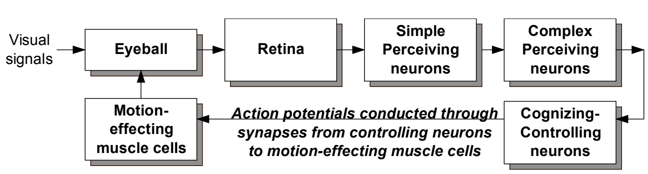

Fig. 2. Visual cognitive system

The two networks naturally are incorporated into an autonomic neural network and cooperate to perceive and move gaze in a repeated mode from a global low resolution to a local high resolution until the system acquire a full cognition of the object.

A. Visual perceiving neural network

This network consists of four parts: retina, simple perceiving neurons, complex perceiving neurons and cognizing-controlling neurons(with reference to Fig. 2). The retina receives original visual signals and transmits them to the perceiving neurons. The perceiving neurons simulate the simple and complex cells in visual cortex, which extract brightness, contrast and some shift-invariant features from their receptive fields on the retina. Fig.4 and 5 illustrates some examples of such receptive fields and their corresponding feature extracting patterns. The cognizing-controlling neurons receive the responses from complex perceiving neurons and recognize the visual input signals.

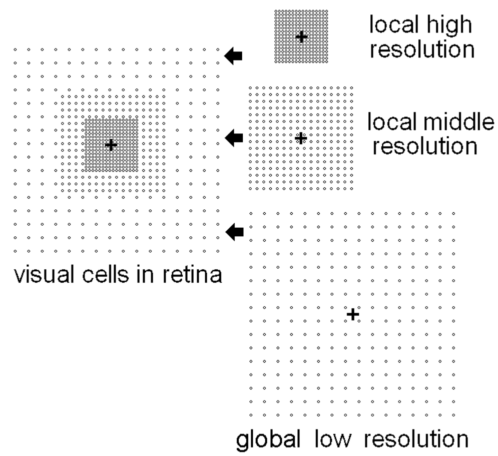

Fig. 3. Visual fields in different scales and their corresponding visual signal receiving cells' resolutions and distributions in retina, where the central cross indicates the position of gaze point

As we know well, human and animals' retina has a significant characteristic: its visual signal receiving cells has a high-density distribution in the central region and a low-density distribution surrounding the central part. An explicit artificial retina has been introduced in reference 15. Here we simulates retina with an overlapping of a group of visual cells arrays or visual fields from global low resolution to local high resolution(with reference to Fig. 3).

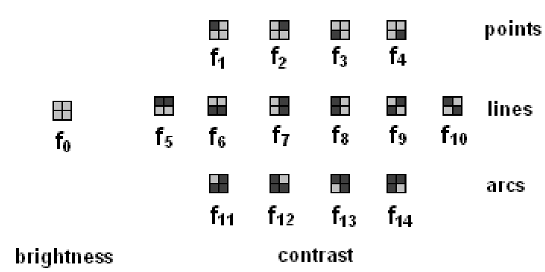

Fig. 4. Simple cells' receptive fields(size=2x2=4 pixels) and corresponding brightness and contrast feature extracting patterns

In Fig. 4, two types of simple cells' feature extracting patterns are given: brightness and contrast. There is one feature pattern for brightness ($\mathbf{f}_0$) and there are 14 patterns for contrast ($\mathbf{f}_1 \sim \mathbf{f}_{14}$) respectively. Among them, the 14 contrast features are actually representing 3 kinds of geometrical features, which are points, line segments and arcs with different positions or orientations, which simulate the simple cells' receptive fields(size=2x2=4 pixels) characteristics described in reference[16].

A gray small box in a feature pattern in Fig. 4 represents an excitatory input with a positive weight and a black box represents an inhibitory input with a negative weight. Thus a feature pattern could be represented by a vector with a group of weights(here are 4 weights). Generally, all weights in each feature vector are normalized to length 1 for unified feature response/similarity computation and comparison.

Let vector $\mathbf{x}_i=(x_{i1}, x_{i2}, x_{i3}, x_{i4})$ represent the i-th local visual signal input pattern and vector $\mathbf{f}_{ij}=(a_{j1}, a_{j2}, a_{j3}, a_{j4})$ represent the j-th feature extracting or perceiving pattern at current location i, then the simple perceiving neuron's feature response $r_{ij}=f_{ij}(\mathbf{x}_i)$, which feature $\mathbf{f}_{ij}$ responds to input $\mathbf{x}_i$, or the similarity $s(\mathbf{x}_i, \mathbf{f}_{ij})$ between $\mathbf{x}_i$ and $\mathbf{f}_{ij}$, could be obtained by orthogonal projection or inner product computation:

$$r_{ij}= f_{ij}(\mathbf{x}_i)=s(\mathbf{x}_i, \mathbf{f}_{ij}) =<\mathbf{x}_i, \mathbf{f}_{ij}>=a_{j1}x_{i1}+a_{j2}x_{i2}+a_{j3}x_{i3}+a_{j4}x_{i4}$$

Generally, a neuron is firing only if its response is larger than a threshold, for example, threshold=0. Thus the real response of a neuron is:

$$r_{ij} = f_{ij}(\vec{x}_i) = \begin{cases} <\vec{x}_i, \vec{f}_{ij}> & \text{if} <\vec{x}_i, \vec{f}_{ij}> \; > 0 \\ 0 & \text{otherwise} \end{cases}$$

Mathematically these feature patterns constitute a set of non-orthogonal bases in local feature vector space for describing local visual input pattern $\mathbf{x}_i$. For example, with reference to Fig. 4, $\mathbf{f}_{i0}=(1,1,1,1)/\sqrt{4}$, $\mathbf{f}_{i1}=(-3,1,1,1)/\sqrt{12}$, $\mathbf{f}_{i5}=(-1,-1,1,1)/\sqrt{4}$, $\mathbf{f}_{i11}=(3,-1,-1,-1)/\sqrt{12}$. Among these feature bases, the brightness feature vector $\mathbf{f}_{i0}$ is orthogonal to any one of the contrast feature vectors $\mathbf{f}_{i1} \sim \mathbf{f}_{i14}$.

According to Hubel and Wisel's outstanding work on simple and complex cells' receptive field characteristics [16], we can simulate complex perceiving cells' shift-invariant feature responding mechanism and explain their nature with mathematical principles.

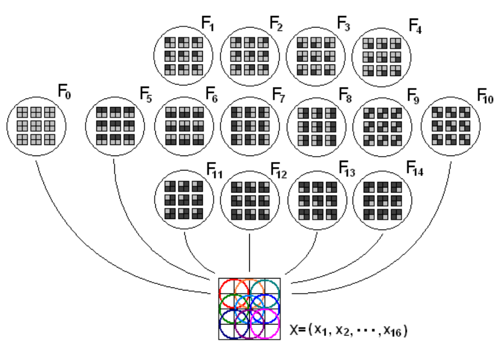

Fig. 5. Complex cells' receptive fields(size=4x4=16 pixels) and corresponding shift-invariant feature extracting patterns

Fig. 5 shows the complex perceiving neurons' receptive fields and their shift-invariant feature extracting patterns. Let vector $\mathbf{X}_i=(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{i9})=(x_{i1}, x_{i2}, ..., x_{i16})$ represent a larger neighborhood i which is composed of 9 adjacent smaller neighborhoods $\mathbf{x}_{ik}(k=1\sim9)$, then the j-th (j=0~14) kind of complex cells' perceiving responses $R_{ij}=F_{ij}(\mathbf{X}_i)$, which feature $\mathbf{F}_{ij}$ responding to neighborhood input $\mathbf{X}_i$, are:

$$R_{i0} = F_{i0}(\vec{X}_i) = \sum_{k=1}^{9} f_{i0}(\vec{x}_{ik})$$

$$R_{i1} = F_{i1}(\vec{X}_i) = \sum_{k=1}^{9} f_{i1}(\vec{x}_{ik})$$

$$\ldots$$

$$R_{ij} = F_{ij}(\vec{X}_i) = \sum_{k=1}^{9} f_{ij}(\vec{x}_{ik})$$

$$\ldots$$

$$R_{i14} = F_{i14}(\vec{X}_i) = \sum_{k=1}^{9} f_{i14}(\vec{x}_{ik})$$

Generally, as a result of irrelative properties between each two features $\mathbf{F}_{ij}$ and $\mathbf{F}_{ik}$ (j,k=0~14, j≠k), the neighborhood visual pattern $\mathbf{X}_i$ can be approximately reconstructed by a sum of weighted $\mathbf{F}_{ij}$(j=0~14), i.e.:

$$\vec{X}_i \approx \sum_{j=0}^{14} c_{ij}\vec{F}_{ij}$$

where c_{ij} is closely relative to the connecting strength (in terms of weight w_{ij}) of a synapse between a complex perceiving neuron and a cognizing neuron. In other words, local neighborhood visual pattern $\mathbf{X}_i$ can be represented by a group of reconstructed coefficients c_{ij} or a group of synapse strengths w_{ij} (j=0~14).

From the point of view of biological visual system, perceptive feature competition is necessary for human and animals to remove noise and insignificant features for efficient object searching and recognition. From the point of view of feature reduction in pattern recognition, the first m features ($F'_{i1}(\mathbf{X}_i)$, $F'_{i2}(\mathbf{X}_i)$, ..., $F'_{im}(\mathbf{X}_i)$) that have the largest responses(R'_{i1}, R'_{i2}, ..., R'_{im}) to the local neighborhood input pattern $\mathbf{X}_i$ could approximately describe or represent the $\mathbf{X}_i$ at the cost of minimum reconstruction error. Generally, m is less than the pixel number or dimension of the local visual input pattern $\mathbf{X}_i$, i.e.:

$$\vec{X}_i \approx \sum_{j=0}^{m} c_{ij}\vec{F}'_{ij}$$

In our implementation, as illustrated in Fig. 5, the size of the receptive fields of complex perceiving neurons is 4x4=16 pixels when an image is used as a visual signal input. Thus the dimension of local neighborhood visual pattern $\mathbf{X}_i$ is 16. For the purpose of reducing features, for example, m could be taken with 10 that is less than the number of pixels of the local visual pattern $\mathbf{X}_i$.

Fig. 6 shows the architecture of a visual perceiving neural network in which the highest-level perceiving neurons, as roll of cognizing-controlling neurons, receive synaptic inputs (weighted with strengths w_{ij}) from the lower perceiving neurons (with response R_{ij}) that perceive brightness and geometrical contrast shift-invariant features in their receptive fields on retina. So the cognizing neuron's response R=F(**X**), for the global visual signal $\mathbf{X}=(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$ which is composed of the local neighborhood visual pattern $\mathbf{X}_i(i=1\sim N)$, is:

$$R=F(\mathbf{X})=F((\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)) = \sum_{i=1}^{N}\sum_{j=1}^{m} w_{ij} F'_{ij}(\vec{X}_i) = \sum_{i=1}^{N}\sum_{j=1}^{m} w_{ij} R'_{ij}$$

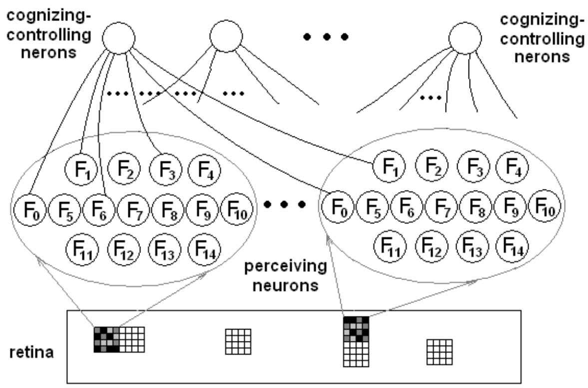

Fig. 6. Visual perceiving neural network

According to universally competitive interaction facts found in the biological perception and cognition systems, the cognizing neuron that has the largest response will win the competition among all the responding cognizing neurons, which means that the winner cognizing neuron represents the visual input signal and recognize an object by behaving as a motion controlling neuron to transfer its response to the motion-effecting muscle cells that directly rotate the eyeball or gaze to the object.

B. *Eyeball-motion controlling neural network*

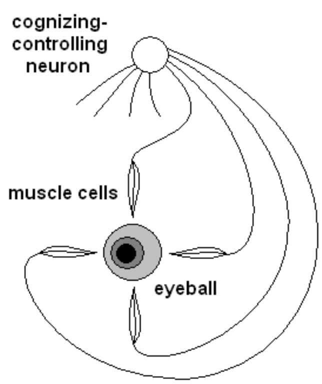

Fig. 7. Eyeball-motion controlling neural network

The eyeball-motion controlling neural network consists of two parts: cognizing-controlling neurons and motion-effecting muscle cells(Fig. 2 and 7). The cognizing-controlling neurons are actually the high-level perceiving neurons that recognize the positions of the attentive visual object. In order to acquire richer information about the object for enhanced recognition, from a global low resolution to a local high resolution, such high-level perceiving neurons transform their recognition of the object into gaze's movement by conducting their responses(R_i) through synapses (weighted with strengths w_{ij}) to four muscle cells that directly rotate eyeball by muscle tightening(with response R_j that is proportional to distances from gaze point to object position in horizontal or vertical directions respectively).

III. Cognizing and Learning Mechanism

The system's cognition include a series of perceiving and controlling procedures, which begin with an initial gaze point and end with a final gaze point. Perception is achieved by a cognizing neuron responding largest among all the cognizing neurons and becoming a winner through competitive interaction. Controlling is achieved by a winner cognizing neuron transferring its response to eyeball-movement controlling muscle cells to change eye's gaze to the position of the object that the perceiving system is attentive to or interested in.

The system's learning experiences are preserved in the system's memory, which is represented by synapses' connecting strengths or weights. Hebbian rule is the fundamental learning rule, i.e., $\Delta w_{ij}=\alpha R_i R_j$, where w_{ij} is connecting synapse's strength; α is learning rate; R_i and R_j are responses of two cells that are connected with the synapse. Learning mechanism is as followed:

1. Given a visual field range with the corresponding resolution of visual signal receiving cells in retina and an gaze position, perceive the attentive object's position (x-distance and y-distance);
2. If perception result is not correct, generate a new cognizing-controlling neuron(let response R=1); else go to 4;
3. Compute synapses' strengths between the new cognizing neuron and lower perceiving neurons and that between the new controlling neuron and motion-effecting muscle cells using Hebbian rule $w_{ij}=\alpha R_i R_j$;
4. Move current gaze point to the position of the object in the current visual field and change the visual field range and the corresponding visual signal receiving cells' resolution to a smaller and a higher ones respectively;
5. Go to 1, until all visual field ranges with corresponding resolutions and all given initial gaze points are learned.

IV. Experiments

The neural network is applied to automatic human facial features searching, such as two eye centers searching and locating in still images.

A. *Design of system*

With reference to Fig. 3 and 8, the retina is designed with a group of 5 different scale visual field(16x16, 32x32, 64x64, 128x128 and 256x256 pixels) layers of visual signal receiving cells at same 16x16 cell arrays, which simulates human and animals' visual cells' distribution. Its physiological meaning is that human and animals look at or watch objects with their visual receiving cells at fitted scale and in specific visual field other than with that at all scales and in whole visual field. Thus human and animals can rapidly react to environmental variations and efficiently memorize the most significant features.

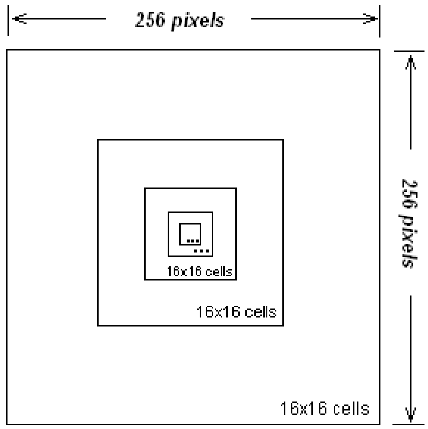

Fig. 8. Retina composed of 5 overlapped different scale visual field layers of visual signal receiving cells at same 16x16 cell arrays

With reference to Fig. 4, 5 and 6, the receptive filed sizes of simple and complex cells are 2x2 and 4x4 pixels respectively, and there are 15 kinds of features for both two types of perceiving neurons. Thus there are totally $15\times5\times[((16/2)\times2-1)]^2=16875$ simple perceiving cells and $15\times5\times[((16/4)\times2-1)]^2=3675$ complex perceiving cells, in which only $3675\times(10/15)=2450$ complex cells(the first m largest responding cells, m=10, see section IIA) win through competitive interactions and transfer their responses R_{ij} through w_{ij}-weighted synapse to the highest-level perceiving cells, i.e., cognizing-controlling neurons. In our experiment the learning rate α is 0.01 when using Hebbian rule.

B. Experiments

Two experiments are done on the face database of the University of Bern, which has total 300 images(320x214 pixels) with 30 people(ten images each person) at different poses.

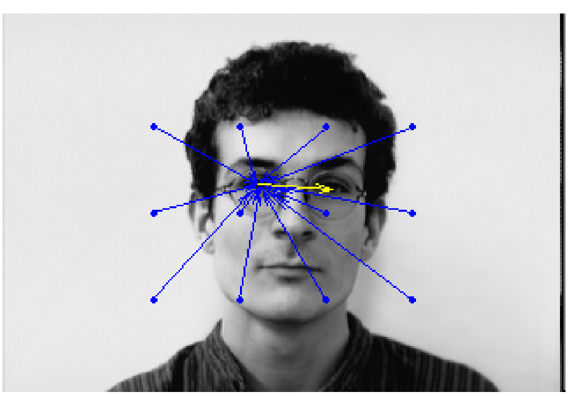

Fig. 9. Training for sequentially searching two eye centers from a group of initial gaze points in even distribution

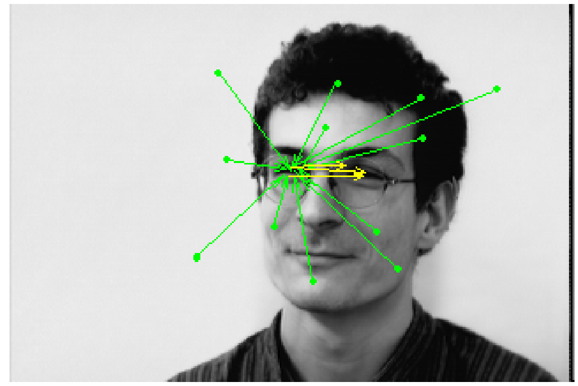

Fig. 10. Testing for sequentially searching two eye centers from a group of initial gaze points in random distribution

As illustrated in Fig. 9 and 10, training is with a group of initial gaze points in even distribution while testing is with a group of initial gaze points in random distribution. Given a initial gaze position, the system is trained or tested to search and locate the left eye center first and then the right eye center on the basis of left eye center's searching results.

In the first experiment, 30 images of 30 people (one frontal image each person) are learned with 368 initial gaze positions on each image, and the rest of 270 images are tested at 48 random initial gaze positions on each image. The average searching error is 5.51 pixels for left eye centers and 7.44 pixels for right eye centers.

In the second experiment, 90 images of 9 people (10 images each one) are learned with 1944 initial gaze positions on each image, and the rest of 210 images are tested at 48 random initial gaze positions on each image. The searching error is 8.82 pixels for left eye centers and 11.10 pixels for right eye centers.

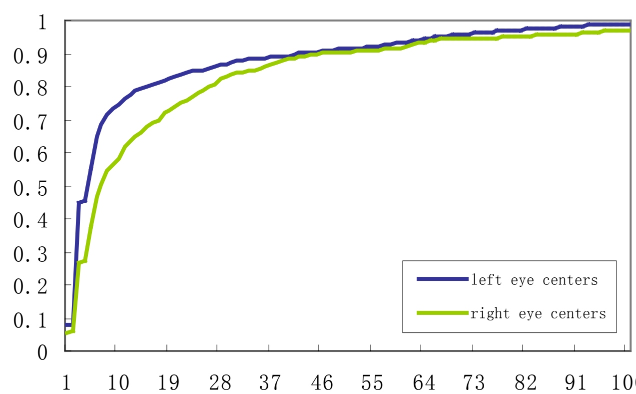

Fig. 11. Test results experiment 1

Fig. 11 and 12 show the statistical results for feature searching, in which the horizontal axis represents the percentage of the distance between searching results and ground truth over the distance between two real eye centers. The vertical axis represents the accumulative correct searching/localization rate. In our experiments, right eye center searching is designed to follow the searching of left eye centers. As a result of such dependence, the performance for left eye centers is better than that for right eye centers. The experiments also show that that the generalizing ability in experiment 1, in which training and testing faces are in different poses from same persons, are a bit better than that in experiment 2, in which training and testing faces are from different persons.

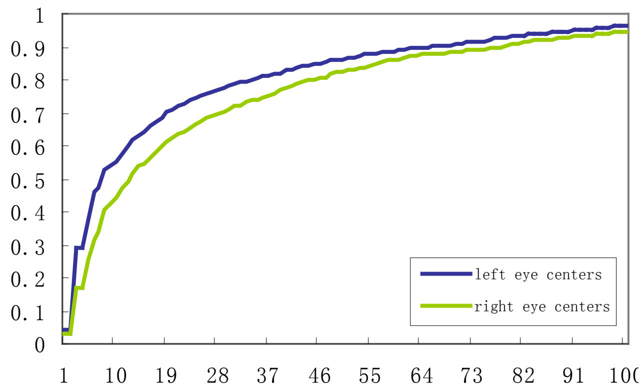

Fig. 12. Test results in experiment 2

V. CONCLUSION

This paper proposed a visual cognitive neural network consisting of perceiving and controlling sub-networks for automatic and active object searching and recognition. The autonomic system is applied to human face features searching and experiments show a promising result. As a result of temporarily lacking rotation-invariant and scale-invariant feature representations and hierarchical neighborhood perception, the cognizing system's generalizing ability needs to be enhanced, which is the next step that we would improve the system with more robust feature representations and more compact cognitive neural network architecture.

ACKNOWLEDGMENT

The authors would like to thank Mr. Qingxu Li, Wenchao Zhang and Bo Cao for technical helps. The authors thank anonymous referees for many valuable comments.

REFERENCES

[1] D. Marr, *Vision: a computational investigation into the human representation and processing of visual information*, W.H. Freeman and Company, San Francisco, 1982.

[2] J. Murdock, Summary of D. Marr 1982 on Vision, in J. Davies (ed.) *Cognitive Science Summaries*, 1997. Available: http://www.cc.gatech.edu/~jimmyd/summaries/marr1982.html.

[3] C.Fermuller and Y. Aloimonos, "Vision and action", *Image and Vision Computing*, vol. 13(10), pp.725-744, 1995.

[4] K. Fukushima, "Cognitron: a self-organizing multi-layered neural networks", *Biological Cybernetics*, vol. 20, pp.121-136, 1975.

[5] K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition", *IEEE-SMC*, vol. 13(5), pp.825-834, 1983.

[6] G. Carpenter and S. Grossberg, "A massively parallel neural network for a self-organizing neural pattern recognition machine", *CVGIP*, vol. 37, pp.54-115, 1987.

[7] G. Carpenter and S. Grossberg, "ART-2: Self-organization of stable category recognition codes for analog input patterns", *Applied Optics*, vol. 26, pp.4919-4930, 1987.

[8] G. Carpenter and S. Grossberg, "ART-3: Hierarchical searching using chemical transmitters in self-organizing pattern recognition architecture ", *Neural Networks*, vol. 3, pp.129-152, 1990.

[9] S. Grossberg, E. Mingolla and D. Todorovic, "A neural network architecture for preattentive vision", *IEEE-BME*, vol. 36, pp.65-84, 1989.

[10] S. Grossberg and L. Wyse, "A neural network architecture for figure-ground separation of connected scenic figures", *Neural Networks*, vol. 4, pp.723-742, 1991.

[11] J. Anderson, and C. Lebiere, *The atomic components of thought*, Mahwah, NJ: Erlbaum, 1998.

[12] J. Anderson, D. Bothell, M. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind", *Psychological Review* vol. 111(4), pp.1036-1060, 2004.

[13] D. Kieras and D. Meyer, "An overview of the EPIC architecture for cognition and performance with application to human-computer interaction", *Human-Computer Interaction*, vol. 12, pp.391-438, 1997.

[14] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex", *Nature Neuroscience*, vol. 2(11), pp.1019-1025, 1999.

[15] S. Balasuriya and P. Siebert, "A biologically inspired computational vision front-end based on a self-organised pseudo-randomly Tessellated Artificial Retina", *Proceedings of International Joint Conference on Neural Networks*, Montreal, Canada, July, 2005, pp. 3069-3074

[16] D. Hubel and T. Wisel, "Receptive fields, binocular interaction and functional architecture in cat's visual cortex", *Journal of Physiology*, vol. 160, pp. 106-154, 1962.

[17] L.O. Chua, *CNN: A paradigm for complexity*. Singapore: World Scientific, 1998

[18] L.O. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Trans. Cir. Syst.*, vol. 35, pp. 1257-1272, 1988.

[19] L.O. Chua and L. Yang, "Cellular neural networks: Applications," *IEEE Trans. Cir. Syst.*, vol. 35, pp. 1273-1290, 1988.

[20] D. Terman and D.L. Wang, "Global competition and local cooperation in a network of neural oscillators," *Physica D*, vol. 81, pp. 148-176, 1995.

[21] D.L. Wang and D. Terman, "Locally excitatory globally inhibitory oscillator networks," *IEEE Trans. Neural Net.*, vol. 6(1), pp. 283-286, 1995.

[22] G. M. Shepherd, *Neurobiology*, 2nd edition. Oxford University Press, 1988

[23] I.A. Rybak, V.I. Gusakova, A.V. Golovan, L.N. Podladchikova, and N.A. Shevtsova, "A model of attention-guided visual perception and recognition", *Vision Research*, vol.8, pp. 2387–2400, 1998

[24] T. Shibata, S. Vijayakumar, J. Conradt, S. Schaal, Humanoid oculomotor control based on concepts of computational neuroscience, Humanoids2001, *Proc. The Second IEEE-RAS International Conference on Humanoid Robots*, 2001