A Visual Search System based on Multi-scale Population Cell Coding

Fang Fang, Jun Miao, Laiyun Qing and Wen Gao

Abstract—**Gaze movement plays an important role in human visual search system. How to simulate such a system to efficiently encode and decode gaze movement for target searching is a meaningful issue. There are two key points that should be addressed for this issue. First, eye movement is affected by the visual context that includes more than one object in images. It is important to study how to encode the spatial relationship between the target and the other objects, and decode the spatial relationship to drive the gaze moving to the target. Second, the human retina has a non-uniform distribution of light sensing neurons, which can be viewed as a composition of neuron arrays in different spatial resolutions. When a system searches a target, how many scales of visual fields should be involved to sense the image stimulus? In this paper, we propose a visual search system using the population cell coding mechanism and the multi-scale visual field as sensing input. As an example, the system is applied to human eye center searching. An experimental comparison of a *Full-scale* visual field coding system and a *Gradual-scale* coding system is carried out. The experiment results show the *Gradual-scale* coding system performs better than *Full-scale* coding system for target searching.**

I. Introduction

Eye gaze movement plays an important role in human vision information acquiring and object searching system. Much research work simulates the gaze variation in bottom-up and top-down modes [2-6]. For example, when one searches specific categories of objects, i.e., pedestrians or vehicles, from images without the knowledge of object scales, orientations and positions, he may moves his gaze to places according to the posterior probability related to the context between objects and environmental features[7,8], which is known as top-down attention or task-driven object searching. When one recognize the located object, he may change his view point from one salient region to another [9,10] to extract key features according to the attraction strength or according to a saliency map[11], which is known as bottom-up attention or feature-driven visual searching.

This research is supported in part by Natural Science Foundation of China (Nos.60673091 ,60702031 and 60970087), Hi-Tech Research and Development Program of China (No.2006AA01Z122), Beijing Natural Science Foundation (Nos.4072023 and 4102013), Beijing Municipal Education Committee (No.KM200610005012), Beijing Municipal Foundation for Excellent Talents(No.20061D0501500211) and National Basic Research Program of China (2009CB320902).

F. Fang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: ffang@jdl.ac.cn).

J. Miao is with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China (e-mail: jmiao@ict.ac.cn).

L. Qing is with the School of Information Science and Engineering, Graduate University of the Chinese Academy of Sciences, Beijing 100049, China (e-mail: lyqing@gucas.ac.cn).

W. Gao is with the Institute for Digital Media, Peking University, Beijing 100871, China (e-mail: wgao@pku.edu.cn).

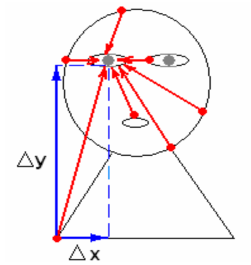


Fig.1. An illustration of visual context: the target (left eye) and the environmental object (left shoulder) plus their spatial relationship (Δx, Δy).

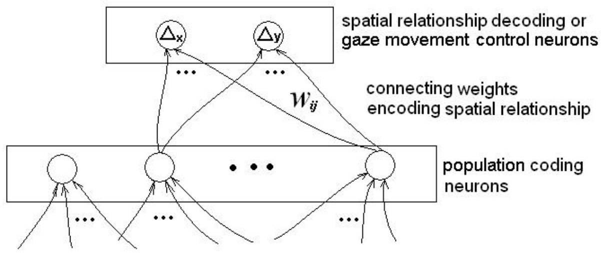


Fig.2. Visual context coding: encoding and decoding spatial relationship for gaze movement control through population cell coding.

Either the top-down or bottom-up methods, if they adopt the decision principle by the largest saliency or the largest probability, is similar to winner-take-all or single-cell-coding mechanism for only one largest response is used to make decision. Single-cell-coding means using one cell (grandmother cell) or one response to represent one object. However, eye movement is affected by the visual context that includes more than one object in images, especially in the case of target searching. From the principle of informatics, single-cell-coding could save large coding quantity at the risk of losing accuracy for recognition or behavior control, while the population coding maintains more stabile and higher accurate at the cost of more coding quantity. In top-down visual object search systems, the context[7, 12-14]between environmental features and targets is usually to be learned or coded for future prediction of the positions of the targets. Visual context is related to the spatial relationship in terms of horizontal and vertical distances (Δx, Δy) between two centers of related objects, as shown in Fig.1. Much more physiological experiments prove that population coding mechanism is widely used in the human brain, vision and movement control system [1, 17]. As shown in Fig.2, we can use population coding neurons to encode and decode the spatial relationship to control the gaze movement in terms of horizontal and vertical shift distances (Δx, Δy).

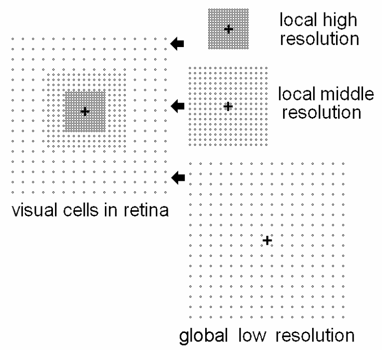


Fig.3. Visual fields in different scales. The corresponding resolutions or distributions of the visual signal receiving cells in retina are simulated, where the central crosses indicate the centers of the visual fields or the positions of gaze points.

The human retina has a non-uniform distribution of light sensing neurons, which can be viewed as a composition of neuron arrays in different spatial resolutions (see Fig.3). When a system searches a target, how many scales of visual fields should be used to sense the image stimulus? To solve the problem, a general principle of high accuracy of target locating with low encoding quantity of training data should be considered. In this paper, we propose a visual search system using the population cell coding mechanism and the multi-scale visual field as sensing input.

An experimental comparison of a *Full-scale* visual field coding system and a *Gradual-scale* coding system is carried out. The experiment results show the *Gradual-scale* coding system performs better than *Full-scale* coding system for target searching.

The following paragraphs are arranged in five respects: (1) Construction of visual search system; (2) Features designed for coding visual context; (3) Multi-scale population cell encoding and decoding gaze movement; (4) Experimental performance comparison and (5) Discussion and Conclusion.

II. A VISUAL SEARCH SYSTEM BASED ON MULTI-SCALE POPULATION CELL CODING

A multi-scale population-cell-coding structure to control gaze movement is designed to implement the visual search system, which is illustrated in Fig.4. The coding structure consists of two parts. The first one is an image content encoding part, including the first three layers: the first layer - input neurons, the second layer - feature neurons, and the third layer –population coding neurons. It inputs a local image from a group of visual fields in different resolutions, then extracts features and encodes the current visual field image in terms of connection weights between the third layer and the second layer. The second one is a spatial relationship coding part, including the last two layers: the third layer - population coding neurons and the fourth layer – gaze movement control neurons. It encodes the spatial relationship between two object centers or between one target and its environmental key points in terms of connecting weights between the third layer and the fourth layer, which corresponds to the horizontal and vertical shift distances (Δx, Δy) from the center position (x, y) in the current visual field to the center of the target. The two parts naturally incorporate into an entire one. They cooperate to encode or decode the image content and the spatial relationship from the current visual field image, then move from the center of the current visual field to the center of target. It runs this spatial relationship encoding (learning/memorizing) or gaze movement decoding (test/ searching) procedure in a repeated mode until the system finds the target and stops the gaze movement.

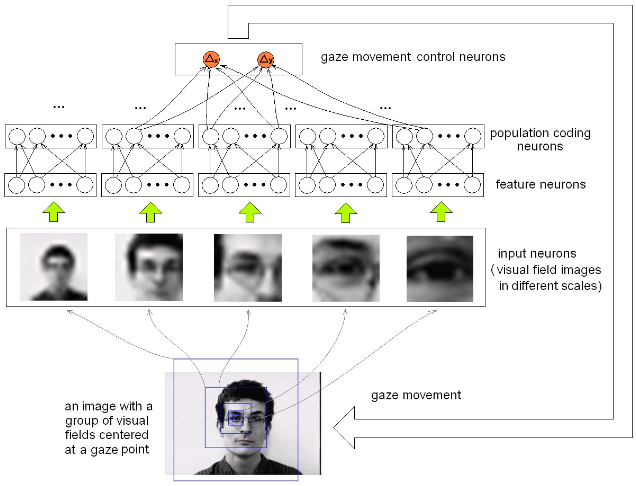


Fig.4. A visual search system implemented with multi-scale population coding and gaze movement controlling mechanisms.

A. Features designed for encoding visual context

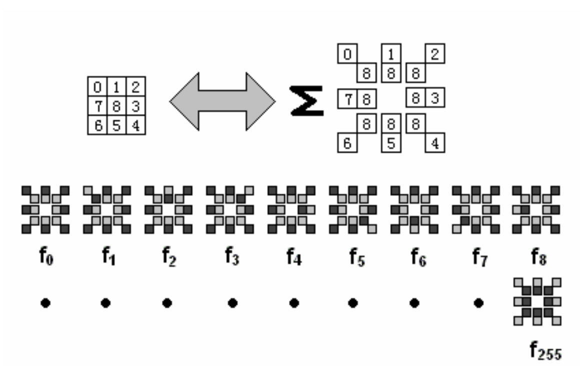


Fig.5. 256 extend LBP features (receptive field=3×3 pixels, each of which is computed by a sum of eight pairs of differences between pixels(labels=0~7) and the central pixel (label=8). The gray box represents weight 1 while the black box represents –1.

A set of features called local binary patterns (LBP) [15] are widely used recently. It is simple and takes into good effects on image feature extraction and classification. We extended the original LBP features to the ones illustrated in Fig.5. LBP is a kind of binary code for representing one of 256 patterns for image blocks of 3×3 pixels. Original LBP code features only output a discrete number from 0~255 to represent an image block pattern instead of producing a continuous comparable value. We extend LBP features by assigning them continuous output with the following function:

$$r_{ij} = f_j(\vec{\mathrm{x}}_i) \approx \sum_{k=0}^{7} (-1)^{b_k} (x_{i8} - x_{ik}) \tag{1}$$

where the vector $\vec{\mathrm{x}}_i = (x_{i0}\ x_{i1}\ ...\ x_{i8})$ represents the i-th image block or receptive field input of 3×3 pixels, the term r_{ij} represents the j-th feature extracted from the i-th image block, and j is a decimal number among 0~255, which corresponds to a 8-bit binary code, i.e. $(j)_{10}$<=>$(b_0b_1...b_k...b_7)_2$, where

$$b_k = \begin{cases} 0 & \text{if } (x_{i8} - x_{ik}) > 0 \\ 1 & \text{otherwise} \end{cases}, \ (k=0\sim7) \tag{2}$$

In our coding system, for each receptive field input $\vec{\mathbf{x}}_i$, there are 256 feature neurons extracting the above extended LBP features. Only the first m neurons with the largest responses win through the mutual competition and produce outputs with their responses $r_{ij'} = f_{j'}(\vec{\mathbf{x}}_i)$, where j'=1~m. To decrease the coding quantity as much as possible, m may be set to 1 for enough sparsity.

B. Encoding visual field image

With reference to Fig.4, for each single-scale visual field input $\vec{\mathbf{x}}$, the k-th coding neuron receives inputs weighted with $w_{k,ij'}$ (j'=1~m)from the ij'-th feature neuron, whose response $r_{ij'}$ represent the j'-th feature for its i-th receptive field input $\vec{\mathbf{x}}_i$. For the visual field image $\vec{\mathbf{x}}=(\vec{\mathbf{x}}_1\vec{\mathbf{x}}_2...\vec{\mathbf{x}}_N)$ which is composed of the receptive field inputs $\vec{\mathbf{x}}_i$ (1≤i≤N), a coding neuron's response is:

$$\begin{aligned} R_k &= F(\vec{\mathbf{x}}) = F((\vec{\mathbf{x}}_1\vec{\mathbf{x}}_2...\vec{\mathbf{x}}_N)) \\ &= \sum_{i=1}^{N}\sum_{j'=1}^{m} w_{k,ij'} f_{j'}(\vec{\mathrm{x}}_i) = \sum_{i=1}^{N}\sum_{j'=1}^{m} w_{k,ij'} r_{ij'} \end{aligned} \tag{3}$$

where the weights $w_{k,ij'}$ is obtained at the encoding or training stage according to the Hebbian rule by one-step updating: $w_{k,ij'}(1)=w_{k,ij'}(0)+\alpha R_k r'_{ij'}$, in which $w_{k,ij'}(0)$=0; α is set as 1 for simplified computation; R_k is set to 1 to represent the response of the k-th coding neuron generated for representing a new visual field image pattern; and $r'_{ij'}$ is the response of the j'-th feature neuron for receptive field input $\vec{\mathbf{x}}_i$ at the training stage; $r_{ij'} = f_{j'}(\vec{\mathbf{x}}_i)$ (j'=1~m) is the response belonging to the first m features that have the largest responses among the total feature responses $\{r_{ij}\}$(j=0~255) at the test stage. The length of the k-th weight vector composed of the weights $w_{k,ij'}$ (i=1~N, j'=1~m) will be normalized to one for unified similarity computation and comparison.

C. Encoding and decoding spatial relationship or gaze movement

Gaze movement control is the key aspect for visual object research, which is implemented in the structure that consists of two layers of neurons: coding neurons and movement control neurons (Fig.4). The movement control neurons, divided into Δx and Δy neurons, represent the target's relative position (Δx, Δy) from the current gaze point (x, y) - the center of the current visual field images. For the visual field image in scale s, the first M_s coding neurons with the largest responses play main roles in activating the movement control neurons. At the test or decoding stage, the responses of gaze movement control neurons could be formulated as:

$$\begin{cases} R_{\Delta x} = \sum_{s=1}^{S}\sum_{k'=1}^{M_s} w_{\Delta x,k',s} R^*_{k',s} \\ R_{\Delta y} = \sum_{s=1}^{S}\sum_{k'=1}^{M_s} w_{\Delta y,k',s} R^*_{k',s} \\ R^*_{k',s} = \dfrac{R_{k',s}}{\sum_{k'=1}^{M_s} R_{k',s}} \end{cases} \tag{4}$$

where S is the number of scales existed in visual fields (with reference to Fig.3 or Fig.4), $R_{k',s}$ (k'=1~ M_s) is the response of the k'-th coding neuron among the M_s coding neurons for the visual field in scale s; $w_{\Delta x,k',s}$ and $w_{\Delta y,k',s}$ are the connecting weights from the k'-th coding neuron to the movement control neurons in x and y directions respectively. They are calculated according to the Hebbian rule by one-step updating:

$$w_{\Delta x,k',s} = 0 + \beta \Delta x_{k'} R'_{k',s}\ ,\ \ w_{\Delta y,k',s} = 0 + \beta \Delta y_{k'} R'_{k',s} \tag{5}$$

where $\Delta x_{k'}$ and $\Delta y_{k'}$ are the responses of two movement control neurons, which equal to the distances from current gaze point to the center of the target when $R'_{k',s}$ is the response of the k'-th coding neuron that is generated for representing a new visual field image pattern with a corresponding spatial relationship ($\Delta x_{k'}$ and $\Delta y_{k'}$) at the encoding stage. The learning rate β and $R'_{k',s}$ can be set to 1 for simplified calculation. Thus, formulae (4) can be rewritten as:

$$\begin{cases} R_{\Delta x} = \sum_{s=1}^{S}\sum_{k'=1}^{M_s} \Delta x_{k'} R^*_{k',s} \\ R_{\Delta y} = \sum_{s=1}^{S}\sum_{k'=1}^{M_s} \Delta y_{k'} R^*_{k',s} \\ R^*_{k',s} = \dfrac{R_{k',s}}{\sum_{k'=1}^{M_s} R_{k',s}} \end{cases} \tag{6}$$

Formulae (6) means the gaze movement distances decoded at the test stage is the sum of spatial relationship encoded at the learning stage, which are weighted by the responses of population coding neurons(S>1 or M_s>1).

When the system searches a target, how many scales of visual fields should be used? There are two extreme cases. The first case is that all the visual field images involved in the encoding or decoding spatial relationship for each gaze movement, which is characterized by formulae (4). We call this case *Full-scale* coding. The second case is that only one visual field image is used each time, for example, the system uses the largest visual field first then uses smaller ones to search the target. We call this case *Gradual-scale* coding. For the second case, formulae (4) can be simplified to formulae (7):

$$\begin{cases} R_{\Delta x,s} = \sum_{k'=1}^{M_s} \Delta x_{k'} R^*_{k',s} \\ R_{\Delta y,s} = \sum_{k'=1}^{M_s} \Delta y_{k'} R^*_{k',s} \\ R^*_{k',s} = \dfrac{R_{k',s}}{\sum_{k'=1}^{M_s} R_{k',s}} \end{cases} \qquad (7)$$

D. *Algorithm description*

The system's encoded visual context is preserved in the weights of the neural coding structure. Hebbian rule is the fundamental learning or coding rule, i.e., $\Delta w_{ij}=\alpha R_i R_j$, where w_{ij} is connecting weights; α is the learning rate; R_i and R_j are responses of two neurons that are connected mutually. The encoding algorithm is described as follows:

1. Input one or a group of visual field image(s) centered at the gaze point, and predict the target's shift distances (Δx, Δy);
2. If predictive error is larger than a threshold, generate a new coding neuron (let response R=1); else go to 4;
3. Encode visual context by computing the connecting weights between the new coding neuron and feature neurons, and those weights between the new coding neuron and two movement control neurons (responses $R_{\Delta x}=\Delta$x and $R_{\Delta y}=\Delta$y) using Hebbian rule $\Delta w_{ij}=\alpha R_i R_j$;
4. Move current gaze point to the center of the target in the current visual field(s);
5. Go to 1, until all given starting gaze points are trained.

Fig.6 describes the target search procedure in terms of gaze movements starting from any given gaze point that is initialized to be the center of the visual field. According to the encoded visual context, the system perceive image input, decode the spatial relationship and move the gaze in a repeated mode until the system ensure that the center of the current visual field(s) is the center of the target in terms of 0 shift distances (Δx=0, Δy=0). The decoding or target search procedure is illustrated in Fig.7.

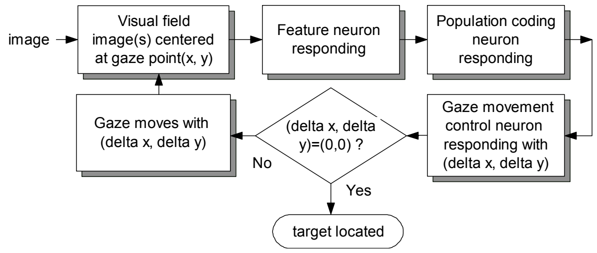
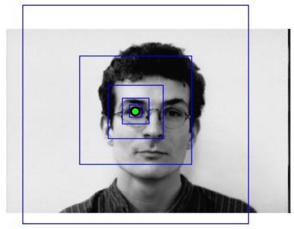
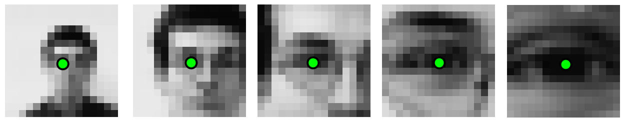


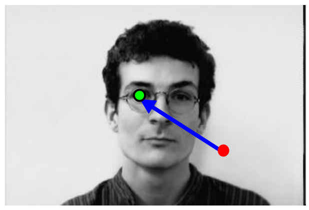
Fig.6. Target search in terms of gaze movement driven by population cell coding.



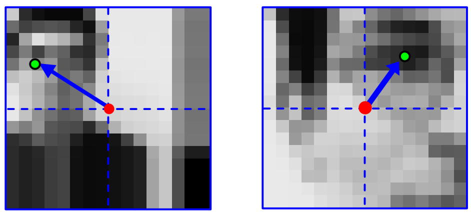
(a) Five visual fields centered at the target center (the left eye center).



(b) Five visual field images (16×16 pixels, scales=5, 4, 3, 2 and 1) sub-sampled from the original image (320×214 pixels).



(c) The spatial relationship between the target center and a given starting gaze point.



(d) encoding the content of visual field images and the spatial relationship between target centers and starting gaze points, or searching the target according to the visual context encoded (here two scales of visual fields are shown).

Fig.7. Illustration of spatial relationship encoding and decoding.

III. EXPERIMENTS ON CODING FOR GAZE MOVEMENT CONTROL IN TARGET SEARCH

To test the system's performance and compare the *Full-scale* coding and *Gradual-scale* coding, two experiments for left eye center searching are carried out on the still face image database of the University of Bern[16], which has total 300 images (320×214 pixels) with 30 people (ten images each person) in ten different poses. Fig.8

illustrates the first ten images.

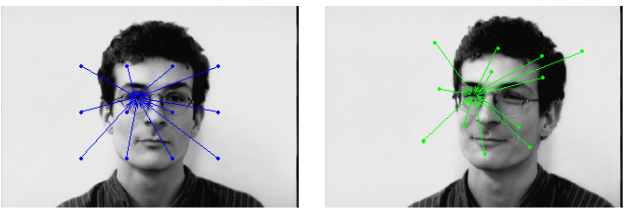
Fig.8. Examples from the face database of the University of Bern (320×214 pixels)

A. Coding Structures

We designed two coding systems using *Full-scale* coding and *Gradual-scale* coding mechanisms. A group of visual fields in five scales (256×256, 128×128, 64×64, 32×32 and 16×16 pixels) are used to input local images from the training and test images (320×214 pixels). For each scale or resolution, there is the same number of 16×16 input neuron with different intervals (16, 8, 4, 2 and 1 pixels). So there are totally 5×16×16=1280 neurons in the first layer of the neural coding structure. With reference to Fig.5, there are 256 kinds of LBP features for five visual fields with different resolutions, and the size of receptive field of each feature neuron is 3×3 pixels, which has 1/2 overlap between neighboring receptive fields, thus there are totally $5\times256\times[16-(3-1)]^2=250880$ feature neurons respectively in two systems, in which only 250880×(1/256)=980 neurons (the first m largest responding feature neurons, m=1 for sparsity, see Section 2) contribute to activate the coding neurons in the third layer. The number of coding neurons in the third layer is dependent on natural categories of visual context patterns that the system learned. The number of gaze movement control neurons in the fourth layer is two, which should, from reasonable expectation, output the value in a range from –8 to 7 to represent 16 positions in x and y directions respectively, corresponding to 16×16 input neuron array for all the five visual fields in the first layer.

B. Experiments on encoding visual context for target search

Two experiments for each system, totally four experiments were carried out on the head-shoulder database of the University of Bern for searching and locating left eye centers.



(a) (b)

Fig.9. Illustration of experiments: (a) Training for encoding visual context between the eye center and a group of initial gaze points in a uniform distribution. (b) Testing for gaze movement for search the eye center from a group of initial gaze points in a random distribution.

As illustrated in Fig.9, encoding was with a group of initial gaze points in a uniform distribution while testing was with a group of initial gaze points in a random distribution. The systems were trained and tested from the two groups of se initial gaze points respectively to encode the context and decode the spatial relationship for searching the left eye centers.

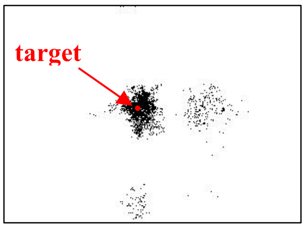
In the first experiment (Exp.1) for two systems, 30 images of 30 people (one frontal image each person) were encoded with 368 initial gaze points on each image, and the rest of 270 images were tested at 48 random initial gaze points on each image. In the second experiment (Exp. 2) for two systems, 90 images of 9 people (10 images each one) were encoded and the rest of 210 images were tested with initial gaze points as Exp.1. The number of total feature neurons in layer 2, the number of total coding neurons in layer 3, the number of connections between feature neurons and coding neurons, and the mean/standard deviation of locating errors are listed in the table below.

TABLE I PERFORMANCES OF THE TWO CODING SYSTEMS (M: MILLION)

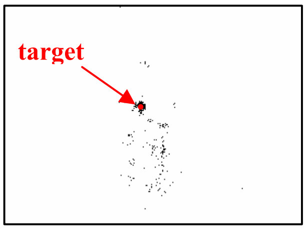
Coding system	number of feature neurons in layer 2	number of coding neurons in layer 3		number of connections between feature neurons and coding neurons (M)		average locating error/standard deviation of locating error (pixels)	
		Exp.1	Exp.2	Exp.1	Exp.2	Exp.1	Exp.2
Full-scale	250880	36609	110497	7.18	21.66	9.98/ 10.72	7.53/ 8.99
Gradual-scale	250880	32961	99549	6.46	19.51	3.00/ 7.45	2.29/ 4.05

The number of population coding neurons to activate movement control neurons is a dynamical value that is decided by the ratio of the sum of the responses of the first M largest responding coding neurons to the sum of the responses of the total coding neurons. According to experimental experience, the best searching accuracy could be obtained when this ratio is set to 1%. The mean radius of the eyeballs in the database is approximately 4.02 pixels. Table 1 shows the average locating error decreases of Exp.1 and Exp.2 are 6.98 pixels (from 9.98 to 3.00 pixels) and 5.24 pixels (from 7.53 to 2.29 pixels) respectively by using two coding mechanisms, which means the average locating positions by using *Full-scale* coding is outside the average borders of the eyeball objects and the average locating positions by using *Gradual-scale* coding is inside the borders of the objects. From the above table, it can be learned that the *Gradual-scale* coding system reached an average locating error that is 69.79% lower than that of the *Full-scale* coding system and cost 10.28% less coding neurons and connections. Simultaneously, the decreases of standard deviations of locating errors for Exp.1 and Exp.2 are 3.27 pixels (from 10.72 to 7.45 pixels) and 4.94 pixels (from 8.99 to 4.05 pixels) respectively, which indicate significant standard deviation decreases of 30.5% and 54.9% respectively. In the respect of two system's encoding quantities, mainly in terms of the numbers of neuron connections, the *Gradual-scale* coding system only account for approximately 90% encoding quantities of the *Full-scale* coding system.

Fig. 10 and Fig. 11 illustrate two pairs of final gaze position distributions that reflect two coding systems' performance difference in target (left eye center) searching for two groups of training-test sets. The first group is composed of 30 training images and 270 test images (Exp.1) and the second group is composed of 90 training images and 210 test images (Exp.2). From Fig. 10 (b) and Fig.11 (b), it can be seen that the *Gradual-scale* coding system has more compact searching results compared to that of the *Full-scale* coding system shown in Fig.10 (a) and Fig.11 (a).



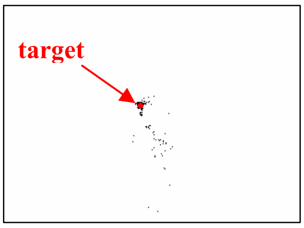
(a)



(b)

Fig.10. The distribution of target locating results for 270 test images in the first experiment using two coding mechanisms. (a) Distribution of locating results using *Full-scale* coding mechanism. (b) Distribution of locating results using *Gradual-scale* coding mechanism.

(a)



(b)

Fig.11. The distribution of target (left eye center) locating results for 210 test images in the second experiment using two coding mechanisms. (a) Distribution of locating results using *Full-scale* coding mechanism. (b) Distribution of locating results using *Gradual-scale* coding mechanism.

IV. Discussion and Conclusion

This paper proposed a visual search system using the population cell coding mechanism and the multi-scale visual field as sensing input. It laid stress on how to efficiently encode and decode gaze movement for target searching. As an example, the system was applied to human eye center searching. An experimental comparison of the *Full-scale* visual field coding system and the *Gradual-scale* coding system is carried out. The experiment results show the *Gradual-scale* coding system performed better than the *Full-scale* coding system in terms of the higher locating accuracy and the lower encoding quantity. It means not all the visual field images in different scales are effective for searching a target each time. Choosing the suitable sequence of visual field scales, e.g. larger scales first and smaller scales in next steps, is more efficient. From another point of view, the experiment verified the reasonability of some efficient searching strategies, for example, the strategy of detecting objects in coarse-to-fine mode.

References

[1] Deliang Wang, The time dimension for scene analysis, *IEEE Trans. NN*, 16(6), pp.1401-1426, 2005.

[2] G. Zelinsky, W. Zhang, B. Yu, X. Chen, and D. Samaras, The role of top-down and bottom-up processes in guiding eye movements during visual search, *Advances in NIPS*, 2006.

[3] M. Mozer and D. Baldwin, Experience-Guided Search: A Theory of Attentional Control, *NIPS*, 2007.

[4] M. Cerf, J. Harel, W. Einhaeuser, and C. Koch, Predicting human gaze using low-level saliency combined with face detection, *Advances in NIPS*, 2007.

[5] J. Wolfe, Guided Search 4.0: Current progress with a model of visual search, in. W. Gray (Ed.), *Integrated Models of Cognitive Systems,*NY: Oxford, 2007

[6] A.Torralba, A.Oliva, M. Castelhano and J. Henderson, Contextual guidance of eye movements and attention in real-world scenes: The role of global features on objects search. *Psych. Rev.*, *113*, 2006.

[7] A. Torralba, Contextual Priming for Object Detection, *IJCV* 53(2): 169-191, 2003.

[8] K. Murphy, A. Torralba, and W.T. Freeman, Using the forest to see the trees: a graphical model relating features, objects, and scenes, *Advances in NIPS,* 2003.

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

[10] T. Shibata, S. Vijayakumar, J. Conradt, S. Schaal, Humanoid oculomotor control based on concepts of computational neuroscience, *Proc. the 2nd International Conference on Humanoid Robots*, 2001.

[11] L. Itti and C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research*, vol. 40, pp. 1489-1506, 2000.

[12] H. Kruppa, M. Santana and B. Schiele, Fast and Robust Face Finding via Local Context, Proc. Joint *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance* (VS-PETS'03), 2003.

[13] L. Paletta and C. Greindl. Context based object detection from video, *LNCS 2626-Proc. International Conference on Computer Vision Systems*, pp. 502-512, 2003.

[14] T. Strat and M. Fischler, Context-based Vision: recognizing objects using information from both 2D and 3D imagery, *IEEE-PAMI* vol.13(10), pp. 1050-1065, 1991

[15] T. Ahonen, A. Hadid, and M. Pietikainen, Face recognition with local binary patterns, *Lecture Notes in Computer Science-Proc.ECCV*, vol. 3021, pp. 469-481, 2004.

[16] The Face Database of the University of Bern, *ftp://iamftp.unibe.ch/pub/Images/FaceImages/*, 2008.

[17] M. Bear, B. Connors, and M. Paradiso, *Neuroscience: Exploring the Brain(3rd)*, Wolters Kluwer Health, 2007.