Single vs. Population Cell Coding: Gaze Movement Control in Target Search

Jun Miao, Laiyun Qing, Lijuan Duan, and Baixian Zou

Abstract—**Gaze movement plays an important role in human visual search system. In literature, the winner-take-all method is wildly used to simulate the controlling of the gaze movement. The winner-take-all is a type of single-cell coding method, which uses one cell (grandmother cell) or one response to represent an object. However, eye movement is affected by the visual context which includes more than one object in images, especially in target search. Therefore, we propose to use the population coding with more than one response rather than the single-cell coding on gaze movement control. The proposed method is supported by the theoretical analysis and experiments on a real image database which show the population-cell-coding improves the target locating accuracy by 44.4% only at the cost of coding 22.4% more information than that of single-cell-coding.**

I. Introduction

Single and population cell coding mechanisms have been an argumentative issue in understanding human brain and vision functions, which has been discussed and debated in the special issue of *Neuron* Vol.24 (1) in detail. Single-cell-coding means using one cell (grandmother cell) or one response to represent one object, in terms of winner-take-all mechanism. Population-cell-coding uses an ensemble of cells or responses to represent an object, which implies relevant cells that encodes similar object categories, are fired with a certain responses. Generally it is concluded that both mechanisms exist in human neural system but play different roles [1].

Eye gaze movement or variation plays an important role in human vision information acquiring and object searching system. Some research work simulates the gaze variation in bottom-up and top-down modes [2-11, 23-26]. For example, when one searches objects in images without the knowledge of object scales, orientations and positions, he may moves his gaze to the places according to the posterior probability with the context between objects and environmental features[7,8,25-26], which is known as the top-down attention or the task-driven object searching. When one recognizes a 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 a saliency map [11] from the most salient to the least salient, which is known as the bottom-up attention or the feature-driven visual searching.

Either the top-down or the bottom-up method, if they use the decision principles by the largest saliency or 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. Generally eye movement is affected by the visual context that includes more than one object in images, especially in the case of target search. So we hypothesize that population coding with more than one response performs better than single-cell-coding in controlling the gaze movement. 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 keeps information representing more stabile and accurate at the cost of more coding quantity.

Experiments that applied single and population coding methods for target search on a real image database verify the hypothesis and the experimental results show that the population coding performs almost twice better than the single-cell-coding.

II. Features designed for coding visual context

In a top-down visual object search system, the context [7, 12-14] between environmental features and targets is usually to be learned or coded for the future prediction of the positions of the targets. Visual context is relative to the spatial relationship in terms of distances (Δx, Δy) between two objects in an image respectively in the horizontal and the vertical directions, as shown in Fig. 1. Generally, learning such context is inevitable to cost a large amount of memory. Choosing or designing concise and efficient features seem quite important. The psychological experimental results show that when human perceive an image (e.g. a human face image) only a few of neurons

This research is partially sponsored by Natural Science Foundation of China (No.60673091 and No.60702031), Hi-Tech Research and Development Program of China (No.2006AA01Z122), Natural Science Foundation of Beijing (No.4072023) and Beijing Municipal Education Committee (No.KM200610005012).

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).

L. Duan is with the College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China (e-mail: ljduan@bjut.edu.cn).

B. Zou is with the Department of Information Science and Technology, College of Art and Science of Beijing Union Technology, Beijing 100083, China (e-mail: zoubx@ygi.edu.cn).

response in his visual cortex [15]. This is the strategy of sparse coding for human's visual neural system. Mathematically, sparse coding means that an image can be approximately represented by a group of sparse coefficients corresponding to a number of bases or features, which obey the super-Gaussian distribution. Many approaches [16-18] have been proposed for finding such spares bases. In addition, independent components of natural scenes are similar to the sparse bases [19], which look like edge filters [20].

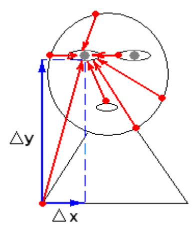


Fig.1. Visual context in terms of distances (Δx, Δy) in horizontal and vertical directions respectively between two objects.

A set of features called local binary patterns (LBP) [21] is widely used recently. Although it has not been proved to be sparse coding feature, it is simple and efficient on image feature extraction and classification. We extended the original LBP features to the new features illustrated in Fig.2.

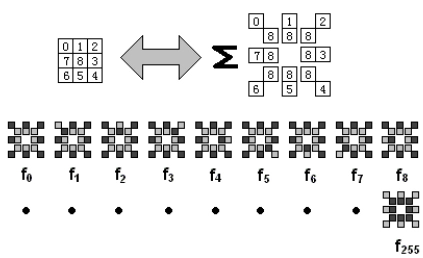


Fig.2. 256 extend LBP features (receptive field=3×3 pixels, each of which is computed by a sum of eight pairs of differences between pixels(labels=0~7) and the central pixel (label=8). The gray box represents weight 1 while the black box represents weight –1.

LBP (local binary patterns) [24] is a kind of binary code for representing one of 256 patterns for image blocks of 3×3 pixels. The original LBP code features only output a discrete number from 0~255 to encode an image block pattern instead of producing a continuous comparable value for a local image pattern. We extend LBP features by assigning them continuous output with the following function:

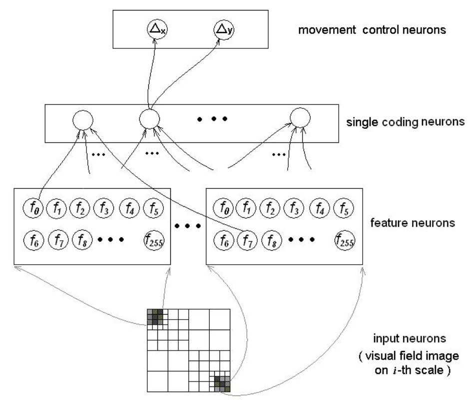
$$r_{ij} = f_j(\vec{x}_i) \approx \sum_{k=0}^{7} (-1)^{b_k} (x_{i8} - x_{ik}) \qquad (1)$$

where the vector $\vec{x}_i = (x_{i0}\ x_{i1}\ \dots\ x_{i8})$ 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 discrete number among 0~255, which corresponds to a 8-bit binary code: $(b_0 b_1 \dots b_k \dots b_7)_2$, where

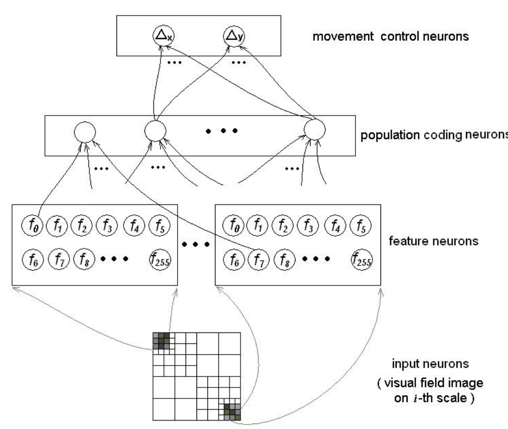
$$b_k = \begin{cases} 0 & \text{if } (x_{i8} - x_{ik}) > 0 \\ 1 & \text{otherwise} \end{cases}, \ (k=0\sim7) \qquad (2)$$

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

III. Single and Population Cell Coding for Gaze Movement Control



(a) Single-cell-coding through winner-take-all in the third layer.



(b) Population-cell-coding in the third layer.

Fig.3. Coding structure of visual context representation and gaze movement controlling

For the convenience of comparison and discussion, a unified neural coding structure is designed for single and population cell coding, and is illustrated in Fig.3. The coding structure consists of two parts. The first part is image content coding, which includes the first three layers: the first layer - input neurons, the second layer - feature neurons, and the third layer – single or population cell coding neurons. The system inputs a local image from a group of visual fields at different resolutions, extracts features and encodes the current visual field image in terms of linking weights between the third layer and the second layer. The second part is the spatial relationship coding part, which includes the last two layers: the third layer - coding neurons and the fourth layer – movement control neurons. It encodes the spatial relation between two objects or between an object and its environmental points in terms of the linking weights between the third layer and the fourth layer, which correspond to the horizontal and vertical shift distances (Δx, Δy) from the center position (x, y) in the current visual field to the center of targets. These two parts naturally incorporate into an entire one. They cooperate to code the image content and the spatial relationship on current visual field image at a larger resolution, and then move from the center of the current visual field to the center of the target. This coding procedure is run in a repeated mode from the largest visual field to the smallest visual field.

A. Coding of visual field image content

The coding neuron(s), whether it is single cell or population cells that win(s) through mutual competition, represent(s) different visual field image patterns. With reference to Fig. 3, the *k*-th coding neuron receives inputs weighted with $w_{k,ij}$ from the *ij*-th feature neuron with the *j*-th feature response r'_{ij} for its *i*-th receptive field image input $\vec{x}_i$. Therefore, a coding neuron's response $R_k=F(\vec{x})$, for the visual field image $\vec{x}=(\vec{x}_1\ \vec{x}_2\ \dots\ \vec{x}_N)$ which is composed of the receptive field inputs $\vec{x}_i$ ($1\leqslant i\leqslant N$), is:

$$R_k=F(\vec{x})=F((\vec{x}_1,\vec{x}_2,\dots,\vec{x}_N)) = \sum_{i=1}^{N}\sum_{j=1}^{m} w_{k,ij} f'_j(\vec{x}_i) = \sum_{i=1}^{N}\sum_{j=1}^{m} w_{k,ij} r'_{ij} \quad (4)$$

where the weights $w_{k,ij}$ is obtained at the coding or training stage according to Hebbian rule $w_{k,ij}=\alpha R_k r_{ij}$, in which 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, α is set as 1 for simplification, r_{ij} is the response of the *j*-th feature neuron for receptive field input $\vec{x}_i$, $r'_{ij}=f'_j(\vec{x}_i)$ belongs to the first *m* features that have the largest responses among the total feature responses $\{r_{ij}\}$(j=0~255) at the testing stage. All the weights $w_{k,ij}$ will be normalized to length 1 for unified similarity computation and comparison.

B. Coding for gaze movement control

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 (see Fig.3). The movement control neurons, divided into Δx and Δy neurons, represent the target position (Δx, Δy) based on the current gaze point (x, y) (the center of the current visual field image). For the current visual field image input, the first M coding neurons which have the largest responses play a main role in activating the movement control neurons. If M=1, it is the single-cell-coding and controlling mechanism, otherwise it is the population-cell-coding mechanism. The responses of gaze movement control neurons can be formulated as:

$$R_{\Delta x}=\sum_{i=1}^{M} w_{x,i}R_i',\quad R_{\Delta y}=\sum_{i=1}^{M} w_{y,i}R_i',\quad R_i'=\frac{R_i}{\sum_{j=1}^{M}R_j} \quad (5)$$

where R_i is the response of the *i*-th coding neuron among the M coding neurons. The terms $w_{x,i}$ and $w_{y,i}$ are the connecting weights from the *i*-th coding neuron to the movement control neurons in x and y directions respectively. Both of them are calculated based on the Hebbian rule:

$$w_{x,i}=\alpha\,\Delta x_i\,R_i,\quad w_{y,i}=\alpha\,\Delta y_i\,R_i \quad (6)$$

where Δx_i and Δy_i are the responses of movement control neurons, which equal to the distances from the current gaze point to the center of the target when R_i is the response of the *i*-th coding neuron that is generated for representing a new spatial relationship at learning or coding stage from the current visual field image. At learning or coding stage, the learning rate parameter α and the response R_i can be set to 1 for simplifying calculation. Thus, formulae (5) can be rewritten as:

$$R_{\Delta x}=\sum_{i=1}^{M}\Delta x_i R_i',\quad R_{\Delta y}=\sum_{i=1}^{M}\Delta y_i R_i' \quad (7)$$

Formula (7) means that the gaze movement distances produced at the perception stage is the sum of the spatial relationship coded at the learning stage, which is weighted by the response(s) of the single coding neuron (M=1) or the population coding neurons(M$\geqslant$1).

C. Algorithm description

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

1. Input a local image from a visual field centered at the gaze point, and perceive the target's shift distances (Δx, Δy);
2. If the perceived result (Δx, Δy) is not correct, generate a new coding neuron (let response R=1); else go to 4;
3. Code or compute connecting weights between the new coding neuron and feature neurons and that between the new coding neuron and two movement control neurons (responses $R_{\Delta x}=\Delta x$ and $R_{\Delta y}=\Delta y$) using the Hebbian rule $w_{ij}=\alpha R_i R_j$;
4. Move current gaze point to the center of the target in the current visual field and change the visual field to a smaller one;
5. Go to 1, until all scales of visual fields and all given initial starting gaze points are coded.

Fig.4 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 at the largest resolution. According to the visual context coding memory, the system perceives image input from the largest visual field to the smallest visual field and moves the gazes in a repeated mode until the system ensures that the center of the visual field at the lowest resolution is the center of the target (i.e. the shift distances are zero (Δ x=0, Δ y=0)). The coding or target search procedure is illustrated in Fig.5.

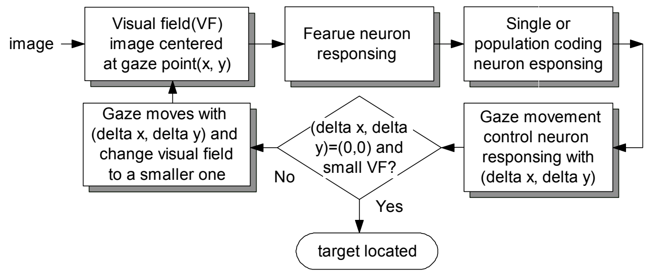
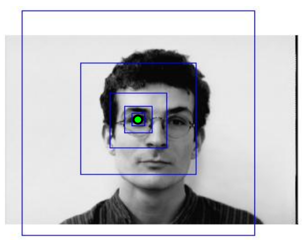
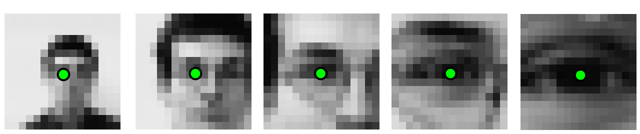


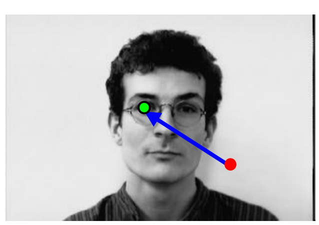
Fig.4. Target search in terms of gaze movement driven by single or population cell coding.



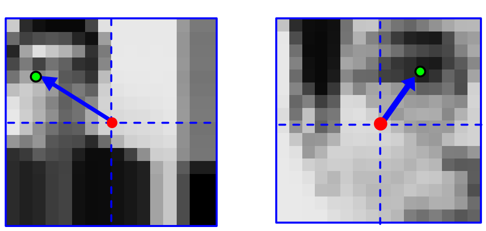
(a) Five visual fields centered at the target center (the eye center).



(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 the any given starting gaze point.

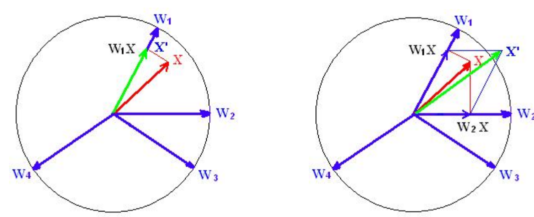


(d) encoding the content of visual field images and the spatial relationship between target centers and gaze points, or searching the target according to the visual context coded, gradually from largest visual field to smaller ones (here two scales of visual fields are shown).

Fig.5. Illustration of gradual spatial relationship coding or target search.

D. Learning properties of single and population coding

From the above coding or learning algorithm, it can be learned that the neural structure codes the "support weight vectors" that distribute on the borders between different visual context categories. With references to Fig.6, there are four visual context classes coded with weight vectors W_1, W_2, W_3, and W_4. A visual field image input X firstly is projected to single or population classes according to its first M largest projections to the four class vectors, and then it is mapped to target positions according to the connections from single or population coding neurons to the movement control neurons. The number of such "support weight vectors" is dependent on the visual context categories in images. For the single-cell-coding, the reconstruction of an input X is equivalent to its largest projection to one of the four class vectors. Obviously, the reconstruction error is larger than the error of the population coding which reconstructs the input X with a group of largest projections and makes the reconstructed X' close enough to the original input X, especially when using the dual-basis method. This learning property theoretically verifies our hypothesis.



(a) Single-cell-coding (b) Population-cell-coding

Fig.6. Comparison of reconstruction effects between two coding mechanisms.

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

Two experiments were carried out to experimentally verify our hypothesis. The results of these two experiments reflect the efficiency and performance of the single and the population coding and are obtained by searching the left eyeball center based on a set of still face images in a database provided by the University of Bern [22]. In this database, totally there are 300 images (320×214 pixels) with 30 people (ten images each person) in ten different poses. The mean radius of the eyeballs of these 30 persons is 4.02 pixels. Fig.7 illustrates the first ten images.

Fig.7. Face database of the University of Bern (320x214 pixels)

A. Coding Structures

We designed two coding systems respectively using the single-cell-coding and the population-cell-coding mechanisms. A group of visual fields at 5 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 array with different intervals (16, 8, 4, 2 and 1 pixels). Thus, there are totally 5×16×16=1280 neurons in the first layer of the neural coding structure. With reference to Fig.2, there are 256 kinds of LBP features for 5 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 single or population coding neurons in the third layer. The number of coding neurons in the third layer is dependent on the natural categories of visual context patterns that the system learned. The number of gaze movement control neurons in the fourth layer is 2. These two control neurons should 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 gaze movement for target search*

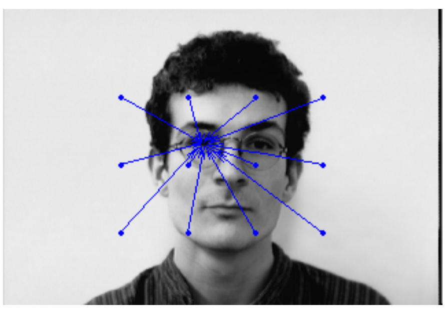


Fig.8. Training for coding visual context between the eye center and a group of initial gaze points in an even distribution.

Fig.9. Testing for gaze movement for search the eye center from a group of initial gaze points in a random distribution.

Two experiments were carried out for each system. Therefore, totally four experiments were done on the face database from the University of Bern for searching and locating left eye centers.

As illustrated in Fig.8 and 9, visual context coding was with a group of initial gaze points in an even distribution while testing was with a group of initial gaze points in a random distribution. The systems were trained or tested from these initial gaze points to code the context or search the eye centers.

In the first experiment (Exp. 1) for two systems, 30 images of 30 people (one frontal image each person) were coded 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 coded 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 average locating error are listed in Table I.

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
Single cell	980	25614	77208	5.02	15.13	4.65/ 7.09	4.82/ 5.67
Population cell	980	32961	99549	6.46	19.51	2.97/ 4.66	2.30/ 2.21

The number of population coding neurons which can 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 our experimental results, the best search 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 I shows the average locating error decreases of Exp.1 and Exp.2 are 1.68 pixels (from 4.65 to 2.97 pixels) and 2.52 pixels (from 4.82 to 2.3 pixels) respectively by using two coding mechanisms, which means the average locating positions by using single coding is outside the average borders of the eyeball objects and that by using population coding is inside the borders of the objects. It can be calculated that the population coding system reached an average locating error that is 44.4% lower than the locating error provided by the single-cell-coding system at the cost of 22.4% more coding neurons and connections. Simultaneously, the decreases of standard deviations of locating error for Exp.1 and Exp.2 are 2.43 pixels (from7.09 to 4.66 pixels) and 3.46 pixels (from 5.67 to 2.21 pixels) respectively by using two coding mechanisms, which indicate significant standard deviation decreases of 34.3% and 61% respectively. The large improvements can be found from the above two groups of experiment comparison in terms of means and standard deviations.

V. DISCUSSION

Winner-take-all or maximum a posteriori principles used in much work is similar to the single-cell-coding mechanism. In this paper, we proposed a hypothesis that population-cell-coding is more suitable for representing the context that contains more than one objects and their spatial relationship and controlling the gaze motion for target search than single-cell-coding mechanism. The hypothesis is verified theoretically and experimentally. Our experimental results indicated that the population-cell-coding made the reconstructed error smaller than the error obtained by using of the single-cell-coding, and it reached a target locating accuracy 44.4% higher than the locating accuracy of single-cell-coding only at the cost of coding 22.4% more information. The significant decreases of standard deviations of locating error, 34.3% and 61%, can be also found from the two experiment comparisons.

Although the experiment was carried out on searching one object, the system can be easily extended to detect multiple objects by training the system to learn the context between any positions and the objects. Similar to the coding features used in references [11, 26], the intensity and orientation features with a receptive field of 2×2 pixels had been adopted to implement the single-cell coding system [25]. Compared with those two kinds of features, the extended LBP feature with a receptive field of 3×3 pixels in this paper, is a kind of contrast features, and has a larger coding range. As a result, a less coding quantity can be achieved. For example, using the extended LBP feature only need 980 coding neurons to achieve the locating error of 5.02 pixels (see Table I), compared with the 2250 coding neurons generated to achieve the locating error of 5.74 pixels [25]. In the future study, both two aspects of the coding quantity and the search accuracy should be enhanced further to make the context coding and visual searching techniques more practical.

REFERENCES

[1] Deliang Wang, The time dimension for scene analysis, IEEE-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] M. Yong, S. Yamane, Sparse Population coding of faces in the inferotemporal cortex, Science, vol.256(1), pp. 1327-1330, 1992.

[16] B. Olshausen and D. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1?, Vision Research, vol.37, pp. 3313-3325, 1997.

[17] A. Hyvarine, P.O. Hoyer, A two-layer sparse coding model learn simple and complex cell receptive fields and topography from natural images, Vision Research, vol.41(18), 2413-2423, 2002.

[18] D. Lee and H. Seung, Learning the parts of objects with nonnegative matrix factorization, Nature, vol.401, pp. 788-791, 1999.

[19] S. Hornillo-Mellado, R. Martín-Clemente, C. Puntonet, and J. Górriz, Connections between ICA and sparse coding revisited, Lecture Notes in Computer Science, Vol. 3512, 1035-1042, 2005.

[20] A. Bell and T. Sejnowski, The independent components of natural scenes are edge filters, Vision Research, vol.37(23), pp.3327-3338, 1997.

[21] 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.

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

[23] C. Guo and Liming Zhang, An attention selection model with visual memory and online learning, Proc. IJCNN, 2007.

[24] M. de Brecht and J. Saiki, A neural network implementation of a saliency map model, Neural Networks, vol.19 (10), 1467-1474, 2006.

[25] J. Miao, L. Qing, L. Duan and W. Gao, Searching eye centers using a context-based neural network, Proc. of the 4th International Symposium on Neural Networks, 2007.

[26] L. Lanyon and L. Denham, A model of active visual search with object-based attention guiding scan paths. Neural Networks, Special Issue: Vision & Brain, 17(5-6): 873-897, 2004.