A Multi-layers Neural Network Model based on Characteristics of Ganglions' Receptive Fields in Retina and an Algorithm for Watchfulness-keeping

Wei Hui and Yang Xian-Bo

Department of Computer Science, Lab of Algorithm for Cognitive Model, Brain Science Research Center,
Fudan University, Shanghai 200433, China

Summary
Machine vision is an active branch of Artificial Intelligence. An important problem in this area is the balance among efficiency, accuracy and huge computing. The visual system of human can keep watchfulness to the perimeter of visual field while at same time their central attention is focused to the center of visual field for fine information processing. This mechanism of computing resource assignment could ease the demand for huge and complex hardware structure. Therefore designing computer model based on biological visual mechanism is an effective approach to resolve problems in machine vision. In this paper a multi-layers neural model was designed based on the features of receptive field of ganglion in retina to simulate multi-scale perceptive fields of ganglion cell. The neural model could maintain the alert to the outer area of whole image while it was capturing and processing more important information locating at the image's central part. It may provide valuable inspiration for the implementation of real-time processing and the avoidance of huge computing in machine vision.
Key words:
Computer vision, Neuro-vision.

1. Introduction: morphology of retina and function

Visual perception is the end result of a two-phases processing system, i.e. the optics processing and the neural processing. The interface of these two phases is retina, which turn photons stimulus into neural impulses. The retina of our human is constituted by three layers of cells, they are receptor cells, bipolar cells and ganglion cells respectively from outer layer to inner one. Between cells belonging to different layers or common layer there are comprehensive neural connections. This is the morphological foundation for optical information processing in early stage [1].

In our human retina there are 125,000,000 rod cells and 7,000,000 cone cells. They all are receptor cells, but their distributions are not identical in all points of retina, in retina's different area the densities of cells are variable regularly. Most rod cells are located in the marginal area of retina, and cone cells are more congregated in central area of retina, in fovea there are cone cells only, more than 150,000 per square millimeter. The characteristic that receptors highly concentrated in fovea makes fovea possess high visual acuity or optical resolution. From the center to the marginal of retina two kinds of receptors mix together and are arranged spokewise, and at the same time their visual acuity are dropping gradually [2]. The rod cells are sensitive to light and dark stimulus, they are specialized in detecting the presence or absence of objects. The cone cells are sensitive to color stimulus, they are specialized in detecting the details of object [3,4]. A ganglion cell possesses a big receptive field of rod cells, and a small even one to one receptive field of cone cells. If consider the ratio between the number of nerve fibers in optic nerve and the amount of receptors in retina (that is 1,000,000 ganglion cells vs. 130,000,000 rod and cone cells), we can see that outside information had been compressed and processed in early stage. From the point of view of sufficiency and efficiency, the variability of ganglions' receptive fields projecting on receptors layer results in both the high visual acuity in focusing area and the watchfulness keeping in all over the marginal area. Processing stimulus falling in the fovea intensively and at same time processing stimulus falling in marginal retina roughly make the computation power of our brain be able to both cover the whole field of vision and respond stimulus rapidly and accurately. The balance between good performance and complex structure is a big problem to biophysics.

The receptive fields of ganglion cells can be divided into two categories: on-center cells and off-center cells[5]. On-center cells respond to lights shone in the central region of their receptive fields with "on" firing and to lights shown in the periphery of their receptive fields with inhibition, followed by "off" firing when the light is turned off. Off-center cells display the opposite pattern, they respond with inhibition and "off" firing in response to lights in the center of their receptive fields and with "on" firing to lights in the periphery of their receptive fields. In another word on-center cells will fire more action potentials when

lights on suddenly, but off-center cells will fire more action potentials when lights off suddenly, and off-center cells will reduce its firing frequency or no fire once lights on. Because the central region and periphery region of receptive field are contrasted each other, when both regions are illuminated diffusely two responses from inner and outer region of receptive field counteract one another, thus both on-center and off-center cells respond very little. Considering from the need of visual information processing, set of a image's boundary lines always is one of the most important and useful clues, this can be proofed by the fact that a caricaturist usually uses only a simple line drawing but does not lack of lifelikeness. A visual information processing system is apt to abstract those meaningful clues existing at the boundaries. This kind of working mechanism of ganglion cells sharpens their responses to boundaries, and reduces their responses to a big area with similar grayness, so as to improve the efficiency of visual information processing system, and be able to lighten the burden of focusing attention and employ computation resource more economically [6].
The research about biological vision mechanism-based artificial vision system always is a very attractive direction, which can be done by simulating one of more phases of visual paths or adopt psychological prototypes in visual cognition, see also [9-18]. The most outstanding advantages of neuro-vision mechanism based computational model are (a) for the using of a kind of neuron-grounded architecture with multi-layers and parallel processing, which is much better at abstracting symbolic descriptions out of numerous pixels, in another word that adapts to make image understanding; (b) this kind of paradigm is much more consistent with substantial base for human cognition, so other intelligent behaviors like memory and learning、knowledge-based reasoning can be merged smoothly without any philosophical difficulty; (c) ability in learning is critical component for vision system, and the biological neural network like model can achieve much higher performance than artificial neural network; (d) there are only neurons or neurons groups in neuro-vision mechanism based computational mode, that is homogeneous either in its structure or in its controlling strategy, no other super power is existing either in composition or in algorithm, so it possesses good performances in robustness, structuralism, simpleness and evolution.

In this paper based on the physiological characteristics of human being's retina, we simulated retina's information processing structure and designed a hierarchical neural network model with radius scalable receptive fields, in order to do fine processing to the stimulus happened to fall in eye center as well as hold sufficient watchfulness to other stimulus fall in marginal area of field of vision.

2. A hierarchical neural network model for information processing level-wisely

2.1 A hierarchical neural network with multi-size receptive fields

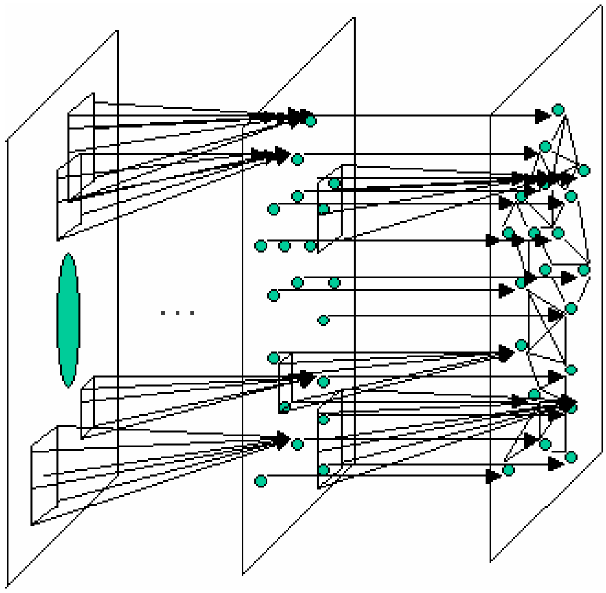


Fig.1 Three-layer model

Rooting from the logical simpleness of retina's structural features and its level-wise signal processing[7], we designed an early vision mode with a three-layer hierarchical neural network. In Figure 1 we can see three layers, from left to right they receptively are: the first layer, a visual stimulus inputting layer that can be an optic sensors array to acquire a bitmap image, the dimension of it decides the scope of visual field, the image projected onto this array is corresponding to an inner image gotten by receptors of retina; the second layer, an artificial neurons array that simulates ganglion cells, any of them has a receptive field distributed on first layer and gets outputs of those sensors that falling in its receptive field; and the third layer, an artificial neuron array gets inputs from the second layer in order to decide whether there is some image changing occurring or not, and judge whether it is necessary to move and replace system's focus on some novel changing, there are localized transverse connections between neurons in order to encourage competitions in this layer and pop out most obvious changing.

The characteristics of the distributions of rod and cone cells in biological retina, in addition the distribution of ganglion cells' receptive fields that projecting to receptors layer make the image that happens to be generated by fovea is finest, and just this part of inner image gains the opportunity of being processed sufficiently, at same time the other part of inner image, i.e. the marginal part of it is processed roughly, and as long as this rough processing can provide enough clues for our mind to hold watchfulness and capture remarkable change occurring in the marginal of visual field then it is deep enough. This proofs that a biological vision system can fulfill coverage as big as possible under limited computational resource. Enlighten by this idea we assigned different ganglion cell-like neurons in the second layer of our model with different radiuses of receptive field, Figure 2 is an example of some ganglion cell-like neurons whose size of receptive fields are different, the circle stands for where and how many sensors a ganglion cell-like neuron covers. We can find that the more nearer a receptive field is away from center, the much smaller the size of it will be. So in center (or in fovea) the interval of sampling is much smaller than that in marginal, this means that (a) there is no information in visual field that had been neglected, (b) much more details in focus area than in around of that area will be reflected upwards. Thus this artificial vision system is not only able to focus its attention to some special position and process stimulus there finely, but also able to perceive and explore a much more expanded area with somewhat low resolution, so that it can afford two types of tasks focusing attention and holding watchfulness simultaneously. It is very worth emphasizing that for the sake of clarity many overlapping between several adjacent receptive fields are omitted in figure 2.

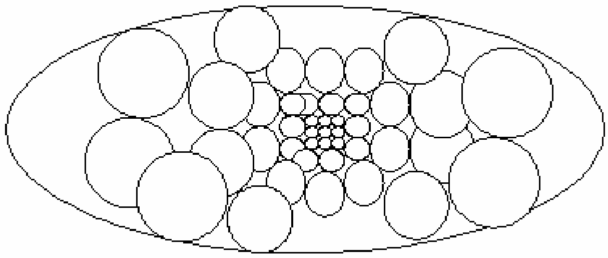


Fig.2 Ganglion cells' receptive fields distribute on sensors layer

In biological retina, from fovea to periphery the sizes of ganglion cells' receptive fields in receptors layer increase rapidly, but visual acuities of them decrease consequently. We see if the distance between two stimuli that we can tell apart is much smaller then our visual acuity is much better. If we explain that telling apart two stimuli is due to they make two diverse neurons be activated, then visual acuity is correlated with the radius of receptive field. Thus if the distance of two stimuli is bigger than all radiuses of receptive fields lying in that area then they must be able to activate two diverse neurons. Because visual acuity is dropping rapidly with the distance of receptive field away from fovea is raising, we can use a function to describe this relationship. If variable D is the radius of receptive field and variable l is the distance between the center of receptive field and the center of retina, then in this paper we choice function $D = a^{\frac{l}{k}}, a > 1$ to depict that relationship, where a and k are two parameters used to control the dropping velocity of visual acuity.

For the sake of convenience for representation and calculation, the shape of receptive field was chosen as a square with odd pixels edge. We build a coordinate system on retina, and set the center of fovea as origin and the unit of two axons is a pixel. Suppose that a ganglion cell-like neuron's position is (x, y), and label $D_{(x,y)}$ as the diameter of its receptive field projected on first layer, we let calculation formula (1) to be $D_{(x,y)} = f(a^{\frac{\sqrt{x^2+y^2}}{k}})$, where function $f(z) = \begin{cases} 2n-1 & 2n-1 \le z < 2n+1 \quad n \in N \\ 1 & 0 < z < 1 \end{cases}$.

Parameter a can be initialized in terms of following factors, such as parameter k 、the desired size of fovea and the size of the most outer receptive field in first layer. For instance if a visual field is a $N \times N$ pixels array, the radius of its fovea is R_f, because all receptive fields in this paper are odd edge square, so the smallest two radiuses are 1 and 3, obviously 1 is the size of receptive fields in fovea (for one to one mapping is one of characteristics of fovea), and 3 can be the smallest size of receptive fields that are most adjacent to fovea, so by formula (1) equations $\begin{cases} f(a^{\frac{r}{k}}) = 1 & 0 \le r < R_f \\ f(a^{\frac{R_f}{k}}) = 3 \end{cases}$ are rational, those make

$$a = 3^{\frac{k}{R_f}}.$$

If let a ganglion cell whose position is $\frac{N}{2}$ away from the center of the visual field has the most outer receptive field and the diameter of that receptive field is D_o, then by formula (1): $a = D_o^{\frac{2k}{N}}$. In general what we decide firstly is the scope of an inner image that need processing finely, i.e. the radius of fovea, after that we can work out the value of parameter a.

The physiological structure of human or other high-level mammals is no doubt the result of natural selection for millions of years. The ganglion cells' receptive fields are greatly intersectant mutually, but what's rule that control the configurations of receptive fields' distribution is unknown now, what we can affirm is that the union of these receptive fields ought to cover all visual field and any point of visual field should better to be covered more than once. The coming algorithm 2.1 is to produce on-center cell-like and off-center cell-like neural computing units.

Step 1: All optical sensors in first layer are labeled as free point.

Step 2: Select a free point $\alpha_{(i,j)}$ in first layer randomly, and add an on-center cell or off-center cell in second layer with equal probability, then label this newly produced neuron as $\beta_{(i,j)}$.

Step 3: In terms of the principle in 2.1.1 compute the receptive field range of neuron $\beta_{(i,j)}$, after that label all sensors covered by this receptive field as no-free points, and joint these sensors and $\beta_{(i,j)}$ together.

Step 4: Iterate algorithm from step 2 until all sensors had been labeled as no-free points.

The producing of topological framework of neural network is a kind of one-off process, once producing finished all neurons and their connections can be stored in a database or a table for future using.

2.2 The characteristics of ganglion cell's receptive field and DOG computational method

The receptive fields of both on-center and off-center cells can be seen as being composed by a center disc surrounded by a concentric ring, the respective characteristics of two components' respondence to light stimulus are opposite, one is excitory and the other is inhibitory, they counteract each other and one of them might predominate. In 1965 Rodieck put forward a mathematical model for receptive field's counteracting, namely a difference of Gaussian (abbr. DOG) model, which is made up of a center function with strong and excitory effect and a more spread periphery function with weak and inhibitory effect, or contrary. Both functions counteract through two opposite directions of positive and negative force, but the spatial distribution of any of two forces is Gaussion alike. There are some other computational models for same purpose [7], because the DOG model has an advantage of low complexity in computation, we can take DOG model as an operator. And from the point of view of parallel computation, many DOG operators would be projected to many receptive fields, then all operators can be implemented conveniently by multi-threads. Another important advantage of selecting an operator like DOG is because that can avoid much spending in treating neuron-differed connections, the number of them is considerable. So in this paper our computational method of ganglion cell-like neurons' receptive fields is DOG model based.

The visual field of our experiments is 256×256 pixels, for the sake of convenience in computation and representation the shape of receptive field is a square with odd pixels edges, for example we chose k =30 and $a = 2$, the radius of fovea can be worked out by formula (1), that is $R_f = \mathrm{int}(k \log_a 3) = \mathrm{int}(30 \times \log_2 3) = 47$, where $\mathrm{int}(x)$ is a floor function, the diameter of most outer receptive field is $D_0 = f(a^{\frac{N}{2k}}) = 19$. The number of ganglion cell-like neurons in second layer produced by Algorithm 2.1 is 19254, in which the number of neurons whose receptive field is confined in fovea is 4906. Figure 3 shows a distribution of receptive fields scattering on visual field, these receptive fields belong to only 25% on-center cells.

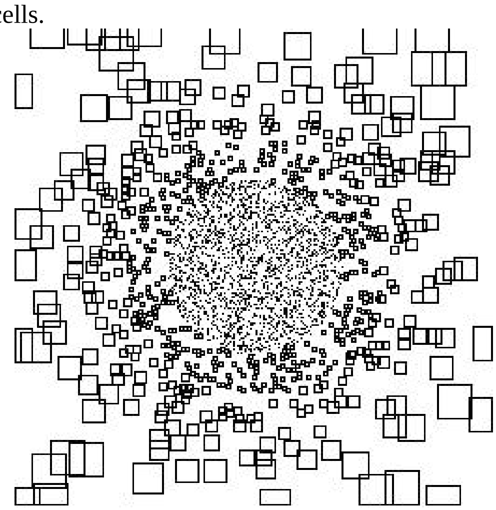


Fig.3 A distribution of some receptive fields scattering on visual field

Because receptive fields of those ganglion cell-like neurons in second layer are intersecting one another, so there are many sensors in first layer covered by some receptive fields more than once. It is easy to make a statistic of sensors' being covered times, because those sensors in fovea are one-to-one connected with equal number of ganglion cell-like neurons whose receptive field's diameter is smallest 1, that means any detail of image can be reflected upwards without any losing, so for economic consideration these sensors in fovea are covered by only one receptive field. Figure 4 shows the relationship between coverage times and position, the low-lying platform-like round in center is fovea that transmits signal from one sensor to one ganglion cell-like neuron.

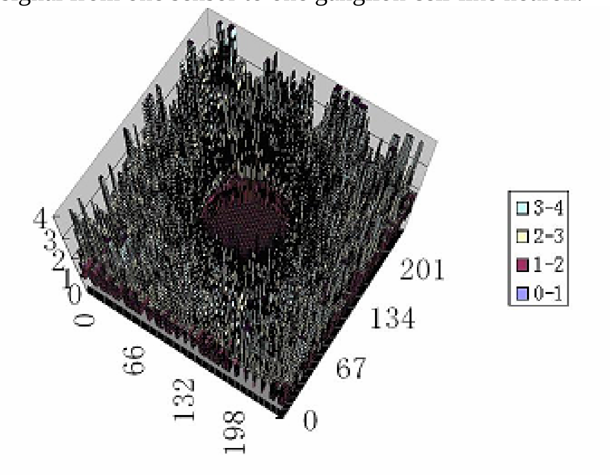


Fig. 4 A coverage statistic of receptive field

DOG model is made up of two Gaussion functions that respond as impulse, they simulate ganglion cell's responding in central and periphery range respectively. A DOG model is

$$DOG(\bar{x}) = \alpha_c G(\bar{x};\delta_c) - \alpha_s G(\bar{x};\delta_s) \qquad (2),$$

where G is a two-dimensional operator, the responding of it on point $\bar{x}(x_1, x_2)$ is

$$G(\bar{x};\delta) = \frac{1}{2\pi\delta^2} e^{-\frac{|\bar{x}|^2}{2\delta^2}} = \frac{1}{2\pi\delta^2} e^{-\frac{x_1^2 + x_2^2}{2\delta^2}},$$

where δ_c and δ_s are standard deviations of center and periphery Gaussion functions, α_c and α_s are two corresponding coefficients about sensitivity. DOG model will show different shape with the changes of δ_c / δ_s and α_c / α_s, as for how these parameters decide DOG model's shape can be found in literature [8].

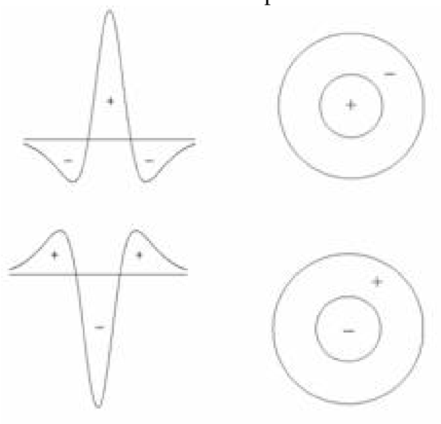
For the sake of the second and third layer of our early visual mode can probe time-continuous change, we adopt continuous time expression of DOG model instead of discrete time one. Let $s(\bar{x}, t)$ be the inputting signal at point $\bar{x}$ and time t, and a ganglion cell-like neuron whose receptive field center is $\bar{x}'(x_1', x_2')$ executes DOG operator, then the responding of this neuron at point $\bar{x}(x_1, x_2)$ and time t is

$$R(\bar{x}, t) = \int_{-\infty}^{+\infty} DOG(\bar{x} - \bar{x}') s(\bar{x}', t) d\bar{x}'.$$

If we define $\bar{x} - \bar{x}' = \bar{\omega}$, then $|\bar{\omega}|$ means the distance between $\bar{x}'$ and $\bar{x}$, thus previous formula can be rewritten as

$$\begin{aligned} R(\bar{x}, t) &= \iint_{|\bar{\omega}|<\infty} DOG(\bar{\omega}) s(\bar{x} - \bar{\omega}, t) dxdy \\ &= \iint_{|\bar{\omega}|<\infty} (\alpha_c G(\bar{\omega};\delta_c) - \alpha_s G(\bar{\omega};\delta_s)) s(\bar{x} - \bar{\omega}, t) dxdy \end{aligned}$$

So base on forgoing theoretic preparation, we can design On-DOG operator with condition $\delta_c < \delta_s$ and Off-DOG operator with condition $\delta_c > \delta_s$, they are corresponding to on-center and off-center cell receptively. Fig. 5 demonstrates these two operators.



$\delta_c < \delta_s$, perspective of vertical and horizontal section

$\delta_c > \delta_s$, perspective of vertical and horizontal section

Fig. 5 On-DOG operator and Off-DOG operator

It is very reasonable to hope that the responding of either On-DOG or Off-DOG operator to a area with equilibrium grayness is zero, in this case any of these two operators should satisfy α_c / α_s =1, because the condition $\iint_{|\bar{\omega}|<\infty} DOG(\bar{\omega}) s(\bar{x} - \bar{\omega}, t) dxdy$ =0 is obligatory. This case is corresponding to both regions of receptive field are illuminated equally and diffusely.

An exception is in fovea, because the size of those receptive fields confined in fovea is 1 sensor, so the outputting of these sensors is transmitted directly to their receptive ganglion cell-like neurons.

Keeping the watchfulness to changes that occur at the marginal area of visual field of is fulfilled by neural network's activities of on-event and off-event exploring. An on-event is like the grayness strengthening in central region of receptive field, and an off-event is like the grayness weakening in that same region. An on-event might be caused by these behaviors like the strengthening of illumination、the emerging of an object with high grayness or the disappearing of an object with low grayness. An off-event might be the weakening of illumination、the disappearing of an object with high grayness or the emerging of an object with low grayness.

We defined $Ron(\bar{x},t)$ to be the respondence of an On-DOG operator to an on-event at position $\bar{x}$. Because an On-DOG operator has $\delta_c < \delta_s$ that make that operator's output is positive at the center of its receptive field and its neighboring region, and in terms of on-event's characteristics the respondence of On-DOG operator at this moment should be greater than that at last moment, that is to say the differential coefficient of $Ron(\bar{x},t)$ should be greater than zero. So we define the conditions under which neural network can judge the occurring of an on-event are:

1. $\dfrac{\partial Ron(\bar{x},t)}{\partial t} > \theta_1$, and
2. $\dfrac{\partial Ron\_c(\bar{x},t)}{\partial t} > \theta_2$,

where $Ron(\bar{x},t) = \iint_{|\bar{\omega}|<\infty} DOG(\bar{\omega}) s(\bar{x}-\bar{\omega},t)dxdy$,

and

$$Ron\_c(\bar{x},t) = \iint_{\{\bar{\omega}|DOG(\bar{\omega})>0\}} DOG(\bar{\omega}) s(\bar{x}-\bar{\omega},t)dxdy$$

.

In order to reduce On-DOG operator's sensitivity to noise and insure the change explored by it is notable, a threshold θ_1 is introduced in condition 1, that means the changing ratio of an On-DOG operator's respondence needs beyond some threshold. In condition 2 $Ron\_c(\bar{x},t)$ is the respondence of an On-DOG operator in its positive output region, a threshold θ_2 is introduced too, thus the changing ratio of that operator's respondence should be greater than θ_2 is in order to insure that there is a notable change emerging in the center of an On-DOG operator's receptive field.

For a same reason, we define $Roff(\bar{x},t)$ to be the respondence of an Off-DOG operator to an off-event at position $\bar{x}$. Because an Off-DOG operator has $\delta_c > \delta_s$ that make that operator's output is negative at the center of its receptive field and its neighboring region, and in terms of off-event's characteristics the respondence of Off-DOG operator at this moment should be greater than that at last moment, that is to say the differential coefficient of $Roff(\bar{x},t)$ should be greater than zero. So we define the conditions under which neural network can judge the occurring of an off-event are:

3. $\dfrac{\partial Roff(\bar{x},t)}{\partial t} > \theta_3$, and
4. $\dfrac{\partial Roff\_c(\bar{x},t)}{\partial t} > \theta_4$,

where $Roff(\bar{x},t) = \iint_{|\bar{\omega}|<\infty} DOG(\bar{\omega}) s(\bar{x}-\bar{\omega},t)dxdy$,

and

$$Roff\_c(\bar{x},t) = \iint_{\{\bar{\omega}|DOG(\bar{\omega})<0\}} DOG(\bar{\omega}) s(\bar{x}-\bar{\omega},t)dxdy$$

.

The reason of introducing threshold θ_3 and θ_4 is same to previous considerations, and $Roff\_c(\bar{x},t)$ is the respondence of an Off-DOG operator in its negative output region.

In order to confine a DOG operator doing convolution only in its receptive field or make it respond weakly enough to those sensors in the outside of its receptive field, we introduce a threshold $\theta_5 > 0$, which is the respondence of a DOG operator at the remotest periphery of its receptive field and should be a positive value and approach to zero. Suppose that the position of the remotest periphery of a receptive field is $\bar{x}_p$, and then formula (3) is like $G(\bar{x};\delta) = \dfrac{1}{2\pi\delta^2} e^{-\frac{|\bar{x}_p-\bar{x}_c|^2}{2\delta^2}} \le \theta_5$, while $\bar{x} \ge \overline{x_p}$.

We now know that if all sensors in a receptive field output equally then their owner ganglion cell-like neuron should output zero, so we introduce another threshold θ_6 to control the difference between a DOG operator's two parts with positive effect and negative effect receptively. If take On-DOG operator as example, the algorithm for selecting δ_c and δ_s is following.

Step 1: For all on-center ganglion cell-like neurons repeat from Step 2 to Step 8;

Step 2: Initialize threshold θ_5 and θ_6;

Step 3: In terms of formula (1) let $|\bar{x}_p - \bar{x}_c|$ equal to the half of receptive field's diameter of an on-center ganglion cell-like neuron, because the curve of $G(\bar{x};\delta)$ is made up of two monotonic and symmetrical sub-curves, so through solving a inequality an appropriate value can be easily selected as δ_s, after that positive effect sub-function $G(\bar{x};\delta_s)$ can be obtained;

Step 4: Suppose that the number of all sensors falling in current receptive field is n, then for all sensors in current receptive field, calculate respondence of $G(\bar{x};\delta_s)$ to all these sensors into variables $rp_1, rp_2, ..., rp_n$;

Step 5: Selected a value less than δ_s as δ_c, after that negative effect sub-function $G(\bar{x};\delta_c)$ can be obtained;

Step 6: Calculate respondence of $G(\bar{x};\delta_c)$ to all those sensors into variables $rn_1, rn_2, ..., rn_n$;

Step 7: If not $|\sum_i^n rp_i - \sum_i^n rn_i| < \theta_6$ then go to Step 5, else go to next step;

Step 8: If all On-DOG operators' parameters are available then end this algorithm, else go to Step 2.

For any Off-DOG operator the way for selecting parameter δ_c and δ_s is similar to foregoing steps, except that $\delta_c > \delta_s$, i.e. the Step 5 should be changed a little.

Table 1 shows some results of On-DOG operator's parameter δ_c and δ_s selecting under different receptive field size.

Table 1 Parameter δ_c and δ_s 's selecting

On-DOG operator	Diameter of receptive field								
	3	5	7	9	11	13	15	17	19
δ_c	0.8	1	1.5	2.2	3	4	5	6	7
δ_s	1	1.2	1.7	2	3.09	4.07	5.06	6.06	7.06
Positive Size	1	3	5	7	9	11	13	15	17
$\lvert\sum_i^n rp_i - \sum_i^n rn_i\rvert$	0.021	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

2.3 Watchfulness holding algorithm

A watchfulness holding algorithm is carried out by the third layer of hierarchy neural network which was demonstrated in Fig. 1. The aim of this algorithm is to explain the ability of balancing, that means while processing subtly the stimulus fallen in the fovea-like area of visual field neural network can at same time explore sudden and outstanding change occurring in non-central area of visual field. Watchfulness holding is achieved by on-event and off-event detecting. The watchfulness holding algorithm is like following.

Step 1: For all ganglion cell-like neurons whose receptive fields' centers are not in fovea, carry out from Step 2 to Step 5;

Step 2: Calculate these measures like $Ron(\bar{x},t), Ron(\bar{x},t-1), Roff(\bar{x},t), Roff(\bar{x},t-1)$, and $Ron\_c(\bar{x},t), Ron\_c(\bar{x},t-1), Roff\_c(\bar{x},t), Roff\_c(\bar{x},t-1)$;

Step 3: Calculate some variances like $\Delta Ron, \Delta Roff, \Delta Ron\_c, \Delta Roff\_c$;

Step 4: If $\Delta Ron > \theta_1$ and $\Delta Ron\_c > \theta_2$ then $R(\bar{x},t) = \Delta Ron$, else $R(\bar{x},t) = 0$;

Step 5: If $\Delta Roff > \theta_3$ and $\Delta Roff\_c > \theta_4$ then $R(\bar{x},t) = \Delta Roff$, else $R(\bar{x},t) = 0$;

Step 6: Any neuron in third layer is similar to neuron in second layer, each of them has a receptive field in second layer and receives that receptive field's outputs as its inputs, these inputs might come from both On-DOG operators and Off-DOG operators. And calculate measure $\prod_{i=1}^n \left|\frac{dR}{dt}\right|$, where n is the number of ganglion cell-like neurons in its receptive field, this measure means to magnify the varying ratio of On-DOG and Off-DOG operators' respondence to this part of visual field, so as to let the part of visual field with great variances be able to pop out. And there are transverse connections existing between third layer neurons to induce competitions among neurons in third layer, thus only the neuron with biggest measure can survive, that neuron would point out the right position or direction that sudden change took place in.

Thus our neural network can keep guard to periphery area of visual field with low hardware and time cost, that position or direction that biggest change took place in is a clue for attention focus moving for the sake of further intensive processing. In our program any ganglion cell-like neuron in second layer is designed to execute both On-DOG and Off-DOG operators, that means this neuron can detect both on-event and off-event, after that transmits its respondence upwards to third layer for decision-making, and third layer chooses most activated neuron as the projecting reference of moving focus.

3. The analysis of experiment results

3.1 The background of applications

Object detection is the most important application of artificial vision system, and the majority of practical usages can be found in military. For instance infrared controlling and guiding can be divided into two types, one is hotspot-guided, and another is imaging-guided. The former is out of date for its easiness of being jammed. The primary component of an infrared guiding instrument is a focal plane array, which is a sensors array and is able to capture the contour of an object instead of a simple spot. A focal plane array cannot be cheated by infrared bait, the sensing distance of it is long, the sensitivity of it is high, and can detect object under tanglesome background. For these advantages large-scale focal plane array can be used widely in fire controlling system, missile's guiding head, scouting system, night watching system, automatic and active defence system etc. Many smart weapons coming into active service will be equipped with this kind of apparatus, such as new generation air-to-air missile IRIS-T of Germany Air Force and ASRAAM in U.K. Air Force, or Gill anti-tank missile of Israel Army and Javelin anti-tank missile of U.S.A Army. Some infrared guiding head still can capture goal even if it had departed the missile's line of sight more than 90 degree. The key technology that casts this advanced "launching then leaving" manner is digging out an any-position-possible goal from a large visual field. Due to many obvious reasons a goal might not always be imaged at the center of artificial retina, it might be projected onto marginal area of artificial retina and continues moving there frequently, the method that we develop in this paper is just aiming at detecting changing happens to emerge in marginal area.

3.2 Experiment results and analysis

We now demonstrate foregoing algorithms and neural network's ability in watchfulness keeping. In Fig. 6 there is a sequence of continues images lasting 0.5 second, in which a paper plane (pointed by a white arrow) is flying from right-up to left-middle across a complicated background.

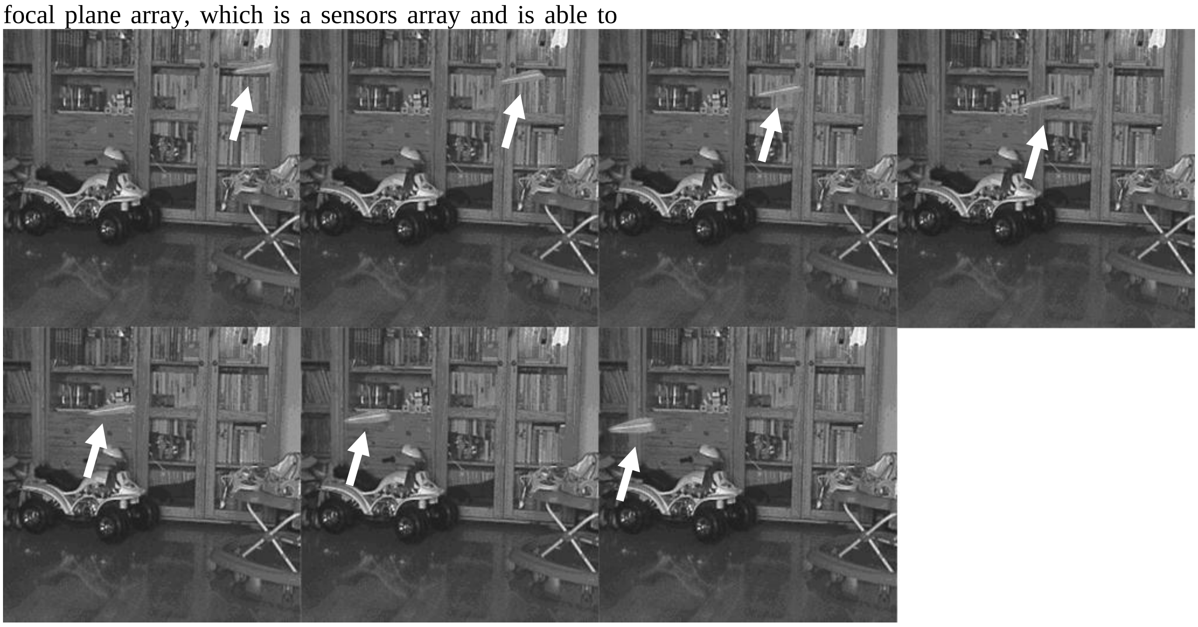


Fig.6 A images sequence of experiment

In paper flying experiment, that flying object did not pass the fovea area, so it was perceived by ganglion cell-like neurons in marginal area. The outputs from on-center and off-center neurons is demonstrated in Fig.7 through those groups (pointed by black arrow) of dots but except any dot in central circle, a dot stands for the ganglion cell-like neuron at that position has non-zero output in terms of algorithms in section 2. We can find that the more adjacently that flying object approaches central area, the more neurons that can perceive it and the more densely dots assemble, and the more vividly flying object's contour is outlined.

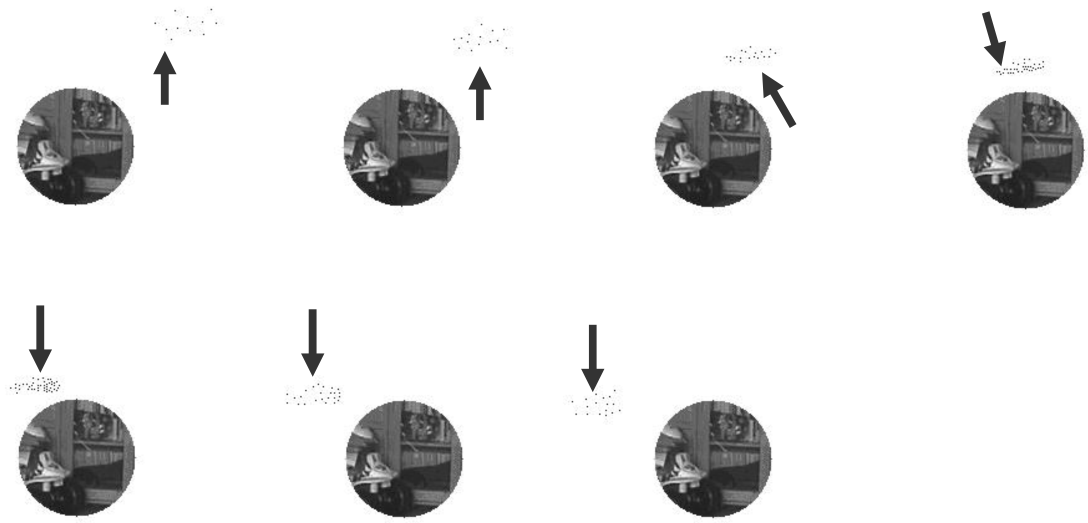


Fig.7 The output sequence of on-center and off-center ganglion cell-like neurons

Based on outputs from second layer, third layer proceeded detecting at which position most outstanding change had taken place through algorithm in section 2.3. The results out from third layer is demonstrated in Fig. 8 one by one frame, each arrow along z axis marked the position of change taking place, and there was no change in central area, so the output of third layer at those points is empty.

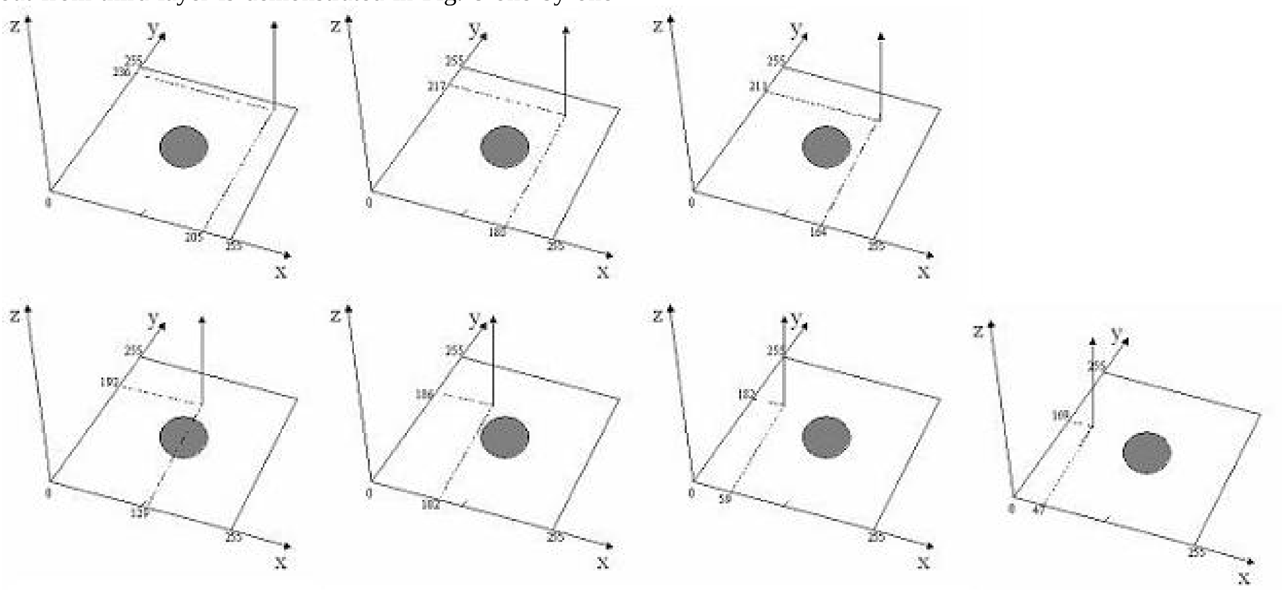


Fig. 8 The position at which change was detected

If we compound those positions in a working memory according to their time orders, then we can acquire an impression of moving track, which is shown in Fig. 9.

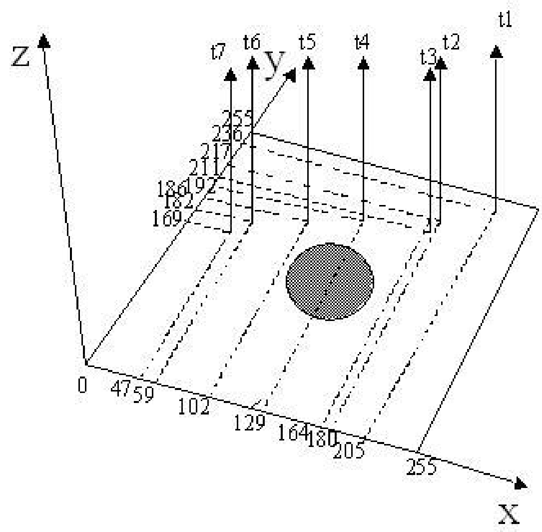


Fig.9 Moving track is stored in a working memory

According to previous experiment results, this artificial early vision model, based on ganglion cells' information processing manner in biological retina, can detect outstanding change taking place in the periphery area of visual field effectively, and it provides a possibility to shift focal attention and trace object's moving continuously through spatial position clue. This artificial early vision model could hold watchfulness to marginal area only using a little computational resource firstly, once object emerged suddenly, it could shift its attention focus rapidly and then process new stimulus elaborately by much denser computational power.

4. Discussion

Real-time image processing applications have tremendous computational workloads and I/O throughput requirements, especially while operation is done in mobile or portable devices with stringent resource limitations (size, weight, and power). Large computational workloads and I/O requirements characterize real-time image processing applications. Typical applications may require 10-1000 Gops/s, which cannot be matched by today's general-purpose microprocessor capabilities (0.2-0.7 Gops/s) [19]. In addition, the traditional sequential access to image data further reduces the available performance.

The design method based on physiological neural information processing structure and mechanism is very promising in parallelization and consistent with hardware simulation. Experiment results support the design choices and suggest that there is the potential for real-time execution of complex, multistage applications. At same time this kind of physiological mechanism based model has a good cognitive significance, it is not only helping computer researchers to design new model but also able to give hints to physiologists on the necessary processes of information acquisition, storage and use.

Acknowledgments

This paper was supported by Open Project of Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, under grant no. IIP2003-3, and Youth Science Foundation, Fudan University, under grant no. 01-13.

Prof. SHOU Tian-De and Prof. GU Fan-Ji from Life Sciences School in Fudan University, and Prof. LI Bao-Ming from Institute of Neural Biology in Fudan University have given us many instructions in neural information processing.

References

[1] Shou TianDe, The brain mechanism of visual information processing, Shanghai: Shanghai Science & Technology Education Publishing Company, 1997, 107-138 (in Chinese)
[2] Schneider A M, Trarshis B., Elements of physiological psychology, McGraw-Hill Inc., 1995, 209-255.
[3] [Kandel E R, Schwartz J H, Jessell T M., Principles of Neural Science, McGraw-Hill Inc., 2000, 507-522.
[4] R. W. Rodieck, The first steps in seeing, Sinauer Associates, Inc., 1997, 87-124.
[5] Kuffler,S.W., Nicholls,J.G., From Neuron to Brain, Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, 1976.
[6] William R. Hendee, Peter N.T. Wells, The perception of visual information, Springer-Verlag New York, Inc., 1997.
[7] Wei Hui, He XinGui, A Hierarchical Model Based On Visual Cortex, Chinese J. Computer, 2000, Vol.23(6):620~628 (in Chinese)
[8] Fleet,D.J., The Early processing of Spatio-Temporal Visual Information, Master's Thesis, University of Toronto,1984.
[9] Aapo Hyvärinen, Patrik O. Hoyer, A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images, Vision Research, 2413-2423.
[10] Bundesen C., A theory of visual attention, Psychological review, vol. 97, 1990, 523-547.
[11] DeAngelis G. C, Ghose G. M, Ohzawa I., Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. Journal of neuroscience, 1999, 4046-4064.

[12] Croner, L. J., Kaplan, E., Receptive fields of P and M ganglion cells across the primate retina, Vision Research, Vol.35(1), 7-24.

[13] B. Skottun, R. De Valois, Classifying simple and complex cells on the basis of response modulation, Vision Research, vol. 31, 1079-1086.

[14] R. Szulborski, L. Palmer, The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells, Vision Research, vol. 30, 249-254.

[15] Culhane S., M, Tsotsos J. k., A prototype for data-driven visual attention, International conference on pattern recognition, vol. A, 1992, 36-39.

[16] DeAngelis G. C, Ghose G. M, Ohzawa I., Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. Journal of neuroscience, 1999, 4046-4064.

[17] Wei Hui, He XinGui, The simulations of function columns in interotemporal cortex by the rational least square approximation, Pattern Recognition & Artificial Intelligence, 1998, Vol.11(2):125-131 (in Chinese)

[18] Yang XianBo, Wei Hui, Focus Image Analysis Based on Multi-size Receptive Field Model, Journal of Computer Engineering, Vol.29(6), 2003, 53~55 (in Chinese)

[19] D. S. Wills, J. M. Baker, H. H. Cat, S. M. Chai, L. Codrescu, J. L. Cruz-Rivera, J. C. Eble, A. Gentile, M. A. Hopper, W. S. Lacy, A. Lòpez-Lagunas, P. May, S. Smith, T. Taha, “Processing Architectures for Smart Pixel Systems”, IEEE Journal of Selected Topics in Quantum Electronics, vol.2 (1), April 1996, p.24-34

WEI Hui was born in November 1971. He received his PhD in Department of Computer Science at Beijing University of Aeronautics and Astronautics in July 1998. From September 1998 to November 2000 he was a postdoctoral in the Department of Computer Science and Institute of Artificial Intelligence at Zhejiang University. In November 2000 he jointed the Department of Computer Science and Engineering at Fudan University as an associate professor. His research interests include artificial intelligence and cognitive science.

YANG Xian-bo was born in April 1975. He got his B.S. degree in computer science from Nanjing Economic College in 1998, and M.S degree in computer science from Fudan University in 2003. His research interest is computer vision.