

有指向性的视觉注意计算机模型

赵宏伟 王 慧 刘萍萍 戴金波
(吉林大学计算机科学与技术学院 长春 130012)
(吉林大学符号计算与知识工程教育部重点实验室 长春 130012)
(zhaohw@jlu.edu.cn)

A Computer Model of Directional Visual Attention

Zhao Hongwei, Wang Hui, Liu Pingping, and Dai Jinbo
(College of Computer Science and Technology, Jilin University, Changchun 130012)
(Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012)

Abstract Attention assigns the limited processing resources to the information which needs processing finely. This feature can enhance the detection ability and response speed of the visual information processing. In the past few years, the allocation mechanism of attention has received a lot of attention and many approaches have been developed. The authors propose a computer model of directional visual attention, which can simulate biological visual attention systems based on their physiological structure. The proposed model has three layers: visual image inputting layer, object encoding layer and control layer. Visual image inputting layer transforms visual image into retina image through imaging mechanisms of biological retina and adjusts the radius of fovea to cover the attention object currently. Object encoding layer encodes objects in retina image by edge detection with maximum gradient and c -means algorithm firstly, then extracts the fundamental information about each object including color, central, set of edge pixels, etc. Control layer transfers the attention focus according to the characteristics of directional objects in knowledge base. In the control layer the space information is used, the object which is nearest from the focus of attention currently in space position has high priority in processing. Experimental results indicate that the attention focus can be transferred to directional objects with the proposed model.

Key words information processing; computer model of visual attention; k -coefficient method; attention; edge detection

摘 要 注意把有限的处理资源优先分配给那些需要精细加工的信息,能提高视觉信息加工中的检测能力和响应速度. 基于生物视觉系统的生理结构特点,建立了模拟生物视觉注意系统的有指向性的视觉注意计算机模型. 模型首先模拟生物视网膜的成像机制,将视场图像转化为视网膜图像;然后将最大梯度边缘检测和 c -均值聚类等方法相结合,对视网膜图像中的目标进行编码,分别提取每个目标的颜色、中心以及边缘点集合等基本信息;最后用知识库中指向性目标的特征来指导注意焦点的转移. 实验结果表明,利用此模型能较好地实现注意焦点的转移.

关键词 信息处理; 视觉注意计算机模型; k -系数法; 注意; 边缘检测

中图法分类号 TP391

人在同一时间从外界获得的信息是大量的,我们不能同时感知周围的一切对象,而只能感知其中少数对象,这种现象叫作注意机制.注意机制的存在使得外界中有用的、我们感兴趣的信息被精细加工,那些无用的信息被抛弃掉.注意把有限的处理资源优先分配给那些需要精细加工的信息,能提高视觉信息加工中的检测能力和响应速度,近几年来,已成为一个热点课题.

当大量信息涌入人脑时,哪些信息是有用的,是需要被处理的,哪些信息是应该被忽略的,取决于此时大脑的特定任务的要求.目前普遍认为视觉注意的选择主要有2种理论观点:基于空间(space-based)的和基于目标(object-based)的视觉注意理论^[1].基于空间的视觉注意理论认为,视觉注意选择的是特定的空间位置,然后才对落在所选空间内的物体进行加工;而基于目标的视觉注意理论认为,视觉注意选择的是整个的物体,而不管物体所处的空间位置^[2].

基于空间的注意计算模型^[3-6],虽然能很好地利用空间信息,但注意焦点的转移是在空间上从一个区域转移到另一个区域,没有考虑到所转移区域中目标的完整性,这种转移机制使得注意焦点有时会转移到没有明确意义的区域^[7].另一方面,一些学者基于格式塔知觉组织理论提出了基于目标的注意计算模型^[8-9],这些模型考虑到了目标的完整性,以单个目标的整体作为处理的对象,但是能有效地应用空间信息的比较少.

本文旨在前人工作的基础上,将基于空间的视觉注意理论和基于目标的视觉注意理论从某种程度上相结合,试图建立更适合视觉信息处理的注意计算模型来模拟有指向性的生物注意转移机制.首先假设指向性目标的基本特征已经存在于知识库当中,用知识库当中的内容来指导注意焦点的转移.

本文的注意计算模型共分为3层:第1层为视觉图像输入层,将视场图像转化为视网膜图像,调整中央凹的半径以把当前注意的目标正好覆盖上;第2层为目标编码层,使用最大梯度边缘检测算法和 c -均值聚类等方法,获得视网膜图像中所有目标的基本特征(其中包括颜色、形状、中心以及边缘点集合);第3层为控制层,控制注意焦点的转移.由于注意资源有限,视场中的目标相互竞争注意资源.在处理时,把空间信息加进来,空间位置上离当前注意焦点近的目标首先被处理,若被处理目标的属性与指向性目标的属性相同,则停止处理,计算空间位置坐

标,转移注意焦点,若不同则处理下个目标直至相同为止.

1 视觉图像输入层

在生物视觉系统中,视觉信息处理的第1步发生在视网膜,视信息在视网膜上形成视觉神经冲动,沿视路将视信息传递到视中枢形成视觉.视网膜的生理结构特点决定了视网膜的分辨力是不均匀的,在视野中心(中央凹),其分辨能力最强,产生最清晰的视觉,当由中央凹逐渐往视野外围测试时,其分辨能力会急剧下降.

当人眼观察一个物体时,人会通过眼动把被观察的物体映像到中央凹处,以获得被观察物体的细节.基于这种想法,把当前注意的目标作为视野中心(中央凹),目标的大小作为中央凹的直径,目标处的图像最清晰,目标以外区域的清晰度随远离中央凹而降低.在图像初次输入系统的时候,没有注意目标,此时以图像中心1/3处作为中央凹进行处理,此后均以第3层反馈的注意目标作为中央凹区域.

设输入的图像为 $M \times N$ 像素的彩色RGB图像,则此时图像的中心点像素为 (x_0, y_0) ,其中 $x_0 = M/2, y_0 = N/2$.由人眼的生理结构特点决定,落在视网膜上的图像是圆形的,因此这一层的输出是以 $\min(M, N)/2$ 为半径以图像中心为中心的圆形视网膜图像.中央凹的半径 r 取决于目标的尺寸,若被注意的目标的边缘点集合为:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\},$$

则

$$r = \max_i \{ \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \}, \quad (1)$$

$$i = 1, 2, \dots, n.$$

(图像初次输入系统的时候,没有注意目标,此时 $r = \min(M/2, N/2)/3$).中央凹内部的图像最清晰,距离中央凹越远的地方图像越不清晰.不清晰表现在目标边缘模糊、颜色不鲜明上,目前大多数模型采用降低分辨率的办法来模拟这一过程,但是分辨率降低会影响处理的效果,因此本文考虑从颜色上着手解决这一问题.RGB的图像中每个像素点的颜色均用 R, G, B 三个分量的值共同表示,每个分量均有256个等级.把从中央凹边缘到图像边缘像素点的 R, G, B 三个分量值分别乘以一个逐步减小的系数 k 来模拟这一过程,这里称作 k -系数法.若用 $(R, G, B)_{(x, y)}$ 代表点 (x, y) 的颜色值,则图像上像素点 (x, y) 的像素值为:

$$(R,G,B)_{(x,y)} = \begin{cases} (R,G,B)_{(x,y)}, & (x-x_0)^2 + (y-y_0)^2 \leq r; \\ (R \times k/256, G \times k/256, B \times k/256)_{(x,y)}, & (x-x_0)^2 + (y-y_0)^2 > r; \end{cases} \quad (2)$$

其中,

$$k = (\min(M,N)/2 - \sqrt{(x-x_0)^2 + (y-y_0)^2} + \eta) \times 256 / (\min(M,N)/2 - r + \eta),$$

η 为可调参数,用于调节视网膜图像周边的明暗程度.

2 目标编码层

为了更有效地对目标进行编码,希望获得的目标边缘信息尽可能少地包含目标内部的细节信息以及背景噪声,同时尽可能充分地保留目标的轮廓边缘信息.传统的边缘检测算法可以很好地得到物体的边缘,但容易得到过多的图像细节,即容易把目标内部的线条误作为边缘;而用颜色聚类的方法对图像进行聚类,对于颜色连续变化的物体可以很好地忽略其中的细节,但却容易把空间上不属于同一个物体的像素点划为一类,用这种方法有时得到的边缘也不准确.本文把2种方法的优点相结合,把从第1层传来的视网膜图像分别用最大梯度边缘检测算法提取出的边缘点集合作为第1组边缘信息;用c-均值聚类算法进行颜色聚类,得到不同颜色的点集合,在每个集合中用8个方向模板进行筛选,得到不同颜色的边缘点集合,把所有颜色的边缘点集合作为另一组边缘信息.把第2组边缘信息中每个颜色的边缘点集合都与第1组边缘点集合取加权子集合作为单个目标的最终边缘点集合.

设视觉场景中共有 N 个目标,它们的颜色各不相同,把它们编号为 $1, 2, \dots, N$. 对视网膜图像使用最大梯度边缘检测算法提取出第1组边缘信息为1个边缘点集合 $A'_1, A'_1 = \{(x'_1, y'_1), (x'_2, y'_2), \dots, (x'_n, y'_n)\}$. 用c-均值聚类算法进行颜色聚类将会得到目标的颜色数目加上背景颜色的 $N+1$ 个颜色点集合,用颜色直方图估计此时输入图像的背景颜色,并在颜色点集合集中去掉这个颜色的集合.这样得到的颜色点集合集中有 N 个集合,分别对应 N 个目标.

设对应于视觉场景中目标 $i (i \in \{1, 2, \dots, N\})$ 的颜色点集合为 C_i ,用图1中的8个方向模板进行筛选,其中“1”代表该点在点集 C_i 中,相反“0”代表

该点不在点集 C_i 中,符合这8个方向模板中的任何一种情况的中心点被标记成边缘点.

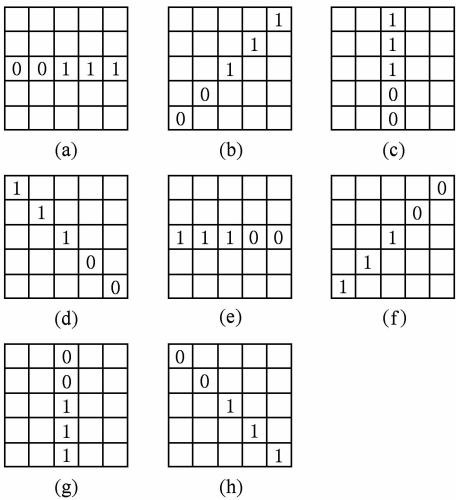


Fig. 1 Eight Oriented Templates. (a) Zero degree oriented templates; (b) Forty five degree oriented templates; (c) Ninety degree oriented templates; (d) One hundred and thirty five degree oriented templates; (e) One hundred and eighty degree oriented templates; (f) Two hundreds and twenty five degree oriented templates; (g) Two hundreds and seventy degree oriented templates; and (h) Three hundreds and fifteen degree oriented templates.

图1 8个方向模板. (a) 0°方向模板; (b) 45°方向模板; (c) 90°方向模板; (d) 135°方向模板; (e) 180°方向模板; (f) 225°方向模板; (g) 270°方向模板; (h) 315°方向模板

之后分别判断每个边缘点的8个方向上外2周的点是否为边缘点(以45°方向为例,若设当前点坐标为 (x, y) ,则判断点 $(x+1, y-1)$ 和点 $(x+2, y-2)$ 是否是边缘点,其中有1个是边缘点则认为该方向是连通的),循环递归,找出图像中所有闭合的边缘.

设用上述方法得到的目标 i 的边缘点集合为 $A'_{i2}, A'_{i2} = \{(x''_{i1}, y''_{i1}), (x''_{i2}, y''_{i2}), \dots, (x''_{im}, y''_{im})\}$,第2组边缘信息为 N 个边缘点集合组成的边缘点集合 $A'_2, A'_2 = \{A'_{12}, A'_{22}, \dots, A'_{i2}, \dots, A'_{N2}\}$.

把第2组边缘信息中每个边缘点集合都与第1组边缘信息点集合取加权子集合作为单个目标的最终边缘点集合.考虑到 A'_1 是整个图像的边缘点集合,而 A'_{i2} 是整个图像中目标 i 的边缘点集合,所以 m 小于 n . 目标 i 的最终边缘点集合 A_i 为用这2种方法得到的边缘点集合中欧氏距离最小的点对 x 和 y 坐标的加权值所组成的点的集合,为排除孤立点的干扰,当最小欧氏距离大于给定阈值时,不作为

边缘点. 这里取平均值, 设满足条件的边缘点对共有 l 对, 则

$$A_i = \{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \dots, (x_{il}, y_{il})\}, \quad (3)$$

其中 $x_{ij} = \frac{x'_j + x''_{ij}}{2}, y_{ij} = \frac{y'_j + y''_{ij}}{2}, j = 1, 2, \dots, l$. 点 (x''_{ij}, y''_{ij}) 距离点 (x'_j, y'_j) 的欧氏距离最小, 即 $O\{(x''_{ij}, y''_{ij}), (x'_j, y'_j)\} = \min\{O\{(x''_{ij}, y''_{ij}), (x'_1, y'_1)\}, O\{(x''_{ij}, y''_{ij}), (x'_2, y'_2)\}, \dots, O\{(x''_{ij}, y''_{ij}), (x'_n, y'_n)\}\},$

并且 $O\{(x''_{ij}, y''_{ij}), (x'_j, y'_j)\} < Threshold$.

通过目标 i 的最终边缘点集合 A_i 可以计算得到目标 i 的目标中心 O_i :

$$O_i = (\sum_{j=1}^l x_{ij}/l, \sum_{j=1}^l y_{ij}/l). \quad (5)$$

目标 i 的颜色为由 c -均值聚类算法迭代计算出的聚类中心. 在整个视野中所有 N 个目标的边缘点集合, 目标中心和目标颜色, 作为下一层的输入.

3 控制层

由于注意资源有限, 视场中的目标相互竞争注意资源. 假设指向性目标的基本特征已经存在于知识库当中, 用知识库中的内容来指导注意焦点的转移. 在处理时, 首先由第 2 层输入的目标中心集合计算出目标中心点距离当前注意焦点的距离集合, 所有目标的颜色, 与焦点的距离和边缘点集合作为初始数据进行处理. 空间位置上距离当前注意焦点近的目标首先被处理, 先判断目标颜色与指向性目标的颜色是否相同, 如果相同则进一步判断形状, 若不同则转到处理下个目标. 若找到指向性目标, 则计算空间位置坐标, 转移注意焦点. 在处理过程中, 考虑到注意的禁止返回机制, 处理过的目标被标记上, 不再进行处理, 本算法中采用删除该目标的信息达到不再被处理的目的. 具体算法如下:

1) 设视场中未处理的目标集合为 $Object$, 目标集合的初值为:

$$(Object)_{[init]} = \{1, 2, \dots, N\};$$

目标的颜色集合为 $Color$:

$$(Color)_{[init]} = \{c_1, c_2, \dots, c_N\},$$

目标中心点距离当前注意焦点的距离集合为 $Distance$:

$$(Distance)_{[init]} = \{d_1, d_2, \dots, d_N\};$$

2) 判断目标集合 $Object$ 是否为空, 若为空, 终止; 否则程序执行, 表示在视场中未找到指向性目

标, 若不为空, 取出距离当前注意焦点最近的目标的编码. 如果此时距离集合中元素的最小值等于 d_i , 则在集合 $Distance$ 中删去 d_i , 转下一步;

3) 从 $Color$ 集合中取出目标 i 的颜色信息 c_i , 与存储在知识库中的指向性目标的颜色比较, 并在集合 $Color$ 中删去 c_i , 若比较结果相同则转下一步, 否则, 在集合 $Object$ 中删去 i , 转 2);

4) 从目标 i 的边缘点集合得到目标 i 的边缘信息, 与知识库中的指向性目标边缘相比较, 若相同, 找到指向性目标, 则根据目标中心坐标, 转移注意焦点, 终止程序执行, 否则, 转 2) 取下一个待处理目标.

4 实验结果

实验采用的原始图为 523×395 像素的彩色图像, 由人眼的生理结构特点决定, 输出的视网膜图像是圆形的, 如图 2 所示. 图 2(a) 为实验原始图, 是世界地图的一部分, 其中各大板块颜色各不相同, 当参数 $\eta = 20$, 以图像中心 $1/3$ 处作为中央凹进行处理时, 模型第 1 层输出的图像为 395×395 像素的视网膜图像, 如图 2(b) 所示. 可以看出图像的中央及其附近的颜色和边缘得到很好的保留, 越往外围颜色越暗淡, 边缘越不清晰, 与人的视网膜生理结构特点相吻合.

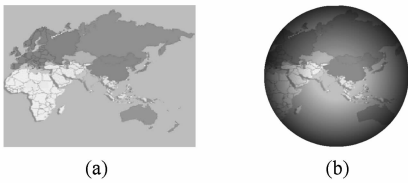


Fig. 2 Experimental result image of retina. (a) Original image and (b) Result image.

图 2 视网膜实验结果. (a) 原图; (b) 实验结果图

图 3(a) 为用最大梯度边缘检测算法对视网膜图像图 2(b) 提取边缘的结果图, 图 3(b) 为对图 3(a) 中像素点值取阈值为 20 时的边缘检测结果图. 可以看出, 虽然目标的边缘都被有效地提取出来了, 但是其中有很多本属于一个目标的内部线条也被误作为了边缘.

用 c -均值聚类算法对视网膜图像进行颜色聚类, 聚类的结果共为 7 组, 分别如图 4 所示. 为便于显示, 把属于一类的像素点置为白色, 其余的置为黑色.

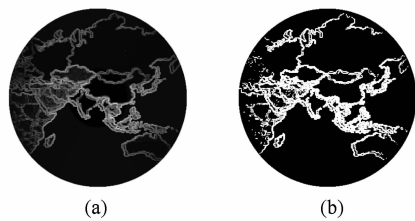


Fig. 3 Experimental result images of edge detection with maximum gradient. (a) Result image of edge detection with maximum gradient to Fig. 2(b) and (b) Result image with threshold equal to 20.

图3 最大梯度边缘检测算法实验结果图. (a) 用最大梯度边缘检测算法对图 2(b)提取边缘的结果图; (b) 阈值为 20 时的边缘检测结果图

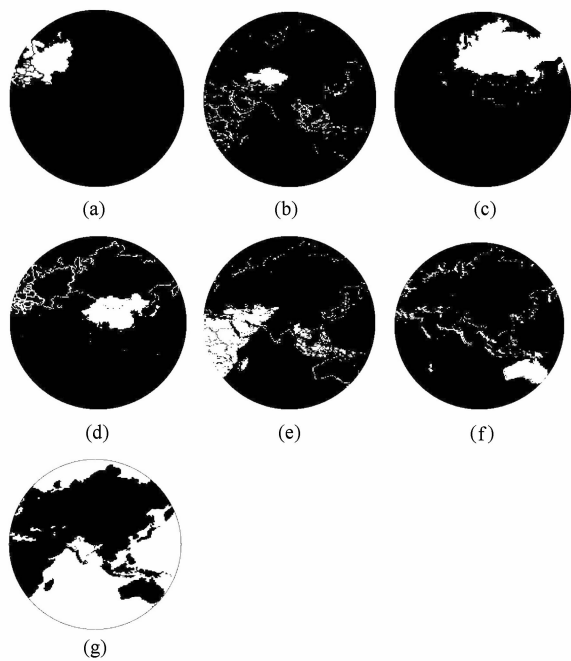


Fig. 4 Experimental result images of *c*-means algorithm. (a)~(g) Each result image.

图4 *c*-均值聚类算法实验结果图. (a)~(g) 为单个聚类结果

用颜色直方图估计此时图像的背景颜色为图 4(g)所对应的类,在颜色点集合集中去掉该图对应的集合. 对于剩下的 6 个目标颜色点集合用文中方法得到的目标的边缘分别如图 5 所示.

实验时设置知识库当中的指向性目标为输入图像中右下角的深色大洋洲板块,按照文中第 3 层所述算法,找到指向性目标,计算目标中心,并实现注意焦点的转移. 实验结果如图 6 所示,利用本文模型能较好地实现注意焦点的转移.

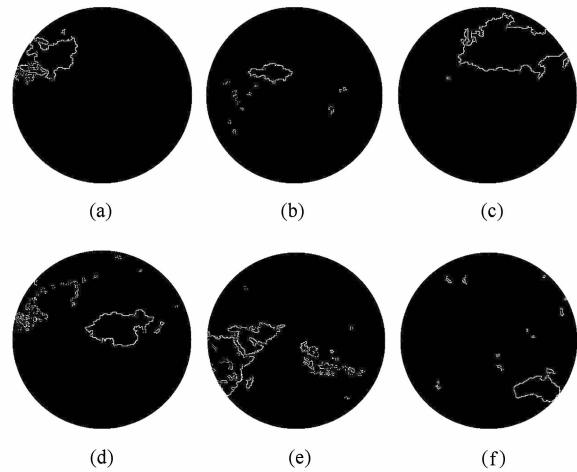


Fig. 5 Experimental result diagram of second layer of proposed model. (a)~(f) Edge of each object.

图5 模型第 2 层实验结果图. (a)~(f)为单个目标边缘

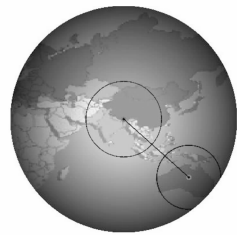


Fig. 6 Experimental result diagram of attention focus transferred.

图6 注意焦点转移实验结果图

5 总 结

本文建立了一种模拟生物视觉注意系统的有指向性的视觉注意计算机模型. 3 个层次的模型分别为视觉图像输入层、目标编码层和控制层. 用世界地图的一部分作为实验原始图,对文中提出的模型进行了仿真实验,给出了各部分的实验结果. 综合实验结果表明,利用本文模型能较好地实现注意焦点的转移. 后续的研究工作将侧重在以下 2 个方面: 1) 模型假设指向性目标的基本特征已经存在于知识库当中,进一步构建能有效模拟人脑知识搜索功能的层次作为模型第 4 层,将更符合生物感知的整个过程; 2) 颜色和形状是确定目标的主要依据,在某些情况下可能会产生局限性,结合纹理特征作分析应是可行的,选择更多的典型实例进行实验,拓展算法的应用范围和普适性将是以后工作的重点.

参 考 文 献

- [1] Harris L R, Jenkin M. Vision and Attention [M]. Berlin: Springer, 2001: 1-14
- [2] Fink G R, Dolan R J, Halligan P W, et al. Space-based and object-based visual attention: Shared and specific neural domains [J]. Brain, 1997, 120(11): 2013-2028
- [3] Rybak I A, Gusakova V I, et al. A model of attention-guided visual perception and recognition [J]. Vision Research, 1998, 38(15/16): 2387-2400
- [4] Itti L, Koch C. Computational modeling of visual attention [J]. Nature Reviews Neuroscience, 2001, 2(3): 194-203
- [5] Treisman A. Features and objects: The fourteenth Bartlett Memorial Lecture [J]. Quarterly Journal of Experimental Psychology, 1988, 40(2): 201-237
- [6] Tian Mei, Luo Siwei, Huang Yaping, et al. Extracting bottom-up attention information based on local complexity and early visual features [J]. Journal of Computer Research and Development, 2008, 45(10): 1739-1746 (in Chinese)
(田媚, 罗四维, 黄雅平, 等. 基于局部复杂度和初级视觉特征的自底向上注意信息提取算法[J]. 计算机研究与发展, 2008, 45(10): 1739-1746)
- [7] Zou Qi, Luo Siwei, Zheng Yu. A computational model of object-based attention using multi-scale analysis and grouping [J]. Acta Electronica Sinica, 2006, 34(3): 559-562 (in Chinese)
(邹琪, 罗四维, 郑宇. 利用多尺度分析和编组的基于目标的注意计算模型[J]. 电子学报, 2006, 34(3): 559-562)
- [8] Sun Y, Fisher R. Object-based visual attention for computer vision [J]. Artificial Intelligence, 2003, 146(1): 77-123
- [9] Deng Y, Manjunath B S. Unsupervised segmentation of color-texture regions in images and video [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(8): 800-810

Research Background

Attention assigns the limited processing resources to the information which needs processing finely. This feature can enhance the detection ability and response speed of the visual information processing. When a large number of information come into brain, which information is useful and needs processing and which information should be ignored depend on the special task of our brains. In the past few years, the allocation mechanism of attention has received a lot of attention and many approaches have been developed. It is known that visual attention has two main methods which are space-based attention and object-based attention at present. But each method has both pros and cons. The authors connect space-based attention with object-based attention to a certain extent, build a model of attention to simulate biological directional visual attention system more rationally. Supposing the basic characteristics of directional object are in knowledge base, the model uses the information in the knowledge base to guide the transference of focus of attention. Experimental results indicate that the attention focus can be transferred to directional objects with the proposed model. Our work is supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20050183032), and the Research Foundation of Education Bureau of Jilin Province, China (No. 2004150).



Zhao Hongwei, born in 1962. Professor and PhD supervisor. His main research interests focus on intelligent information system and thinking computation.

赵宏伟, 1962年生, 博士, 教授, 博士生导师, 主要研究方向为智能信息系统。



Wang Hui, born in 1982. PhD candidate in the College of Computer Science and Technology, Jilin University. Her research interests focus on intelligent information system and thinking computation.

王 慧, 1982年生, 博士研究生, 主要研究方向为智能信息系统。



Liu Pingping, born in 1979. PhD candidate in the College of Computer Science and Technology, Jilin University. Her main research interests focus on intelligent information system and thinking computation.

刘萍萍, 1979年生, 博士研究生, 主要研究方向为智能信息系统、思维计算。



Dai Jinbo, born in 1971. PhD candidate in the College of Computer Science and Technology, Jilin University. Her main research interests focus on intelligent information system and thinking computation.

戴金波, 1971年生, 博士研究生, 主要研究方向为智能信息系统、思维计算。