

## 人类视觉注意机制在目标检测中的应用

桑 农<sup>1</sup>, 李正龙<sup>1</sup>, 张天序<sup>2</sup>

- (1. 华中科技大学 图像识别与人工智能研究所, 湖北 武汉 430074;
2. 华中科技大学 图像信息处理与智能控制教育部重点实验室, 湖北 武汉 430074)

**摘要:**根据人类视觉感知理论,在介绍了两种比较有代表性的视觉注意模型的基础上,采用 bottom-up 控制策略的预注意机制和 top-down 控制策略的注意机制,提出了一种适用于自动目标识别的目标检测算法。从输入图像出发,采用 Gabor 算子建立多尺度、多方位的多通道图像,通过全波整流和各通道间的对比度增益控制,得到多尺度、多方位的方位特征图,这些特征图的线性组合则为显著性图。给出了仅采用 bottom-up 控制策略的船舶目标检测实验结果,待检测目标在显著性图中得到明显增强,有利于检测的实现。

**关键词:** 注意机制; 目标检测; 显著性图

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 1007-2276(2004)01-0038-05

## Applications of human visual attention mechanisms in object detection

SANG Nong<sup>1</sup>, LI Zheng-long<sup>1</sup>, ZHANG Tian-xu<sup>2</sup>

(1. Institute for Pattern Recognition and Artificial Intelligence,

Huazhong University of Science and Technology, Wuhan 430074, China;

2. Key Laboratory of Ministry of Education for Image Processing and Intelligent Control,

Huazhong University of Science and Technology, Wuhan 430074, China)

**Abstract:** Two models of visual attention which are consistent with human visual perception are introduced. Based on these two models, an ATR algorithm employing attention mechanisms with bottom-up and top-down control strategies is developed. A multi-channel representation of the input image was obtained by a bank of Gabor filters, corresponding to multiple scales and multiple orientations. Full-wave rectification of each channel and contrast gain control among the channels were performed sequentially to get orientation feature maps with multiple scales and multiple orientations. The so-called salient map was the linear combination of these orientation feature maps. The experimental result with ship detection is given by use of bottom-up control strategy only. The objects for detection are effectively enhanced on the result map.

**Keywords:** Attention mechanisms; Object detection; Saliency maps

收稿日期:2003-06-20; 修订日期:2003-09-10

作者简介:桑农(1968-),男,四川重庆人,教授,博士,从事图像分析和模式识别领域的研究。

## 0 引言

人类的视觉系统既要求其具有处理大量输入信息的能力,又要求有准实时反应能力,两者实际上是相互矛盾的。视觉心理学研究表明,在分析复杂的输入景象时,人类视觉系统采取了一种串行的计算策略,即利用选择性注意机制,根据图像的局部特征,选择景象的特定区域,并通过快速的眼动扫描,将该区域移到具有高分辨率的视网膜中央凹区,实现对该区域的注意,以便对其进行更精细的观察与分析。这可看作是将全视场的图像分析与景象理解通过较小的局部分析任务的分时处理来完成。可见,选择性注意机制是人类从外界输入的大量信息中选择特定感兴趣区域的一个关键技术,与自动目标识别中的目标检测有着相似之处,因此,研究人类选择性注意机制在目标检测中的应用有着重要意义。

视觉心理学研究表明,人类视觉系统选择性注意机制主要包括两个子过程:(1)快速的、采用 bottom-up 控制策略的预注意机制,该机制是基于输入景象的显著性计算的,属于低级的认知过程。(2)慢速的、采用 top-down 控制策略的注意机制,它通过调整选择准则,以适应外界命令的要求,从而达到将注意力集中于特定目标的目的,属于高级的认知过程。这两种过程均是基于显著性计算的,即通过计算输入景象的显著性图,突出感兴趣区域。不同之处在于,过程(1)仅以输入景象为依据,计算其显著性,而过程(2)则还引入了外部命令的作用,将视觉系统的注意力导向由外部命令所定义的感兴趣区域。目前视觉心理学对过程(1)的研究较多,提出了许多计算模型<sup>[1~11]</sup>,在某些模型中还讨论了 bottom-up 策略的生理实现机理<sup>[11]</sup>,而对过程(2)的研究较少,其原因主要在于研究外界命令如何参与注意的计算方面存在较大的困难。与之对应,注意机制在计算机视觉中的研究也较多地集中于 bottom-up 控制策略的应用。

## 1 采用 bottom-up 控制策略的注意模型

该模型是 Itti 等人于 1998 年提出的<sup>[1]</sup>,模型框图见图 1,它利用生物的中央周边滤波器结构,在若

干个空间尺度上提取低级特征,通过组合得到的特征图生成 3 个特征显著性描述,分别对应于颜色、灰度和方位,而显著性图则为这 3 个特征显著性描述的线性组合。一个赢者取全神经网络用于将注意力集中于当前最突出的图像区域,瞬时的抑制性反馈则抑制了当前所注意区域,使得注意力能移至下一个最突出的区域。

模型中各金字塔具有高斯金字塔形式,而每一特征的计算采用了与生物感受野类似的线性中央周边操作。中央周边操作是指典型的视觉神经元对位于其中央的小区域最敏感,而其中央区周围的更广、更弱区域内的刺激将抑制该神经元的响应。显然,这样的对局部空间不连续性的敏感结构特别适用于检测相对于其局部周围突出的区域,这也是视网膜、外膝体及视皮层内采用的一般计算原则。模型中的中央周边操作是通过计算小尺度和大尺度的差来实现的。

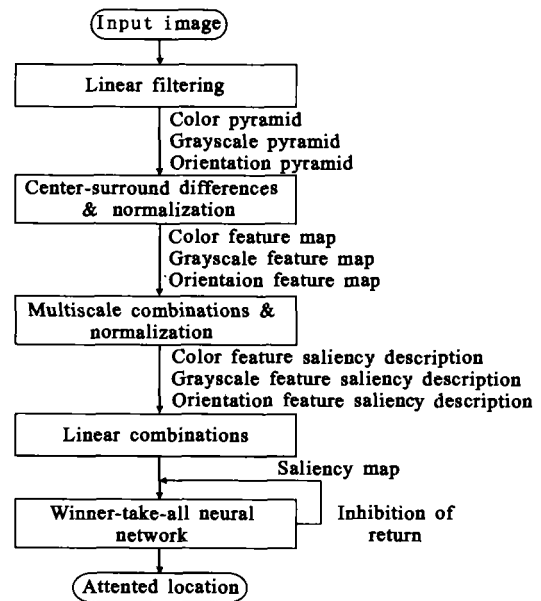


图 1 基于 bottom-up 控制策略的注意模型

Fig. 1 Attention model with bottom-up control strategy

设  $r, g, b$  分别对应于输入图像的红、绿、蓝信道,于是可得灰度图  $I = (r + g + b)/3$ , 以及高斯金字塔  $I(\sigma)$ 。为了从强度中分离出色度信号,使用  $I$  来归一化  $r, g, b$  信道。因为亮度非常低的色度变化是觉察不到的,因此也是不显著的,故归一化仅在灰度大于全图最大的  $1/10$  位置上进行,而其他位置的  $r, g, b$

值则被赋为 0。根据归一化后的  $r, g$  和  $b$ , 建立 4 个宽调谐的颜色通道: 红色  $R=r-(g+b)/2$ , 绿色  $G=g-(r+b)/2$ , 蓝色  $B=b-(r+g)/2$  和黄色  $Y=(r+g)/2-|r-g|/2-b$ , 负值则置为 0。进一步地, 根据这些颜色通道, 可建立 4 个高斯金字塔  $R(\sigma), G(\sigma), B(\sigma)$  和  $Y(\sigma)$ , 由  $I$  可得具有方向选择性的实 Gabor 金字塔  $O(\sigma, \theta)$ 。

考虑 3 种特征, 灰度、颜色和方位。若记中央周边差操作为  $\Theta$ , 则灰度特征图可由下式求得:

$$I(c, s) = |I(c) \Theta I(s)| \quad (1)$$

在人类视觉系统中, 该特征是由对暗中央亮周边或亮中央暗周边敏感的神经元来检测的。

第 2 个特征与颜色有关, 在视皮层中, 它用颜色拮抗 (Color-Double-Opponent) 系统来表达, 即在感受野的中央, 神经元被一种颜色 (如红色) 所激励, 并被另一种颜色所抑制 (如绿色), 而在感受野的周边则相反。在人类的视皮层中, 共有 4 种空间和颜色拮抗, 分别为红/绿、绿/红、蓝/黄和黄/蓝颜色对。本模型中, 对应于红/绿和绿/红拮抗的特征图  $RG(c, s)$  可由下式同时得到:

$$RG(c, s) = |(R(c) - G(c)) \Theta (G(s) - R(s))| \quad (2)$$

而对应于蓝/黄和黄/蓝拮抗的特征图由下式求得:

$$BY(c, s) = |(B(c) - Y(c)) \Theta (Y(s) - B(s))| \quad (3)$$

第 3 个特征是关于方位的, 计算公式为:

$$O(c, s, \theta) = |O(c, \theta) \Theta O(s, \theta)| \quad (4)$$

对输入图像的每一像素, 显著性图使用一个标量刻画其显著性, 并基于显著性的空间分布来指导注意点的选择, 特征图的组合提供了显著性图的 bottom-up 输入。组合特征图的困难在于它们表达的特征是不可比的, 各自具有不同的动态范围和提取机制。而且, 由于所有特征图都被组合在一起, 仅仅在某些图中出现的显著目标, 可能被图中大量的噪声或不太显著的目标所淹没。

在 Itti 等人最初的论文中, 针对无 top-down 外部命令的情况, 提出了一个特征图归一化算子

$N(\cdot)$ , 以增强显著峰较少的特征图, 而削弱存在大量显著峰的特征图。对每一特征图, 该算子的操作包括: (1) 归一化该特征图至一固定范围  $[0, \dots, M]$  内, 以消除依赖于特征的幅度差别; (2) 计算除全局最大外所有局部极大的均值  $\bar{m}$ ; (3) 用  $(M - \bar{m})^2$  乘该特征图。

仅考虑局部极大可使  $N(\cdot)$  对特征图中有意义的区域进行比较, 而忽略均匀区域。全局最大与所有局部极大均值之差反映了最感兴趣区域与平均感兴趣区域间的差别。若该差值很大, 最感兴趣区域就会突出出来, 若该差值较小, 表明特征图中不含任何具有独特性质的区域。  $N(\cdot)$  的生物学依据在于: 它近似地表达了皮层的侧抑制机制, 即近邻相似特征通过特定的联结相互抑制<sup>[12]</sup>。

特征图被组合成 3 个特征显著性描述, 即灰度特征显著性描述  $\bar{I}$ 、颜色特征显著性描述  $\hat{C}$  和方位特征显著性描述  $\hat{O}$ , 如公式 (5)~(7) 所示。

$$\bar{I} = \bigoplus_{c,s} N(I(c, s)) \quad (5)$$

$$\hat{C} = \bigoplus_{c,s} [N(RG(c, s)) + N(BY(c, s))] \quad (6)$$

$$\hat{O} = \sum_{\theta} N(\bigoplus_{c,s} N(O(c, s, \theta))) \quad (7)$$

式中  $\bigoplus$  表示逐点求和。

建立 3 个特征显著性描述  $\bar{I}, \hat{C}$  和  $\hat{O}$  并各自进行归一化的原因是基于如下假设: 在显著性方面, 相似特征竞争最强, 而不同特征将各自独立地对显著性图起作用。这 3 个特征显著性描述被进一步地归一化, 并相加得到显著性图  $S$ , 如公式 (8) 所示, 其最大值点则为下一注意点。

$$S = \frac{1}{3} [N(\bar{I}) + N(\hat{C}) + N(\hat{O})] \quad (8)$$

在 Itti 等人后来的工作中<sup>[2]</sup>, 比较了多种特征组合策略, 发现归一化每一特征图至固定动态范围并相加将在检测复杂景象中的突出目标时表现出较差的检测性能, 改进性能的可能途径是通过学习欲检测目标的样本得到特征图的线性组合权。该方法虽然可极大地提高检测性能, 但却带来对不同目标的检测任务, 及需使用不同模型 (即不同的权) 的困难。在参考文献 [3] 中, Itti 等人给出了一种新的特征组合策略,

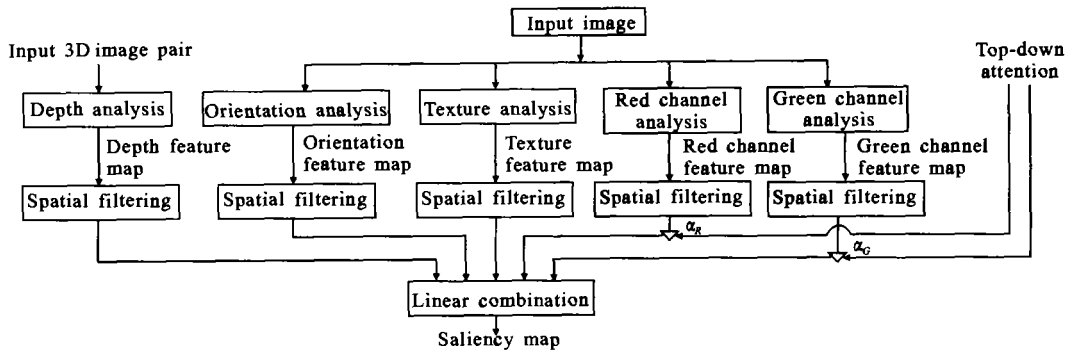


图 2 基于 top-down 控制策略的注意模型

Fig. 2 Attention model with top-down control strategy

即在归一化的基础上,通过各特征图内的局部非线性竞争增强显著峰较少的特征图,而削弱存在大量显著峰的特征图。这种特征内竞争模式与由电生理学所观察到的非经典抑制相互作用类似<sup>[13]</sup>。

## 2 采用 top-down 控制策略的注意模型

该模型是 Blaser 等人在研究三阶运动时提出的<sup>[1]</sup>,其框图见图 2。

按流程,外界景象将在不同的通道中被分析,得到多个特征图,图中示出了深度、方位、纹理及颜色等通道,各通道的输出之和构成了显著性图。但模型中并未给出各特征图的计算方法,因此可结合其他技术,如 Itti 等人的算法得到。该模型的重点在于与任务有关的注意命令的运用,图中示出了对红色和绿色进行注意的命令是如何作用于两颜色通道的,即通过分别增加两通道的输出增益至  $1 + \alpha_R$  和  $1 + \alpha_G$  来得到。

## 3 基于注意计算的目标检测

### 3.1 基于注意计算的目标检测算法

在自动目标识别系统中,出于系统成本、运算速度等实际考虑,通常使用仅提供灰度信息的单一传感器(不考虑多传感器的情形),无颜色、深度等信息可供应用,因此文中仅考虑利用方位特征。由于使用单一特征,避免了不同特征不可比的难题。此外视觉心理学研究并不支持 Itti 等人的模型中所使用的归一化技术,不同频率和方位通道之间的相互作用也并非

中央周边差形式,而是表现为对对比度增益控制<sup>[14]</sup>。据此提出如图 3 所示的注意模型。

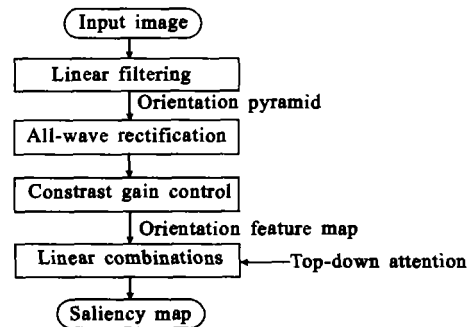


图 3 文中采用的注意模型

Fig. 3 Attention model developed by us

### 3.2 实验结果

实验中在 4 个尺度上计算方位特征<sup>[15,16]</sup>,图 4 为原始图像,图 5~8 为经过全波整流和对比度增益控制后,4 个尺度上的方位特征图,而图 9 为在不存在外部注意命令的情况下,图 5~8 线性组合后的显著性图。可以看出,处于河中船舶的显著性是最强的。实验表明,仅根据 bottom-up 控制策略,图中的船舶相对于其他目标是最显著的。如果能将关于船舶的某些知识,如大小、空间频率等转化为相应的外部注意命令,用于控制各方位特征图的线性组合,相信能进一步突出船舶目标。类似地,如能建立外部注意命令与特定的目标检测任务间的联系并使数学模型化,也可实现对其他目标的检测。如何根据目标检测任务确定形式化的外部注意命令还需要进一步研究。

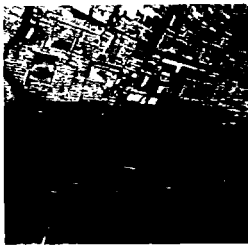


图 4 原始图像  
Fig. 4 Original image



图 5 尺度 1 上的方位特征图  
Fig. 5 Orientation feature map  
at selected scale 1



图 6 尺度 2 上的方位特征图  
Fig. 6 Orientation feature map  
at selected scale 2

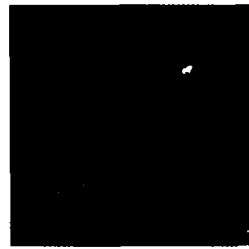


图 7 尺度 3 上的方位特征图  
Fig. 7 Orientation feature map  
at selected scale 3



图 8 尺度 4 上的方位特征图  
Fig. 8 Orientation feature map  
at selected scale 4



图 9 显著性图  
Fig. 9 Saliency map

## 4 结 论

提出了一种适用于自动目标识别的目标检测算法。从输入图像出发,采用 Gabor 算子建立多尺度、多方位的多通道图像,通过全波整流和各通道间的对比度增益控制,得到多尺度、多方位的方位特征图,这些方位特征图的线性组合则为显著性图。给出了仅采用 bottom-up 控制策略的船舶目标检测实验结果。尽管还没有更多的实验结果支持,但可以期望使用与目标检测任务有关的 top-down 注意命令进行各方位特征图的线性组合,欲检测的目标将在显著性图中被突出出来,以便于检测,而不仅限于由 bottom-up 控制策略所得到的结果。

## 参考文献:

- [1] Itti Laurent, Koch Christof, Niebur Ernst. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on PAMI, 1998, 20(11): 1254-1259.
- [2] Itti Laurent, Koch Christof. A comparison of feature combination strategies for saliency-based visual attention [A]. Proc SPIE[C]. 1999.3644. 473-482.
- [3] Itti Laurent. Visual attention and target detection in cluttered natural scenes [J]. Optical Engineering, 2001, 40(9): 1784-1793.
- [4] Blaser Erik, Sperling George, Lu Zhonglin. Measuring the amplification of attention[A]. Proceedings of the National Academy of Sciences of the United States of America[C]. 1999. 96. 11681-11686.
- [5] Tsotsos J K, Culhane S M, Wai W Y K, et al. Modelling visual attention via selecting tuning [J]. Artificial Intelligence, 1995, 78(1-2): 507-545.
- [6] Niebur E, Koch C. Computational architecture for attention. Parasuraman R. The Attention Brain [M]. MA: MIT Press. 1998. 163-186.
- [7] Olshausen B A, Anderson C H, Van Essen D C. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information [J]. The Journal of Neuroscience, 1993, 13(11): 4700-4719.
- [8] Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry [J]. Human Neurobiology, 1985. 4: 219-227.
- [9] Milanese R, Gil S, Pun T. Attentive mechanisms for dynamic and static scene analysis[J]. Optical Engineering, 1995. 34(8): 2428-2434.
- [10] Baluja S, Pomerleau D A. Expectation-based selective attention for visual monitoring and control of a robot vehicle [J]. Robotics and Autonomous Systems, 1997. 22(3-4): 329-344.
- [11] Li Zhaoping. A saliency map in primary visual cortex [J]. TRENDS in Cognitive Science, 2002. 6(1): 9-16.
- [12] Cannon M W, Fullenkamp S C. A model for inhibitory lateral interaction effects in perceived contrast [J]. Vision Research, 1996. 36(8): 1115-1125.
- [13] Silito A M, Grieve K L, Jones H E, et al. Visual cortical mechanisms detecting focal orientation discontinuities [J]. Nature, 1995. 378: 492-496.
- [14] Lu Zhonglin, George Sperling. Contrast gain control in first- and second-order motion perception [J]. Journal of the Optical Society of America(A), 1996, 13(12): 2305-2318.
- [15] 田金文,柳健,张天序. 变窗 Gabor 变换理论及其在图像处理中的应用[J]. 红外与激光工程, 1998, 27(4): 1-5.
- [16] 陈勇,皮德富,周士源. 基于小波变换的红外图像融合技术研究 [J]. 红外与激光工程, 2001, 30(1): 15-17.