

中图法分类号: TP391.9 文献标志码: A 文章编号: 1006-8961(2010)11-1650-08

检索信息: 田明辉, 万寿红, 岳丽华. 自然场景中的视觉显著对象检测[J]. 中国图象图形学报, 2010, 15(11): 1650-1657

自然场景中的视觉显著对象检测

田明辉, 万寿红, 岳丽华

(中国科学技术大学计算机科学与技术系, 合肥 230027)

摘要: 显著对象检测是视觉注意机制的一个重要应用基础研究, 对于图像检索、场景分析、图像标注与对象识别都有着重要的研究意义。基于 Treisman 特征整合理论和 Koch 计算框架, 提出一种自然场景中视觉显著对象的检测方法。该方法首先建立适用于彩色自然场景的视觉显著度模型, 计算多种不同特征的显著度, 然后在融合不同特征的综合显著度图中提取显著对象。实验结果表明, 与经典的 Itti 模型相比, 这种方法不仅检测快速而且更准确地将视觉显著对象从背景中分离出来, 更符合人眼的真实视觉注意过程。

关键词: 显著对象检测; 视觉显著度; 视觉注意; 自然场景

Visual salient objects detection in natural scenes

TIAN Minghui, WAN Shouhong, YUE Lihua

(Department of Computer Science and Technology, University of Science and Technology of China, Hefei 230027)

Abstract: Visual salient objects detection is an important fundamental application research of visual attention mechanism. It plays an important role in image retrieval, scene analysis, image annotation and object recognition. This paper proposes a novel approach for visual salient objects detection in natural scenes based on Treisman's feature integration theory and Koch's framework. In this approach, a visual saliency model for colored natural scenes is proposed and different feature saliencies are considered and computed. Then an effective method is given to extract salient objects from an integrated saliency map which is combined by different feature saliency maps. Comparing with Itti's model, the experimental results indicate that not only the detection speed of this approach is faster, but also this approach can separate visual salient objects from their backgrounds more accurately and more efficiently. From this aspect, the approach in this paper is more similar to human's real visual attention process than Itti's model.

Keywords: salient object detection; visual saliency; visual attention; natural scene

0 引言

当前, 随着信息技术的发展, 图像数据的规模变得越来越大, 面对如此庞大的图像数据, 如何能够快速而准确地完成各种图像分析任务已经成为人们研究的热点。传统的图像分析方法将图像中所有区域都被赋予相同的优先级, 然而很多图像分析任务(诸如图像检索^[1]、图像语义标注、场景分析与理

解^[2]、目标识别^[3]等)所关心的内容通常仅占图像中很小一部分, 因此, 这种全面加工不但增加了分析过程的复杂度, 而且带来了不必要的计算浪费。近年来, 许多研究学者发现人类视觉系统(HVS)在面对一个复杂场景时, 能够迅速将注意力集中在少数几个显著的视觉对象上, 对其进行优先处理, 而该过程则被称为视觉注意^[4]。显然, 将这种机制引入到图像分析领域是非常必要的, 它可以提供容易引起观察者注意的图像区域信息, 帮助制定合理的计算

基金项目: 国家自然科学基金项目(60833005)。

收稿日期: 2009-04-07; 改回日期: 2009-07-08

第一作者简介: 田明辉(1981—), 男, 现为中国科学技术大学计算机科学与技术系知识与数据工程实验室博士研究生。主要研究方向为视觉显著度建模、计算机视觉及图像分析等。E-mail: mhtian@mail.ustc.edu.cn。

资源分配方案,从而极大地提高现有图像分析系统的工作效率。

显著对象检测是视觉注意机制中的一项关键技术,它是将图像中的显著对象快速而准确地定位并提取,进而进行优先处理与分析,为后续的图像分析提供指导和决策层的支持。目前,对于自然场景图像,已有的显著对象检测方法可分为 3 类:基于像素的、基于频域分析和基于区域的方法。基于像素的方法主要以像素为单位计算不同特征的显著度图,在显著度图中定位显著对象,此类方法的特点是计算特征较容易提取、计算量较小、较为快速,但对于噪声较为敏感、鲁棒性较差,代表性的文献有 [2-3,5];基于频域分析的方法主要是寻找不同特征在频域中的特点,构建频域滤波器来定位显著对象,此类方法鲁棒性较好,但精确度不高,只能得到显著对象的大致位置形状,代表性的文献有 [6-8]。基于区域的方法主要利用分割将图像划分为不同的区域,然后通过区域竞争定位显著对象,此类方法基于心理学研究基础,是目前发展趋势,但由于对分割算法过度依赖不利于对象语义的完整性,并且计算量较大,鲁棒性有待提高,代表性的文献有 [9-10]。本文综合基于像素与基于区域两类方法的优点,通过融合不同特征显著度图定位显著对象,再利用区域竞争模拟视点转移。

1 相关工作

近年来,对于视觉注意建模的研究一直是计算视觉和多媒体方向研究的热点之一,也涌现出了一些较为经典的理论和框架。其中两个重要的概念是 Treisman 的特征整合理论 (feature integration theory)^[11] 和 Koch 与 Ulman 提出的神经生物学结构框架^[12]。特征整合理论认为视觉信息是由不同的特征图整合且并行处理的。这些特征图由人眼视野中的不同位置组织起来。每一个视觉特征都对应一个特征图。基于特征整合理论,一些研究学者设计了用来模拟人类视觉注意机制的框架。其中较为经典的是 Koch 和 Ulman 提出的神经生物学结构框架。这个框架的主要思想是计算一个表达图像中每个位置显著度的显著度图。如图 1,该框架给出了一个通用的视觉显著度计算模型,主要包括早期视觉特征抽取、特征图构建及特征图融合 3 个阶段。

基于 Koch 与 Ulman 的神经生物学框架,一些仿

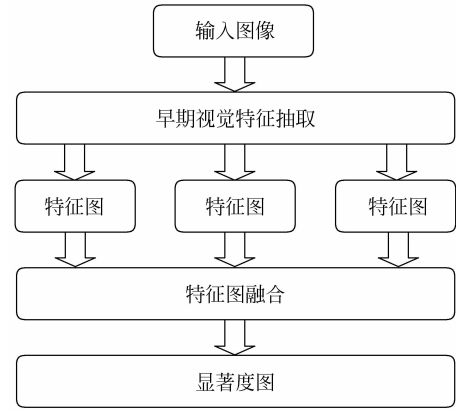


图 1 Koch 和 Ulman 提出的神经生物学框架

ig. 1 The neural architecture proposed by Koch and Ulman

真计算模型逐渐被研究学者们提出。在这些计算模型中,最具代表性的是 Itti 与 Koch 在 1998 年提出的 Itti 模型^[2]。Itti 模型提供了一种自底向上、各个特征图可并行计算的机制,并且将多种特征在多尺度下进行融合,最后合成一张综合的视觉显著度图。该视觉显著度图描述了在无先验信息指导下图像自身特质对人眼的刺激程度,量化了图像中各个位置在多种特征综合考虑下的显著性,并根据显著度的分布指导被注意的区域。已有实验结果表明,该模型对噪声、对比度和亮度具有很好的鲁棒性^[13]。

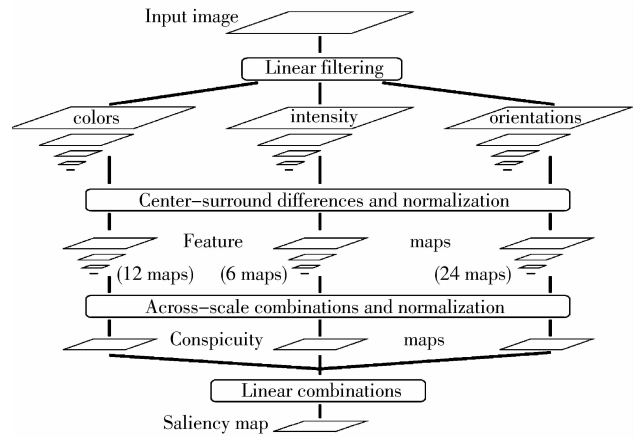


图 2 Itti 视觉显著度模型

Fig. 2 Itti's visual saliency model

虽然 Itti 模型较为被学者认可,但在该模型中仍存在两个较为明显的问题。一是由于 Itti 模型中采用多尺度特征计算,以及在显著度计算中采用了较为复杂的 center-surround 算子,故使得 Itti 模型的整体计算量较大,违背了自底向上视觉注意过程中快速的特点;二是由于多尺度特征图的融合,显著对

象的边缘信息已经丢失,使得在最终的显著度图中只能通过 center-surround 算子进行固定尺寸的圆形区域来模糊地描述视觉显著对象的大致位置,而无法准确地表达显著对象的外形轮廓,这对于许多图像分析任务是不够精确的。本文基于特征整合理论与 Koch 与 Ulman 的神经生物学框架,针对 Itti 模型的上述不足,提出一种适用于自然场景彩色图像的视觉显著度模型,并在此模型基础之上结合分割方法给出一种视觉显著对象的检测算法。该算法不但速度较快,而且能够较为准确地表达出显著对象的位置与外形轮廓,较好地显著对象与其背景分离,更符合人眼的真实视觉注意过程。

2 视觉显著度模型

如图 3 所示,本文中的视觉显著度模型是基于 Treisman 的特征整合理论和 Koch 与 Ulman 的神经生物学框架体系,整个模型分为 3 个阶段。第 1 阶段是特征提取,在早期视觉系统中选取颜色、方向和纹理 3 种特征来描述一幅图像对人眼的刺激通道^[14];第 2 阶段是显著度计算,在这个阶段中,提出一种新颖的显著度计算方法代替 Itti 模型中的 center-surround 算子来计算每个特征的显著度图;第 3 阶段为特征整合阶段,首先对不同特征进行指数

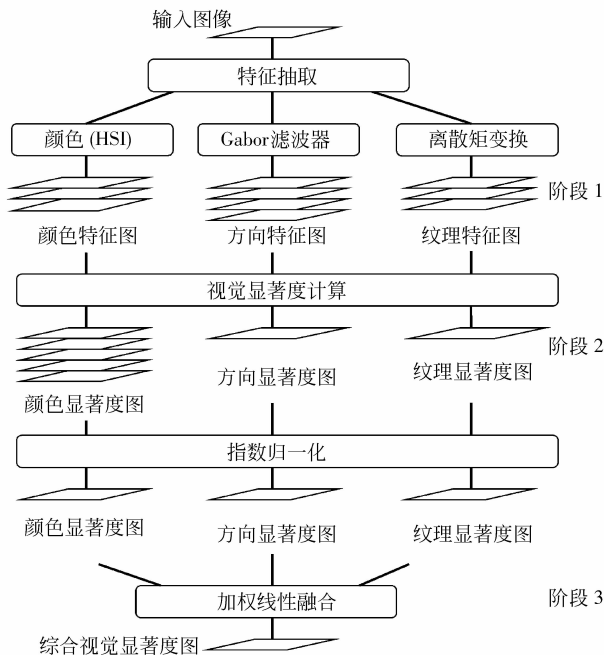


图 3 本文的视觉显著度模型

Fig. 3 Visual saliency model proposed in this paper

归一化,然后再进行线性加权融合生成最后的显著度图。

2.1 特征提取

根据神经生物学理论,人类在前期的快速预注意过程中通常对于颜色、方向和纹理 3 类特征较为敏感^[14]。因此,在特征抽取阶段选择颜色、方向和纹理 3 类特征作为外界图像向人眼反馈刺激的特征通道,如图 4 所示。对于人眼来说,由于色调 (hue)、饱和度 (saturation) 和亮度 (intensity) 信息所形成的刺激要远比红绿蓝 (RGB) 三色所形成的刺激来得直观^[15],因此对于颜色特征,选用 HSI 颜色空间,不同于 Itti 模型的 RGB 颜色空间。对于方向特征的表达,采用 Gabor 滤波器^[16]对已经除去颜色的灰度图像进行 0°、45°、90° 和 135° 4 个方向的空间滤波,获得 4 幅方向特征图。具体计算公式如下:

$$g(\theta, x, y) = K \exp(-\pi\sigma^2(x^2 + y^2)) \times \left(\exp(j\sqrt{2}\pi\sigma^2(x\cos\theta + y\sin\theta)) - \exp\left(-\frac{\sigma^2}{2}\right) \right) \quad (1)$$

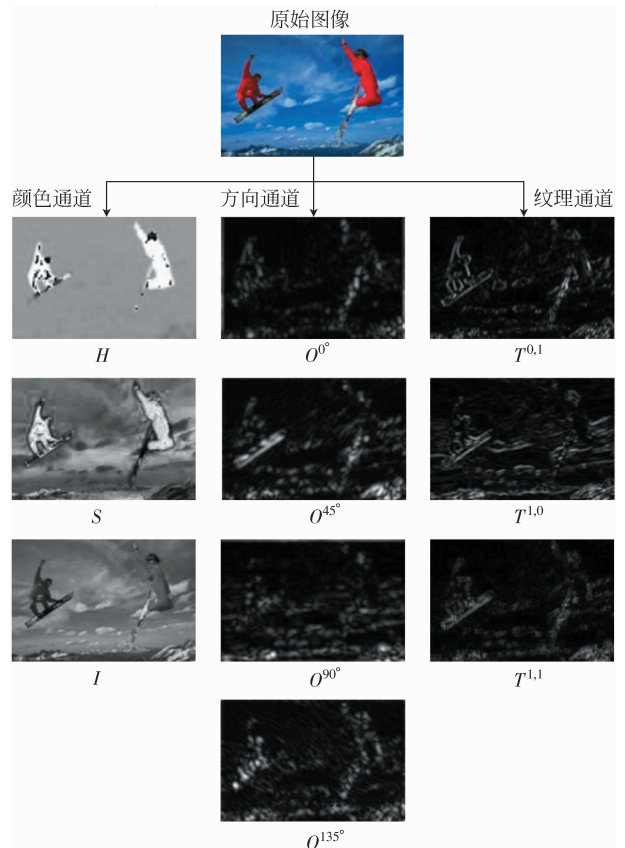


图 4 特征提取示例

Fig. 4 An example of feature extraction

式中 K 是常量,表示高斯卷积的幅值; σ 也是常量,用于描述高斯卷积的两个轴的尺度; θ 是高斯卷积的旋转方向,表示 Gabor 滤波器的滤波方向。根据上述滤波器 g ,生成的方向特征图为

$$O^\theta(x, y) = \|L(x, y) \otimes g(\theta, x, y)\| \quad (2)$$

对于纹理特征,采用离散矩变换(discrete moment transform)^[17]来具体描述纹理特征。文献[16]表明离散矩变换可以很好地描述纹理特征。如果窗体大小为 $2k + 1$,那么离散矩的变换形式为

$$T^{p,q}(x, y) = \sum_{r=-k}^{r=k} \sum_{s=-k}^{s=k} L(x-r, y-s) r^p s^q \quad (3)$$

考虑计算复杂性和性能权衡,文中仅计算 $T^{0,1}$ 、 $T^{1,0}$ 和 $T^{1,1}$ 3 个一阶离散矩值来描述纹理特征。

2.2 显著度计算

心理学研究表明,图像中那些可以令人眼产生更多刺激或者新奇刺激的视觉对象,或人们较为期待的视觉对象,较容易引起人们的注意^[4,18]。视觉显著度即是描述图像中不同位置空间或者视觉对象对人眼产生刺激的程度。根据显著性的成因,显著度分为自显著性和互显著性^[19]。自显著性描述显著对象由于自身的某些特征优势可以单独地对人眼产生刺激,而互显著性描述显著对象的特征需要与周围背景进行对比才能对人眼产生刺激的因素。综合自显著和互显著特性,主要考虑以下几个关于颜色对比性^[14,20]和特征增益^[21]方面的因素:

1) 色调对比 图像中不同色调在颜色环上的角度差异可以对人眼产生刺激。大的色调角度差显然可以形成较多的色调对比。在色调环中,两个最大的角度差异为 180° 。

2) 饱和度对比 由图像中高低不同的颜色饱和度形成了对比。饱和度对比的强弱由颜色的饱和度和差异决定。

3) 亮度对比 图像中的明暗变化也可以对人眼产生刺激。亮度的差异越大,则对比越强烈。

4) 暖色优势 色调环中的暖色,如红色、黄色和橙色等颜色较其他颜色可以对人眼产生更多的刺激。这些颜色的色调角度一般小于 45° 。

5) 亮度与饱和度优势 高亮度和高饱和度更容易吸引人眼的注意^[21],因此较亮色的对象和较饱和的颜色相对于其他对象具有更多的优势。

6) 方向性 图像中视觉对象呈现的方向性对人眼具有较为明显的刺激^[14]。方向性越强,刺激越强。

7) 纹理性 图像中视觉对象的纹理在某些情况下也可以对人眼产生一定程度的刺激^[14]。纹理性越强,显著性越大。

根据上述对视觉显著性的分析,对于其中的颜色互显著性,即基于对比的显著性(1)、(2)、(3),提出一种快速的全局显著度度量方法。这种度量可以计算图像空间每个位置相对于整幅图像的视觉显著度。具体计算公式如下:

$$D_i(x, y) = |F_i(x, y) - \bar{F}_i| \quad i = 1, 2, 3 \quad (4)$$

$$S_i(x, y) = 1 - \exp\left\{-\frac{D_i(x, y)}{\bar{D}_i}\right\} \quad (5)$$

式中, D_i 为颜色特征 F_i (分别为 H, S, I) 的特征差异图像, S_i 是颜色特征 F_i 的显著度图。

对于颜色自显著性(4)、(5),即暖色优势和亮度与饱和度优势带来的增益,提出如下计算公式描述:

$$S_4(x, y) = \begin{cases} I(x, y) S(x, y) \cos(\theta) \delta \frac{\sqrt{2}}{2} \leq \cos(\theta) \leq 1 & (6) \\ 0 & \text{其他} \end{cases}$$

$$S_5(x, y) = I(x, y) S(x, y) \delta \quad (7)$$

式中 $I(x, y)$ 和 $S(x, y)$ 分别表示图像中位置 (x, y) 像素的亮度与饱和度, θ 为图像中位置 (x, y) 像素的色调 H 偏角, δ 为放大系数。此处将 I 和 S 归一到 $[0, 1]$, 而将色调偏角归一到 $[0, 2\pi]$ 。

如图 5 所示,综合色调对比、饱和度对比、亮度对比、暖色优势、亮度与饱和度优势 5 个方面的分

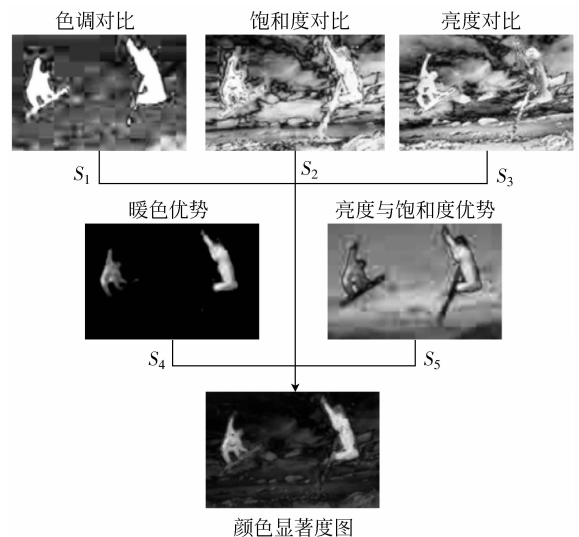


图 5 颜色显著度图

Fig. 5 Color saliency maps

析,采用如下计算将 5 幅颜色显著度图合称为 1 幅颜色显著度图(w_i 为不同因素权值):

$$S_c = \sum_{i=1}^5 w_i S_i \quad (8)$$

对于方向性和边缘纹理性(6)、7),其显著性呈现的方式与颜色不同,故采用与文献[1]类似的公式进行进算,将描述各个方向的特征图和描述不同纹理的特征图合成为方向显著度图与纹理显著度图,具体公式如下:

$$S_o(x, y) = \frac{1}{w_{00^\circ} + \exp\left(-\frac{O^{0^\circ}(x, y)}{O_{avg}^{0^\circ}}\right)} \times \frac{1}{w_{045^\circ} + \exp\left(-\frac{O^{45^\circ}(x, y)}{O_{avg}^{45^\circ}}\right)} \times \frac{1}{w_{090^\circ} + \exp\left(-\frac{O^{90^\circ}(x, y)}{O_{avg}^{90^\circ}}\right)} \times \frac{1}{w_{0135^\circ} + \exp\left(-\frac{O^{135^\circ}(x, y)}{O_{avg}^{135^\circ}}\right)} \quad (9)$$

$$S_T(x, y) = \frac{1}{w_{T01} + \exp\left(-\frac{T^{01}(x, y)}{T_{avg}^{01}}\right)} \times \frac{1}{w_{T10} + \exp\left(-\frac{T^{10}(x, y)}{T_{avg}^{10}}\right)} \times \frac{1}{w_{T11} + \exp\left(-\frac{T^{11}(x, y)}{T_{avg}^{11}}\right)} \quad (10)$$

式中, w_o 和 w_T 分别为方向特征图和纹理特征图的权值,这里所有 w_o 和 w_T 均设置为 0.5。 S_o 和 S_T 分别为方向显著度图和纹理显著度图(图 6)。

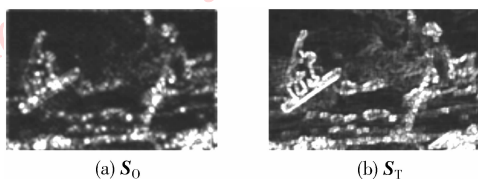


图 6 方向显著度图和纹理显著度图

Fig. 6 Orientation saliency map & texture saliency maps

2.3 特征融合策略

根据 Treisman 特征融合理论,不同特征显著度图最后要经过统一的归一化与融合,形成一张综合视觉显著度图。采用文献[2]中的放大系数和文

献[22]中的指数归一化计算,将不同特征显著图中显著对象区域与背景区域的差异加大。通过指数运算来增强显著度较高的视觉对象区域,从而降低显著度较低的背景区域对显著对象分割造成的影响,便于后续显著对象的提取。

如图 7 所示,不同特征之间采用线性加权进行融合,具体公式如下:

$$N_{exp}(S_i) = (M - \bar{m})^2 (S_i)^\gamma \quad j = C, O, T \quad (11)$$

$$S_f(x, y) = \sum_{j=C, O, T} w_j N(S_j)(x, y) \quad (12)$$

$$\sum_{j=C, O, T} w_j = 1 \quad w_j \geq 0 \quad (13)$$

这里 S_j 为第 i 类特征对应的显著图, M 和 m 分别为 S_j 的全局最大值和局部最大值。 w_j 是第 j 类特征的权重,且满足约束式(13)。

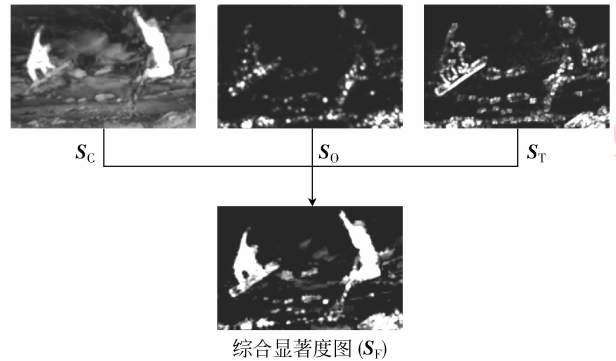


图 7 增强后的不同特征显著度图融合

Fig. 7 Enhanced saliency maps fusion of different features

3 显著对象检测算法

根据所构建的视觉显著度模型,对 1 幅自然场景彩色图像进行显著度计算,生成 1 幅该彩色图像对应的综合视觉显著度图。那么如何在这幅显著度图中将视觉显著对象与其背景分离以达到提取的目的,具体分为 3 个阶段:综合显著度图分割、区域筛选和显著对象间的竞争。

3.1 综合显著图分割

由视觉显著度模型生成视觉显著度图,经过归一化处理,将显著度图归一为 L 级的灰度图像,显著对象与背景之间以具有较为明显的灰度级差异,其中显著对象(SO)区域呈现较高的灰度值,而背景区域呈现较低的灰度值。因此,在对综合显著度图进行中值滤波后,采用二值分割算法中较为经典的最小误差准则分割方法^[23]对该视觉显著图进行分割,

可以很好地将显著对象分割出来。

考虑噪声和斑点因素的影响,在分割之前首先对综合显著度图进行 3×3 的均值滤波以平滑显著度图,便于寻找分割阈值。假设分割阈值为 t , $h(g)$ 为综合显著度直方图分布函数,那么有

$$\text{背景比例} \quad P_b(t) = \sum_{g=0}^t h(g) \quad (14)$$

$$\text{SO 比例} \quad P_o(t) = \sum_{g=t+1}^{L-1} h(g) \quad (15)$$

$$\text{背景均值} \quad \mu_b = \frac{\sum_{g=0}^t h(g) \cdot g}{P_b(t)} \quad (16)$$

$$\text{SO 均值} \quad \mu_o = \frac{\sum_{g=t+1}^{L-1} h(g) \cdot g}{P_o(t)} \quad (17)$$

$$\text{背景均方差} \quad \sigma_b^2(t) = \sum_{g=0}^t (g - \mu_b(t))^2 \cdot \frac{h(g)}{P_b(t)} \quad (18)$$

$$\text{SO 均方差} \quad \sigma_o^2(t) = \sum_{g=t+1}^{L-1} (g - \mu_o(t))^2 \cdot \frac{h(g)}{P_o(t)} \quad (19)$$

根据 Kittler 和 Illingworth 的最小误差分割思想^[23],给出如下评价计算公式:

$$J(t) = 1 + 2[P_b(t) \ln \sigma_b(t) + P_o(t) \ln \sigma_o(t)] - 2[P_b(t) \ln P_b(t) + P_o(t) \ln P_o(t)] \quad (20)$$

则最佳分割阈值为

$$t^* = \operatorname{argmin} J(t) \quad 1 < t < L - 1 \quad (21)$$

因而,如图 8 所示,其显著对象分割结果图 B 为

$$B(x, y) = \begin{cases} 0 & S(x, y) < t \\ 255 & S(x, y) \geq t \end{cases} \quad (22)$$

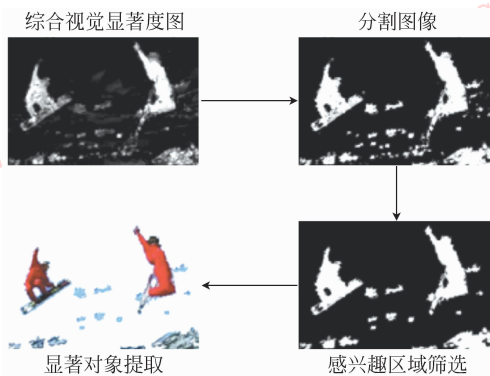


图 8 综合显著度图分割

Fig. 8 Segmentation of integrated saliency map

3.2 区域筛选

在一幅自然场景图像中,通常图像要表达的关

键视觉对象或核心视觉对象是图像中具有有一定大小的区域,所以一般可以认为一幅图像中所表达的主体内容一定不会是面积较小的区域。因此,对于分割得到的显著对象区域,首先采用数学形态学的膨胀、腐蚀和孔洞填充及凹形修补等常用图像处理手段去除少量斑点及噪声,然后采用面积过滤的方法去除面积小于某个阈值 $T (T > 0)$ 的非对象无意义区域,或根据面积由大到小的顺序取前面的 $N (N \geq 1)$ 个显著对象区域。

3.3 显著对象间的竞争

由于各个显著对象间存在着某种心理学上的联系,这种联系就是视觉对象之间的竞争,它是视点转移过程中很重要的一个成因^[24]。对于这种竞争,提出如下几方面计算:

1) 尺寸大小 显然,视觉对象的尺寸越大,则越容易吸引人眼的注意,故用如下公式来描述此种竞争因素

$$\varepsilon_1(A_i) = \frac{A_i}{A_{\text{Image}}} \quad (23)$$

式中 A_i 为显著对象 i 的面积大小, A_{Image} 为图像的面积大小。

2) 相互性 显著对象之间通常会相互影响,根据 HVS 的邻近优先性^[24],显著对象对其距离较近的其他显著对象的影响比距离较远者大。因此,采用式(24)来描述,其中 $L_{i,j}$ 为显著对象 i 与 j 之间的欧氏距离并归一到 $[0, 1]$:

$$\varepsilon_2(L_{i,j}) = 1 - \exp(-L_{i,j}/2\sigma^2) \quad (24)$$

3) 中心效应 位置居于图像中心的显著对象较容易吸引人眼的注意^[7,24]。因而,采用式(25)来描述此种竞争因素,其中 P_i 是显著对象 i 的重心到图像中心点的距离并归一到 $[0, 1]$:

$$\varepsilon_3(P_i) = 1 - \exp\left(-\frac{P_i^2}{2\sigma^2}\right) \quad (25)$$

综合考虑以上竞争因素,则各个显著对象的综合竞争力为

$$V_i = w_{\text{in}} \cdot \bar{S}^i \cdot \varepsilon_1(A_i) \cdot \varepsilon_3(P_i) + w_{\text{out}} \sum_{j=1}^{N-1} (\bar{S}^j \cdot \varepsilon_1(A_j) \cdot \varepsilon_2(L_{i,j})) \quad (26)$$

式中, \bar{S}^i 和 \bar{S}^j 分别为显著对象 i 与 j 的平均显著度值, w_{in} 和 w_{out} 分别为显著对象竞争内部因素和外部影响的权重常数,且满足 $w_{\text{in}} + w_{\text{out}} = 1$ 。

根据各个显著对象的综合竞争力大小,可以得出人眼视点在各个显著对象间进行转移的次序。竞

争力较大的视觉显著对象较容易吸引入眼的注意,而竞争力相对较小的显著对象则排在序列的后面。通过这种视觉对象之间的竞争,模拟视点转移的过程,如图 9 所示。



图 9 视点转移
Fig.9 Location shift of FOA

4 实验仿真与分析

为验证上述算法的有效性,实验在 AMD Althon™ XP 2600 + 处理器,1 G 内存的硬件环境下进行。操作系统为 Microsoft Windows XP Service Pack 2,实验开发环境是 Matlab-7.0.4。

实验中采用的图像部分来源于文献[1,3,6],其余采用标准的 Coral Draw 图像库中的部分图像,共计 100 幅图像,原始图像大小均设定为 384 × 256。这里采用一种较为有效的检测效果评估方法^[8],即检测的命中率(RH)和错误率(RFA)。对于一幅输入图像 $L(x,y)$,通过人工二值标记出图像中的第 k 个显著对象,标记图像为 $Q_k(x,y)$,其中 1 表示对象区域,0 表示背景区域。那么在一幅给定的综合显著度图 $S(x,y)$ 中,命中率(R_H)和错误率(R_{FA})分别为

$$R_H = E(\prod_k Q_k(x,y) \cdot S(x,y)) \quad (27)$$

$$R_{FA} = E(\prod_k (1 - Q_k(x,y)) \cdot S(x,y)) \quad (28)$$

$$Q_k(x,y) = \begin{cases} 1 & \text{目标} \\ 0 & \text{其他} \end{cases} \quad (29)$$

图 10 为本文检测算法对视觉显著对象的检测结果图,图 11 为本文检测算法与 Itti 模型^[2]相比较的结果,包含了视点转移的过程。从图 10、图 11 可以直观地看出本文算法在检测结果上相比经典的 Itti 模型有一定的提高,尤其在视觉显著对象轮廓和语义的完整性上具有较为明显的改善。而在视点转移的阶段,本文算法与 Itti 模型相比,具有完整的对象语义性,更接近人眼的真实注意过程。

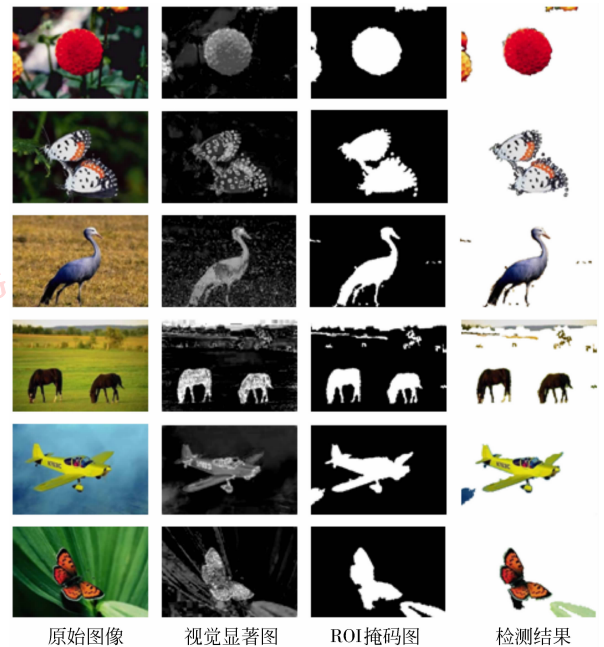


图 10 本文算法检测结果

Fig.10 Detection results of the approach in this paper



图 11 本文算法与 Itti 模型结果比较

Fig.11 The results of this approach in comparison with Itti's model

在具体的评估指标上,如表 1 所示,经 100 幅实验图像的统计数据,与经典的 Itti 模型相比较,在相同命中率(R_H)下,本文算法的错误率较低,而相同错误率(R_{FA})下,命中率相对较高。在平均检测时间上,本文算法也相对较少。

表 1 本文算法与 Itti 模型的性能比较

Tab.1 Performance of Itti's model and our approach

	本文算法	Itti 模型
R_H	0.714 1	0.328 1
R_{FA}	0.154 2	0.154 2
R_H	0.513 7	0.513 7
R_{FA}	0.102 6	0.293 1
平均检测用时间/s	6.223	15.405

5 结 论

本文基于 Treisman 特征整合理论和 Koch 神经生物学框架,构建了一种多特征综合视觉显著度模型,并在此模型基础之上提出一种彩色自然场景中视觉显著对象的检测方法。实验结果表明,本文的显著对象检测方法在彩色自然图像上获得了较为满意的效果。与经典的 Itti 模型相比,本文提出的方法在检测效果和检测速度上都有较为明显的提高,尤其是在显著对象的轮廓及语义完整性上。在人眼的视点转移特性上,本文算法则比 Itti 模型更具有完整的视觉对象语义,更接近于真实的人类视觉注意过程。

由于人类视觉注意的生物学机理仍未能彻底弄清,对于各个特征之间关系尚不明确。不可否认,本文算法对于某些图像仍存在检测错误的情况,尚有改进的空间。在未来的工作中,还需要深入探讨人眼的神经生物学机理,试图寻找更适合的数学计算模型来模拟人类视觉注意过程;深入研究各个特征之间的相互关系;进一步探讨自然图像中的视觉对象的抽象语义,在视觉对象的语义上进行深入研究,构建图像对象的语义模型。

参考文献 (References)

- [1] Zhou Qiang, Ma Limin, Celenk Mehmet, et al. Content-based image retrieval based on ROI detection and relevance feedback [J]. *Multimedia Tools and Applications*, 2005, 27:251-281.
- [2] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(11): 1254-1259.
- [3] Walther D, Itti L, Riesenhuber M, et al. Attentional selection for object recognition—a gentle way [J]. *Lecture Notes in Computer Science*, 2002, 2525(1): 472-479.
- [4] James W. *The Principles of Psychology* [M]. Oxford: Harvard University Press, 1890.
- [5] Sun Y, Fischer R. Object-based visual attention for computer vision [J]. *Artificial Intelligence*, 2003, 146(2): 77-123.
- [6] Hu Y, Rajan D, Chia L T. Robust subspace analysis for detecting visual attention regions in images [C]//*Proceedings of the 13th Annual ACM International Conference on Multimedia*, New York: ACM, 2005: 716-724.
- [7] Meur O L, Callet P L, Barba D, et al. A coherent computational approach to model bottom-up visual attention [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(5): 802-816.
- [8] Hou Xiaodi, Zhang Liqing. Saliency detection: a spectral residual approach [C]//*Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, New York: IEEE, 2007: 1-8.
- [9] Lu Z, Lin W, Yang X, et al. Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation [J]. *IEEE Transactions on Image Processing*, 2005, 14(11): 1928-1942.
- [10] Zhao Xunpo, Wang Lu, Hu Zhanyi. A perceptual object based attention mechanism for scene analysis [J]. *Journal of Image and Graphics*, 2006, 11(2): 281-288. [赵训坡,王璐,胡占义.一种基于感知物体的场景分析注意机制[J]. *中国图象图形学报*, 2006, 11(2): 281-288.]
- [11] Treisman A M, Gelade G. A feature-integration theory of attention [J]. *Cognitive Psychology*, 1980, 12(1) 97-136.
- [12] Koch C, Ulfman S. Shifts in selection in visual attention: toward the underlying neural circuitry [J]. *Human Neurobiology*, 1985, 4(4):219-227.
- [13] Itti L, Koch C. Computational modeling of visual attention [J]. *Nature Reviews Neuroscience*, 2001, 2(3): 194-230.
- [14] Treisman A, Gormican S. Feature analysis in early vision: evidence from search asymmetries [J]. *Psychology Review*, 1995(3):15-48, 1988.
- [15] Rafael C Gonzalez, Richard E Woods, Steven L Eddins. *Digital Image Processing Using MATLAB* [M]. Beijing: Publishing House of Electronics Industry, 2005. [Rafael C Gonzalez, Richard E Woods, Steven L. 数字图像处理: MATLAB 版 [M]. 阮秋琦等译. 北京: 电子工业出版社, 2005.]
- [16] Greenspan H, Belongie S, Goodman R, et al. Overcomplete steerable pyramid filters and rotation invariance [C]//*Proceeding of IEEE Computer Vision and Pattern Recognition*, Seattle, USA: IEEE Computer Society Press, 1994: 222-228.
- [17] Gesu V D, Valent C, Strinati L. Local operators to detect regions of interest [J]. *Pattern Recognition Letter*, 1997, 18(11-13): 1077-1081.
- [18] Wang Su, Wang Ansheng. *Cognitive Psychology* [M]. Beijing: Beijing University Press, 1992. [王甦,汪安圣. 认知心理学 [M]. 北京: 北京大学出版社, 1992.]
- [19] Zhang Peng, Wang Runsheng. A survey of detecting regions of interest in a static image [J]. *Journal of Image and Graphics*, 2005, 10(2): 142-148. [张鹏,王润生. 静态图像中的感兴趣区域检测技术 [J]. *中国图象图形学报*, 2005, 10(2): 142-148.]
- [20] Itten J. *The Elements of Color* [M]. New York: John Wiley & Sons Inc., 1961.
- [21] Mahnke F. *Color, Environment and Human Response* [M]. Detroit: Van Nostrand Reinhold, 1996.
- [22] Nabil Ouerhani, Alexandre Bur, Heinz Hugli. Linear vs. nonlinear feature combination for saliency computation: a comparison with human vision [J]. *Lecture Notes in Computer Science*, 2006, 4174: 314-323.
- [23] Kitterler J, Illingworth J. Minimum error thresholding [J]. *Pattern Recognition*, 1986, 19(1): 41-47.
- [24] Zabrodsky H, Peleg S. Attentive transmission [J]. *Journal of Visual Communication and Image Representation*, 1990, 1(2): 189-198.