

## 基于贝叶斯理论和统计学习的视觉显著性检测

戴花\*, 王建平

(长沙航空职业技术学院 计算机与信息工程系, 长沙 410014)

(\*通信作者电子邮箱 daihua\_18@126.com)

**摘要:**图像处理技术依赖于高质量的视觉显著图才能获得较好的处理结果,现有的视觉显著性检测方法通常只能检测得到粗糙的视觉显著性属性图,严重影响了图像处理的最佳效果。为此,提出一种采用贝叶斯理论和统计学习的视觉显著性检测方法检测图像的视觉显著性属性。该方法基于贝叶斯理论的静态图像的自上而下的显著性和整体显著性,将自上而下的知识和由下向上的显著性进行结合;针对特征整合问题,利用线性模型的加权线性组合方法和正规化神经网络相结合的非线性加权方法来研究与所有因素相关的权值参数。根据自下而上的视觉显著性模型在两个标准数据集中采用 ROC 曲线来进行定量评价,结果表明非线性组合效果优于线性组合。

**关键词:**视觉显著性;视觉显著性图;贝叶斯理论;线性模型;神经网络

**中图分类号:** TP391.41 **文献标志码:** A

### Visual saliency detection algorithm based on Bayes theorem and statistical learning

DAI Hua\*, WANG Jian-ping

(Department of Computer and Information Engineering, Changsha Aeronautical Vocational and Technical College, Changsha Hunan 410014, China)

**Abstract:** Image processing technology depends on the quality of the visual saliency map to obtain better results. The existing visual saliency detection method usually can only detect and get rough visual saliency attribute graph, seriously affecting the image processing results. This paper put forward a method of using Bayes theorem and statistical learning of visual saliency detection to detect the visual saliency property of image. The method was based on Bayes theorem of static image top-down significant and overall significance, and combined the top-down knowledge and the down-top significance. For the synthetic integration of characteristics, all the factors related to the weight parameter were studied by using linear model with the linear weighting combination method and regularized neural network combined with nonlinear weighting method. The ROC curves of the bottom-up visual saliency model in two fixed data set for quantitative evaluation, show that the effect of non-linear combination is better than that of linear combination.

**Key words:** visual saliency; visual saliency map; Bayes' theorem; logistic model; neural network

## 0 引言

图像分割系统需要从周围环境中实时选定重要性区域,其中重要性区域包含大量的可视信息。视觉显著性是一种有效的突出视觉重点机制,它能从周围复杂环境中找到重要区域,从而快速吸引人们的视觉注意力。视觉显著性的优势在于利用数据驱动处理自下而上进行提示,将显著区域从周围环境中“区别”出来。视觉显著性可以通过自上而下和任务处理因素进行显著调整。视觉显著性广泛应用于许多计算机视觉的应用领域,比如图像处理、图像压缩、目标检测、目标识别、自动控制和场景识别等。

Koch 等<sup>[1]</sup>基于生物视觉机制的启示提出“视觉显著图”的概念,利用拓扑结构图来表示一种场景的视觉显著性;之后,Itti 等<sup>[2]</sup>首次提出了计算显著性模型,他运用从中心向外扩散的机制贯穿了三种通用换算的图像特征途径;Harel 等<sup>[3]</sup>运用图表方式来计算视觉显著性;Bruce 等<sup>[4]</sup>基于信息最大化来计算视觉显著性;Gao<sup>[5]</sup>基于判别式的显著性假设提出了一种新的视觉显著性判断方法;Zhang 等<sup>[6]</sup>运用自然统计学为视觉显著性提出基于贝叶斯理论的检测框架;Torralba 等<sup>[7]</sup>提出了基于全局特征值的视觉显著性概率模型;Judd

等<sup>[8]</sup>运用低、中、高层次的特点来进行研究视觉显著性。之前相关研究人员提出的方法均没有考虑将贝叶斯分类理论加入视觉显著性的检测过程中。因此,本文提出一种基于贝叶斯理论和统计学学习的检测方法研究视觉显著性,该方法的贝叶斯理论模型与 Zhang 等<sup>[6]</sup>、Torralaba 等<sup>[7]</sup>的研究工作相关。

Zhang 等<sup>[6]</sup>研究中提出的兴趣概率是在场景中的目标概率,即在一个场景中的单点概率,Torralaba 等<sup>[7]</sup>研究的是在全局特征中图像条件作用下的目标概率,而本文关注的概率则属于显著级别点的概率。在不同特征和不同定位的角度,在静态图像上应用基于自下而上的显著性和综合显著性效果更明显。此外,运用两种改善特征相结合的方法来研究权值参数。利用公制接收机工作特征(Receiver Operator Characteristic, ROC)<sup>[9]</sup>值在 MIT 和 Berkeley 数据集上将本文方法和其他三种先进的显著性模型进行比较,定量评价结果表明非线性组合优于线性组合。

## 1 基于贝叶斯的视觉显著性新型检测方法

本文方法的期望目标是完成区分两种类的经典问题,它能够区分图像上的每个像素点是显著性类别还是非显著性类

收稿日期:2012-02-21;修回日期:2012-03-20。

基金项目:湖南省教育厅科研基金资助项目(11C0009);湖南省科技计划项目(2011GK3086,2011SK3079)。

作者简介:戴花(1980-),女,湖南株洲人,讲师,硕士,主要研究方向:图像识别、信息处理;王建平(1966-),男,湖南湘乡人,副教授,硕士,主要研究方向:网络技术。

别。应用该方法来评估该像素点的显著性程度,并将该点定义为属于该点的视觉特征和在现场的位置显著性类别的概率。在本章中,将通过贝叶斯理论获得视觉显著性检测公式。

假设  $x$  代表场景中的某个像素点,二元随机变量  $S_x$  表示属于显著性类或不属于显著性类的像素点,随机变量  $F$  表示一个像素点的特征,随机变量  $L$  表示一个像素点的平面位置。由此,点  $x$  的显著性程度  $SD_x$  就可以表示为概率  $p(S_x = 1 | F = f_x, L = l_x)$ 。其中:  $f_x$  代表点  $x$  的特征值,  $l_x$  代表  $x$  的坐标。利用贝叶斯分类原理,式(1)可以计算出该点的显著性程度。

$$SD_x = p(S_x = 1 | F = f_x, L = l_x) = \frac{p(F = f_x, L = l_x | S_x = 1)p(S_x = 1)}{p(F = f_x, L = l_x)} \quad (1)$$

为简化计算,本文假设功能和位置都是独立的,而有条件独立给出  $S_x = 1$ ,式(1)可简化为式(2)。

$$SD_x = \frac{p(F = f_x | S_x = 1)p(L = l_x | S_x = 1)p(S_x = 1)}{p(F = f_x)p(L = l_x)} = \frac{p(F = f_x | S_x = 1)}{p(F = f_x)} \frac{p(L = l_x | S_x = 1)}{p(L = l_x)} p(S_x = 1) = \frac{1}{p(F = f_x)} p(F = f_x | S_x = 1)p(S_x = 1 | L = l_x) \quad (2)$$

为计算简便,最好使用对数,式(2)可变为式(3)。

$$\log SD_x = -\log p(F = f_x) + \log p(F = f_x | S_x = 1) + \log p(S_x = 1 | L = l_x) \quad (3)$$

式(3)中的第一项  $(-\log p(F = f_x))$  表明该显著点具有“稀有性”特征,如果一个像素点的功能有一个小概率,那么该像素点将会提供大量信息,具有较高的显著性等级。此项仅仅依赖于观察该点的特征,因此,它是属于纯粹的自下而上的因素。

式(3)中的第二项  $(\log p(F = f_x | S_x = 1))$  表明一个包含自下而上和自上而下的因素。当没有指定任务的时候,此项代表该点具有“特殊性”特征。可视化领域具有一种架构:视觉神经对中心位置是最为敏感的,而呈现在周围的视觉刺激将会抑制神经元的反应。本文所提属性,使得该点位置从周围的“特殊性”中脱颖而出。此外,当给出了具体的任务时,此项也包括自上而下的知识,并有利于目标对象的特征值。本文将这个属性称为“连贯性”。

式(3)中的第三项  $(\log p(S_x = 1 | L = l_x))$  提出位置优越性,它是独立的视觉特征。根据 Judd 等<sup>[8]</sup>、Tatler<sup>[10]</sup> 和 Zhao 等<sup>[11]</sup> 的研究工作,中心偏差目前在视觉显著性检测中起着重要作用。所以,本文认为当人们自由查看现场时,位置优越性即中心位置。当给出了观察者具体的视觉任务时,位置之前是其中的目标有可能出现的地方。

下面将针对静态图像关注它的自下而上显著性,并且没有给出具体的任务。然后,式(3)中的第二项只会显示“特殊性”属性位置。因此,一个像素点的显著性程度具体包括三个因素:“稀有性”、“特殊性”和“中心偏差”。

## 2 检测方法具体说明

本章将阐述视觉显著性检测公式和两种权值相结合方法的算法执行细节。

### 2.1 三种控制因素

“稀有性”因素:“稀有性”因素是通过计算操控图像中每一点的自身信息,这种自身信息在特征颜色、强烈程度和方向

上跨越了不同的比例。利用类似方法建立高斯金字塔  $R(\sigma)$ 、 $G(\sigma)$ 、 $B(\sigma)$ 、 $Y(\sigma)$  和  $I(\sigma)$  来融合 Itti 等<sup>[2]</sup> 的研究。方位信息可通过利用可操作的金字塔  $O(\sigma, \theta)$  来得到。当  $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$  时,方向可表示为首选方向,在所有的金字塔中,选取四种比例进行更深层次计算,并让  $\sigma \in [0, 1, 2, 3]$  代表输入图像的尺寸大小。

在每个子图像中,可以为每一个点计算功能的自我信息,具体过程如下:首先,用 100 柱直方图来计算子图像在每一柱中的频率;然后运用插值计算法来计算子图像中每个点的特征值的概率;接着在子图像中用公式  $h(x) = -\log p(x)$  来计算自身信息;最后,将每个金字塔中不同大小的图像组成一个特征映射。

“特殊性”因素:有许多方法可以用来实现“特殊性”属性,Itti 等<sup>[2]</sup> 运用从中心向周围的机制,Goferman 等<sup>[12]</sup> 使用不同补丁之间的不同像素点来实现“特殊性”。在实施过程中,本文利用图形的视觉显著性<sup>[3]</sup> 来构成该像素点的“特殊性”特征模型。

“中心偏差”因素:人们在拍照时会很自动自觉地将他们感兴趣的物体“框”在整个图像的中心位置。当他们随意地观察图像时,起始点往往都是由整幅图像的中心位置开始的。当没有特殊情况说明时,中心位置即原来位置<sup>[13]</sup>。

可以使用径向基函数法来模拟中心偏差。中心偏差函数是实值函数,其价值只取决于该点和中心点之间的欧氏距离。当该点距中心点的距离较远时,函数值会较小。本文采用高斯核函数<sup>[14]</sup>,因此中心偏差函数可表达为式(4)。

$$\mathcal{G}(x, y) = \exp\left(-\frac{(x - x_c)^2 + (y - y_c)^2}{2\delta^2}\right) \quad (4)$$

其中  $\delta$  是比例参数,不同的  $\delta$  相当于不同的中心偏差调整,  $\delta$  的值较大则精确率就较低;反之,则复检比率较低。可从标准数据集中得到最佳的  $\delta$  比例参数<sup>[15]</sup>。

### 2.2 两种权重相结合的方法

线性组合的最优权重:为了评估在视觉显著性检测问题中各因素的重要性,用线性模型从标准数据集中学习权重参数,并采用梯度下降算法来解决优化问题。

基于神经网络的非线性组合:在大脑中的神经元是高度非线性且结构复杂的人体单元。要实现神经元属性,在本文中进一步非线性整合基于正规化神经网络的不同因素。

本文采用了三个层次的神经网络:一个输入层次,一个隐藏层次和一个输出层次。在输入层次中,节点的数量对应的特征点的数量。在标准数据集中有大量的图像,每个图像又有成千上万个像素点,因此,训练样本点的数量可能会很大,但是每个像素点的特征数量却不多。为了得到更好的输出结果,在隐藏层次的节点数量要比较大。在执行过程中,指定隐藏层次中节点的数量为 70,用 BP 算法规则来研究最佳模式。为了避免出现过拟合问题,在神经网络中增加了正规化术语,同时应用验证集来选择正规化术语参数。

## 3 实验结果与分析

本文使用 MIT<sup>[5]</sup> 和 Berkeley<sup>[8]</sup> 图像分割数据集来测试本文所提出的方法。

### 3.1 MIT 数据集实验结果

MIT 数据集记录来自于 15 位观察者随意观察的 1003 幅室内和室外自然图像。在线性一体化策略逻辑模式中,随机地从数据设置中选取了 100 张训练图像和 200 张测试图像;

在正规化神经网络的非线性组合策略中,随机选取 100 张训练图像、100 张确认图像和 200 张测试图像。为了得到更精确的结果,对每种模式进行了 5 次实验后取其平均值。实验中采用验证图像来选择最佳正规化参数。

图 1 是本文方法的处理过程示意,即输入图像首先可以被分解为三个因素:“稀有性”、“特殊性”和“中心偏差”,然后将这些因素通过逻辑模式和规则的神经网络整合在一起。

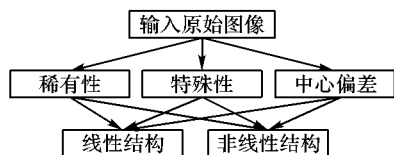


图1 本文方法功能结构

图 2 的横坐标为图像内部对象的 ROC 值,纵坐标为比较的各个算法的 ROC 值(ROC 值的计算方法参见文献[3])。由于本文方法将中心偏差当作先前位置,因此简称本文方法为 CB;GBVS 表示的是 Harel 等的方法[3];SUN 表示的是 Zhang 等的方法[6];Itti 表示的是 Itti 等的方法[2]。根据图 2 的结果,随着图像内部对象的 ROC 值的增长,各个算法的 ROC 值都会有不同程度的递增,而且本文的 CB 算法递增速度较 GBVS、SUN、Itti 三种主流方法均有提高。

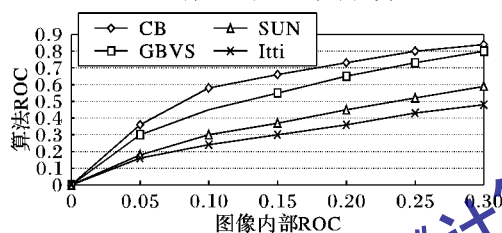


图2 MIT数据集上的实验结果

### 3.2 Berkeley 数据集实验结果

Berkeley 数据集包含来自 11 位观察者随意观察的 120 幅室内和室外场景的彩色图像数据,在线性一体化策略逻辑模式中,从数据集中随机选择了 30 张训练图像和 50 张测试图像;在正规化神经网络的非线性组合策略中,随机选取了 30 张训练图像、30 张确认图像和 30 张测试图像。同样对每种模式进行了 5 次实验后取其平均值作为实验结果。

表 1 是使用两种改进的特征一体化决策在位置前或不在位置前的模型的性能。从中可以看到:1)有中心偏差作为先前位置的模型性能表现要优于没有位置的,这证明在随意观察图像时会倾向于看图像的中心;2)对于整合功能来说,正规化神经网络与逻辑模型相比是更好的方法。

表1 Berkeley 数据集上的查找准确率比较

| 线性一体化策略逻辑模式 |        | 正规化神经网络的非线性组合策略 |        |
|-------------|--------|-----------------|--------|
| 无先验概率       | 有先验概率  | 无先验概率           | 有先验概率  |
| 0.7650      | 0.7904 | 0.8261          | 0.8332 |

在 Berkeley 数据集上的 ROC 值实验结果见图 3,可以看出:随着图像内部对象的 ROC 值的增长,各个算法的 ROC 值都会有不同程度的递增,而本文的 CB 算法递增速度较 GBVS、SUN、Itti 三种主流方法均有提高。

## 4 结语

本文基于贝叶斯理论和统计学学习原理,提出一种视觉显著性检测方法来计算图像中每个像素点的显著性程度。通过引入贝叶斯公式来得到视觉显著性检测方法中的三个重要因素;同时,运用两种改进特征一体化的方法来整合不同的因

素。实验结果表明本文的方法拥有更好的性能;而且定量评价结果还表明,非线性组合优于线性组合。在今后工作中,我们打算进一步研究全方位的视觉显著性,同时将本文方法应用到动态视频检测过程中来判断动态视觉显著性。

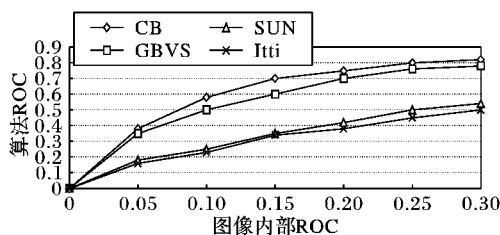


图3 Berkeley数据集上的实验结果

### 参考文献:

- [1] KOCH C, ULLMAN S. Shifts in selective visual attention: towards the underlying neural circuitry[J]. *Human Neurobiology*, 1985, 4(9): 219-227.
- [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] HAREL J, KOCH C, PERONA P. Graph-based visual saliency[EB/OL]. [2012-02-15]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.2254>.
- [4] BRUCE N D B, TSOTSOS J K. Saliency, attention and visual search: an information theoretic approach[J]. *Journal of Vision*, 2009, 9(3): 1-24.
- [5] CAO D, HAN S, VASCONCELOS N. Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(6): 989-1005.
- [6] ZHANG LINGYUN, MARKS T K, TONG M H, et al. SUN: A Bayesian framework for saliency using natural statistics[J]. *Journal of Vision*, 2008, 10(1): 1-20.
- [7] TORRALBA A, OLIVA A, CASTELHANO M, et al. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search[J]. *Psychological Review*, 2006, 113(4): 766-786.
- [8] JUDD T, EHINGER K, DURAND F, et al. Learning to predict where humans look[C]// ICCV 2009: IEEE International Conference on Computer Vision. Piscataway: IEEE, 2009: 8-20.
- [9] HOU XIAODI, ZHANG LIQING. Saliency detection: a spectral residual approach[C]// CVPR '07: IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2007: 1-8.
- [10] TATLER B W. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions[J]. *Journal of Vision*, 2007, 7(14): 1-17.
- [11] ZHAO Q, KOCH C. Learning a saliency map using fixated locations in natural scenes[J]. *Journal of Vision*, 2011, 11(3): 1-15.
- [12] GOFERMAN S, ZELNIK-MANOR L, TAL A. Context-aware saliency detection[C]// CVPR 2010: IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010: 2376-2383.
- [13] 黄志勇, 何发智, 蔡贤涛, 等. 一种随机的视觉显著性检测算法[J]. *中国科学 F 辑: 信息科学*, 2011, 41(7): 863-874.
- [14] 杨明川, 吕学斌, 周群彪. 不完全 K-means 聚类与分类优化结合的图像分割算法[J]. *计算机应用*, 2012, 32(1): 248-251.
- [15] 杨华勇, 余正红, 郑晨. 结合多空间特征的多尺度马尔可夫随机场彩色图像分割[J]. *计算机应用*, 2011, 31(12): 3378-3381.