

一种显著性区域提取的新方法

叶 聪¹ 沈金龙²

(1. 苏州信息职业技术学院计算机科学与技术系, 江苏 苏州 215200)

(2. 南京邮电大学计算机科学与技术系, 江苏 南京 210046)

[摘要] 提出一个基于贝叶斯理论和统计学习理论的显著性提取算法. 该方法基于贝叶斯理论分别阐明图像中不同特征信息、自下而上显著性和全局显著性不同位置的先验信息. 本文针对特征融合问题分别使用加权线性组合 Logistic 模型和基于加权的非线性组合方法的正则化的神经网络来学习权值并获得所有因子. 2 个定位数据集的受试者工作特征(ROC)曲线的实验结果表明, 我们的方法得到的显著图比其他先进的显著性模型效果更好. 扩展的定量评价也证明了非线性组合优于线性组合的策略.

[关键词] 视觉显著性, 贝叶斯理论, 中心位置偏倚, 神经网络

[中图分类号] TP391 **[文献标志码]** A **[文章编号]** 1001-4616(2012)03-0134-04

A Novel Saliency Region Detection Algorithm

Ye Cong¹, Shen Jinlong²

(1. Department of Computer Science, Suzhou College of Information Technology, Suzhou 215200, China)

(2. Department of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

Abstract: A visual saliency detection method based on Bayes' theorem and statistical learning is proposed in this paper. The method clarifies different feature likelihood and different location prior for bottom-up saliency and overall saliency in static images based on Bayes' theorem. For feature integration problem, a weighted linear combination method using logistic model, and a weighted non-linear combination method using regularized neural network are used to learn the weight parameters to combine all factors. Experimental results demonstrate that our method's bottom-up saliency maps perform better than other state-of-the-art saliency models in two fixation data sets by the metric receiver operator characteristic (ROC) curve. And the extensive quantitative evaluation also demonstrates that the non-linear combination outperforms the linear combination strategy.

Key words: visual saliency, Bayes' theorem, central bias, neural network

经过研究, 人们发现图像中最重要的信息往往集中在一些小的关键区域. 人们把这些关键区域称为显著区域或感兴趣区域. 如果能够准确地提取这些显著区域, 我们就能有效且快速地处理图像信息, 从而大大地减少计算的复杂度和减少计算的时间. 目前, 显著性分析在图像分割、图像目标提取、自适应图像压缩和编码等很多领域中都有非常成功的应用^[1].

人们发现显著区域能够引起人的视觉注意, 于是认为图像的显著区域即是图像在视觉上吸引注意的区域. 视觉注意是人类视觉的一项重要的心理调节机制, 心理学及神经生物学上将视觉注意机制分为 2 个方面^[2]: 自底向上的选择性注意和自顶向下的选择性注意. 为了从图像提取显著区域, 人们提出了很多基于视觉注意的计算模型.

1998 年 Itti 和 Koch 提出了一个改进了的注意计算模型(Itti 模型)^[3], 率先将其应用到真实场景图像的分析过程中, 并获得了与人类视觉感知较为接近的实验结果. Harel 等人使用图论方法计算视觉显著性^[4], 该方法简称 GBVS 方法. Bruce 和 Tsotsos 基于信息最大化来计算视觉显著性^[5]. Gao 等人提出一个

收稿日期: 2012-07-27.

基金项目: 国家自然科学基金(61100135).

通讯联系人: 叶 聪, 硕士, 讲师, 研究方向: 图形图像处理. E-mail: fengyun_2012@126.com

新的视觉显著性方法,取名叫判别显著性假设^[6]. Zhang 等人基于自然图像统计特性提出了一种贝叶斯框架来计算显著性^[7],该方法简称 SUN 方法. Torralba 等人基于全局特征值提出一种视觉显著性概率模型^[8]. Juddy 使用低级、中级、高级特征去学习视觉显著性^[9].

本文提出了一种基于贝叶斯理论和统计学习的视觉显著性检测算法. 本文工作的贝叶斯模型与 Zhang 等人的工作^[7]有一定关联. 但是 Zhang 等人^[7]的工作关注的是目标场景中作为一个点的概率,而本文所关心的概率是该点属于显著类的可能性. 并且本文对于静态图像的自底而上显著性和全局显著性明确了不同特征的可能性和不同位置的先验信息. 此外,本文使用 2 种改进的特征融合策略来学习权值. 把提出的方法和其他 3 种最先进的显著性模型在 2 个定位数据集上,分别使用受试者工作特征(ROC)曲线进行比较. 同时扩展的定量评价,也证明了因子的非线性组合优于线性组合.

1 基于贝叶斯理论的视觉显著性检测方法

本文算法的目标是 2 类的分类问题. 将图像的每一个点分类为显著类或非显著类. 因此,在给定场景的视觉特征和位置信息情况下,以该方法估计每个点的显著度. 在这里,显著度定义为每个点属于显著类的概率. 本节将基于贝叶斯定理推导视觉显著性检测的公式.

x 表示场景中的点,可以表示为 1 个像素、1 个区域或 1 个目标. 二值变量 S_x 表示该点属于显著类还是非显著类,变量 F 表示该点的特征,变量 L 表示该点的位置. 因此,点 x 的显著度 SD_x 用公式表示为概率 $p(S_x = 1 | F = f_x, L = l_x)$,其中 f_x 是点 x 的特征,而 l_x 对应点 x 的坐标. 然后利用贝叶斯定理,该点的显著度可以计算得到:

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

为了简单表示,特征和位置对于 $S_x = 1$ 都假定是独立的,那么(1)可以写成:

$$\begin{aligned} SD_x &= \frac{p(F = f_x | S_x = 1) p(L = l_x | S_x = 1) p(S_x = 1)}{p(F = f_x) p(L = l_x)} = \\ &= \frac{p(F = f_x | S_x = 1)}{p(F = f_x)} \frac{p(L = l_x | S_x = 1)}{p(L = l_x)} \frac{p(S_x = 1)}{1} = \frac{1}{p(F = f_x)} p(F = f_x | S_x = 1) p(S_x = 1 | L = l_x). \end{aligned} \quad (2)$$

为方便计算,对于(2)式左右两边分别采取 \log 变换,得到:

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

第 1 项 $-\log p(F = f_x)$ 表示显著点具有稀有性性质. 如果点的特征对应较小的概率,那么该点提供大量的信息,因此具有很大的显著度. 这一项只取决于观察到的点的特征,因此是一个纯粹的自下而上的因子.

第 2 项 $\log p(F = f_x | S_x = 1)$ 包含了自下而上和自上而下的因子. 如果没有指定的任务,这一项代表点的独特性性质. 视觉认知领域存在一个原理,就是视觉神经元对图像中心最敏感,而周围的刺激会一直刺激神经元的反应. 我们称该独特性使点的位置从周围中突出来. 此外,当给一个特定的任务时,该项还包括自顶向下的知识和目标对象的特征值. 此时我们称这个属性为一致性. 举个例子,当给定的任务是找到你的穿黑色夹克的朋友,那么对于黑色点这一项的值将非常大.

第 3 项 $\log p(S_x = 1 | L = l_x)$ 提供了位置的先验信息并且对视觉特征是独立的. 根据 Juddy 等人的工作^[9],人们的视觉对图像中心最敏感. 因此本文假定,当人们随意浏览图像时,位置的先验信息就是图像中心的位置. 当观察者被赋予特定任务时,位置的先验信息就是目标可能出现的地方.

本文重点关注无特定任务时静态图像的自下而上的显著性. 那么第 2 项只显示独特性,第 3 项提供中央位置作为位置的先验信息. 因此,点的显著度 SD_x 包含 3 个因素:稀有性、独特性和位置的先验信息.

2 算法具体实现

本节主要提供显著性检测的 3 个因子的具体实现和 2 个权重结合方法.

2.1 显著性检测的3个因子

稀有性: 该稀有性因子由在不同尺度下对图像的每个像素点颜色、亮度和方向的自信息量3个特征计算得到. 本文采用与 Itti^[2] 类似的方法建立高斯金字塔 $R(\sigma)$, $G(\sigma)$, $B(\sigma)$, $I(\sigma)$ 和 $L(\sigma)$. 方向信息由金字塔 $O(\sigma, \theta)$, $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ 来求得. 对于所有金字塔, 选择4个不同的尺度 $\sigma \in [0, 1, 2, 3]$.

在每一个子图像, 我们计算每一点特征的自信息量. 具体过程如下: 首先我们使用含有100种灰度的直方图来计算每种灰度在子图像中出现的频率, 然后我们利用差值来计算子图像每点特征值的概率, 那么子图像的自信息量使用计算公式 $h(x) = -\log p(x)$ 来得到, 最后, 将不同尺度的每一个金字塔图像合并为1个特征图.

独特性: 目前有很多方法来计算独特性. 比如 Itti 等人^[2] 使用中间-周边差方法, 而 Goferman 等人^[10] 计算不同块的非相似性. 本文采用基于图论的显著性提取方法^[3] 来衡量点的“独特性”.

中心位置偏倚: 当人们拍照时, 他们自然地把感兴趣的目标设置在中央位置, 并且, 当他们自由地观赏一幅图像时, 关注点通常开始于图像的中心位置. 由于中心位置偏离, 所以经常会产生上述情况. 因此, 当没有指定任务时, 我们建议中心位置作为位置的先验信息.

本文采用径向基函数来建立中心偏倚模型. 中心偏倚函数是一个实数函数, 它只取决于该点与中心点之间的距离. 当该点远离中心点时, 该函数值变小. 有不同类型的径向基函数, 其中包括高斯核函数、反二次函数等等. 本文选用高斯核函数, 所以中心偏倚函数是:

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

δ 是尺度参数. 不同的 δ 对应不同的中心偏倚. δ 取值很大时, 就对应比较低的准确率; 而 δ 取值很小时, 就对应比较小的召回率. 本文通过从定位数据集中学习最好的尺度参数 δ .

2.2 2种权值组合算法

线性组合学习最优权值: 为了评价显著性检测问题的每个因素的重要性, 我们用 Logistic 模型从定位数据集中学习权值, 梯度下降算法被用来求解该优化问题.

基于神经网络的非线性组合学习: 大脑神经元是高度非线性和复杂的设备. 为了模拟神经元的性质, 本文基于正则化神经网络来进一步非线性地整合不同因素.

在本文中, 我们利用3层网络: 1个输入层、1个隐层和1个输出层. 输入层节点的数量对应该点的特征的数量. 在定位数据集里有数百幅图像, 在每一幅图像里有成千上万个点, 所以训练样本点的数目可能会很大, 但是每一点特征的数目并不多. 为了使结果更好, 在隐层节点的数量必须大. 在本文的试验中, 我们指定隐层节点的数量为70. 我们用反向传播算法来学习优化模型. 为了避免过拟合的问题, 在神经网络里添加正则项, 我们使用验证集来选择正则项的参数.

图1显示的是本文模型算法流程图. 首先是输入的图像分解成3个因子: “稀有性”、“独特性”和“中心位置偏倚”. 然后, 这些因子通过 Logistic 模型和正则化神经网络进行整合, 我们也可以发现非线性组合比线性组合效果更好.

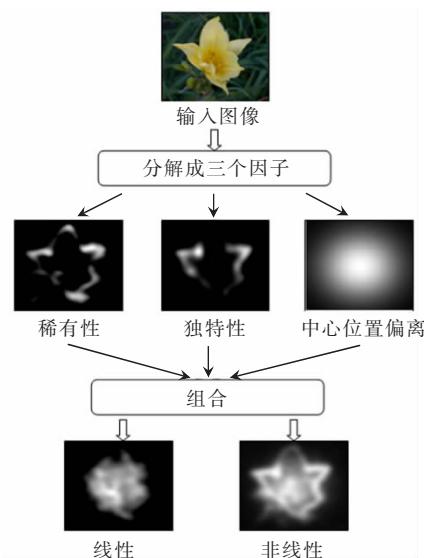


图1 本文模型算法流程图

Fig.1 The flow chart of our algorithm

3 实验比较

测试使用的机器(PIV, CPU 3.0 G, 2 G 内存), 操作系统为 Microsoft Windows XP Service Pack 2, 软件平台采用 Matlab 7.0, 在 MIT^[5] 和 Toronto 数据集^[9] 上进行测试.

3.1 MIT 数据集

MIT 数据集记录了15个观测者自由地观测1003幅室内和室外的自然图像的定位情况. 我们首先比较了2个特征融合策略模型在 MIT 数据集的表现. 为了达到更精确的结果, 我们对每个模型进行5次试验

来获得平均性能. 为集合 Logistic 模型的策略, 我们从数据集中随机选择 100 个测试图像和 200 个训练图像, 并且为了基于正则化神经网络的非线性整合模式, 我们随机选择 100 个训练图像、100 个验证图像和 200 个测试图像来验证正则化参数的最优选择.

我们也可以将本文算法和目前其他 3 种流行算法进行比较, 结果如图 2 所示. 因为本文算法是使用“中心位置偏倚”(Central bias)作为位置先验信息, 所以用缩写 CB 来表示. 从图 2 可以看出, CB 方法与其他 3 种方法相比, 命中率更高, 假阳性率更低.

3.2 Toronto 数据集

我们也在 Toronto 数据集比较其他不同方法的性能来验证结论. 这个数据集包含 11 个观众自由观看的数据, 即从 120 幅户外和室内场景彩色图像得到的数据. 对于 Logistic 模型整合的模式, 我们从数据集中随机选择 30 幅训练图像和 50 幅测试图像. 而对于基于正则化神经网络的非线性整合模式, 我们随机选择 30 幅训练图像、30 幅验证图像和 30 幅测试图像. 我们依然对每个模型进行 5 次试验来获得平均性能.

表 1 显示, 分别在有和没有位置先验信息的情况下, 使用 2 种改进的特征整合策略的性能比较. 我们观察发现: (1) 具有位置先验信息的模型比无位置先验信息的模型效果更好. 这证实当自由浏览图像时, 我们很容易注视这幅图的中心. (2) 正则化神经网络比 Logistic 模型能够更好地整合特征. 这也证实人类大脑神经元确实是非线性、复杂的设备.

图 3 显示本文算法和其他 3 种先进算法在 Toronto 数据集上的性能比较. 可以看出, CB 方法与其他 3 种方法相比, 命中率更高, 假阳性率更低.

4 总结

我们提出了一种基于贝叶斯方法和统计学习理论的视觉显著性算法来计算场景里各点的显著度. 通过推导贝叶斯公式, 我们得到视觉显著性的 3 个重要因子. 我们使用 2 种改进集成策略来结合不同的 3 个因子. 结果表明了本文算法的先进性. 扩展的定量评价也证明了非线性组合优于线性组合.

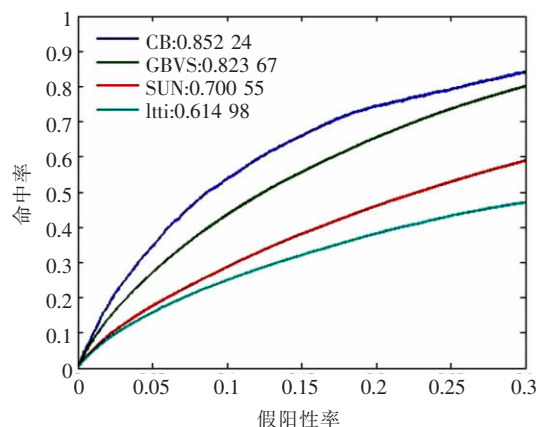


图 2 MIT 数据集上与其他方法性能比较

Fig.2 Performance in MIT Data Set comparing to other methods

表 1 Toronto 数据集的性能比较

Table 1 Performance in Toronto Data Set

方法	Logistic 回归		神经网络	
	无先验信息	有先验信息	无先验信息	有先验信息
AUC	0.765 0	0.790 4	0.826 1	0.833 2

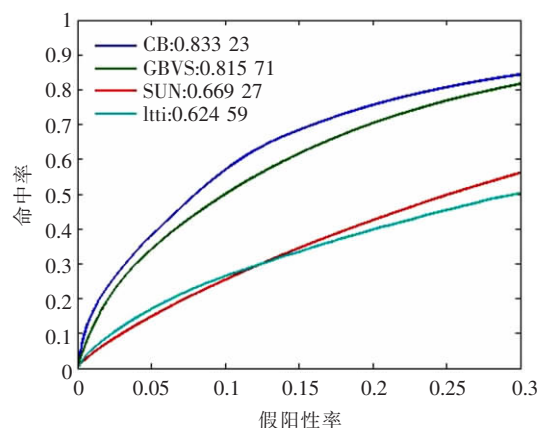


图 3 与其他方法在 Toronto 数据集上的性能比较

Fig.3 Performance in Toronto Data Set comparing to other methods

[参考文献]

- [1] 余映, 王斌, 张立明. 基于脉冲余弦变换的选择性视觉注意模型[J]. 模式识别与人工智能, 2010, 23(5): 616-623.
- [2] 韩成美, 吕皖丽, 罗斌. 基于 Tchebichef 矩的感兴趣区水印[J]. 计算机工程与设计, 2011, 32(5): 1585-1588.
- [3] 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, 22(11): 1254-1260.
- [4] Harel J, Koch C, Perona P. Graph-based visual saliency[J]. Neural Information Processing Systems, 2006, 19: 545-552.
- [5] Bruce N D B, Tsotsos J K. Saliency, attention and visual search: An information theoretic approach[J]. Journal of Vision, 2009, 9: 1-24.

(下转第 142 页)

[参考文献]

- [1] Feng D. Network storage key technology of research and progress [J]. Mobile Communications, 2009, 33(11): 35-39.
- [2] Chuan Y, Xu J P. Recommendation algorithm combining the user-based classified regression and the item-based filtering [C]// Processing of the International Conference on Electronic Commerce. Melbourne, 2006: 574-578.
- [3] 边根庆, 高松, 邵必林. 面向分散式存储的云存储安全架构 [J]. 西安交通大学学报: 自然科学版, 2011, 45(4): 41-45.
- [4] 蔡文, 杨春燕, 林伟初. 可拓工程方法 [M]. 北京: 科学出版社, 2000: 130-136.
- [5] 张光卫, 康建初, 李鹤松, 等. 基于云模型的全局最优化算法 [J]. 北京航空航天大学学报: 自然科学版, 2007, 33(4): 486-491.
- [6] 张光卫, 李德毅, 李鹏, 等. 基于云模型的协同过滤推荐算法 [J]. 软件学报, 2007, 18(10): 2403-2411.
- [7] 代劲, 何中市, 胡峰. 基于云模型的连续属性决策表简化算法 [J]. 南京大学学报: 自然科学版, 2009, 45(5): 638-644.

[责任编辑: 黄 敏]

(上接第137页)

- [6] Gao 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.
- [7] Zhang L, Tong M H, Marks T K, et al. SUN: A bayesian framework for saliency using natural statistics [J]. Journal of Vision, 2008, 8(7): 1-20.
- [8] Torralba A, Oliva A, Castelano M S, 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): 766-86.
- [9] Judd T, Ehinger K, Durand F, et al. Learning to predict where humans look [J]. ICCV, 2009: 8.
- [10] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2376-2383.

[责任编辑: 黄 敏]