高技术通讯 2013 年 第23 卷 第3 期:263 ~269

doi:10.3772/j.issn.1002-0470.2013.03.007

基于 Tobii 眼动和多尺度条件随机场的视觉显著性模型①

赵志诚②*** 白雅娟*** 周仁来*** 蔡安妮*

(*北京邮电大学信息与通信工程学院 北京 100876)
(**北京市网络系统与网络文化重点实验室 北京 100876)
(***北京师范大学认知神经科学与学习国家重点实验室 北京 100875)

摘 要 提出了一种新的基于分类的视觉显著性计算模型。运用频谱残差、全局亮度和颜色对比度分别检测图像的显著区域,随后将显著性检测看作一个图像标注问题,提出一种基于多尺度条件随机场(CRFs)的显著性融合算法,以产生标注结果。CRFs 的参数以 Tobii 眼动跟踪结果为依据,通过最大似然估计算法学习出来。实验结果表明,该模型优于当前的8 种典型算法,与心理学实验结果有较好的一致性。

关键词 Tobii 眼动,条件随机场(CRFs),视觉注意模型,视觉显著性,图像标注

0 引 言

视觉选择性注意是人眼具备的从复杂环境中快速搜索到感兴趣目标的机制。普遍认为这一搜索过程包含两个交互协同的阶段:早期自底向上(数据驱动)的预注意和后期主动的自顶向下(任务驱动)的集中注意[1]。前者由观察场景中对象的显著性(刺激量)决定,而后者取决于搜索任务、注意偏向、观察者的先验知识以及记忆。大量的心理学实验表明,当任务和先验不同时,注视点会不同,语义理解将存在巨大的差异,这直接导致自顶向下的注意难以用一种通用模型来量化和计算,已提出的模型仅限于少数特殊任务和领域知识[2-6]。因而自底向上的视觉显著性模型仍然是当前认知心理学、神经生理学和计算机视觉及交叉学科的研究重点,它帮助人将有限的感知资源快速指向有意义的目标和显著区域,降低了场景分析的复杂度。从这一角度出发,本研究提出了一种基于 Tobii 眼动和多尺度条件随机场(conditional random fields, CRFs)的视觉显著性模型,并对其进行了详细的剖析。

1 相关研究

显著性源于我们感知的多种视觉刺激,如图像中的颜色、亮度、纹理、形状、边缘、梯度等。当信息加工开始后,视觉刺激被整合在一起,通过抑制返回(inhibit of return,IOR)和竞争机制,显著目标被突显(pop-out)[7,8]出来。脑认知实验显示:当刺激物之间出现高对比度时就能激起感受野细胞的空间重组织,从而引发观察者的注意[9,10],即发生突显,进而完成对象的分割和识别。因此,采用对比度衡量显著性是自底向上模型建立的一种合理途径。

基于上述实验结果,结合 Treisman 等的特征整合理论[11],Mishkin 等人提出的两条通路理论[12],Itti 等[13]和 Wolfe 等[14]提出了代表性的自底向上注意模型,实现了视觉资源的定量计算,并应用在目标识别[2]、图像分割[15]、自适应图像缩放[16]、广告设计[17]、图像检索[18]等领域。

最近几年,一些有别于 Itti 模型的模型也被提出来,例如 Torralba 等提出的基于空间结构的模型[19]、Hou 等的基于频谱残差的注意模型[20]、Cheng 等的全局颜色对比度显著模型[21]、Achanta 等的基于频率统计的方法[22,23]、Goferman 等的基于上下文的方法[24]、Harel 等的基于子图的检测算法[25]、Ma 等的模糊生长[26]、Zhai 等的基于时空线索的方法[27],它们都在实验中显示出了一定的有效性。但是上述模型仍存在明显的不足,即模型的建立脱离了眼动实验的支持,因而不具备通用性。例

① 国家自然科学基金重大研究计划(90920001)和国家自然科学基金(61101212)资助项目。
② 男,1976 年生,博士;研究方向:语义视频搜索,视觉认知,模式识别;联系人,E-mail: zhaozc@bupt.edu.cn
(收稿日期:2012-03-20)

如图 1 中显示的 Itti 模型和真实眼动存在较大差异。

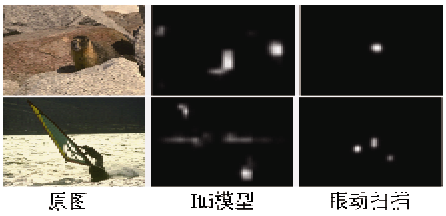

图 1 Itti 注意模型和眼动扫描的差异

针对以上问题，本文提出一种新的选择性注意模型，其性能优于当前的 8 种典型算法，与心理学实验结果有较好的一致性。本文的主要贡献如下：(1) 利用 Tobii 眼动仪进行了无任务条件下注视点汇聚和变化的眼动实验，将所获得的眼动数据应用到选择性注意模型的建立上；(2) 针对多数现有方法强调局部对比度的不足，提出了亮度直方图及全局颜色对比度两种新的显著性检测子，并结合频谱残差，在空域和频域分别提取这三者的高阶矩作为训练特征，使得所检出的显著对象具有较好的整体性和连续性；(3) 将显著性检测看作图像分类问题，提出了一种基于多尺度 CRFs 的显著度标注算法；CRFs 的参数则以眼动数据为依据采用最大似然估计法学习，从而提高了注意模型的综合辨识能力。

2 注意模型框架

图 2 显示了本文提出的视觉显著性模型框架。框架主要包括 4 个部分：(1) 利用 Tobii 眼动仪进行眼部跟踪实验，记录注视点的转移及热区图；(2) 分别提取图像的频谱残差、亮度和全局颜色对比度显著图；(3) 基于显著图，提取多尺度特征送入 CRFs 模型进行最优参数训练；(4) 利用 CRFs 进行图像的显著性标注，给出最终结果。

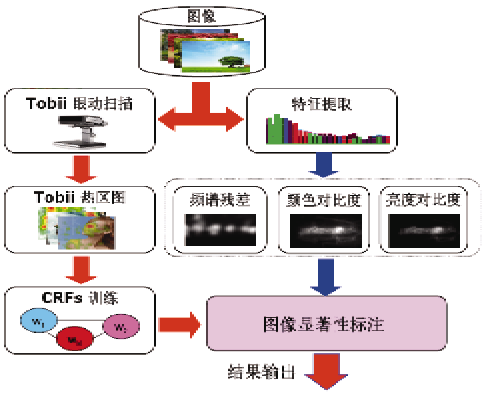

图 2 注意模型框架

3 Tobii 眼动实验

3.1 实验对象

参与实验的对象共 60 名大学生，其中男生 42 名，女生 18 名，平均年龄 22.3 岁，均为右利手。裸眼或矫正视力 5.0 以上，色觉正常。

3.2 实验仪器与刺激材料及流程

实验采用 Tobii T120 眼动仪，采样率为 120Hz，分辨率为 1280 × 1024。实验流程如图 3 所示。实验中，从 Corel 图像库中选取人物、动物、植物以及无明显主题的自然场景和城市共 200 幅作为实验样本。实验对象双眼正视屏幕，采用自由观看范式（无指导语）进行眼动实验，在 500ms 的注视点校准后，每隔 150ms 依次呈现图像和黑屏，通过眼动仪观测并记录注视点的转移及热区图。实验对象双眼在一点（误差范围 0.5° 内）的注视时间达到 100ms 就记为一次注视。实验结束后，将 60 名被试的热区图进行累积平均获得眼动显著图以及注视点转移图。

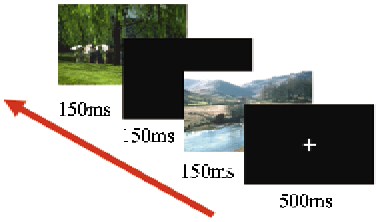

图 3 眼动实验流程

4 频谱残差及颜色亮度显著图

大量的心理学实验表明，人眼在只受单纯的刺激驱动，即在无任务指引时，产生的选择性注意区域较非显著区域而言，与邻域的颜色、亮度、纹理等的对比度更高。利用这一共性，我们可以分别在空域和频域提取不同的显著性检测子。

4.1 基于频谱残差的显著图

自底向上的选择性注意能够根据感受视野的各种突变信息如闪烁、锐变边缘等提取感兴趣物体，而对无关或不重要的细节进行大量的信息压缩。统计观察显示，图像中的信息包括两方面：高频度出现的信息和较稀疏的信息。前者一般属于背景及噪声，应该被压缩，而后者通常是有用的前景对象，需要被突显出来。因此，注意模型的建立要保证在压缩图像中高频度特征的同时，对稀疏特征保持较高的敏感度[10]。

Hou 等人将这一原则应用到频域中,提出了一种基于频谱残差的显著度检测算法[20]。该算法分析出图像中环境噪声的表现形式,采用频谱残差的方法将其滤除,从而得到稀疏特征的位置,即图像的显著性区域。本文将算法扩展到图像的多个尺度空间,改进了算法性能:

$$S_f = \frac{1}{N}\sum_{k=1}^{N} G_T \cdot [\ln(|FFT(I_k)|) - F \otimes \ln(|FFT(I_k)|)] \quad (1)$$

式中,N 为金字塔图像 I_k 的个数,G_T 为高斯加权模板,F 为一平滑滤波器。

4.2 基于亮度和全局颜色对比度的显著图

我们注意到,基于频谱残差的方法能够较准确检测纹理变化明显、细节丰富的目标区域,但忽略了对象的整体性,如图 4 所示。因此,我们提出利用全局颜色和亮度对比度来弥补上述不足。即为图像中颜色或亮度相似的区域分配相同的显著性,均匀地突出目标。

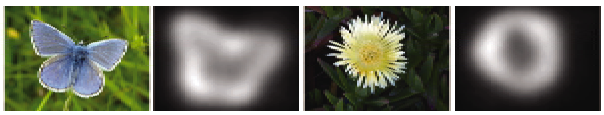

图 4 频谱残差的显著图

4.1.1 基于亮度直方图对比度的显著性模型

为了优先响应高亮度对比度的刺激,我们提出了基于亮度对比度的显著性计算方法,即依据与其它像素的亮度差异来分配当前像素的显著值。为了简化计算,我们将图像的亮度值均匀量化到 32 级,然后根据下式来计算每种亮度的对比度:

$$S_I(I_h) = \sum_{k=1}^{n=32} p_k D(I_h, I_k) \quad (2)$$

式中,$D(I_h, I_k)$ 代表两种亮度之间的距离,I_h 为当前像素的亮度值,p_k 表示亮度 I_k 在图像中出现的概率。与频谱残差一样,我们将 $S_I(I_h)$ 扩展到图像的多个尺度空间进行叠加,G_T 为高斯加权模板。

$$S_I = \frac{1}{N}\sum_{k=1}^{N} G_T \cdot S_I(I_k) \quad (3)$$

4.1.2 基于全局颜色对比度的显著性模型

颜色的突显效应较亮度更加明显,且颜色相似的区域具有相近的显著性;图像中的显著目标具有整体性,能较均匀地被突显出来而与周围环境分离。基于此观察,本文提出了基于全局颜色对比度的显著性检测方法,算法主要包括 5 个步骤:

(1)将图像变换到 LUV 颜色空间,并利用矢量量化方法[28]将原图像量化为 I_q,颜色总数为 C_N,实验中,我们选 $C_N = 50$。

(2)统计 I_q 的颜色直方图 H_q 并计算第 k 种颜色 c_k 在图像出现的概率 q_k。

(3)参考文献[21]计算每种颜色的对比度:

$$S_c(c_h) = \sum_{k=1}^{C_N} q_k D(c_h, c_k) \quad (4)$$

$D(c_h, c_k)$ 是颜色 c_h 和 c_k 在 LUV 颜色空间中的距离。

(4)采用式(5)对显著性进行平滑,将每种颜色的显著值替换为相近颜色显著值的加权平均,从而使得目标均匀地突显出来。实验中,我们选择与当前颜色 c_h 邻近的 N_c 种颜色进行叠加,如下式所示:

$$S_c'(c_h) = \frac{1}{(N_c - 1)V_T}\sum_{k=1}^{C_N} [V_T - D(c_h, c_k)] S_c(c_k) \quad (5)$$

式中,V_T 为 c_h 和它邻近的 N_c 种颜色的距离和。

(5)将 $S_c'(c_h)$ 扩展到多个尺度空间进行加权平均,从而获得颜色显著值 S_c,

$$S_c = \frac{1}{N}\sum_{k=1}^{N} G_T \cdot S_c'(c_k) \quad (6)$$

G_T 为高斯加权模板。

5 基于条件随机场的显著性融合

在显著性融合框架中,我们将显著区域的检测当作为一个图像标注问题,即显著区域作为前景而非显著区域作为背景来进行标注。因此,我们提出基于多尺度条件随机场(CRFs)的标注算法,来实现多种显著特征 S_f,S_p 和 S_I 的融合。

5.1 条件随机场(CRFs)模型

条件随机场[29]是一种图模型,已被应用于序列以及人造目标的标注[30]、交互式图像分割[31]、显著对象检测等[32]。本文则提出新的基于 CRFs 的显著性标注算法。首先将图像分成互不重叠 4×4 像素的块 $(x_i)_{4\times4}$,然后建立 CRFs 模型。图5 显示了本

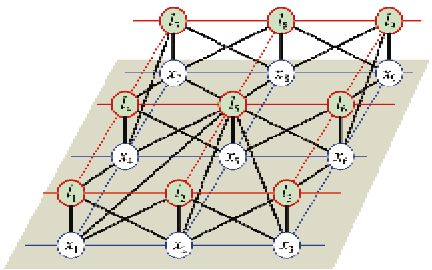

图 5 CRFs 模型用于图像标注

文提出的两层无向图模型：下层为图像块 x_i，上层代表对应的标签 l_i，l_i 与 x_i 之间的连线表示随机场中图像块及其邻域 N_i 对标注结果的影响。

对于一幅输入图像 I，令 $X = \{x_i\}_{i \in S}$ 为观测图像块序列，i 是块集合 S 中的第 i 个块；$L = \{l_i\}_{i \in S} = \{1, -1\}$ 为像素 i 对应的标签。两者的关系用条件概率分布 $P(L/X)$ 描述为

$$P(L \mid X) = \frac{1}{Z}\prod_S P_S(L \mid X) \tag{7}$$

其中，$Z = \sum_L \prod_S P_S(L \mid X)$ 是归一化因子。$P_S(L \mid X)$ 定义为独立势函数 k_i 和交互势函数 g_{ij} 的线性组合[31]：

$$P_S(L \mid X) = \exp\left(\sum_{i \in S} k(l_i, x_i) + \sum_{i \in S}\sum_{j \in N_i} g(l_i, l_j, x_i, x_j)\right) \tag{8}$$

独立势函数假定图像块之间是相互独立的，标签值 l_i 完全由观测序列 X 的特征向量 f_i 决定：

$$k(l_i, x_i) = \ln \frac{1}{1 + \exp(-l_i \boldsymbol{\theta}_1^{\mathrm{T}} f(x_i))} \tag{9}$$

交互势函数反映的是邻域内图像块显著性的依赖关系，如式

$$g(l_i, l_j, x_i, x_j) = \ln \frac{1}{1 - \exp(-\delta_{ij} \boldsymbol{\theta}_2^{\mathrm{T}} D(x_i, x_j))} \tag{10}$$

所示。其中，$D(x_i, x_j) = \| f(x_i) - f(x_j) \|$ 表示图像块特征向量的距离。当 $l_i = l_j$，$\delta_{ij} = 1$，否则 $\delta_{ij} = -1$。$\boldsymbol{\theta}_1^{\mathrm{T}}$ 和 $\boldsymbol{\theta}_2^{\mathrm{T}}$ 为待估计的参数向量。当相邻两个图像块被标记为不同的标签时，交互特征表现为一个惩罚项。即对两个块，它们的颜色值越相近，就越不可能分属两个标签，该特征项保证了一个显著性物体内部的像素块同样被标记为显著点。

5.2 CRFs 模型参数估计

我们用 Tobii 眼动图像的注视点和显著区域来建立样本集，进而提取频域、亮度和颜色特征，以此训练 CRFs 的模型参数。

5.2.1 训练样本集及特征提取

为了有效地将眼动数据结合进注意模型，我们对第 2 节中获得的眼动注视点数据与覆盖区域进行了统计分析。根据分析结果，对目标数目不同的图像采取了不同的处理方法：多目标图像采用除第一注视点外的所有注视点生成显著图，而单目标图像则采用所有注视点。然后以注视点的中心为圆心、半径为 30 像素的区域作为显著区域，并用邻域为 15 像素的高斯模板进行平滑，如图 6 所示。

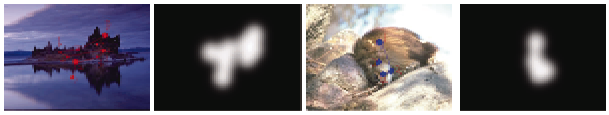

图 6 训练图像产生

根据第 3 节，我们分别产生基于频谱残差、亮度对比度和全局颜色对比的显著图，分成大小为 4×4 的块后，计算 3 幅显著图中对应图像块的一阶、二阶和三阶矩，并级联起来构成局部势函数中的 $f(x_i)$。参考文献[30]，我们又提取了图像另外两个尺度空间上的特征矢量，即当前图像块 x_i 的 8 邻域、16 邻域上求 $f(x_j)$，并按照式(10)计算 D_{ij}。

5.2.2 参数估计

基于最大似然准则，本文采用循环置信传播(loopy belief propagation，LBP)算法[33]，对 K 幅训练样本图像进行 CRFs 参数估计：

(1)构造似然函数：

$$L(\boldsymbol{\theta}) = \prod_{k=1}^{K}\prod_{i \in S} P_S(l_i^k / x_i^k, \boldsymbol{\theta}) \tag{11}$$

(2)求似然函数的最大值：

$$\boldsymbol{\theta}^* = \arg\max_{\theta} L(\boldsymbol{\theta}) \tag{12}$$

此时 $\boldsymbol{\theta}^* = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ 即为最大似然准则下的最优参数。

(3)根据 $\boldsymbol{\theta}^*$，计算后验概率，将最大值对应标签作为显著图像标注结果：

$$l = \arg\max_{l} P(l_i \mid x_i, \boldsymbol{\theta}^*) \quad l_i \in \{1, -1\} \tag{13}$$

6 实验及比较

为了验证算法的有效性，我们选择了现有的 8 种典型算法：Itti[13]、Ma[26]、Harel[25]、Hou[20]、Achanta[22]、Goferman[24]、Achanta[23]、Zhai[27] 进行性能比较。

6.1 人造图像的实验对比

为了评估算法对于形状、颜色、朝向的显著性辨识能力，我们构造了 37 幅具有明确心理学显著性的布局和稀疏度不同的“红花-绿叶”测试图像，加入随机噪声后，对本文提出的算法和 8 种算法进行了测试。图 7 显示了部分结果。从图中我们可以看到，当目标对象较小且分布紧密时，8 种现有算法均能在一定程度上突显出对象，但当红花绿叶(红叶绿花)区域较大，同时出现或颜色相同(都是红色或绿色)时，现有算法的性能下降比较明显，而本文算法

的检测结果则和心理学实验保持了较好的一致性。这是由于:(1)本文采用的全局对比度(亮度和颜色)特征和 CRFs 中使用的多尺度交互势函数改善了现有算法在局部对比度上的不足,避免了仅在轮廓、高亮度或高颜色对比度附近产生较高显著性,使得检出的显著对象具有整体性,并且在对象分布稀疏时,仍然能够对形状和颜色差别具有较好的辨识能力;(2)我们的模型参数是通过眼动数据训练的。通过眼动数据的学习,聚焦了显著区域范围,去除了计算出来的大量无关显著区域,由于人的眼动数据是形状、颜色、朝向、亮度等感知显著性综合辨识的结果,因此学习后的模型表现出了综合辨识能力的改善。

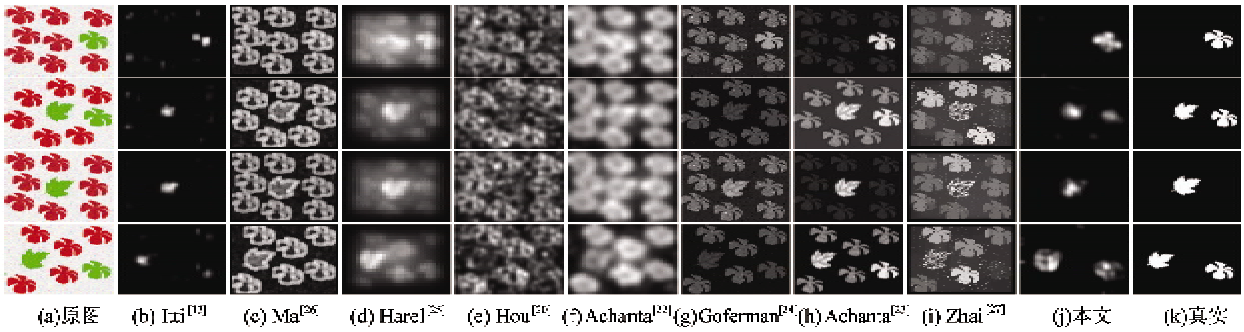

图 7　和现有 8 种算法的显著性检测性能比较 1

6.2　自然图像的实验对比

为了考察算法的综合性能,我们在自然图像显著性测试集上[23]测试了本文算法,图 8 显示了部分结果。可以看到,频谱残差、亮度对比度和全局颜色对比度 3 种显著特征具有互补性,而 CRFs 实现了有效地融合。

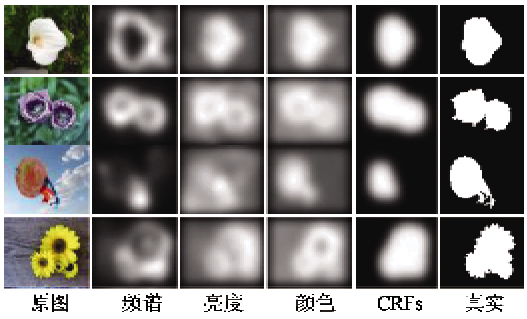

图 8　融合算法和基于频谱残差、亮度对比度、全局颜色对比度算法的比较

图 9 是本文算法与 8 种现有算法在自然图像集[23]上的部分比较结果。图 10 则显示了本文算法和上述 8 种算法以真实标注结果为基准计算的准确率、召回率以及 F_β[21]。结果显示,本文的算法由于利用了 Tobii 眼动数据作为训练样本,并采用多尺度 CRFs 模型进行多种显著特征的融合,无论从主观评价还是客观评价上都优于上述 8 种算法。图 9 还显示,只采用简单的阈值法,本文的算法就能有效地分割出显著目标(倒数第 2 列),因此,本文算法除了能有效检测图像的显著区域外,还能应用于显著图像的检索等领域。

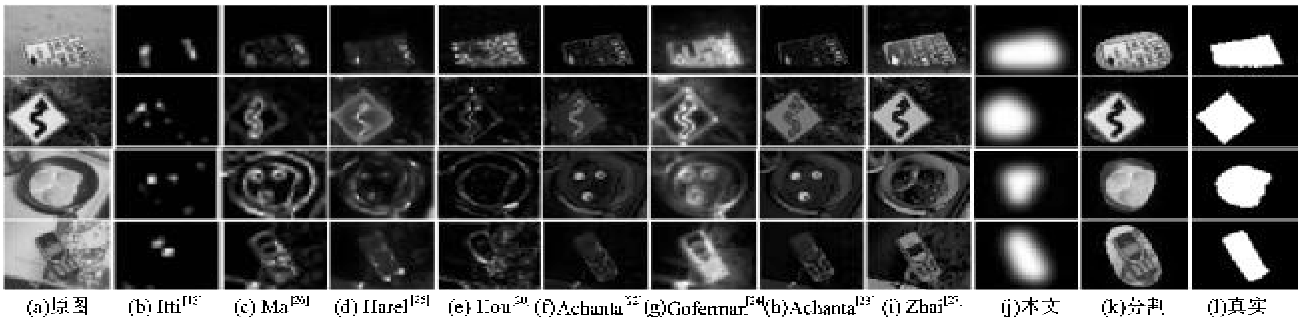

图 9　和现有 8 种算法的显著性检测性能比较 2

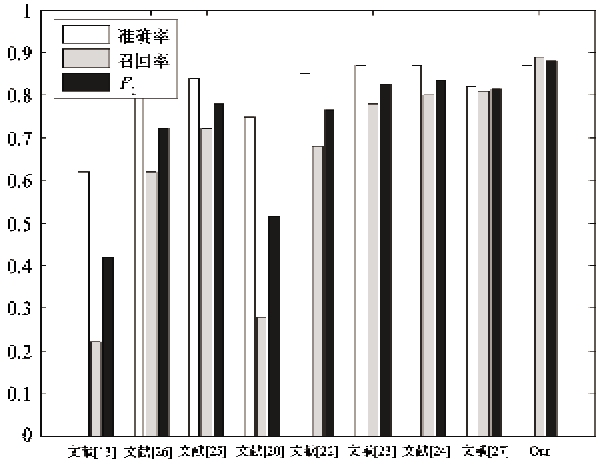

图 10 准确率、召回率以及 F_1

7 结 论

本文提出了一种新的基于学习的图像显著性检测方法。运用频谱残差、亮度、全局颜色对比度分别检测图像的显著区域,随后采用一种基于多尺度条件随机场的显著性融合算法产生最终结果。CRFs 的参数以 Tobii 眼动仪的眼部跟踪结果为依据,通过最大似然估计法学习出来。实验结果表明,本文提出的模型在人造图像集以及自然图像集上均优于现有的 8 种典型算法。

参考文献

[1] Marr D. Vision. New York: Freeman W H, 1982

[2] Forssén P E, Meger D, Lai K, et al. Informed visual search: combining attention and object recognition. In: Proceedings of International Conference on Robotics and Automation, Pasadena, USA, 2008. 953-942

[3] Mozer M C, Shettel M, Vecera S. Top-down control of visual attention: a rational account. In: Proceedings of the Neural Information Processing Systems. Vancouver, Canada, 2005. 923-930

[4] Gao D S, Han S, Vasconcelos N. Discriminant saliency, the detection of suspicious coincidences and applications to visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(6): 989-1005

[5] Lee T S, Mumford D. Hierarchical bayesian inference in the visual cortex. *Journal of the Optical Society of America*, 2003, 20(7): 1434-1448

[6] Chikkerur S, Serre T, Tan C, et al. What and where: a bayesian inference theory of attention. *Vision Research*, 2010, 50(22): 2233-2247

[7] Cave K R, Wolfe J M. Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 1990, 22(2): 225-271

[8] Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 1985, 4(4): 219-227

[9] Solomon J A, Morgan M J. Facilitation from collinear flanks is cancelled by non-collinear flanks. *Vision Research*, 2000, 40(3): 279-286

[10] Mareschal I H, Andrew J, Shapley R M. A sychophysical correlate of contrast depencent changes in receptive field properties. *Vision Research*, 2002, 42(15): 1879-1887

[11] Treisman A M, Gelade G. A feature-integration theory of attention. *Cognitive Psychology*, 1980, 12(1): 97-136

[12] Ungerleider L G, Mishkin M. Two Cortical Visual Systems. Cambridge: MIT Press, 1982

[13] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(11): 1254-1259

[14] Wolfe J M, Cave K R, Franzel S L. Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology Human Perception and Performance*, 1989, 15(3):419-433

[15] Han J W, Ngan K N, Li M J, et al. Unsupervised extraction of visual attention objects in color images. *IEEE Transactions on Circuits and Systems for Video Technology*, 2006, 16(1): 141-145

[16] Wang Y S, Tai C L, Sorkine O, et al. Optimized scale-and-stretch for image resizing. *ACM Transactions on Graphics*, 2008, 27(5): 1-8

[17] Itti L. Models of bottom-up and top-down visual attention. [Ph. D Dissertation]. California Institute of Technology, 2000

[18] Chen T, Cheng M M, Tan P, et al. Sketch2photo: internet image montage. *ACM Transactions on Graphics*, 2009, 28(5): 1-10

[19] Torralba A, MurPhy K P, Freeman W T. Contextual models for object detection using boosted random fields. In: Proceedings of the Neural Information Processing Systems. Vancouver, Canada, 2005. 1401-1408

[20] Hou X D, Zhang L Q. Saliency detection: a spectral residual approach. In: Proceedings of the 20th conference on Computer Vision and Pattern Recognition, Minneapolis, USA, 2007. 1-8

[21] Cheng M M, Zhang G X. Global contrast based salient region detection. In: Proceedings of the 24th conference on Computer Vision and Pattern Recognition, Colorado Springs, USA, 2011. 409-416

[22] Achanta R, Estrada F. Salient region detection and segmentation. In: Proceedings of International Conference in Computer Vision Systems, Santorini, Greece, 2008. 66-75

[23] Achanta R, Hemami S. Frequency-tuned salient region detection. In: Proceedings of the conference on Computer Vision and Pattern Recognition, Miami, USA, 2009. 1597-1604

[24] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. In: Proceedings of the conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010. 2376-2383

[25] Harel J, Koch C, Perona P. Graph-based visual saliency. *Advances in neural information processing systems*, 19, 2007: 545-552

[26] Ma Y F, Zhang H J. Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of the ACM International Conference on Multimedia, Berkeley, USA, 2003. 374-381

[27] Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the ACM International Conference on Multimedia, Santa Barbara, USA, 2006. 815-824

[28] 赵志诚, 蔡安妮. 图像颜色矢量量化算法. 北京邮电大学学报, 2007, 30(5): 131-134

[29] Lafferty J D, McCallum A, Pereira F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the International Conference on Machine Learning, Williamstown, USA, 2001. 282-289

[30] Kumar S, Hebert M. Man-Made structure detection in natural images using a causal multiscale random field. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, Madison, USA, 2003. 119-126

[31] Boykov Y, Jolly M P. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In: Proceedings of the International Conference on Computer Vision, Vancouver, Canada, 2001. 105-112

[32] Liu T, Sun J, Zheng N N, et al. Learning to detect a salient object. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, 2007. 1-8

[33] Weiss Y. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 2000, 12(1): 1-41

A new visual saliency model based on Tobii eye tracking and multiscale conditional random fields

Zhao Zhicheng[*,**], Bai Yajuan[***], Zhou Renlai[***], Cai Anni[*]

([*] School of Information and Communication Engineering, Beijing University of Posts and Telecommunication, Beijing 100876)

([**] Beijing Key Laboratory of Network System and Network Culture, Beijing 100876)

([***] National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875)

Abstract

A new visual saliency model based on images' labeling was proposed. The spectral residual, global luminance and color contrast were first used to detect images' salient regions, and then, a saliency fusion approach based on multiscale conditional random fields (CRFs) was present to generate the final result. According to Tobii eye tracking data, the optimal parameters of CRFs were trained by the maximum-likelihood estimation (MLE) method. The experimental results show that the proposed model is better than the eight state-of-the-art models.

Key words: Tobii eye tracking, conditional random fields (CRFs), visual attention model, visual saliency, image labeling