

## 利用层次先验估计的显著性目标检测

徐威<sup>1</sup> 唐振民<sup>1</sup>

**摘要** 有效的显著性目标检测在计算机视觉领域一直是具有挑战性的问题. 本文首先对图像进行树滤波处理, 采用 Quick shift 方法将其分解为超像素, 再通过仿射传播聚类把超像素聚集为代表性的类. 与以往方法不同, 本文提出根据各类中拥有的超像素的类内和类间的空间离散程度及其位于图像边界的数目, 自适应地估计先验背景, 并提取条状背景区域; 由目标性度量 (Objectness measure) 粗略地描述前景范围后, 通过与各类之间的空间交互信息, 估计先验前景; 再经过连通区域优化前景与背景信息. 最后, 综合考虑各超像素与先验背景和前景在 CIELab 颜色空间的距离, 并进行显著性中心加权, 得到显著图. 在 MSRA-1000 和复杂的 SOD 数据库上的实验结果表明, 本文算法能准确、完整地检测出显著性目标, 优于 21 种 State-of-the-art 算法, 包括基于部分类似原理的方法.

**关键词** 显著性目标检测, 层次估计, 先验背景和前景, 显著性中心加权, 仿射传播聚类

**引用格式** 徐威, 唐振民. 利用层次先验估计的显著性目标检测. 自动化学报, 2015, 41(4): 799–812

**DOI** 10.16383/j.aas.2015.c140281

### Exploiting Hierarchical Prior Estimation for Salient Object Detection

XU Wei<sup>1</sup> TANG Zhen-Min<sup>1</sup>

**Abstract** Effective salient object detection is still a challenging problem in computer vision. In this paper, images are processed by tree filter firstly. Then quick shift is adopted to decompose images into perceptually homogeneous superpixels. This is followed by using affinity propagation clustering to aggregate all the superpixels into representative clusters. Different from previous methods, this paper proposes a novel adaptive background prior estimation strategy. The intra-cluster and inter-cluster spatial variances of superpixels owned by some cluster are calculated, and the numbers of superpixels located along the image boundary are counted to complete the processing. Also, the strip regions of background are extracted. The objectness measure is employed to get a coarse foreground scope, which is then used to compute the spatial interactive information with all the clusters, so as to get the foreground prior. After the optimization based on connected regions, A final foreground and background prior are confirmed. A saliency map is generated by measuring the differences of CIELab color space between all the superpixels and the background and foreground prior, enhanced by salient center weighting. Experimental results on MSRA-1000 and complicated SOD databases show that the proposed method can accurately detect the whole salient object. It is superior to the 21 state-of-the-art methods, including the methods partially based on similar principles.

**Key words** Salient object detection, hierarchical estimation, background and foreground prior, salient center weighting, affinity propagation clustering

**Citation** Xu Wei, Tang Zhen-Min. Exploiting hierarchical prior estimation for salient object detection. *Acta Automatica Sinica*, 2015, 41(4): 799–812

人类视觉能够准确、快速地识别图像中最感兴趣的重要区域. 而如何模拟这种视觉注意机制, 即显著性检测, 已被心理学、认知神经学和计算机视觉领域的专家广泛研究<sup>[1]</sup>. 早期的显著性倾向于预测人眼视觉关注的焦点<sup>[2]</sup>, 而显著性目标检测由于能够得到较为完整的整体目标, 在图像分割、目标识别和

图像检索等领域有着广泛的应用, 因此受到更多的关注. 经典理论将视觉显著性分为快速的、自底向上的、数据驱动的显著性提取以及慢速的、自顶向下的、目标驱动的注意机制<sup>[3]</sup>. 本文研究的则是自底向上的显著性目标检测方法, 可分为基于局部和基于全局两大类.

基于局部的方法通常假设某区域与其邻域存在某种可量化的区分性. Itti 等<sup>[4]</sup>融合多尺度的颜色、亮度、方向特征, 计算跨尺度的局部中央-周边差异, 获取显著图. Ma 等<sup>[5]</sup>根据局部分块在 CIELuv 空间的颜色对比度结合模糊膨胀的方法分配各点的显著值. 类似地, Achanta 等<sup>[6]</sup>采用多尺度双窗口的形式, 以内外窗口的像素点在 CIELab 空间的距离

收稿日期 2014-05-04 录用日期 2014-12-03  
Manuscript received May 4, 2014; accepted December 3, 2014  
国家自然科学基金 (61473154) 资助  
Supported by National Natural Science Foundation of China (61473154)  
本文责任编辑 封举富  
Recommended by Associate Editor FENG Ju-Fu  
1. 南京理工大学计算机科学与工程学院 南京 210094  
1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094

作为显著性度量. 而 Rahtu 等<sup>[7]</sup> 采用贝叶斯框架, 计算局部中心与外围边缘像素点之间正规化的颜色对比度, 得到显著图. 但仅仅基于局部特征, 往往突出物体边缘或较小的局部物体, 目标内部显著值缺失.

多数方法从全局角度刻画显著目标的独特性. 谱残余<sup>[8]</sup> 在统计自然图像的 log 频谱规律后, 认为谱线中的奇异点, 表明了图像中的显著区域. 频率调谐方法<sup>[9]</sup> 提出在频域分析图像, 将高斯滤波后的各像素点与图像整体颜色均值的差异, 作为显著值. Cheng 等<sup>[10]</sup> 使用稀疏直方图简化图像颜色, 采用基于图的分割方法划分图像后, 通过测量某区域与其余区域的颜色对比度来计算显著性. 同样基于稀疏直方图, 文献 [11] 采用高斯混合模型将颜色相似的像素聚为区域, 综合考虑各区域的颜色对比度和空间分布, 以概率模型生成显著图. Shen 等<sup>[12]</sup> 将图像表示成某特征空间下的低秩矩阵 (非显著性区域) 和稀疏噪声 (显著性区域) 的合成, 并结合高层先验知识, 获取显著区域. Perazzi 等<sup>[13]</sup> 采用 SLIC<sup>[14]</sup> 将图像分解为超像素, 通过高维的高斯滤波估计全局显著值. 同样不以像素为单位, 文献 [15] 将图像分块, 定义各块经过主成分分析 (Principle component analysis, PCA) 获取主成分后之间的差异为模式可区分性, 结合全局颜色对比度, 得到显著程度. 这些方法都取得不错的效果, 但自然图像中目标变化极大, 一般性的目标检测是个不适定的难题, 应该多角度、多层次地考虑各种显著性假设.

因此, 文献 [16] 不从前景目标出发, 而是提出大部分图像边界区域属于背景. 本文也以此为初步依据, 但并不依赖已有的较为简单的显著性算法<sup>[17]</sup> 确定真正的背景区域 (显著性目标也可能接触到图像边界), 而是采用仿射传播聚类将图中超像素按 CIELab 空间的颜色差异分类, 然后依靠各类中超像素类内和类间的空间离散程度和其位于图像边界的数目, 自适应地确定最有可能属于背景类, 并提取条状背景区域. 接着估计先验前景信息, 但不同于

传统的中心偏置 (Center-bias) 假设<sup>[18]</sup>, 也不提取显著点并结合凸包 (Convex-hull)<sup>[19]</sup> 粗略地划分前景区域, 而是利用目标性度量<sup>[20]</sup> 的统计信息, 生成前景概率图, 阈值分割后获取粗略的前景范围. 再由各类与该前景范围的空间交互信息, 得到精确的先验前景, 并通过连通区域对其优化. 与融合基于图的分割、目标性度量、上下文感知的显著性目标检测方法<sup>[21]</sup> 不同, 本文的显著值由各超像素与预估前景和背景在 CIELab 颜色空间的距离计算得到, 并进行显著性中心加权. 图 1 显示了几种使用部分类似原理的方法与本文方法的视觉效果对比, 本文能更为准确、完整地突出显著性整体目标.

本文组织结构如下: 第 1 节主要介绍图像预处理过程, 包括树滤波、超像素分解和仿射传播聚类; 第 2 节主要介绍先验背景和前景的估计以及基于连通区域的前景和背景的优化; 第 3 节主要总结图像的层次划分以及显著性计算过程; 第 4 节的内容为实验结果及分析; 第 5 节为结论.

## 1 图像预处理

自然图像的颜色较为生动和丰富, 如图 2(a) 所示花朵和背景, 局部颜色存在细微变化, 虽然人类视觉更倾向于将其忽略, 但这些变化的累积容易在局部产生不必要的高显著值. 但采用稀疏直方图<sup>[10-11]</sup> 将自然图像的颜色数目减少到 85 个左右, 虽然能确保覆盖 95% 以上的像素, 却难免丢失部分图像细节. 因此本文采用树滤波 (Tree filtering)<sup>[22]</sup> 平滑图像, 使相似的像素点颜色趋于一致, 但保留各区域的边缘信息 (防止前景和背景的边界变模糊). 树滤波综合考虑了像素点间的空间距离、颜色差异以及连接性, 以概率的形式度量像素点间的相似性. 最重要的连接性由最小生成树模型刻画. 其效果如图 2(b) 所示, 背景的颜色更加均匀, 目标内部的颜色也更为平滑.

超像素能够将颜色和空间接近的区域作为局部

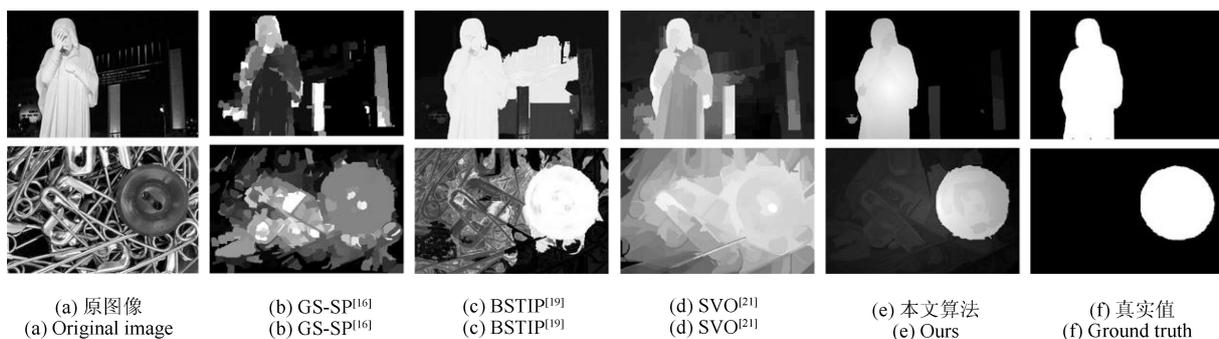


图 1 显著图的视觉效果对比

Fig. 1 Visual comparison of saliency maps

单元统一处理, 比基于像素级的显著性计算<sup>[6-9]</sup>更为有效, 且符合视觉注意机制. Quick shift<sup>[23]</sup>采用核密度估计, 充分考虑各像素点间的空间和颜色一致性, 且搜索快速, 得到的超像素更为合理和稳定. 图 2(c) 为超像素分解结果, 各超像素颜色由其所含有的像素点的颜色均值代替.

更进一步, 将相似的超像素聚集为更大范围的区域, 有助于完整地分离显著性目标与背景. 但自然场景变化复杂, 各种图像的颜色分布并不确定, 传统的  $K$  均值聚类由于需要预先假设分类数目, 并不适用. 而仿射传播聚类<sup>[24]</sup>弥补了这样的缺陷. 该方法将所有的超像素作为潜在聚类中心, 构造各超像素间的相似度矩阵, 以某超像素在相似矩阵中对角线上的数值(参考度)作为评判标准, 通过迭代不断更新每一个超像素的吸引度和归属感, 自适应产生若干个高质量的聚类中心. 本文中, 超像素  $sp_i$  和  $sp_j$  的相似度为

$$S(i, j) = - \sum_{k \in \{L, a, b\}} (sp_i^k - sp_j^k)^2 \quad (1)$$

即为 CIELab 空间中各通道的颜色距离之和的负数, 据此构造相似度矩阵. 聚类效果见图 2(d).

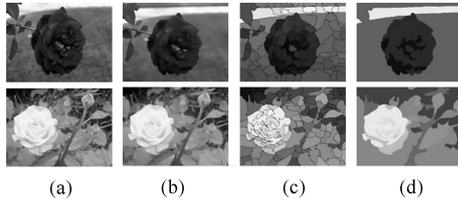


图 2 预处理结果 ((a) 原图; (b) 树滤波图; (c) Quick shift 超像素分解; (d) 仿射传播聚类)

Fig. 2 The results of pre-processing ((a) Original image; (b) Tree filter image; (c) Quick shift superpixel decomposition; (d) Affinity propagation clustering.)

## 2 先验背景和前景的估计

### 2.1 先验背景估计

图像背景比较分散和杂乱, 也正因为如此, 如果能预先排除部分背景, 有助于突出显著性前景目标. 文献 [16] 根据成像规律, 提出大部分图像边界属于背景, 但前景目标有时也会接触到图像边界, 如图 1 中第一幅图像. 虽然可采用经典的显著性算法<sup>[17]</sup>进行预判, 但总体显著性检测效果不甚理想, 如图 1 中文献 [16] 的检测效果. 因此, 本文充分考虑背景的空间分布, 只确定那些最有可能成为背景的区域.

假设经过仿射传播聚类后, 图像中第  $k$  个类为  $C_k$ , 其中包含  $n$  个超像素. 计算  $C_k$  的空间聚类中心: 即统计属于  $C_k$  的所有像素点的二维坐标, 取均值, 得到中心  $\mu_k$ .  $C_k$  中第  $i$  个超像素的空间中心为

$sp_i^s$ , 则  $C_k$  的空间离散程度为

$$V(C_k) = \frac{1}{n} \sum_{i=1}^n (sp_i^s - \mu_k)^2 \quad (2)$$

而由于各超像素中含有的像素数目不同, 因此参照文献 [10], 加入权重  $w_i$ , 为第  $i$  个超像素中所含有的像素点总数.  $|C_k|$  为类  $C_k$  中含有的像素点总数, 得到:

$$V(C_k) = \frac{1}{|C_k|} \sum_{i=1}^n w_i (sp_i^s - \mu_k)^2, \quad |C_k| = \sum_{i=1}^n w_i \quad (3)$$

计算时, 图像空间坐标标准化为  $[0, 1]$ .  $C_k$  的空间离散程度越大, 越有可能是背景.

前景类中超像素的空间分布较为集中, 其空间聚类中心位于非图像边界的区域, 倾向于靠近图像中央(见图 3(a)、(c)的中心); 而背景类中的超像素空间分布跨度范围大, 且有较高概率接触图像边界(见图 3(b)、(d)中的背景). 因此, 这些背景类中彼此分散的超像素与前景类的空间聚类中心的平均距离也相对较大. 而正由于背景类中的超像素围绕着前景目标, 分布广泛, 其空间聚类中心也一定程度上倾向于靠近前景类附近的区域, 甚至图像中心附近(图 3(b)、(d)的中间区域), 造成前景类中彼此靠近的超像素与背景类的空间聚类中心的平均距离相对较小. 进而可知, 相对前景类而言, 背景类的类间相对空间离散程度较大.

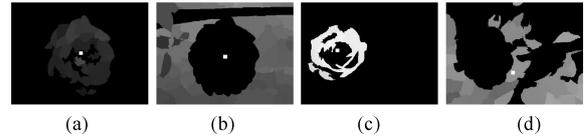


图 3 各类及其空间聚类中心 ((a)、(c) 分别为某前景类及其空间聚类中心, (b)、(d) 分别为某背景类及其空间聚类中心)

Fig. 3 Different clusters and the corresponding spatial clustering centers ((a), (c) are the foreground clusters and the corresponding spatial clustering centers respectively, (b), (d) are the background clusters and the corresponding spatial clustering centers, respectively.)

受文献 [25] 启发, 假设类  $C_l$  的空间中心为  $\mu_l$ , 则  $C_k$  相对于  $C_l$  的空间离散程度:

$$VR(C_k, C_l) = \frac{1}{|C_k|} \sum_{i=1}^n w_i (sp_i^s - \mu_l)^2 \quad (4)$$

设聚类数目为  $N$ ,  $C_k$  最终的相对离散程度为

$$VR(C_k) = \frac{1}{N-1} \sum_{k \neq l} VR(C_k, C_l) \quad (5)$$

图 3(a) 中前景类和图 3(b) 中背景类的相对离散度分别为 0.33 和 0.83, 而图 3(c) 中前景类和图 3(d) 中背景类的相对离散度分别为 0.98 和 2.16, 说明了其有效性.

最后再考虑类  $C_k$  中的超像素位于图像边界的数目  $Num_k$ , 与文献 [16] 一样,  $C_k$  中超像素包含的像素点与图像边界的最短距离在 10 个像素以内, 认为该超像素接触到图像边界 (见图 4(a) 中黑色的超像素).

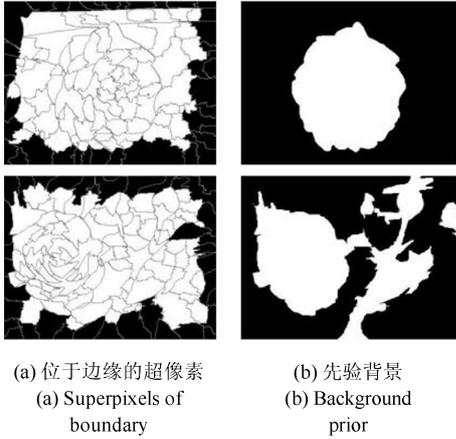


图 4 图 2 的先验背景估计

Fig. 4 The background prior estimation of Fig. 2

综合考虑以上三个要素, 如果  $C_k$  满足如下条件:

$$V(C_k) > V_\mu, VR(C_k) > VR_\mu, Num_k > Num_\mu \quad (6)$$

则初步认定  $C_k$  属于背景类.  $V_\mu$ 、 $VR_\mu$  和  $Num_\mu$  分别为所有类的类内空间离散度均值、类间空间离散度均值和接触到图像边界的超像素数目的均值. 这些背景类组成先验背景  $B$ .

除此以外, 自然图像也存在这样一些背景: 水平或者垂直贯穿于整体图像. 如图 2 中第一幅图像所示的道路, 以及海平线、绿化带等. 如果仅考虑颜色对比度或者空间离散程度, 容易混淆于前景目标. 因此从形状角度分析, 分别统计  $C_k$  中优化的横向和纵向的投影, 计算其比例. 设  $M$ 、 $N$  分别为图像的长宽,  $C_k$  中横向第  $i$  个位置的投影数目为

$$H_i^k = \sum_{j=1}^N (p(i,j) \in C_k) \quad (7)$$

即固定横坐标  $i$ , 统计属于  $C_k$  的像素点  $p(i,j)$  的个数, 其中  $1 \leq i \leq M$ . 同样的, 纵向第  $j$  个位置的投

影数目为

$$V_j^k = \sum_{i=1}^M (p(i,j) \in C_k) \quad (8)$$

其中,  $1 \leq j \leq N$ . 如果是计算  $C_k$  的整体横向投影  $H^k$  和纵向投影  $V^k$  的比例, 则有:

$$H^k = \frac{1}{N_H} \sum_{i=1}^M (H_i^k > 0), V^k = \frac{1}{N_V} \sum_{j=1}^N (V_j^k > \frac{1}{2}M) \quad (9)$$

式中,  $N_H$  表示横向投影中  $H_i^k$  大于 0 的个数,  $N_V$  表示纵向投影中  $V_j^k$  大于  $M/2$  的个数.  $V^k$  的设置是由于类  $C_k$  的形状并不规则, 并非严格的长方形, 如图 2 第一幅图像所示的道路, 应排除那些空间上明显位于类内边缘部分的点. 同样的,  $C_k$  的整体纵向投影  $\bar{V}^k$  和横向投影  $\bar{H}^k$  的比例也如此计算:

$$\bar{V}^k = \frac{1}{N_{\bar{V}}} \sum_{j=1}^N (V_j^k > 0), \bar{H}^k = \frac{1}{N_{\bar{H}}} \sum_{i=1}^M (H_i^k > \frac{1}{2}N) \quad (10)$$

当满足

$$H^k < \alpha V^k \quad (11)$$

$$\bar{V}^k < \alpha \bar{H}^k \quad (12)$$

这两者中的任一个, 则认为类  $C_k$  也属于背景类,  $\alpha$  为比例系数, 本实验中设置为 1/6.

由图 4(b) 知 (黑色表示背景), 本文方法仅将类内和类间空间离散度高、接触到边界多的超像素所属的类以及明显的条状区域确定为先验背景, 如果背景简单 (如图 2 中的花朵), 可明显突出前景目标 (图 4(b) 第一幅图); 而如果背景较为复杂 (如图 2 中第二幅图), 可经由先验前景估计, 进一步确定显著性目标区域.

## 2.2 先验前景估计

自然图像中往往存在着中央偏置, 一些显著性方法将此作为先验知识, 采用简单的高斯平滑, 认为像素点越接近图像中心, 显著性越大<sup>[18]</sup>. 但前景目标并非总是严格地位于图像中心, 于是文献 [19] 先提取图中显著点, 去除接近边缘的部分点后, 采用凸包的形式将筛选后的点连接成闭合的区间, 作为粗略的先验前景. 但此方法得到的前景有时并不准确, 部分背景在显著图中也较为突出 (如图 1(c)). 而目标性度量<sup>[20]</sup> 综合考虑了多尺度的显著性、颜色对比度、边缘密度等因素, 依靠训练的方法用贝叶斯框架将这些要素融合, 计算一系列的矩形窗口中含有完

整的前景目标的概率. 但由于无法将不规则的目标从背景中完全分离, 本文并不采用能量函数的形式将目标性度量与基于上下文的显著性算法融合<sup>[21]</sup>, 而仅通过目标性度量粗略估计前景范围.

假设图像  $I$  通过目标性度量产生  $W$  个矩形窗口, 每个矩形窗口  $w_k$  含有完整前景目标的概率为  $p(w_k)$ . 如果像素点  $p_i$  位于  $w_k$  内, 则其属于前景的概率为  $p(w_k)$ ; 反之其概率为 0. 计算所有  $W$  个矩形窗口的概率, 得到像素点  $p_i$  的最终目标性度量<sup>[20]</sup>:

$$O(p_i) = \frac{1}{W} \sum_{k=1}^W P(w_k) \quad (13)$$

如果  $p_i \in w_k$ , 有  $P(w_k) = p(w_k)$ ; 否则  $P(w_k) = 0$ . 由文献 [20] 知,  $W$  设为 100 就已足够. 图 5(a) 为目标性度量图, 矩形框越亮表明其含有前景的概率越大, 图 5(b) 为目标性度量的前景概率图, 标准化为  $[0, 1]$ , 像素点越亮表明其属于前景的概率越大, 取阈值  $\beta$  分割前景概率图, 得到粗略的前景范围, 如图 5(c) 所示. 实验中  $\beta$  取值为 0.5.

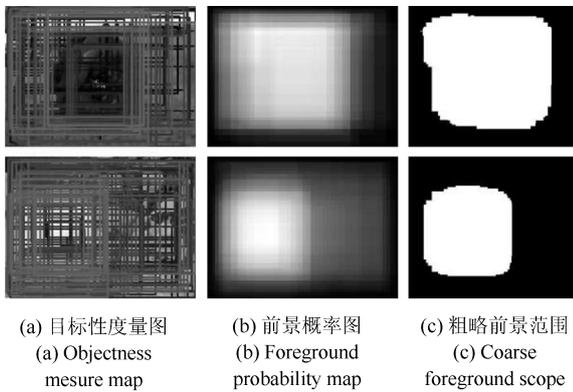


图 5 图 2 的先验前景估计

Fig. 5 The foreground prior estimation of Fig. 2

接着结合各类的空间分布信息, 得到较为精确的先验前景. 设类  $C_k$  的面积为  $A_k$ , 如果  $C_k$  不属于先验背景  $B$ , 且其与粗略前景范围的面积  $AF$  相交的区域满足:

$$\frac{A_k \cap AF}{A_k} > \gamma \quad (14)$$

则认为该类属于先验前景  $F$ .  $\gamma$  为面积比例阈值, 本文中设置为 0.65. 某些图像的  $AF$  可能过小, 无法满足此阈值, 则取与  $AF$  相交区域最大的类为先验前景, 即:

$$F = \max_k \frac{A_k \cap AF}{A_k} \quad (15)$$

对比图 4(b) 和图 5(c) 第二幅图的白色区域, 将其按照式 (14) 计算可知, 对于背景较为复杂的图像, 本文方法能够滤除干扰, 准确地估计出先验前景.

### 2.3 基于连通区域的先验前景和背景的优化

某些类虽然被判断属于先验前景  $F$ , 但该类中部分超像素并不位于目标性度量得到的粗略前景范围内, 如图 6 所示, 图像中央的枫叶所属的类 (见图 6(b)) 的大部分与粗略前景范围 (图 6(c)) 相交, 属于先验前景  $F$ , 但该类中部分超像素位于图像边缘 (图 6(d)). 如果不加区分地将该类整体当做先验前景, 显然会在图像边缘产生不必要的高显著值, 必须调整. 如果类  $C_k$  属于先验前景  $F$ ,  $C_k$  中共含有  $R$  个连通区域, 第  $i$  个连通区域  $r_i$  的面积为  $Ar_i$ , 如果满足:

$$Ar_i \cap AF = 0 \quad (16)$$

则认为该连通区域中所有的超像素不属于真正的先验前景. 如图 6(d), 该类中共含有 4 个连通区域, 有 3 个不在前景范围内, 优化后的先验前景见图 6(e).

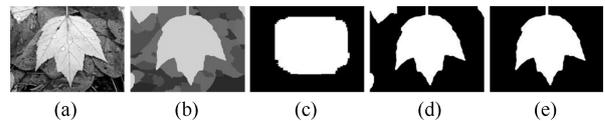


图 6 先验前景优化示意图 ((a) 原图; (b) 仿射传播聚类结果; (c) 粗略前景范围; (d) 属于先验前景的类; (e) 优化后的先验前景)

Fig. 6 The schematic diagram of foreground prior optimization ((a) Original image; (b) Affinity propagation clustering result; (c) Coarse foreground scope; (d) The cluster included in foreground prior; (e) The optimized foreground prior.)

某些属于先验背景的类中, 也有可能有一部分超像素位于粗略前景范围内, 如图 7 所示. 图 7(b) 中的天空和衣服聚为一类, 该类整体被认为属于先验背景 (图 7(c) 黑色部分); 但粗略的前景范围 (图 7(d)) 又与白色的衣服大面积重叠, 如果不加以调整, 得到的先验前景 (图 7(e)) 会丢失大部分重要信息. 因此, 如果类  $C_k$  属于先验背景  $B$ ,  $C_k$  中第  $i$  个连通区域  $r_i$  的面积为  $Ar_i$ , 如果满足:

$$Ar_i \cap AF \geq \lambda \cdot AF \quad (17)$$

则认为该连通区域中所有的超像素属于真正的先验前景,  $\lambda$  一般取 0.5. 优化后的效果见图 7(f).

## 3 生成显著图

本文将图像划分为 5 个不同层次的组成结构. 第 1 层包含所有的像素点, 第 2 层为超像素级的图像分解, 第 3 层为连通区域级的图像表示, 第 4 层将图像划分为不同的类, 第 5 层为先验背景、前景以及未确定的图像区域. 估计先验背景 (第 5 层) 时,

先以类(第4层)为基础,计算类内和类间的超像素(第2层)空间离散程度以及位于图像边缘的数目;估计先验前景(第5层)时,以类(第4层)中含有的所有像素点(第1层)的面积参与计算;而先验背景和前景又经过连通区域(第3层)优化。



图7 先验背景和前景优化示意图((a)原图;(b)仿射传播聚类结果;(c)先验背景(黑色);(d)粗略前景范围;(e)未优化的先验前景;(f)优化后的先验前景)

Fig. 7 The schematic diagram of background and foreground prior optimization ((a) Original image; (b) Affinity propagation clustering result; (c) Background prior (black); (d) Coarse foreground scope; (e) The foreground prior before optimization; (f) The optimized foreground prior.)

人类视觉倾向于将大尺度感知相似的部分当作整体处理<sup>[11]</sup>,因此本文在预估先验背景与前景时,以高层的类为处理单元.但考虑到目标内部各超像素在吸引视觉注意时也略有不同,则以超像素为单位计算显著值,并将其分配至各像素,得到全分辨率的显著图。

图像的第5层,分为先验背景 $B$ 、先验前景 $F$ 以及未确定的图像区域 $U$ .设 $B$ 中共含有 $K$ 个类,第 $k$ 个类 $C_k$ 中共含有 $L$ 个超像素,第 $l$ 个超像素为 $sp_l$ .由于在先验背景和前景的估计过程中,已经隐含着空间分布信息,因此仅考虑颜色对比度.计算图像中第 $i$ 个超像素 $sp_i$ 与属于背景的超像素在CIELab颜色空间的距离.取其中的最小值,如果该值较大,表明相对于先验背景 $B$ , $sp_i$ 越显著,即有:

$$S_i^{Bsp} = \min_l \{d(sp_i, sp_l)\}, sp_l \in C_k, C_k \in B \quad (18)$$

式中 $1 \leq l \leq L, 1 \leq k \leq K$ .考虑到某些超像素的颜色可能与属于先验背景的类中的个别超像素较为接近,而与属于先验背景的类的整体颜色相差较大,因此引入 $S_i^{BC}$ :

$$S_i^{BC} = \min_k \{d(sp_i, C_k)\} \quad (19)$$

即为 $sp_i$ 与属于 $B$ 的类 $C_k$ 在CIELab颜色空间的最小距离.则 $sp_i$ 最终相对于先验背景的显著性 $S_i^B$ 为

$$S_i^B = \frac{1}{2}(S_i^{Bsp} + S_i^{BC}) \quad (20)$$

相对的,设 $F$ 中共含有 $K$ 个类,第 $k$ 个类 $C_k$ 中共含有 $L$ 个超像素,第 $l$ 个超像素为 $sp_l$ .计算 $sp_i$ 与

属于先验前景的超像素在CIELab颜色空间的距离.其中的最大值越小,表明 $sp_i$ 与先验前景 $F$ 越接近,也越显著。

$$S_i^{Fsp} = \max_{l, l \neq i} \{d(sp_i, sp_l)\}, sp_l \in C_k, C_k \in F \quad (21)$$

同样的,引入 $sp_i$ 与属于先验前景的类的颜色对比度,得到:

$$S_i^{FC} = \max_k \{d(sp_i, C_k)\} \quad (22)$$

$sp_i$ 最终相对于先验前景的显著性 $S_i^F$ 为

$$S_i^F = \frac{1}{2}(S_i^{Fsp} + S_i^{FC}) \quad (23)$$

需要注意的是, $S_i^F$ 越小,表明与前景越接近,该超像素越显著。

计算所有超像素的显著值,得到先验背景显著图 $S^B$ 和先验前景显著图 $S^F$ ,将两者都标准化为 $[0, 1]$ .综合考虑这两者的影响,超像素 $sp_i$ 的显著性度量为

$$S_i = \frac{1}{2}(S_i^B + (1 - S_i^F)) \quad (24)$$

一般而言,目标性度量得到的粗略前景范围能达到95%以上的正确率<sup>[20]</sup>,而图像中的未确定区域 $U$ 按上式计算得到的显著值有时也略大(见图7第一排的显著图).因此,如果在粗略的前景范围可靠的情况下, $sp_i$ 的显著性度量可如下计算:

$$S_i = \begin{cases} \max\{S_i^B, (1 - S_i^F)\}, & sp_i \in F \\ \min\{S_i^B, (1 - S_i^F)\}, & sp_i \notin F \end{cases} \quad (25)$$

即可更为突出先验前景中超像素的显著性.将各个超像素中的显著值分配到具体的像素点,得到全分辨率的显著图.与传统的中央偏置<sup>[18]</sup>不同,本文通过显著性中心加权,进一步凸显目标区域。

设图像长宽为 $M$ 、 $N$ ,像素点 $p(x, y)$ 的显著值为 $S(x, y)$ ,则显著性中心的坐标 $(cx, cy)$ 为

$$cx = \frac{\sum_{x=1}^M \sum_{y=1}^N S(x, y) \cdot x}{\sum_{x=1}^M \sum_{y=1}^N S(x, y)} \quad (26)$$

$$cy = \frac{\sum_{x=1}^M \sum_{y=1}^N S(x, y) \cdot y}{\sum_{x=1}^M \sum_{y=1}^N S(x, y)} \quad (27)$$

点 $p(x, y)$ 经过显著性中心加权后,最终显著值为

$$S(x, y)' = S(x, y) \cdot \exp\left(-\frac{\sqrt{(x - cx)^2 + (y - cy)^2}}{\max(M, N)}\right) \quad (28)$$

将所有像素点的显著值标准化为  $[0, 1]$ , 得到显著图. 如图 8 所示, 第一排为式 (24) 结合显著性中心加权产生的显著图, 第二排为式 (25) 结合显著性中心加权得到的显著图. 两者都能完整地突出显著性目标整体, 抑制杂乱无章的背景.

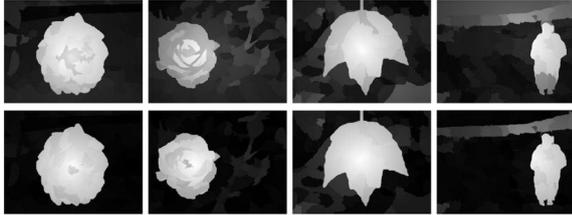


图 8 图 2、6 和 7 中的图像对应的显著图  
Fig. 8 Saliency maps corresponding to the images of Fig. 2, 6 and 7

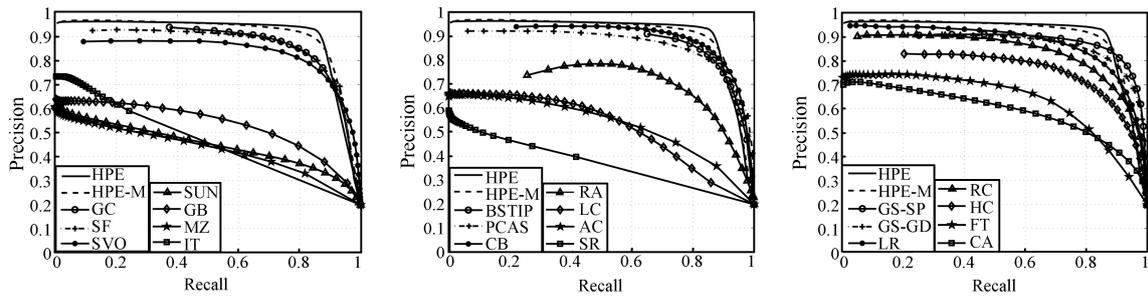
### 4 实验结果与分析

本文在两个标准数据库上进行对比实验, 分别为 MSRA-1000 和 SOD. MSRA-1000 是评价显著性目标检测效果优劣最常用的数据库<sup>[9]</sup>, 大部分图像仅含单个目标. SOD 由 300 幅图像组成, 来自于 Berkeley 图像分割数据库, 每幅图像含有一个或者多个显著性目标, 且目标尺寸差别极大, 被大多数文章认为是最具挑战性的数据库<sup>[18, 26]</sup>. 这两个数据库中所有的图像都由人工标注出真实值 (Ground truth, GT), 以便客观地评估性能. 本文将所提方法与其余 21 种 Stateofheart 显著性方法比较, 包括:

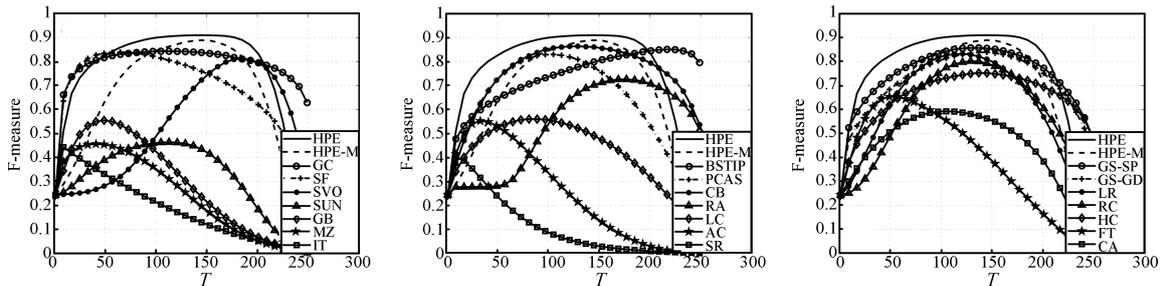
IT<sup>[4]</sup>、MZ<sup>[5]</sup>、AC<sup>[6]</sup>、RA<sup>[7]</sup>、SR<sup>[8]</sup>、FT<sup>[9]</sup>、HC<sup>[10]</sup>、RC<sup>[10]</sup>、GC<sup>[11]</sup>、LR<sup>[12]</sup>、SF<sup>[13]</sup>、PCAS<sup>[15]</sup>、GS-GD<sup>[16]</sup>、GS-SP<sup>[16]</sup>、CA<sup>[17]</sup>、BSTIP<sup>[19]</sup>、SVO<sup>[21]</sup>、CB<sup>[27]</sup>、SUN<sup>[28]</sup>、GB<sup>[29]</sup> 和 LC<sup>[30]</sup>. 有些方法的显著性结果由原作者给出, 有些则是通过运行原作者公开的源代码得到, 具体可参考综述性文献 [1]. 本文将按式 (24) 计算的方法定义为 HPE-M (Hierarchical prior estimation-median), 按式 (25) 计算的方法定义为 HPE (Hierarchical prior estimation). 具体参数设置见第 4.3 节.

#### 4.1 MSRA-1000 数据库实验比较

与文献 [9–11] 一样, 本文首先采用 P-R 曲线 (Precision-recall curve) 比较算法性能. 准确率 (Precision) 衡量算法检测到的且属于 GT 的部分占检测到的所有部分的比例, 查全率 (Recall) 衡量算法检测到的且属于 GT 的部分占 GT 的比例. 将各算法得到显著图标准化为  $[0, 255]$  的范围, 然后阈值从 0 至 255 依次变化分割显著图, 对每幅阈值分割后的二值图像参照 GT 分别计算准确率和查全率, 根据各算法对于 MSRA-1000 中所有图像在各个阈值处的平均准确率和查全率, 绘制 P-R 曲线图, 见图 9 (a). 观察可知, 本文算法在 P-R 曲线上明显优于其余算法, 尤其是与具有部分相似原理的方法: BSTIP、GS-SP、GS-GD 和 SVO 相比, 准确率和查全率有着显著提高. 当查全率低于 0.9 时, 在同样的查全率处, HPE 和 HPE-M 比其余所有算法的准



(a) P-R 曲线图  
(a) P-R curves



(b) F-measure vs 阈值变化图  
(b) The diagram of F-measure vs threshold

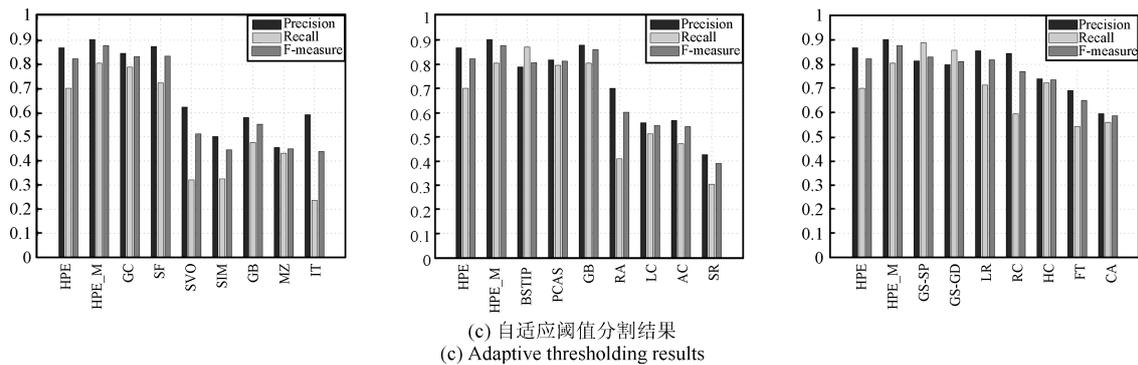


图 9 MSRA-1000 数据库上的准确率、查全率和 F-measure 量化评价

Fig. 9 The quantitative performance of precison, recall and F-measure on MSRA-1000 database

准确率都高. HPE 在查全率约为 0.86 时, 准确率达到 0.92, HPE-M 在查全率约为 0.84 时, 准确率达到 0.9, 即查全率和准确率之和在某一阈值处大于其余所有算法, 表明本文算法结果与 GT 更为接近. 而由式 (25) 知, HPE 抑制了非先验前景区域的显著值, 减少了非显著性目标取得较大显著值的可能. 因此比 HPE-M 有更好的表现.

值得注意的是, 当查全率在 1~0.9 之间时, 本文算法的准确率比近年来几种具有最优表现的算法 GC、PCAS、BSTIP、LR、SF、GS-SP、GS-GD、CB 和 SVO 略低; 当查全率从 1 变化为 0.9 时, 准确率一直以很陡的斜率攀升, 而不像其余算法, 准确率变化逐渐平缓. 这主要是由于本文算法将图像划分为先验背景、先验目标以及未确定三部分, 而未确定部分也拥有一定的显著值, 但小于真正目标; 当查全率在 [0.9, 1] 区间时, 阈值较小, 此时容易将未确定的部分划分为目标, 造成准确率不高; 而随着阈值变大, 即查全率变低时, 本文方法的准确率急剧变高, 在查全率位于 [0.8, 0.9] 时, 此处的阈值能正确划分出大部分真正的目标, 从而达到稳定状态, 准确率缓慢变化, 维持最高水平. 这也从侧面说明本文算法采用简单的阈值分割策略后, 可得到非常优秀的分割结果.

本文也比较了各算法的 F-measure:

$$F_{\beta} = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (29)$$

同文献 [9–11, 31] 一样, 设置  $\beta^2 = 0.3$ , 使得准确率权重高于查全率. F-measure 随着绘制 P-R 曲线时的阈值 [0, 255] 变化的曲线图如图 9(b) 所示, 与其余 21 种方法相比, 本文两种方法在不同的阈值  $T$  处所能达到的最大 F-measure 值是所有算法中最高的, 并且阈值在相当大的范围内变化时, 如 HPE 算法中阈值位于区间 [100, 175], HPE-M 算法中阈值位于区间 [125, 175], 对应的 F-measure 值都位于最高值附近, 证明了本文算法的稳定性.

各算法在何阈值  $T$  处取得最高 F-measure 的数值对比参见表 1 (从高到低排列), HPE 和 HPE-M 分别达到了 0.9111 和 0.8902, 且对应的阈值处于区间 [0, 255] 的中段, 而不是较为极端的边界附近, 较为合理.

为了使各算法在同等条件下进行性能评价, 采用自适应阈值分割策略<sup>[9–11]</sup>, 即将阈值  $T$  自动设置为图像平均显著值的两倍, 然后比较分割后所有图像的平均准确率、查全率和 F-measure 值, 见图 9(c).

表 1 各算法最高的 F-measure 值及对应的阈值  $T$ Table 1 The highest F-measure and the corresponding threshold  $T$  of all methods

方法名称	HPE	HPE-M	CB	GS-SP	BSTIP	GC	GS-GD	SF	LR	PCAS	SVO	RC
F-measure	0.9111	0.8902	0.8655	0.8581	0.8515	0.8443	0.8398	0.8395	0.8355	0.8309	0.8138	0.8031
$T$	143	146	125	131	222	106	140	59	120	102	185	130

方法名称	HC	RA	FT	CA	LC	GB	AC	SUN	IT	MZ	SR
F-measure	0.7535	0.7272	0.6550	0.5930	0.5605	0.5550	0.5528	0.4639	0.4633	0.4563	0.4016
$T$	141	173	53	105	86	49	34	113	3	42	10

HPE 和 HPE-M 算法的 F-measure 值最高, 分别为 0.8809 和 0.8781, 均优于近年来表现最优的算法: GC(0.8319)、PCAS(0.8132)、BSTIP(0.8072)、LR(0.8187)、SF(0.8341)、GS-SP(0.8305)、GS-GD(0.8117)、CB(0.8612)、RC(0.7705) (按发表时间顺序排列).

正如文献 [13] 所说, 准确率和查全率没有考虑被正确标记为非显著性目标的点, 这两种指标使那些对 GT 中的目标 (白色部分) 能够正确分配更高显著值的方法取得更好的结果, 并不全面. 因此, 对整幅显著图计算平均绝对误差 (Mean absolute error, MAE):

$$\text{MAE} = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N |S(x, y) - \text{GT}(x, y)| \quad (30)$$

式中,  $M$  和  $N$  分别为图像长和宽,  $S(x, y)$  和  $\text{GT}(x, y)$  分别为点  $(x, y)$  在显著图和 Ground truth 中对应的值. 所有算法对于 MSRA-1000 中图像的平均 MAE 结果见图 10 (按 MAE 值大小从低到高排列).

本文 HPE 算法的 MAE 值最小, 仅为 0.0976, 说明 HPE 产生的显著图与 GT 整体 (包括背景部分) 的相似性最强. HPE-M 算法的

MAE 值为 0.1772, 略大于 GC、GS-SP、GS-GD、SF、CB、PCAS 和 HC. 这是由于 HPE-M 采用平均值的策略 (见式 (24)), 使得未确定部分  $U$  的显著值略高, 而本文算法中大部分的  $U$  属于非显著性目标区域, 因此造成了 MAE 计算时, 取值较大, 但依然比其余 State-of-the-art 方法的 MAE 值小.

最后, 本文采用重叠评分 (Overlap score) 衡量算法的优劣程度<sup>[32]</sup>:

$$O = \frac{|GT \cap S_c|}{|GT \cup S_c|} \quad (31)$$

其中,  $S_c$  为经过自适应阈值分割<sup>[9-11]</sup> 后的二值图像. 该指标越高, 表明算法越能容易地提取出正确的目标区域. 各算法对 MSRA-1000 中所有图像的平均重叠评分值见表 2. 本文 HPE 算法取得最高的重叠评分值, 为 0.786, HPE-M 比 GS-SP 的重叠评分值略低, 为 0.737, 位列第三, 优于其余算法.

各算法产生的显著图的视觉效果对比见图 11 (按发表时间从左至右排序, Org. 表示原图像). 本文算法的显著图能够较为均匀地突出目标整体, 非目标区域的显著值较低, 与 Ground truth 最为一致. 对于近年来最好的几种方法中效果不佳的图像 (图 11 中第一行和第二行), 也能整体明显地突出显著性目标. 对于背景纹理性较强的图像 (图 11 中第 4 行、第 7 行和第 9 行), 也能完全排除干扰.

表 2 各算法的重叠评分

Table 2 The Overlap Score of all methods

方法名称	HPE	GS-SP	HPE-M	CB	BSTIP	GC	GS-GD	PCAS	SF	LR	HC	RC
重叠评分	0.7863	0.7393	0.7367	0.7216	0.7020	0.7012	0.7002	0.6676	0.6673	0.6312	0.5878	0.5201

方法名称	FT	CA	LC	RA	GB	AC	SVO	MZ	SUN	SR	IT
重叠评分	0.4476	0.3950	0.3890	0.3762	0.3550	0.3503	0.3024	0.2800	0.2313	0.2088	0.1966

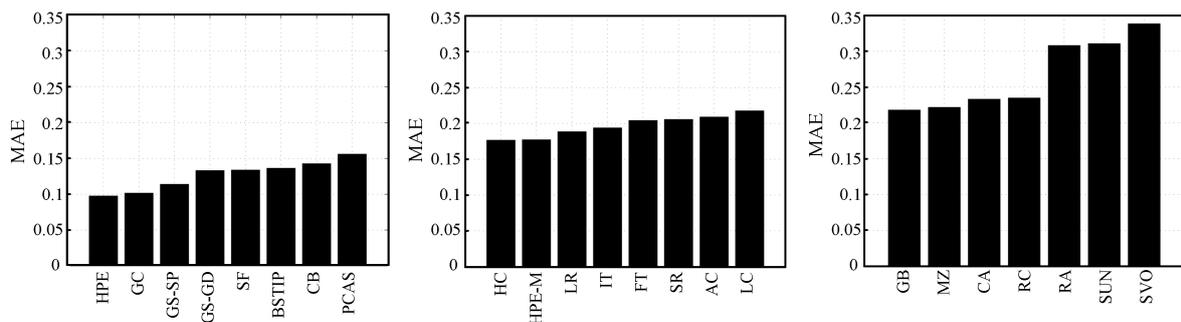


图 10 MAE 柱状图

Fig. 10 MAE histograms

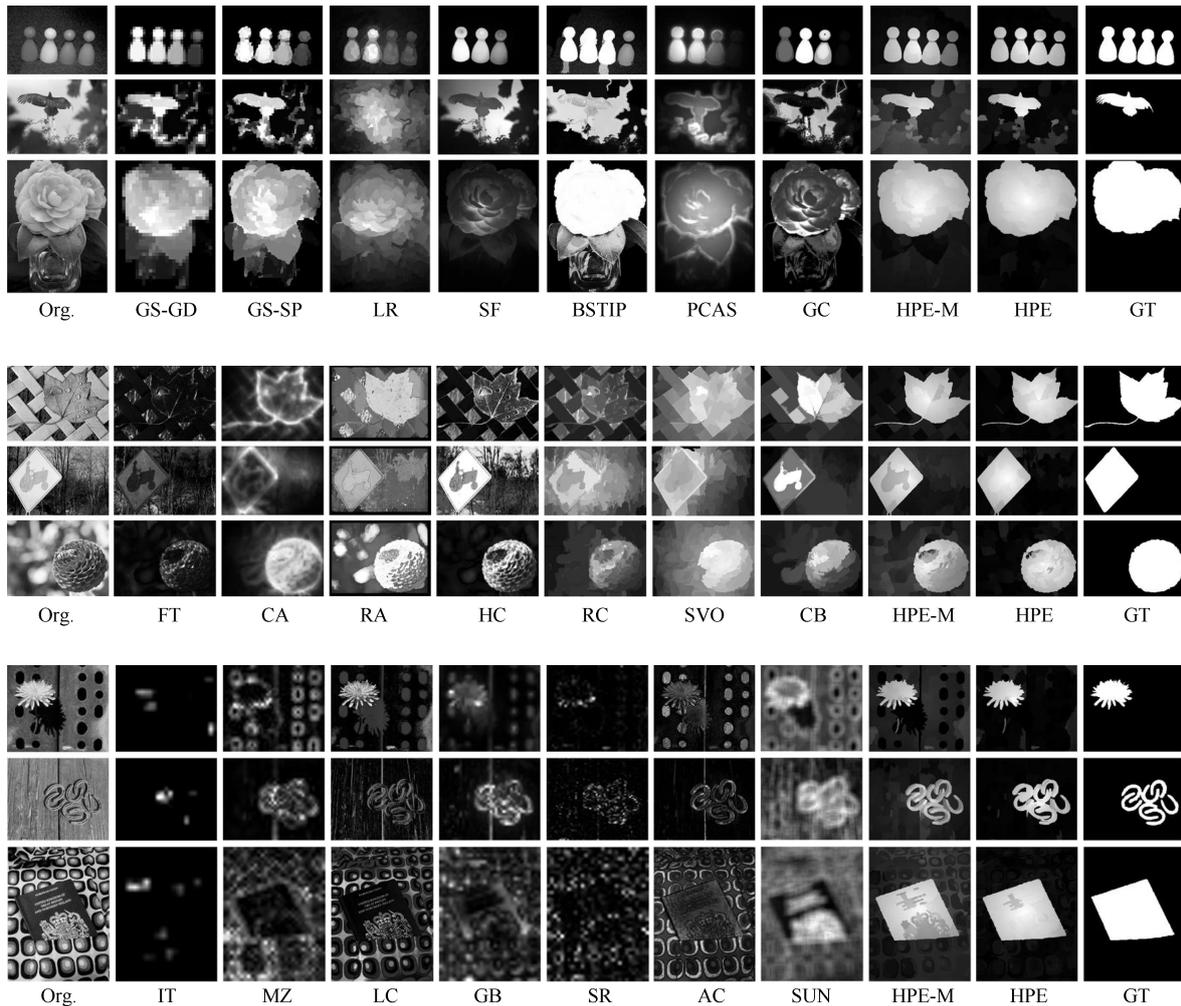


图 11 各算法在 MSRA-1000 数据库上的部分显著图

Fig. 11 Some saliency maps of all methods on MSRA-1000 database

#### 4.2 SOD 数据库实验比较

由于该数据库复杂性高,所有算法产生的显著图都不太理想,准确率较低.实验选择了表现较好的 9 种 State-of-the-art 算法与本文算法 HPE、HPE-M 进行比较.分别为 RA<sup>[7]</sup>、FT<sup>[9]</sup>、HC<sup>[10]</sup>、RC<sup>[10]</sup>、GC<sup>[11]</sup>、LR<sup>[12]</sup>、PCAS<sup>[15]</sup>、GS-SP<sup>[16]</sup> 和 CB<sup>[27]</sup>.其 P-R 曲线图见图 12(a),自适应阈值分割后的平均准确率、查全率和 F-measure 见图 12(b).

本文算法的 P-R 曲线与表现最好的 GS-SP 方法较为接近,当查全率小于 0.55 时,HPE 和 HPE-M 的准确率是所有曲线中最高的;在查全率位于区间 [0.4, 0.55] 时,HPE 的准确率明显高于近年来最好的 GS-SP;而与 MSRA-1000 中的原因一样,当查全率位于区间 [0.55, 1] 时,HPE 和 HPE-M 的准确率略低于近年来表现最好的几种算法.

自适应阈值分割后,HPE-M 的 F-measure 值为 0.5482,处于中等水平,但其准确率为 0.7087,是所

有算法中最高的.HPE 的 F-measure 值为 0.6286,略低于 GS-SP (0.6324),但明显优于其余几种近年来表现最好的算法.

HPE-M 和 HPE 与其余算法在 SOD 上的视觉效果对比见图 13.无论目标尺寸较小(图 13 中第 2 行的图像)还是很大(图 13 中第 4 行的图像),HPE 和 HPE-M 的显著图虽然有少量目标部分缺失,但目标整体较为突出.这表明本文算法对于多目标、场景复杂的图像,也有着较好的显著性检测效果.

#### 4.3 算法运行时间和参数选择

本文实验环境为 Intel Core i5-2410M CPU、4G 内存的计算机.所提算法和 21 种 State-of-the-art 方法的运行时间见表 3(按时间长短排列).在 MSRA-1000 和 SOD 数据库上,本文算法的预处理过程(树滤波、超像素分解和仿射传播聚类)平均耗时 4.45 秒;先验背景和前景的估计以及优化平均花费约 5.25 秒,其中由于先验前景的目标

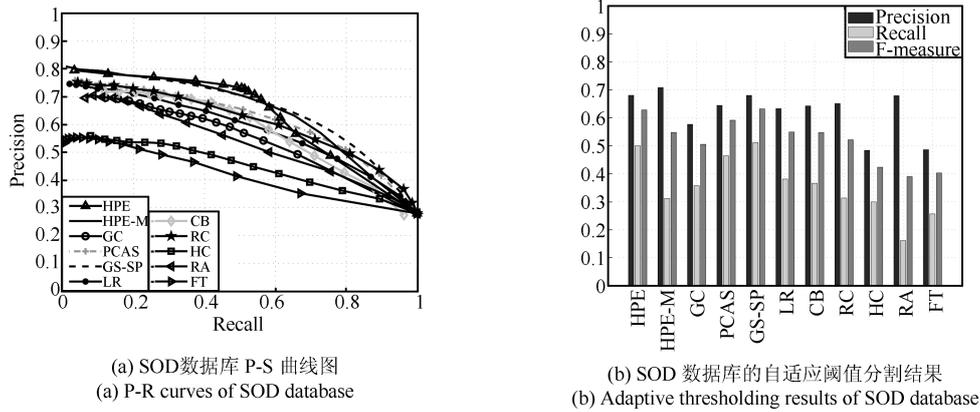


图 12 SOD 数据库的各算法量化性能评价

Fig. 12 The quantitative performance of SOD database

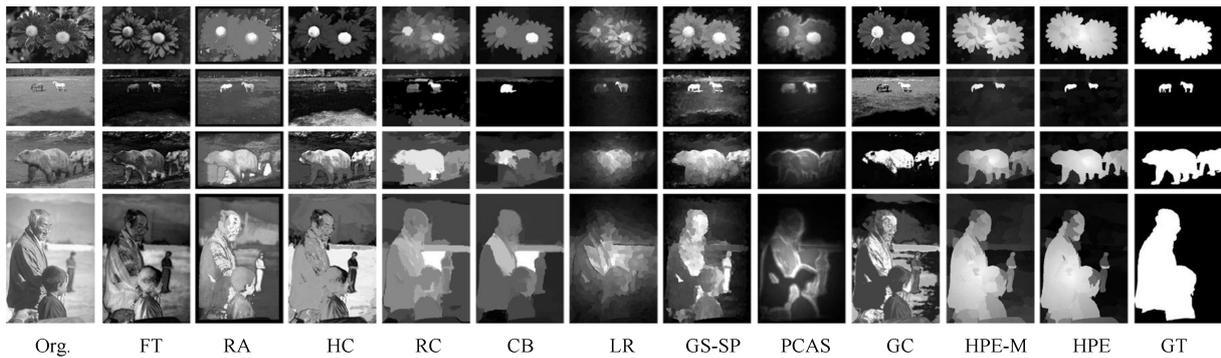


图 13 各算法在 SOD 数据库上的部分显著图

Fig. 13 Some saliency maps of all methods on SOD database

表 3 各算法的平均运行时间

Table 3 The average running time of all methods

方法名称	GS-GD	LC	HC	SR	MZ	AC	GC	RC	SF	FT	IT	GB	CB
时间 (秒)	0.002	0.01	0.01	0.02	0.04	0.06	0.09	0.13	0.16	0.18	0.34	1.34	2.85
代码类型	C++	C++	C++	Matlab	C++	C++	C++	C++	C++	Matlab	Matlab	Matlab	Matlab

方法名称	PCAS	GS-SP	RA	HPE-M	HPE	LR	SUN	CA	SVO	BSTIP
时间 (秒)	6.27	7.42	9.51	10.10	10.11	22.55	24.56	54.57	130.90	158.29
代码类型	Matlab	C++	Matlab							

性度量较为复杂, 平均需要 4.54 秒; 虽然本文算法生成显著图时的算法复杂度为  $O(n^2)$ , 但其以超像素为单位, 总数为 200 个左右, 计算速度较快, 平均为 0.41 秒. 总之, 本文算法 HPE 和 HPE-M 的总体平均运行时间约为 10.11 秒和 10.10 秒. 早期的显著性方法, 如 LC、SR 等, 虽然运行时间极快, 但各种性能评价指标表现较差; 本文方法虽然慢于一些近年来以运算速度见长的算法, 如 RC、GC, 但也快于近年来表现较好的部分算法, 如 LR、BSTIP. 再

加上本文算法的量化评价指标占优, 因此综合而言, 仍然具有一定优势. 且本文算法由 Matlab 编写, 如果转化为 C++, 运行速度仍具有提升空间.

本文算法 HPE 和 HPE-M 共有 4 个需要设置的参数, 分别为式 (11) 和 (12) 中的比例系数  $\alpha$ 、对按照式 (13) 计算得到的前景概率图进行分割的阈值  $\beta$ 、式 (14) 中的面积比例阈值  $\gamma$  以及式 (17) 中的阈值  $\lambda$ ; 对应的最佳值分别为 1/6、0.5、0.65 以及 0.5, 其中最重要的参数为  $\beta$  和  $\gamma$ . 将其余参数固定为最

佳值, 这两个参数在 MSRA-1000 和 SOD 数据库上不同取值对应的自适应阈值分割后的准确率、查全率和 F-measure 变化曲线图分别见图 14 和 15. 当  $\beta$  取值过小时, 本文算法会误将部分图像区域划分为先验前景, 导致准确率较低; 而当  $\beta$  取值过大时, 粗略前景范围过小, 本文算法仅将最有可能为前景的类作为先验前景, 从而导致查全率有较大下降, 而此时的准确率却维持在较高水平 (有些图像中最有可能为前景的类并非真正的前景, 准确率略有下降), 见图 14. 参数  $\gamma$  因为同样的原因也有类似的表现, 见图 15, 当  $\gamma$  取值在  $[0.65, 0.7]$  区间时, F-measure 维持在最高值附近.

参数  $\alpha$  在合理的范围内才可有效地提取图像中的条状背景区域, 以 HPE 在 MSRA-1000 数据库的表现为例, 其不同的取值对应的自适应阈值分割后的准确率、查全率和 F-measure 见表 4.  $\alpha$  过小, 无法有效地筛选出条状背景, 造成该区域的显著值偏

大, 从而准确率略有下降;  $\alpha$  过大, 则条状约束失去了意义, 有较大概率将各种图像区域划分为背景, 显著性目标查全率降低. 但由于算法中先验前景以及其余先验背景条件的约束, 这些量化评价指标变化幅度并不大, 这也从侧面说明了本文算法的稳定性.

参数  $\lambda$  控制着先验前景的优化效果, HPE 在 MSRA-1000 上随不同  $\lambda$  值的量化性能评价的变化见表 5.  $\lambda$  过大, 条件过于严苛, 有较高概率无法将误分为背景的区域正确地纠正为目标前景, 导致查全率的降低; 而  $\lambda$  过小, 又会错误地将部分非目标区域划分为前景, 导致准确率下降.

#### 4.4 失败实例

由于本文所提方法采用了层次的背景和前景估计, 这是一种基于统计特征的结果, 难免存在不满足该假设原理的图像, 如图 16 所示. 当前景和背景中颜色极为接近时, 算法容易失效.

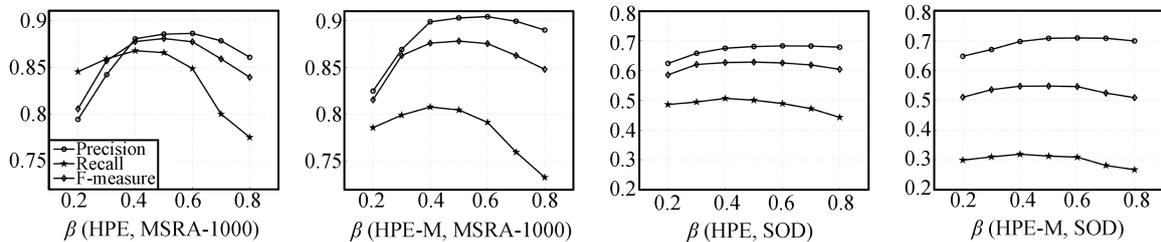


图 14  $\beta$  vs 自适应阈值分割后的准确率、查全率和 F-measure 曲线图

Fig. 14  $\beta$  vs precision, recall and F-measure after adaptive thresholding

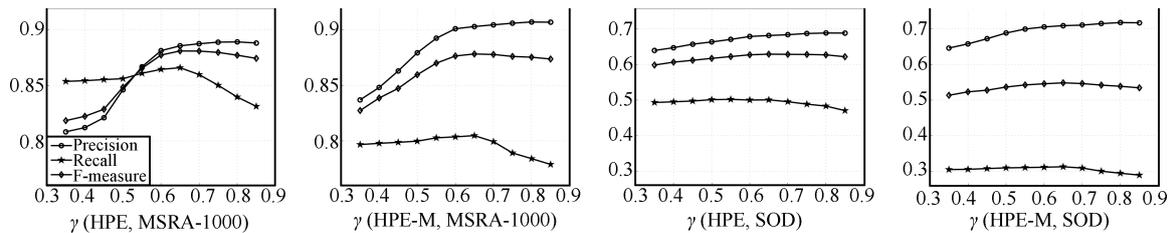


图 15  $\gamma$  vs 自适应阈值分割后的准确率、查全率和 F-measure 曲线图

Fig. 15  $\gamma$  vs precision, recall and F-measure after adaptive thresholding

表 4 MSRA-1000 上  $\alpha$  vs 自适应阈值分割后的准确率、查全率和 F-measure

Table 4  $\alpha$  vs precision, recall and F-measure after adaptive thresholding on MSRA-1000

$\alpha$ 取值	1/3	1/4	1/5	1/6	1/7	1/8
准确率	0.8848	0.8851	0.8857	0.8855	0.8831	0.8806
查全率	0.8591	0.8623	0.8641	0.8658	0.8662	0.8655
F-measure	0.8787	0.8797	0.8806	0.8809	0.8796	0.8771

表 5 MSRA-1000 上  $\lambda$  vs 自适应阈值分割后的准确率、查全率和 F-measure

Table 5  $\lambda$  vs precision, recall and F-measure after adaptive thresholding on MSRA-1000

$\lambda$ 取值	0.30	0.40	0.50	0.60	0.70
准确率	0.8387	0.8621	0.8855	0.8858	0.8855
查全率	0.8624	0.8667	0.8658	0.8607	0.8562
F-measure	0.8441	0.8477	0.8809	0.8799	0.8786

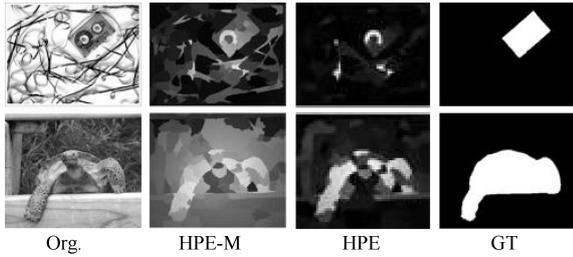


图 16 部分失败的显著图

Fig. 16 Some failure saliency maps

## 5 结论

本文将图像分为 5 个不同层次的组成成分, 各层之间协同合作, 得到较为精确的先验背景与前景, 以此为依据计算显著图。

树滤波使感知上相似的像素点的颜色更为一致, 同时保留边缘等图像结构信息; 而仿射传播聚类无需预设类的数目, 比固定类数的方法<sup>[19]</sup>更具有包容性, 也更为准确。

本文提出一种新的根据类内和类间超像素的空间离散程度以及其位于图像边界的数目, 自适应地确定最有可能属于真正背景的图像区域的方法, 并提取明显的条状背景区域。比简单假设位于图像边界的超像素大部分属于背景<sup>[16]</sup>的方法, 更为有效与稳定。本文并不将目标性度量融入到显著性计算中<sup>[21]</sup>, 而是将其作为粗略的前景估计, 通过与各类的空间交互信息, 确定先验前景。并且本文利用连通区域对先验背景与前景进行优化, 使得两者更加准确。

本文充分考虑了各超像素与先验前景和背景的颜色对比度, 并引入基于类的对比度调整, 使显著图更加稳定可靠。而显著性中心加权, 使得目标整体更为突出。在 MSRA-1000 上的实验结果表明, 本文方法的 5 种指标优于 21 种 State-of-the-art 方法, 而在复杂的 SOD 数据库上, 本文方法也明显优于近年来表现最好的几种算法, 部分指标是最好的。

下一步本文考虑将所提显著性方法运用于具体的应用任务, 如目标跟踪<sup>[33]</sup>, 并综合考虑更为丰富的显著性线索, 如多尺度区域显著性特征与显著性边缘<sup>[34]</sup>以及基于机器学习显著性框架<sup>[35]</sup>, 去改善那些颜色可区分性较差的图像的显著性结果。

## References

- 1 Borji A, Itti L. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(1): 185–207
- 2 Borji A, Sihite D, Itti L. Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Transactions on Image Processing*, 2013, **22**(1): 55–69
- 3 Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 1985, **4**(4): 219–227
- 4 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
- 5 Ma Y F, Zhang H J. Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of the 11th ACM International Conference on Multimedia. Berkeley, USA: ACM, 2003. 374–381
- 6 Achanta R, Estrada F, Wils P, Süsstrunk S. Salient region detection and segmentation. In: Proceedings of the 6th International Conference on Computer Vision Systems. Santorini, Greece: Springer, 2008. 66–75
- 7 Rahtu E, Kannala J, Salo M, Heikkilä J. Segmenting salient objects from images and videos. In: Proceedings of the 11th European Conference on Computer Vision. Heraklion, Greece: Springer, 2010. 366–379
- 8 Hou X D, Zhang L Q. Saliency detection: a spectral residual approach. In: Proceedings of the 2007 IEEE International Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007. 1–8
- 9 Achanta R, Hemami S, Estrada F, Süsstrunk S. Frequency-tuned salient region detection. In: Proceedings of the 2009 IEEE International Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 1597–1604
- 10 Cheng M M, Zhang G X, Mitra N J, Huang X, Hu S M. Global contrast based salient region detection. In: Proceedings of the 2011 IEEE International Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2011. 409–416
- 11 Cheng M M, Jonathan W, Lin W Y, Zheng S, Vineet V, Crook N. Efficient salient region detection with soft image abstraction. In: Proceedings of the 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013. 1529–1536
- 12 Shen X, Wu Y. A unified approach to salient object detection via low rank matrix recovery. In: Proceedings of the 2012 IEEE International Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 853–860
- 13 Perazzi F, Krähenbühl P, Pritch Y, Hornung A. Saliency filters: contrast based filtering for salient region detection. In: Proceedings of the 2012 IEEE International Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 733–740
- 14 Achanta R, Shaji A, Smith K. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(11): 2274–2282
- 15 Margolin R, Tal A, Zelnik-Manor L. What makes a patch distinct? In: Proceedings of the 2013 IEEE International Conference on Computer Vision and Pattern Recognition. Sydney, Australia: IEEE, 2013. 1139–1146
- 16 Wei Y C, Wen F, Zhu W J, Sun J. Geodesic saliency using background priors. In: Proceedings of the 2012 European Conference on Computer Vision. Florence, Italy: Springer, 2012. 29–42
- 17 Goferman S, Zelnik-Manor L, Tal A. Content-aware saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(10): 1915–1926

- 18 Borji A, Sihite D N, Itti L. Salient object detection: a benchmark. In: Proceedings of the 2012 European Conference on Computer Vision. Florence, Italy: Springer, 2012. 414–429
- 19 Xie Y L, Lu H C, Yang M H. Bayesian saliency via low and mid level cues. *IEEE Transactions on Image Processing*, 2013, **22**(5): 1689–1698
- 20 Alex B, Deselaers T, Ferrari V. Measuring the objectness of image windows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(11): 2189–2202
- 21 Chang K Y, Liu T L, Chen H T, Lai S H. Fusing generic objectness and visual saliency for salient object detection. In: Proceedings of the 2011 IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 914–921
- 22 Bao L C, Song Y B, Yang Q X, Yuan H, Wang G. Tree filtering: efficient structure-preserving smoothing with a minimum spanning tree. *IEEE Transactions on Image Processing*, 2014, **23**(2): 555–569
- 23 Vedaldi A, Soatto S. Quick shift and kernel methods for mode seeking. In: Proceedings of the 2008 European Conference on Computer Vision. Marseille, France: Springer, 2008. 705–718
- 24 Frey B J, Dueck D. Clustering by passing message between data points. *Science*, 2007, **315**(5814): 972–976
- 25 Xu D, Tang Z M, Xu W. Salient object detection based on regional contrast and relative spatial compactness. *KSI Transactions on Internet and Information Systems*, 2013, **7**(11): 2737–2753
- 26 Jiang B W, Zhang L H, Lu H C, Yang C, Yang M H. Saliency detection via absorbing Markov chain. In: Proceedings of the 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013. 1665–1672
- 27 Jiang H Z, Wang J D, Yuan Z J, Liu T, Zheng N N, Li S P. Automatic salient object segmentation based on context and shape prior. In: Proceedings of the 2011 British Machine Vision Conference. Dundee, UK: BMVA, 2011. 110.1–110.12
- 28 Zhang L, Tong M H, Marks T K, Shan H, Cottrell G W. Sun: a Bayesian framework for saliency using natural statistics. *Journal of Vision*, 2008, **8**(7): 32.1–32. 20
- 29 Harel J, Koch C, Perona P. Graph-based visual saliency. In: Proceedings of the 12th Conference on Neural Information Processing Systems. Vancouver, Canada: MIT, 2007. 545–552
- 30 Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th ACM International Conference on Multimedia. Santa Barbara, USA: ACM, 2006. 815–824
- 31 Ma Ru-Ning, Tu Xiao-Po, Ding Jun-Di, Yang Jing-Yu. To evaluate salience map towards popping out visual objects. *Acta Automatica Sinica*, 2012, **38**(5): 870–876  
(马儒宁, 涂小坡, 丁军娣, 杨静宇. 视觉显著性凸显目标的评价. *自动化学报*, 2012, **38**(5): 870–876)
- 32 Li X, Li Y, Shen C H, Dick A, Van Den Hengel A. Contextual hypergraph modelling for salient object detection. In: Proceedings of the 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013. 3328–3335
- 33 Guo Ying-Chun, Yuan Hao-Jie, Wu Peng. Image saliency detection based on local and regional features. *Acta Automatica Sinica*, 2013, **39**(8): 1214–1224  
(郭迎春, 袁浩杰, 吴鹏. 基于 Local 特征和 Regional 特征的图像显著性检测. *自动化学报*, 2013, **39**(8): 1214–1224)
- 34 Jiang Xiao-Lian, Li Cui-Hua, Li Xiong-Zong. Saliency based tracking method for abrupt motions via two-stage sampling. *Acta Automatica Sinica*, 2014, **40**(6): 1098–1107  
(江晓莲, 李翠华, 李雄宗. 基于视觉显著性的两阶段采样突变目标跟踪算法. *自动化学报*, 2014, **40**(6): 1098–1107)
- 35 Mai L, Niu Y Z, Liu F. Saliency aggregation: a data-driven approach. In: Proceedings of the 2013 IEEE International Conference on Computer Vision and Pattern Recognition. Sydney, Australia: IEEE, 2013. 1131–1138



**徐 威** 南京理工大学计算机科学与工程学院博士研究生. 2009 年获得南京理工大学计算机科学与技术学院学士学位. 主要研究方向为图像处理, 计算机视觉.  
E-mail: xuwei904@163.com

(**XU Wei** Ph.D. candidate at the School of Computer Science and Engineering, Nanjing University of Science and Technology. He received his bachelor degree from the School of Computer Science and Technology, Nanjing University of Science and Technology in 2009. His research interest covers image processing and computer vision.)



**唐振民** 南京理工大学计算机科学与工程学院教授. 2002 年获得南京理工大学计算机系模式识别与智能系统方向博士学位. 主要研究方向为智能机器人, 模式识别, 图像处理与智能信息系统. 本文通信作者. E-mail: tzm.cs@njust.edu.cn

(**TANG Zhen-Min** Professor at the School of Computer Science and Engineering, Nanjing University of Science and Technology. He received his Ph.D. degree in the major of pattern recognition and intelligent system from the Computer Department, Nanjing University of Science and Technology in 2002. His research interest covers intelligent robot, pattern recognition, image processing, and intelligent information system. Corresponding author of this paper.)