

# 基于视觉特征的不规则形状目标分割方法

李雄飞<sup>1,2</sup>,赵浩宇<sup>1</sup>,陈霄<sup>3</sup>,赵宏伟<sup>1,2</sup>

(1. 吉林大学 软件学院,长春 130012;2. 吉林大学 计算机科学与技术学院,长春 130012;3. 吉林农业大学 信息技术学院,长春 130118)

**摘要:**该方法利用视觉底层特征颜色、方向、强度和轮廓构建显著图,通过最大熵估计方法获得显著特征分割蒙板;利用中层视觉特征对图像进行超分割,其中在聚类时考虑特征向量的空间信息,并依据显著性自动分配初始参数,使分割后的超像素与目标轮廓更接近;最后将底层视觉特征和中层视觉特征融合,通过底层特征分割蒙板判定图像的超像素归类,将不规则目标从背景中分离。实验结果表明:本文分割方法受复杂背景和光照的影响较小,分割目标轮廓准确,实现了不规则显著目标与复杂背景的有效分离。

**关键词:**计算机应用;超像素;图像分割;视觉显著特征;显著图

**中图分类号:**TP391    **文献标志码:**A    **文章编号:**1671-5497(2014)04-1140-05

**DOI:**10.13229/j.cnki.jdxbgxb201404036

## Irregular shape object segmentation based on visual feature

LI Xiong-fei<sup>1,2</sup>, ZHAO Hao-yu<sup>1</sup>, CHEN Xiao<sup>3</sup>, ZHAO Hong-wei<sup>1,2</sup>

(1. College of Software, Jilin University, Changchun 130012, China; 2. College of Computer Science and Technology, Jilin University, Changchun 130012, China; 3. College of Information Science and Technology, Jilin Agricultural University, Changchun 130118, China)

**Abstract:** A salient object segmentation method based on low-level visual feature and middle-level visual cues was proposed. First, the low-level visual feature of the original image was extract via color, intensity, orientation and local energy feature channels to build the saliency map. The salient feature mask was acquired via the maximum information entropy principle. Then In middle-level, the visual cues were applied for over-segmentation of an image into superpixels. In clustering, the spatial information of the feature vector was taken into consideration according to the salient intensity, and the initial parameters were automatically set. Thus, the superpixels after segmentation accurately approach the object contour. Finally, for segmenting the irregular object from background, the superpixels were classified using the salient feature mask, and the low-level and middle-level features were fused. The experiment results demonstrate that the proposed method is less sensitive to complex illumination and background, and can be used to segment contour accurately. Moreover, it can be applied to segment irregular objects from complex background.

**Key words:** computer application; superpixels; image segmentation; visual salient features; saliency map

**收稿日期:**2013-04-07.

**基金项目:**国家自然科学基金项目(61101155);吉林省科技发展计划项目(20140101184JC).

**作者简介:**李雄飞(1963-),男,教授,博士生导师.研究方向:数据智能处理. E-mail:lxf@jlu.edu.cn

**通信作者:**赵宏伟(1962-),男,教授,博士生导师.研究方向:智能信息系统. E-mail:zhaohw@jlu.edu.cn

## 0 引言

将图像中不规则目标从复杂背景中分离一直是计算机视觉与图像处理领域研究的热点。人们在针对目标分割的应用中通常采用两种方法,一种是基于阈值的自适应分割方法<sup>[1]</sup>,另一种为基于显著图的分割方法<sup>[2]</sup>。基于阈值的自适应分割方法计算简单、实时性高、鲁棒性强,如 Otsu、最大熵或最小交叉熵等分割算法。该分割方法从图像全局出发对目标背景进行分析,复杂背景下会对分割效果产生影响。相对于全局分析,先将图像分块再逐一分割的方法有更理想的分割效果。Chou<sup>[3]</sup>和 Pai<sup>[4]</sup>分别提出了一种基于图像分块并逐块分割的方法。基于分块区域的分割方法优于基于全局的分割方法,可以应用于文本分割<sup>[4]</sup>,但应用范围较窄。基于显著图的目标分割方法主要分为两个方向:一种是基于数学方式构建显著图,如利用统计颜色直方图等方式。另一种为基于视觉生理特性构建显著图,如 Itti 视觉选择注意模型<sup>[2]</sup>。近些年基于超像素的图像分割方法逐渐成熟,研究者们将超像素对图像的描述看作中层视觉特征<sup>[5]</sup>,中层视觉特征可以很好地反映图像中目标形状、轮廓等特征。

本文从视觉特征出发,利用底层视觉显著特征,从复杂背景中检出目标,通过中层视觉特征对图像进行超分割,用超像素描述图像中的背景和目标,最后将两种视觉特征融合,实现了不规则目标与背景的分离。

## 1 基于底层视觉特征的目标分割

视觉认知模型中,视觉显著性的直接表现为显著图,本文构建的显著图基于视觉生理特性,将图像金字塔并行通过底层视觉特征通道(颜色、方向、强度和局部能量),然后标准化并融合每个通道下不同尺度的特征图,获取图像的显著图。为了在显著图上更完整地描述目标轮廓,本文将局部能量特征引入到构架显著图的通道中,依据生理学研究<sup>[6]</sup>,局部能量是视皮层感知的重要特征,并且对图像中显著特征有良好的响应,局部能量公式为:

$$E(x) = \sqrt{I^2(x) + H^2(x)} \quad (1)$$

$I(x)$  和  $H(x)$  是二维 Gabor 函数的偶变换和奇变换。获得基于视觉底层特征的显著图后需要依据显著性强度对图像进行预分割,利用最大

熵估计方法选取显著分割蒙板。获取显著分割蒙板的流程如图 1 所示。

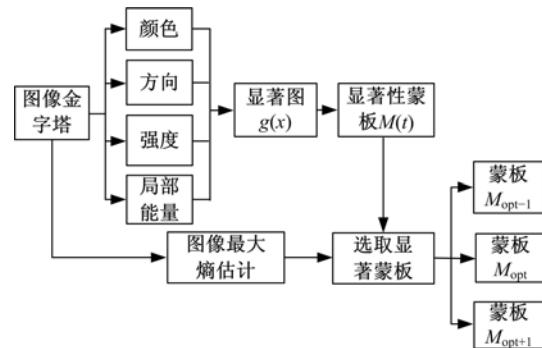


图 1 视觉底层分割

Fig. 1 segmentation method based on low-visual feature

首先根据最大熵估计方法获得图像中的目标熵。图像中的目标区域  $O$  和背景区域  $B$  的熵分别表示为:

$$H_O(t) = \ln \sum_i P_i \ln \frac{P_i}{P_O(t)} \quad (2)$$

$$H_B(t) = \ln \sum_{i=t+1}^{255} P_i \ln \frac{P_i}{P_B(t)} \quad (3)$$

式中:  $P_i$  是灰度  $i$  在图像中出现的概率,表示为  $P_i = \frac{n_i}{M \times N}$ ,  $M \times N$  为原图像的像素数量,图像的灰度集合  $Gray = \{0, 1, \dots, L-1\}$ ,并且  $i \in Gray$ ,  $P_O = \sum_{i=0}^t p_i$  表示目标区域的灰度概率,并且满足  $P_O(t) + P_B(t) = 1$ 。图像总体的熵表示为:  $H(t) = H_O(t) + H_B(t)$ 。当  $H(t)$  达到最大时,  $H_O(t')$  值为目标区域的熵。

基于视觉底层特征获得的四通道显著图用  $g(x)$  表示,显著图中灰度值的大小反映出原图  $f(x)$  中对应位置的显著性强弱。将  $g(x)$  图像按照灰度大小排列,划分 99 个等级,获得 99 个显著性二值蒙板  $M_i (i = 1, 2, \dots, 99)$ ,  $i$  表示显著区域像素数量占据整幅图像的比例。利用公式(4)将图像中显著区域与背景区域分离。

$$f(M) = f(x) M_i \quad (4)$$

$f(M)$  表示分割后图像,计算  $f(M)$  的熵  $H(M)$ :

$$H(M) = - \sum_i^n p(i) \log p(i) \quad (5)$$

当  $H(M)$  和  $H_O(t')$  近似相等时,对应蒙板  $M_i$  为选取的显著特征蒙板,并命名为  $M_{opt}$ ,选取与显著特征蒙板  $M_{opt}$  相邻的两幅显著蒙板记作  $M_{opt-1}, M_{opt+1}$ ,三幅蒙板作为基于底层视觉特征对目标的初级分割,融合中级视觉特征后,分割会变得更为准确。

## 2 基于中层视觉特征的图像表示

本文对 SLIC 算法<sup>[7]</sup>进行了改进,根据显著区域大小设置超像素大小。SLIC 方法首先定义参数  $k$ ,  $k$  趋近于超像素的数量,可以估计超像素大小  $N/k$ ,  $N$  为图像总像素个数。

本文针对参数  $k$  的自适应性做了改进,显著目标所占比例可以由前文中的最大估计获取,当  $H(M)$  和  $H_O(t')$  近似相等时获取显著蒙板  $M_{opt}$ ,此时  $opt$  表示显著性强度较高区域在图像中所占比例。实验表明:当  $opt < 10$ , 意味着图像中显著目标所占比例小于 10, 设置  $k = 300$ , 如果  $10 < opt < 25$ , 则  $k = 200$ ; 如果  $25 < opt < 50$ , 则  $k = 150$ ; 如果  $opt > 50$ , 则  $k = 100$ 。这种判断设置使超像素的大小可以根据显著目标区域大小自适应选取。

获得初始参数  $k$  后, SLIC 算法利用 k-means 聚类生成超像素, 使用距离度量方法, 计算一个像素与聚类中心的颜色特征距离  $d_{lab}$  和空间位置特征距离  $d_{xy}$ 。在 CIELAB 颜色空间中, 图像的像素  $i$  与聚类中心向量  $C_i$  的颜色距离为  $d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2}$ , 空间位置距离为  $d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2}$ , 综合距离为:

$$D_s = \sqrt{d_{lab}^2 + \left(\frac{d_{xy}}{s}\right)^2} \quad (6)$$

式中:  $m$  是权重参数, 当  $m$  增大, 空间距离对聚类影响大, 超像素排列更为紧密; 相反  $m$  减小, 颜色距离相对重要, 超像素的排列更接近图像边缘。本文选用较小的参数  $m$ , 以获得更好的形状描述。

SLIC 算法在聚类时还对聚类的搜索范围做了规定, 该算法聚类时限制在  $2S \times 2S$  范围内汇聚相同的超像素,  $S = \sqrt{N/k}$  约为超像素的边长, 超像素空间上所占据的区域大小接近  $S \times S$ , 图 2 所示为基于显著性强度的自适应超分割结果。

图 2(a) 为基于底层视觉特征的显著分割图,



(a)  $opt=13$  (b)  $k=200$  (c)  $k=100$  (d)  $k=300$

图 2 自适应超分割对比

Fig. 2 Compare results of adaptive over-segmentation

参数  $opt=13$ ; 图 2(b) 为基于显著性判断的超分割图像,  $k=200$ ; 图 2(c) 和(d) 的分割参数分别为  $k=100$  和  $k=300$ , 可见, 图 2(c) 超像素过大对显著目标描述不够准确(如左侧长颈鹿的颈部), 图 2(d) 超像素太小, 容易将完整目标割裂(如右侧长颈鹿头部与耳朵被割裂)。

将超分割图像  $IM_s$  表示为超像素集合形式,  $IM_s = \{SP_1, SP_2, SP_3, \dots, SP_k\}$ , 图像  $IM_s$  由  $k$  个超像素组成, 每个超像素由自然像素集合  $P_i$  构成。在底层视觉特征融合过程中, 需要判定每个超像素  $SP_n$  所归属的类。

## 3 视觉特征融合分割方法

超像素图像  $IM_s$  表示为超像素集合  $\{SP_1, SP_2, SP_3, \dots, SP_k\}$ 。显著性蒙板作用于超像素图像  $IM_s$ 。

$$\begin{cases} S_1 = IM_s M_{opt+1} \\ S_2 = IM_s M_{opt} \\ S_3 = IM_s M_{opt-1} \end{cases} \quad (7)$$

$S_1$ 、 $S_2$  和  $S_3$  是超分割图像与基于底层视觉构建显著性蒙板初始融合的结果, 融合后,  $S_1$ 、 $S_2$  和  $S_3$  中的超像素被分为 3 类: 背景类  $SP_B = 0$ , 目标类  $SP_O = 1$  和待选类  $SP_U$ 。

根据预分割蒙板  $M_{opt+1}$ 、 $M_{opt-1}$  和  $M_{opt+1}$  的不同设计  $S_1$ 、 $S_2$  和  $S_3$  三种归类机制, 如图 3 所示。

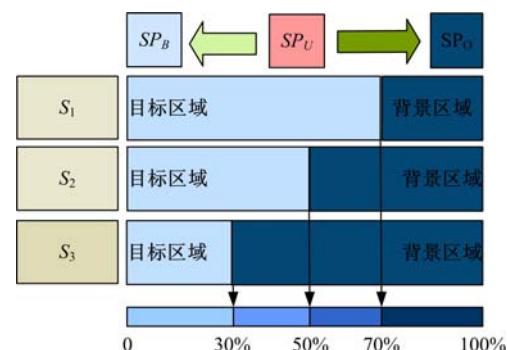


图 3 超像素分类策略

Fig. 3 classification method of superpixels

在  $S_1$  情况下, 超分割图像  $IM_s$  与显著性蒙板  $M_{opt+1}$  融合获得  $S_1$ , 对于  $S_1$  中的  $SP_U$ , 如果超过 70% 的像素被目标区域占据, 则此超像素归为  $SP_O$  类中, 如果超像素中不足 70% 的像素被目标区域占据, 则将此超像素化归为  $SP_B$  类, 归类后重新构成超像素蒙板  $S'_3$  在  $S_3$  情况下, 获得  $S_3$  后,  $SP_U$  中的超像素, 如果有超过 30% 的像素被目标区域覆盖, 则将此超像素归为  $SP_O$  类中; 不

足30%的像素被目标区域覆盖,则将此超像素归为 $SP_B$ ,并获得超像素蒙板 $S'_3$ 。

在S2情况下,如果 $S_3$ 中的超像素 $SP_U$ 超过50%的像素被目标区域覆盖,将此超像素归为 $SP_O$ 类,反之归为 $SP_B$ 类,获得超像素蒙板 $S'_2$ 。

最后,融合超像素蒙板 $S'_1$ 、 $S'_2$ 和 $S'_3$ 。将超分割图像中对应的每个超像素看做一个单元 $SP_i$ ,此时图像 $S'_1$ 、 $S'_2$ 和 $S_3$ 中只有 $SP_B = 0$ 和 $SP_O = 1$ 两类超像素,累加图像 $S'_1$ 、 $S'_2$ 和 $S'_3$ ,对应每个单元的超像素 $SP_i$ 可能获得0、1、2、3四个值。如果对应的超像素 $SP_i = 0$ 或1,则令超像素 $SP_i = 0$ ;如果 $SP_i = 2$ 或者3,则令超像素 $SP_i = 1$ ,最终由超像素构建出二值分割蒙板,也构成了不规则目标的分割蒙板。

## 4 实验与评价

本文从Grab Cut分割图像库<sup>[8]</sup>中选取图像作为实验数据,从自建的图像数据集中选取工业零件、植物、昆虫、动物等不规则形状目标作为实验数据,总计200幅。每幅图片都配备相应的标注用于验证目标检测结果。软件环境Window操作系统、Matlab 2010a,硬件环境双核Intel P8700、2.53 GHz、4 G内存。

### 4.1 与基于显著图的分割模型的比较

利用查准(Precision)、召回(Recall)和正确率(Accuracy)三个评价指标评价分割结果。

$$\left\{ \begin{array}{l} \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{Recall} = \frac{\text{TN}}{\text{FP} + \text{TN}} \\ \text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{FP} + \text{TN} + \text{TP} + \text{FN}} \end{array} \right. \quad (8)$$

TP(True Positive)表示为正确分类的阳性样本数目;FN(False Negative)表示错误地归类为阴性的阳性样本数目;TN(True Negative)表示正确分类的阴性样本数目;FP(False Positive)表示错误地归类为阳性的阴性样本数目。查准率评价图像中阳性样本归类,召回率评价阴性样本归类,准确率对分割效果总体评价。对比实验结果见表1。表中显示结果由200幅实验样本图片获得。

本文算法在查准率、召回率和准确率三个性能参数上均超出另外两种对比方法。图4所示为本文分割方法与Itti显著模型、文献[6]分割方法的实验结果对比。图4(a)为原图像,图4(e)为

Ground-truth图像,图4(b)(c)(d)分别为Itti模型、文献[6]和本文分割结果,可以看出本文方法对不规则目标轮廓的分割相对更准确。

表1 对比实验结果

Table 1 Performance comparison of different methods

	Itti 模型	文献[6]	本文方法
Precision	0.815	0.864	0.873
Recall	0.674	0.797	0.891
Accuracy	0.701	0.807	0.824

### 4.2 与阈值分割方法的比较

在与最小交叉熵和二维Otsu分割方法对比实验中,采用误分率作为客观评价指标,它是现今使用较多的一种图像分割性能测试标准<sup>[9]</sup>。ME值反映了背景像素被错误划分到前景区域的比例,或者是前景像素被错误划分到背景区域的比例,计算式如下:

$$ME = \frac{B_G \cap B_O + F_G \cap F_O}{B_G + F_G} \quad (9)$$

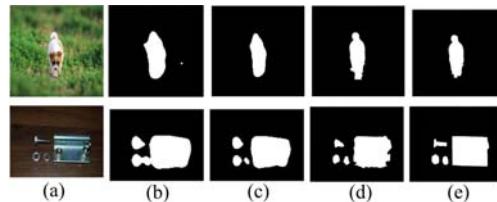


图4 分割结果对比

Fig. 4 Segmentation result comparison of different methods

式中: $B_G$ 与 $F_G$ 表示标准Ground truth图像中的背景与前景,是由研究者经观察人工制作而成, $B_O$ 与 $F_O$ 为分割结果中的背景与前景,∩为取交操作, $B_G$ 、 $F_G$ 、 $B_O$ 与 $F_O$ 是统计出的各部分像素点个数。ME值主要用于衡量分割结果与标准Ground truth图像相差的比率,该值越低分割效果越好,基于阈值的分割方法分割后图像会出现背景区域为1与目标区域为0的结果,此情况下取反后再与Ground-truth图像进行比较。

实验结果表明,在误分率性能上,基于显著图的本文方法( $ME=0.191$ )优于基于阈值分割方法(最小交叉熵的 $ME=0.331$ ,Otsu的 $ME=0.309$ ),图5(a)(b)(c)(d)分别是原图像、最小交叉熵、Otsu和本文方法分割结果。

基于阈值的分割方法获得的分割结果在目标区域内含有光照引起的躁点,如果目标处于复杂背景下,基于阈值分割方法很难清晰地将目标从背景中分离,从图5中看,本文方法在光照不均和

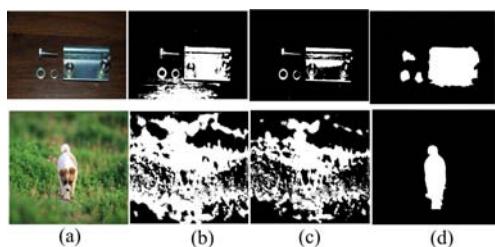


图 5 本文方法与最小交叉熵和 Otsu 方法对比结果

Fig. 5 Compared with the minimum cross entropy and Otsu

背景复杂的状况下,对目标分割效果均较好。

图 6 为本文分割方法与基于显著图和基于阈值分割方法的综合对比,图 6(a)和(f)分别为原图和 Ground-truth,图 6(b)(c)(d)(e)分别是文献[4]、最小交叉熵、Otsu 和本文方法的分割结果。可以看出,本文分割方法对于复杂背景以及光照影响有较好的鲁棒性,并且分割目标的轮廓准确,实现了将不规则显著目标完整地从复杂场景图像中分离的目的。



图 6 综合对比分割实验

Fig. 6 Compared with four methods

## 5 结束语

本文通过对视觉特征的研究,以分割不规则显著目标为目的,提出了基于显著分割与图像超分割结合的方法。通过融合局部能量强化目标轮廓特征,改进底层视觉显著图,利用最大熵估计实现显著目标的初级分割。根据超像素对图像目标

描述准确的特点,将其作为中层视觉特征与底层视觉特征融合,从而使显著目标的描述更准确。在利用超像素分割过程中,还利用显著性判断选取初始分割参数,使算法更适用于本文研究。实验结果验证了本文方法优于传统方法,适用于不规则目标的显著性分割。

## 参考文献:

- [1] Chou Chien-Hsing, Lin Wen-Hsiung, Fu Chang. A binarization method with learning-build rules for document images produced by cameras[J]. Pattern Recognition, 2010, 43(4): 1518-1530.
- [2] 马儒宁,涂小坡,丁军娣,等.视觉显著性凸显目标的评价[J].自动化学报,2012,28(5):876-879.  
Ma Ru-ning, Tu Xiao-bo, Ding Jun-di, et al. To evaluate salience map towards popping out visual objects [J]. Acta Automatica Sinica, 2012, 28(5):876-879.
- [3] Chou C H, Li W H, Chang F. A binarization method with learning-build rules for document images produced by cameras [J]. Pattern Recognition, 2010, 43(4):1518-1530.
- [4] Pai Y T, Chang Y F, Ruan S J. Adaptive thresholding algorithm: Efficient computation technique based on intelligent block detection for degraded document images[J]. Pattern Recognition, 2010, 43(9): 3177-3187.
- [5] Yu Jin-gang, Tian Jin-wen. Saliency detection using midlevel visual cues[J]. Optics Letters, 2012, 27(23):4994-4996.
- [6] 赵宏伟,陈霄,刘萍萍,等.视觉显著目标的自适应分割[J].光学精密工程,2013,21(2): 531-538.  
Zhao Hong-wei, Chen Xiao, Liu Ping-ping, et al. Adaptive segmentation for visual salient object[J]. Optics and Precision Engineering, 2013, 21(2): 531-538.
- [7] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state of the art superpixel method [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11):2274-2282.
- [8] Rother C, Kolmogorov V, Blake A. GrabCut interactive foreground extraction using iterated graph cuts[J]. ACM Transactions on Graphics, 2004, 23(3): 309-314.
- [9] Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation[J]. Journal of Electronic Imaging, 2004, 13(1): 146-168.