

文章编号:1007-2780(2016)01-0117-07

基于区域特征融合的 RGBD 显著目标检测

杜 杰*, 吴 谨, 朱 磊

(武汉科技大学 信息科学与工程学院, 湖北 武汉 430081)

摘要:为了对各类自然场景中的显著目标进行检测,本文提出了一种将图像的深度信息引入区域显著性计算的方法,用于目标检测。首先对图像进行多尺度分割得到若干区域,然后对区域多类特征学习构建回归随机森林,采用监督学习的方法赋予每个区域特征显著值,最后采用最小二乘法对多尺度的显著值融合,得到最终的显著图。实验结果表明,本文算法能较准确地定位 RGBD 图像库中每幅图的显著目标。

关 键 词:目标检测;深度信息;区域特征;随机森林;监督学习

中图分类号:TP394.1;TH691.9 **文献标识码:**A **doi:**10.3788/YJYXS20163101.0117

RGBD salient object detection based on regional feature integration

DU Jie*, WU Jin, ZHU Lei

(College of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China)

Abstract: In order to detect all kinds of salient object in nature scene, we present a detection method which adds regional depth information of image to saliency computation, to apply to object detection. Firstly, we get several region through multi-level image segmentation. Then, we build the regression random forest by learning varieties of regional features, and use the supervised learning approach to map the regional feature vector to a saliency score. Finally, we fuse the saliency scores across multiple levels by least square method, yielding the saliency map. Experiments show our method can accurately locate the salient objects from RGBD images.

Key words: object detection; depth information; regional feature; random forest; supervised learning

1 引 言

视觉注意是人类处理周围环境信息的一种内部特征,该特征能将有限的处理资源快速而准确地分配到显著的视觉区域上。显著目标检测是计

算机模仿人类视觉注意机制,提取出人类对图像中感兴趣的目标区域,在目标检测与识别,图像压缩,图像检索,图像分割等方面,有着广泛的应用^[1-2]。

目前模拟视觉注意机制的计算模型主要分为两类:一类是快速的、由数据驱动的自底向上的模

型,如 Itti 在 1998 年提出的经典 Itti 模型^[3],建立了颜色、亮度、方向 3 个特征通道,利用中心-邻域对比度算子得到多个尺度的特征图,再将这些图归一化到一个尺度,进行线性叠加得到一幅显著图。文献[4]提出了一种基于全局对比度的显著目标检测方法,通过比较每个区域与图中所有其他的区域来计算全局对比度算子从而得到显著图。文献[5]提出了一种基于区域对比度的方法,将颜色和空间分布特征引入到区域显著值的计算中。另一类则是慢速的、由先验知识指导的自顶向下的模型,其中比较有代表性的是基于辨别显著性的检测方法,将提取的像素块特征聚类作为先验知识来模拟人眼对不同物体的辨别能力^[6]。可见现有的许多方法是提取区域的颜色、纹理等低层特征,并由此获得区域间的对比度。在此基础上利用“中心-邻域对比度”原则来计算每个区域的显著值。但是基于局部对比度的方法对于信号的高频部分或噪声过于敏感,而基于全局对比度的方法则对于包含相似前景和背景的图片,检测效果不够理想。针对此不足,DRFI 算法^[7]则是用对比度向量来表示区域的显著特征,并引入了一个新的区域特征向量来表示图像的背景。

深度信息在人类视觉系统中同样起着重要的作用,它反映了物体和观察者之间的距离,被用于图像显著性检测。Zhang 等^[8]设计了一种基于多种感知刺激的立体视觉注意算法,用来提取 3D 视频中的显著区域。Charmaret 等^[9]提出了一种提取感兴趣区域的算法,用于自适应的 3D 图像显示。以上的算法都是直接将深度图和 2D 显著图线性加权,得到最终的显著图。而另一种 3D 显著性检测的方法,是通过将深度显著图融入传统的 2D 显著图来实现。Lang 等^[10]通过在 2D 和 3D 图像中进行眼动实验,用以深度显著性分析,从而提出了一种 3D 显著性检测模型。Ciptadi 等^[11]用颜色和深度特征设计了一个 3D 显著性检测模型,应用于图像分割领域。上述两种方法均是自行定义各显著图的融合权值,缺点是针对不同背景图像中的显著目标,检测结果不够理想。

本文在 DRFI 算法的基础上,将图像的深度信息引入区域的显著性计算。首先对图像进行多尺度的分割得到若干区域;然后对区域多类特征学习构建回归随机森林分类器,采用监督学习的方法赋予每个区域特征显著值;最后用最小二乘

法融合这些多尺度的显著值。

2 算法概述

本文的算法是在 DRFI 算法框架的基础上,将图像的深度信息引入区域显著性计算中。算法包括 3 个主要步骤:多尺度图像分割,区域显著性计算,多尺度融合。算法框架如图 1 所示。

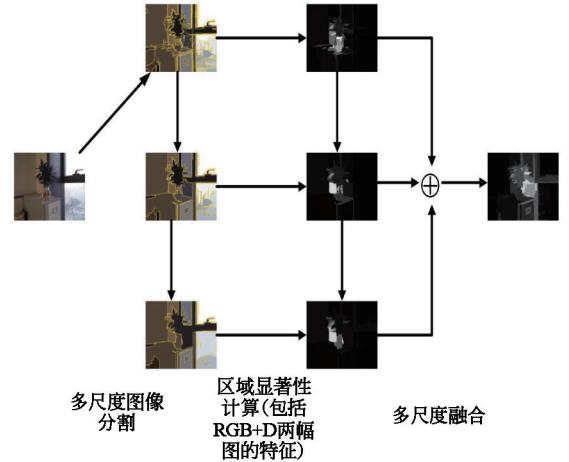


图 1 算法框架图

Fig. 1 Framework of algorithm

首先使用基于图的方法^[12]对原 RGB 图 I , 进行多尺度图像分割。定义无向图 $G = (V, E)$, 顶点 $v_i \in V$, 两个相邻区域的公共边 $(v_i, v_j) \in E$ 。该公共边有一个非负的权值 $w((v_i, v_j))$, 用来衡量这两个区域的不相似性。定义一个区域 $R \subseteq V$ 的内部差异性为 $\text{Int}(R)$, $\text{Int}(R)$ 为该区域最小生成树 $\text{MST}(R, E)$ 的最大权值:

$$\text{Int}(R) = \max_{e \in \text{MST}(R, E)} w(e), \quad (1)$$

定义两个区域 $R_1, R_2 \subseteq V$ 之间的差异性为这两个区域公共边的最小权值:

$$\text{Dif}(R_1, R_2) = \min_{v_i \in R_1, v_j \in R_2} w((v_i, v_j)), \quad (2)$$

定义两个区域的最小内部差异性 $M\text{Int}$:

$$M\text{Int}(R_1, R_2) = \min \left(\begin{array}{l} \text{Int}(R_1) + \tau(R_1) \\ \text{Int}(R_2) + \tau(R_2) \end{array} \right), \quad (3)$$

$$\tau(R) = k / |R|, \quad (4)$$

其中: $|R|$ 表示区域 R 的大小, k 是一个常数。阈值函数 τ 控制两个区域间的内部差异性。当 $M\text{Int}(R_1, R_2) < \text{Dif}(R_1, R_2)$ 时, 这两个区域就可以合并成一个区域。

定义数组 $S = \{S_1, S_2, \dots, S_m\}$ 表示一组 m 尺度的分割图。每一个分割图 S_m 是由 I 分割而成,它包含 K_m 小区域。 S_1 是经过最精确分割得到的,它包含最多的区域,而 S_m 则是经过最粗略分割得到,它包含最少数量的区域。定义 $S_1 = \{R_1^1, R_2^1, \dots, R_{K_1}^1\}$ 表示分割图 S_1 中所有的区域。其它的分割图 $\{S_2, \dots, S_m\}$ 是基于 S_1 计算出来的, S_2 是通过合并 S_1 中的区域得到,以此类推, S_m 则是通过合并 S_{m-1} 中的区域得到的。

分割尺度增多,检测效果会提高,但所需时间亦会增加。综合考虑时间消耗和检测效果,取 $m = 3$ 。

3 区域显著性计算

本文引入了图像的区域深度信息,用 3 种类型的特征来描述每个区域的显著性:区域对比度,区域属性,区域背景对比度。前两种是区域的局部对比度,而背景对比度则是区域的全局对比度。

3.1 区域对比度描述子

首先为每个区域定义一个特征向量 v , 包括多维的颜色、纹理和深度特征。对于一个区域 $R \in S_m$, 分别用 v^R 和 v^N 表示该区域及其邻域的特征向量,则 R 的区域对比度描述子为 $\text{diff}(v^R, v^N)$, 如表 1 所示。特征向量中直方图特征之间计算 χ 距离:

$$\chi^2(h_1, h_2) = \sum_{i=1}^b \frac{2(h_{1i} - h_{2i})^2}{h_{1i} + h_{2i}}, \quad (5)$$

其中 b 代表直方图的维度。而其它特征之间计算绝对差值:

$$D(x_1, x_2) = (|x_{11} - x_{21}|, \dots, |x_{1d} - x_{2d}|), \quad (6)$$

其中 d 代表向量 x_1 和 x_2 中元素的个数。表 1 定义栏中, c_i 表示区域对比度每一维的特征。

3.2 区域属性描述子

除了区域对比度,区域的固有属性也需被考虑,它包含形状和几何特征。区域属性描述子如表 2 所示, p_i 表示区域属性每一维的特征。形状特征包含了一个区域颜色和纹理的分布,它描述了显著目标和背景的一般属性。几何特征则包含了一个区域的大小和位置,它描述了显著目标和背景的空间分布。

3.3 区域背景对比度描述子

图像的背景判定依赖于整幅图的信息。拥有相似形状特征的区域,在一幅图中可能属于背景,而在其它图中则可能属于显著目标。因此,仅仅使用区域属性特征来判定一个区域属于背景或是显著目标是不够的。

本文提取“背景”,并计算其中每个区域的背景描述子作为参考。图像 4 条边 15 像素宽的区域被定义为“背景”区域 B , 用 v^B 表示背景区域的特征向量。按照“区域对比度描述子”中的特征以及计算方式,计算 R 的区域背景对比度描述子 $\text{diff}(v^R, v^B)$, 如表 1 所示, b_i 表示背景对比度每一维的特征。

表 1 区域对比度和背景对比度描述子
Tab. 1 Regional contrast and regional background descriptors

符号	特征描述	定义	区域	背景	维度
a_1	RGBD 均值	$D(a_1^R, a_1^S)$	$c_1 \sim c_4$	$b_1 \sim b_4$	4
a_2	$L * a * b$ 均值	$D(a_2^R, a_2^S)$	$c_5 \sim c_7$	$b_5 \sim b_7$	3
a_3	HSV 均值	$D(a_3^R, a_3^S)$	$c_8 \sim c_{10}$	$b_8 \sim b_{10}$	3
r_1	LM 滤波器绝对响应(RGB)	$D(r_1^R, r_1^S)$	$c_{11} \sim c_{25}$	$b_{11} \sim b_{25}$	15
r_2	LM 滤波器绝对响应(D)	$D(r_2^R, r_2^S)$	$c_{26} \sim c_{40}$	$b_{26} \sim b_{40}$	15
r_1	LM 滤波器最大响应(RGB)	$D(r_1^R, r_1^S)$	c_{41}	b_{41}	1
r_2	LM 滤波器最大响应(D)	$D(r_2^R, r_2^S)$	c_{42}	b_{42}	1
h_1	$L * a * b$ 直方图	$\chi^2(h_1^R, h_1^S)$	c_{43}	b_{43}	1
h_2	色度直方图	$\chi^2(h_2^R, h_2^S)$	c_{44}	b_{44}	1
h_3	饱和度直方图	$\chi^2(h_3^R, h_3^S)$	c_{45}	b_{45}	1
h_4	纹理直方图(RGB)	$\chi^2(h_4^R, h_4^S)$	c_{46}	b_{46}	1
h_5	纹理直方图(D)	$\chi^2(h_5^R, h_5^S)$	c_{47}	b_{47}	1

表 2 区域属性描述子
Tab. 2 Regional property descriptor

特征描述	属性	维度	特征描述	属性	维度
平均归一化 x 坐标	p_1	1	平均归一化 y 坐标	p_2	1
10%归一化 x 坐标	p_3	1	10%归一化 y 坐标	p_4	1
90%归一化 x 坐标	p_5	1	90%归一化 y 坐标	p_6	1
归一化周长	p_7	1	边界框的宽高比	p_8	1
归一化面积	p_9	1	相邻区域归一化面积	p_{10}	1
RGB 方差	$p_{11} \sim p_{14}$	3	$L * a * b * \text{方差}$	$p_{15} \sim p_{17}$	3
HSV 方差	$p_{18} \sim p_{20}$	3	D 方差	p_{21}	1
LM 滤波器响应方差	$p_{22} \sim p_{36}$	15	LM 滤波器响应方差(D)	$p_{37} \sim p_{51}$	15

$$\Phi(x_p,\theta)\in\{1,0\}, \tag{8}$$

这个函数将当前节点的训练数据集分为两个子集,分给它的两个子节点。在当前分割测试中,找到参数 θ 来最大化信息增益 $G(\theta)$:

$$G(\theta)=H(S_j)-\sum_{k\in\{L,R\}}\frac{|S_j^k|}{|S_j|}H(S_j^k), \tag{9}$$

其中: S_j 表示节点 j 上的训练数据集, $S_j^L = \{(x_p,y_p)\in S_j|\Phi(x_p,\theta)=1\}$ 表示训练数据集分配到左边的子节点,而 $S_j^R = S_j\backslash S_j^L$ 表示训练数据集被分配到右边的子节点, $|\cdot|$ 表示数据集的大小, $H(\cdot)$ 表示 y_p 的香农熵或基尼不纯度^[13]。通过反复分裂一棵树的节点,直到达到叶子节点中训练数据的最小量或者一棵树深度的最大值。每一棵树都要在输入的图像块中被反复评估,直到到达树的一个叶节点。

随机森林的优势在于:其一能够处理高维数据,且不需选择特征,直接将区域的多维特征输入随机森林训练即可;其二能够在训练完成后给出哪些特征比较重要。

随机森林中某个特征 x 的重要性的计算方法如下:首先对于随机森林中的每一棵决策树,使用相应的袋外数据 OOB 来计算它的袋外数据误差,记为 errOOB1;然后对袋外数据所有样本的特征 x 随机地加入干扰噪声,这样就能改变样本在特征 x 处的值,再次计算它的袋外数据误差,记为 errOOB2;最后假设随机森林中有 N 棵决策树,计算特征 x 的重要性 $\text{Im}(x)$:

$$\text{Im}(x)=\sum(\text{errOOB2}-\text{errOOB1})/N, \tag{10}$$

若某个特征加入噪声之后,袋外的准确率大

4 多尺度融合

通过区域显著值的计算,每个区域 $R_n^m \in S_m$ 都有了一个显著值 a_n^m 。对于每一个尺度的分割图,将其中每个区域的显著值都赋予这个区域所包含的所有像素点。这样就能生成 m 张显著图 $\{A_1,A_2,\dots,A_m\}$,然后将它们用函数 $A = g(A_1,A_2,\dots,A_m)$ 融合到一起,得到最终的显著图 A 。

4.1 随机森林

随机森林是由 LeoBreiman 在 2001 年提出来的,它利用重抽样方法从原始样本中抽取多个样本,对每个样本进行决策树建模,然后组合多棵决策树的预测,通过投票得出最终的预测结果。

定义 $I_p \in \mathbf{R}^{\omega \times \omega \times 3}$ 是以像素点 p 为中心,大小为 $\omega \times \omega$ 的图像 I 中的一个块, x_p 为块 I_p 的一个特征向量,决策树 $f_{\Theta}(x_p)$,被 Θ 参数化,它将特征 x_p 分入二元标记 $y_p \in \{0,1\}$ 中。随机森林是由 T 棵这样的决策树构成,每一棵树都有一个独立的参数 Θ_i 。 i 决策树叶子节点上的后验分布为 $P_i(y_p|x_p)$,随机森林的输出 $F(x_p)$ 就是最后的分类标记 y_p^* :

$$y_p^* = \underset{y_p}{\operatorname{argmax}} \frac{1}{T} \sum_{i=1}^T P_i(y_p|x_p), \tag{7}$$

在训练的过程中,每一棵决策树都是从训练数据集 $S = \{(x_p,y_p)\}$ 随机采样出的子集中,被独立构建的。对于每棵树上的决策节点,特征向量 x_p 中的每一个元素都是从二元测试中被选取的。带有参数 θ 的分割函数被定义为:

幅度降低,则说明这个特征对样本的分类结果影响很大。

决策树越多,训练效果越好,但是所耗时间也越长。综合考虑时间消耗和训练效果,取 $N = 200$ 。

4.2 区域显著性回归器学习

区域显著性的估计量是通过学习一组训练样本获得。在每个尺度上,对分割训练图像而获得的每一个区域 R ,在人工标记结果图上寻找对应区域 G 。若 G 中所含像素的标记有 80% 属于前景(背景),则 R 标记为前景(背景), R 的标记值 a 被设为 1(0)。训练样本中包含一组已标记的区域 $\mathcal{R} = \{R_1, R_2, \dots, R_Q\}$,和对应的标记值 $A = \{a_1, a_2, \dots, a_Q\}$ 。

如前文所述,每个区域都被一个特征向量 x 描述,向量中包含 3 种类型的描述子。利用标记的训练样本 $\Psi = \{x_1, x_2, \dots, x_Q\}$ 和它们的标记值 $A = \{a_1, a_2, \dots, a_Q\}$,通过对区域的多类特征的学习来构建标准的回归随机森林,得到区域显著性回归器 f ,然后在实际检测中利用该分类器对每个区域进行回归分析。该回归器能够自行组合特征,并发现其中较为重要和显著的特征。

4.3 多尺度显著图融合学习

设每个训练图像得到的多尺度显著图为 $\{A_1, A_2, \dots, A_m\}$,对应的人工标记图为 T 。算法需要通过学习得到一个函数 $g(A_1, A_2, \dots, A_m)$ 来将它们融合,得到最终的显著图 A 。这个问题在现有的算法中已经有解决的途径,例如条件随机域模型^[14]。本文是用最小二乘的方式,训练多尺度线性融合权重 ω_m :

$$\underset{\omega_m}{\operatorname{argmin}} \left\| T - \sum_{m=1}^M \omega_m A_m \right\|^2, \quad (11)$$

那么最终的显著图 A 为:

$$A = \sum_{m=1}^M \omega_m A_m, \quad (12)$$

与现有许多算法自行定义融合权值不同,本文是采用学习的方法得到 ω_m 。通过最小二乘法,学习得到与人工标记图差异较小的显著图,赋予它较大的权值,反之与人工标记图差异较大的显著图,则赋予它较小的权值。这样,融合以后的显著图就会更加接近人工标记图,从而满足人类的视觉特性。

5 仿真实验结果与分析

RGBD 图像库^[15]包含 1000 幅具有深度数据的图片,这些图片是用 Kinect 设备从不同的场景下拍摄的,包含有办公室、超市、校园、街道等,比起其他现有的 3D 图像库,该库数据量大且类型丰富。另外,人工标记结果的生成包括两个步骤:首先由多位测试者对图像中显著的目标进行框选,然后再通过 Photoshop 等图像编辑软件在框内对目标进行精确的分割。

随机选取 RGBD 图像库中的 500 幅图片训练随机森林回归器,剩下的 500 幅作为测试的数据样本。

从 500 幅样本中随机选取六幅图片,用本文算法及其它 6 种较新的算法:DRFI^[7]、GBMR^[16]、HSD^[17]、STD^[18]、GS^[19]、RARE^[20],做对比实验,检测结果如图 2 所示。

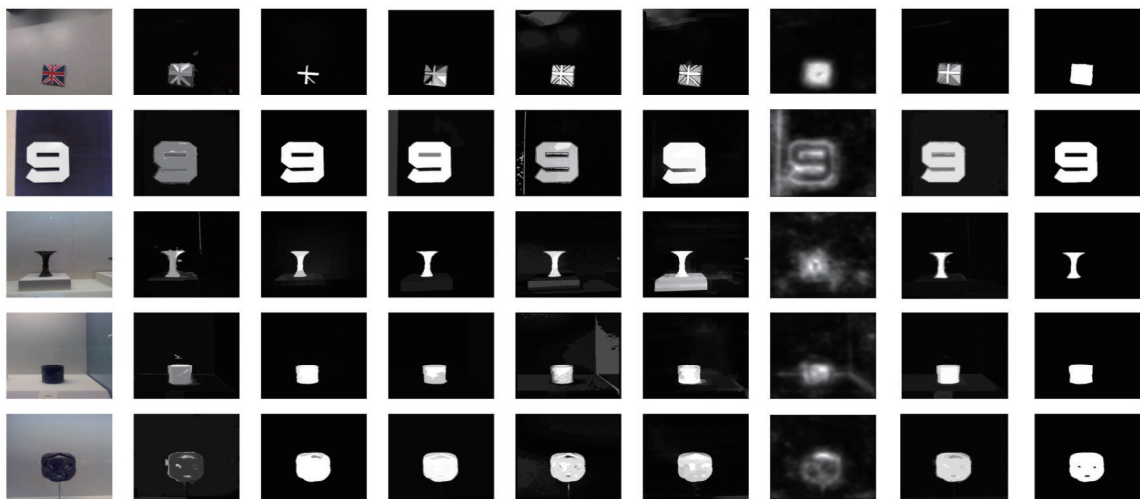




图 2 显著图对比图
Fig. 2 Visual comparison of saliency maps

由图 2 可以看出,本文的算法能够准确地定位原图中的显著目标,显著图较之其他 6 种方法更加接近人工标记的结果。特别地,与 DRFI 算法的效果相比,加入了区域深度信息的本文算法,检测结果更加准确,尤其在前景目标和背景的深度距离有一定差别时,该算法能够定位显著区域,并突出前景目标。

用本文方法与上述 6 种方法分别获得测试样

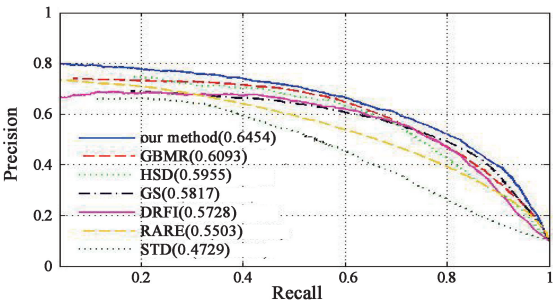


图 3 精度 - 查全率曲线
Fig. 3 Precision-Recall curves

本中的 500 幅图片的显著图,再对比显著图和 Ground truth 计算精度和查全率,精度 - 查全率曲线如图 3 所示。

由图 3 可知,用本文的算法进行显著目标检测,精度和查全率明显高于其它 6 种方法,AUC 值为 64.54%,说明本文算法具有较好的检测效果。

6 结 论

本文在 DRFI 算法的基础上,将图像的深度信息加入到包含局部和全局对比度的 3 种描述子中,进行区域的显著性计算。首先对图像进行多尺度的分割,然后通过对区域多类特征的学习构建回归随机森林,采用监督学习的方法赋予每个区域特征显著值;最后用最小二乘法融合这些多尺度的显著图。实验结果表明,与其他几种算法相比,本文的算法能取得较好的检测效果。

参 考 文 献:

[1] 张辉,王耀南,周博文,等. 医药大输液可见异物自动视觉检测方法及系统研究[J]. 电子测量与仪器学报,2010,24(2):125-130.
ZHANG H, WANG Y N, ZHOU B W, *et al.* Research on automatic visual inspection method and system for foreign substances in medicine transfusion liquid[J]. *Journal of Electronic Measurement and Instrument*, 2010, 24(2): 125-130. (in Chinese)

[2] 张新龙,汪荣贵,张璇,等. 基于视觉区域划分的雾天图像清晰化方法[J]. 电子测量与仪器学报,2010,24(2): 125-130.
ZHANG X L, WANG G R, ZHANG X, *et al.* Foggy image enhancement based on regions of human visual[J]. *Journal of Electronic Measurement and Instrument*, 2010, 24(2): 125-130. (in Chinese)

[3] Itti L, Koch C, Niebu E, *et al.* A model of saliency-based visual attention for rapid scene analysis[J]. *IEEE TPA-MI*, 1998, 20(11): 1254-1259.

[4] Cheng M M, Zhang G X, Mitra N J, *et al.* Global contrast based salient region detection[C]. CVPR, 2011: 409-416.

[5] Perazzi F, Krahenuhl P, Pritch Y, *et al.* Saliency filters: Contrast based filtering for salient region detection[C]. CVPR, 2012: 733-740.

[6] LIU G C, LIN Z C, TANG X O, *et al.* Unsupervised object segmentation with a hybrid graph model[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 910-924.

- [7] JIANG H Z, WANG J D, YUAN Z Y, *et al.* Salient object detection: A discriminative regional feature integration approach[C]. IEEE Conference on CVPR, 2013:208-209.
- [8] ZHANG Y, JIANG G, YU M, *et al.* Stereoscopic visual attention model for 3d video[C]. Proc. 16th Int. Conf. Adv. Multimedia Model. , 2010:314-324.
- [9] CHAMARET C, GODEFFROY S, LOPEZ P, *et al.* Adaptive 3D rendering based on region-of-interest[C]. Proc. SPIE, Stereo-scopic Displays and Application XXI, 2010:75240V.
- [10] LANG C, NGUYEN T V, KATTI H, *et al.* Depth matters: Influence of depth cues on visual saliency[C]. EC-CV, 2012:101-115.
- [11] CIPTADI A, HERMANS T, REHG J M, *et al.* An in depth view of saliency[C]. BMVC, 2013:1-11.
- [12] FELZENSZWALB P F, HUTTENLOCHER D P. Efficient graph-based image segmentation[J]. *International Journal of Computer Vision*, 2004,59(2):167-184.
- [13] BREIMAN L, FRIEDMAN J, STONE C J, *et al.* Classification and regression trees[M]. NY, USA:Chapman and Hall, 1984.
- [14] LIU T, YUAN Z, SUN J, *et al.* Learning to detect a salient object[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011,33(2):353-367.
- [15] PENG H W, LI B, XIONG W H, *et al.* RGBD salient object detection: A benchmark and algorithms[C]. EC-CV, 2014:92-109.
- [16] YANG C, ZHANG L H, LU H H, *et al.* Saliency detection via graph-based manifold ranking[C]. Computer Vision and Pattern Recognition, 2013:3166-3173.
- [17] YAN Q, XU L, SHI J P, *et al.* Hierarchical saliency detection[C]. IEEE Conference on CVPR, 2013: 1155-1162.
- [18] WONG A, FERGANI K, ZELEK J S, *et al.* Statistical textural distinctiveness for salient region detection in nature images[C]. IEEE Conference on CVPR,2013:979-986.
- [19] WEI Y, WEN F, ZHU W, *et al.* Geodesic saliency using background priors[C]. ECCV, 2012:29-42.
- [20] MANCAS M, GOSSELIN B, DUTOIT T. RARE: A new bottom-up saliency model[C]. IEEE International Conference on Image Processing (ICIP), 2012:641-644.

作者简介:杜杰(1990—),男,湖北武汉人,硕士,主要从事图像处理与显著目标检测方面的研究。E-mail:515303239@qq.com

吴谨(1967—),女,安徽芜湖人,教授,博士生导师,主要从事图像处理与模式识别方面的研究。E-mail:471854385@qq.com

朱磊(1982—),男,湖北武汉人,博士,讲师,主要从事图像处理与机器学习方面的研究。E-mail:179901529@qq.com