

中图法分类号: TP391 文献标志码: A 文章编号: 1006-8961(2011)02-0244-06

论文索引信息: 刘娟妮, 彭进业, 李大湘, 王平. 基于谱残差和多分辨率分析的显著目标检测[J]. 中国图象图形学报, 2011, 16(2): 244-249

基于谱残差和多分辨率分析的显著目标检测

刘娟妮¹⁾, 彭进业¹⁾, 李大湘¹⁾, 王平²⁾

¹⁾(西北大学信息科学与技术学院, 西安 710127) ²⁾(中国人民解放军 63628 部队, 北京 101601)

摘要: 根据人类视觉系统的特点, 提出一种融合谱残差和多分辨率分析的显著目标检测方法。该方法通过在不同尺度上计算图像的亮度、颜色以及方向特征的谱残差, 构建多分辨率显著性图谱序列, 然后用线性插值方法将不同分辨率的特征显著图叠加得到 3 个特征显著图, 再利用 k 均值聚类算法将每个特征显著图聚为两类, 选择聚类中心距离最大的特征显著图作为最终的显著图, 最后经过动态阈值处理获得图像的显著目标区域。基于自然图像的显著目标检测实验结果表明, 该方法具有较强的稳定性和实用性, 得到较为满意的检测结果。

关键词: 视觉注意; 目标检测; 谱残差; 多分辨率分析; k 均值聚类

Detecting salient objects based on spectral residual and multi-resolution

Liu Juanni¹⁾, Peng Jinye¹⁾, Li Daxiang¹⁾, Wang Ping²⁾

¹⁾(College of Information Science and Technology, Northwest University, Xi'an 710127 China)

²⁾(The 63628th Troops of People's Liberation Army, Beijing 101601 China)

Abstract: According to the characteristics of human visual system, a salient object detection method based on spectral residual and multi-resolution is proposed. We first compute the spectral residual of three features i. e. intensity, color and orientation under different scales to build a series of multi-resolution saliency maps, which can be combined through linear interpolation to generate three feature-saliency maps. Then we use k-means clustering for binary clustering and select the feature-saliency map with the largest distance between two centroids. Finally we apply dynamic threshold segmentation to get salient regions in an image. The experimental results on natural images show that the new algorithm is stable and practical, and we achieve satisfied results.

Keywords: visual attention; objects detection; spectral residual; multi-resolution analysis; k-means clustering

0 引言

随着多媒体信息技术的发展, 图像已经成为一种重要的信息载体, 日益增多的图像数据给人工处理带来了很大困难。在图像处理和分析中, 人们所关心的区域通常仅占图像中的一小部分, 称为显著区域。所谓显著区域, 通常包含人类感兴趣的重要目标, 最能表达图像的内容, 是人的视觉能够在较短时间内将注意力集中到图像中某个能引起人们关注

的区域。如果对于不同的图像区域赋予不同的处理优先级, 不仅能降低分析过程的复杂度, 而且能提高分析计算的效率。

研究者发现, 在复杂的场景中, 人类视觉系统(HSV)能够迅速地将注意力停留在少数几个显著的目标上, 优先对其进行处理, 这个过程就是视觉注意。这种注意机制使视觉感知过程具有了选择能力, 这正好符合计算机图像分析的要求。因此, 将 HSV 中的视觉注意机制引入到计算机图像分析过程中, 优先处理容易引起观察者注意的区域, 可以极

收稿日期: 2009-07-28; 修回日期: 2009-10-12

第一作者简介: 刘娟妮(1985—), 女。西北大学信息科学与技术学院通信与信息系统专业硕士研究生, 主要研究方向为图像处理和模式识别。E-mail: ljn322@yahoo.com.cn。

大地提高现有图像分析系统的工作效率。显著目标检测正是在这种思想的基础上提出并发展起来的。

如何选择和提取一组能够准确衡量候选区域视觉显著性的图像特征是整个显著目标检测过程中的关键技术。目前的方法大致可以分为以下两类。

1) 基于局部特征的方法。这类方法认为视觉显著性是由视觉对象自身所具有的能够激发观察者兴趣的某种特殊属性产生的,往往对某些特定的目标或图像效果比较好。例如,Kadir^[1]提出的基于像素邻域复杂度的显著性检测算法和 Gesu^[2]提出的基于像素邻域对称性的显著性检测算法。

2) 基于视觉对比的方法。这类方法通过计算目标区域和周围区域的差异来表示显著性。最具代表性的是 Itti^[3]提出的算法。该算法借鉴了视觉心理学中有关人类视觉注意机制的研究成果,通过 Center-surround 算子比较目标区域和周围区域在多种特征、多种尺度上的差异从而获得一幅显著图。这幅显著图表明了图像中各个像素点的显著性,越亮表明该点的显著性越大。

提出了一个融合图像底层特征(亮度、颜色、方向)谱残差和多分辨率分析的显著目标检测算法。首先,提取图像的亮度、颜色、方向特征,然后进行金字塔式分解,将每个特征分为3层,在每个尺度层上分别计算特征的谱残差,即对特征进行傅里叶变换后所得幅度的对数值与其进行均值滤波后的差;其次,用此幅度差值和原相位值重构可得不同分辨率的显著性度量图;再者,利用线性插值将各层显著性度量图进行叠加得到3个特征显著图;最后利用k均值聚类算法将每个显著图聚为两类,选择聚类中心距离最大的特征显著图作为最终的显著图,并进一步对其进行动态阈值处理提取出图像中的显著目标区域。

1 谱残差模型

Barlow^[4]提出的有效编码假说指出,感知系统的信息处理过程与外界信号的统计特性密切相关。初级感知系统的作用就是去除输入信号的统计冗余。也就是说,大脑之所以能够对外界环境自适应,是因为复杂的外界刺激存在着冗余,而大脑的神经元能够有效地去除这些冗余,从而可以利用较少的

资源尽可能有效地表达更多的信息。

从信息论的角度来看,有效编码假说将图像信息 H 分解为两部分:

$$H = H_1 + H_2$$

式中 H_1 代表图像中与众不同的部分; H_2 代表背景,是多余的信息,可以在频率域用图像的统计不变性^[5]来描述。通过去除图像的背景信息,就可以获得图像与众不同的部分,即显著目标。

1.1 log 谱表示

自然图像的统计特性具有变换不变性:即将图像从原来的空间坐标变换到频率坐标系中后,图像在空间中具有的统计特性在频域中仍然保留,这种不变性恰好保证了采用能量谱来刻画自然图像空间相关性的可靠性。

在自然图像的统计特性中尺度不变性是最经典也是研究最广泛的特性,这种特性也被称为 $1/f$ 法则^[5],即自然图像集合的平均傅里叶谱的幅值 $A(f)$ 服从下式的分布,如图1(a)(b)所示。

$$E\{A(f)\} \propto 1/f \quad (1)$$

图1所示反映了自然图像的 log-log 谱和 log 谱的区别。由图可见,log-log 曲线近似为一条直线,而 log 曲线基本符合 $1/f$ 法则。由于单幅图像不具有尺度不变性^[6]以及 log-log 谱的数据分布不均衡(低频部分数据跨度大,高频部分数据汇集)^[7],所以本文采用 log 谱 $L(f)$ 表示图像。

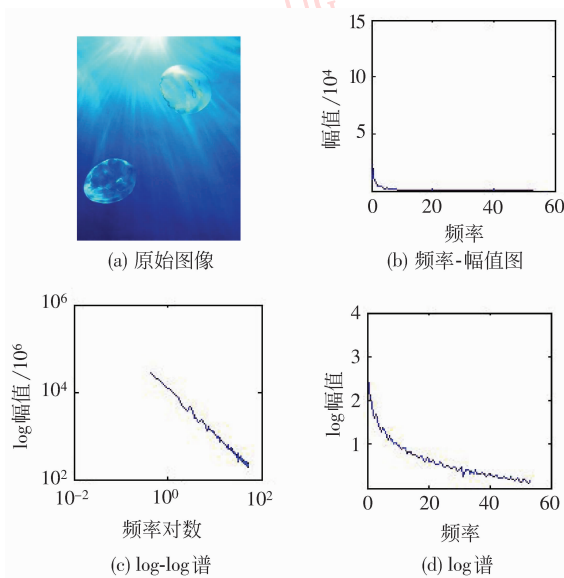


图1 自然图像及其傅里叶频谱曲线

Fig.1 Fourier spectrum curve of natural image

1.2 谱残差描述

对于一个系统模型而言,考虑输入刺激信号的统计相似性可以极大地减少多余的视觉信息,这是因为相似意味着冗余。不同的图像数据,其 log 谱却有着相似的分布趋势,而且曲线满足局部线性条件^[6]。因此,在不同的 log 谱中我们只需要关注其差异部分忽略相似部分(即冗余信息),差异部分对应于原始图像中变化剧烈的区域,也就是我们认为的显著区域。

给定一幅图像 $I(x)$ 首先计算其 2 维离散傅里叶变换,将其从空间域转换到频域,对幅值取对数后得到 log 谱 $L(f)$:

$$A(f) = |F[I(x)]| \quad (2)$$

$$P(f) = \varphi(F[I(x)]) \quad (3)$$

$$L(f) = \log(A(f)) \quad (4)$$

式中 F 代表 2 维离散傅里叶变换, $|\cdot|$ 代表其幅值, φ 代表其相位。由于 log 曲线满足局部线性条件,所以用局部平均滤波器 $h_n(f)$ 对其进行平滑,获得 log 谱的大致形状:

$$V(f) = L(f) * h_n(f) \quad (5)$$

式中 $h_n(f)$ 是一个 $n \times n$ 的矩阵,定义为

$$h_n(f) = \frac{1}{n^2} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \quad (6)$$

实验中, n 取 3。因此谱残差就是 log 谱和其进行均值滤波后的差,可按下面的式子计算:

$$R(f) = L(f) - V(f) \quad (7)$$

谱残差能够描述一幅图像中的异常区域,因此可以用来进行显著目标检测。将谱残差和相位 $P(f)$ 进行 2 维离散傅里叶反变换:

$$S(x) = |F^{-1}[\exp\{R(f) + iP(f)\}]|^2 \quad (8)$$

由此可以重构出一幅图像,用来表示原图像各像素的显著性,称为显著图(saliency map)。

2 模型建立

2.1 特征选择

根据人类视觉原理,视网膜对亮度、颜色、空间方位、运动等具有敏感性。由于在文献[6]中只考虑了图像的亮度特征,这样对于彩色图像来说,并没有有效地利用图像的彩色信息,而且在有些情况下仅使用亮度特征检测显著性会失效,如图 2 所示。

从图 2 可以清楚地看到使用亮度特征并没有检测出视觉上认为的显著目标光圈,而使用颜色特征可以解决这个问题。

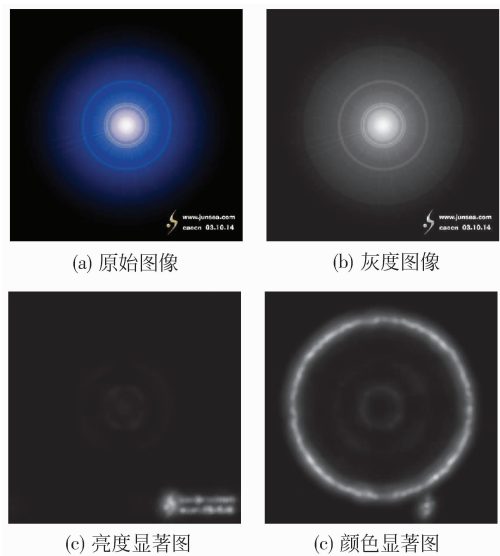


图 2 图像特征选择实例

Fig. 2 An example of image feature selection

因此使用了一种自动选择特征的方法^[8]。首先计算输入图像的亮度、颜色和方向 3 个特征显著图,取值均在 $[0,1]$ 之间,然后利用 k 均值聚类算法将每幅显著图聚为两类,即显著点(初始聚类中心为 1)和非显著点(初始聚类中心为 0),最后选择聚类中心距离最大的特征作为对显著性检测最有效的特征。这是因为假设显著性在能够被有效检测出来前,显著区域和非显著区域间的对比差应该最大。

$$\text{有效特征} = \arg \max_x (|\text{centroid1}_x - \text{centroid2}_x|) \quad (9)$$

式中 x 为亮度、颜色或方向特征。

2.2 多分辨率分析

人类用眼睛观察物体时,随着物体和观察者之间的距离不断变化,视网膜也将感知到不断变化的图像信息。例如,在远处看一栋建筑,可以看到建筑全部,走近一些,可以看到建筑的某一层,再走近一些,可以看到建筑某层的物体,也就是说,在不同分辨率下我们关注的显著目标不同。如图 3 所示,在低分辨率下我们关注的是整片树叶,而在高分辨率下我们会将注意力集中在小虫上,如图 3(b)(c)。

分析和综合这些不同分辨率下的视觉信息可以获得被观察物体的本质特征,这种视觉多分辨率分析的方法既兼顾了图像中大目标和小目标的处理,

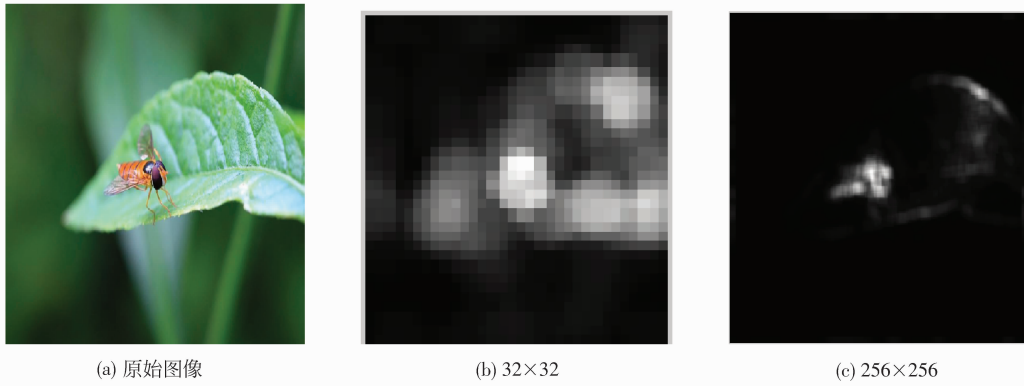


图3 不同分辨率得到的显著图

Fig. 3 Saliency maps at different resolution

又减少了后续处理中的计算量。目前显著性检测算法^[3]常采用多分辨率分析方法。

基于以上考虑,构建下面的显著目标检测模型。

首先,提取图像的亮度、颜色、方向特征。亮度特征

$$I = (r + g + b) / 3 \quad (10)$$

颜色特征 C 有 2 个分量,分别表示“红/绿”和“蓝/黄”色彩通道的对比,定义为^[9]

$$RG = (r - g) / \max(r, g, b) \quad (11)$$

$$BY = (b - \min(r, g)) / \max(r, g, b) \quad (12)$$

方向特征 O 是利用 Gabor 滤波器和亮度卷积得到的,共 4 个分量,Gabor 滤波器表达式如下:

$$G(x, y, \theta) = \exp\left(-\frac{x'^2 + y'^2}{2\delta^2}\right) \cos\left(2\pi \frac{x'}{\lambda}\right) \quad (13)$$

$$x' = x \cos(\theta) + y \sin(\theta) \quad (14)$$

$$y' = -x \sin(\theta) + y \cos(\theta) \quad (15)$$

式中角度 $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, 标准差 $\delta = 7/3$, 纵横比 $\gamma = 1$, 波长 $\lambda = 7$, x' 和 y' 通过计算 θ 得到, 如式(14)(15)。其次,对每一特征分量建立多分辨率金字塔。将特征图的大小统一调整为 256×256 (作为金字塔的第 1 层), 然后采用高斯滤波和隔点采样进行金字塔式分解, 每个特征分量分解为 3 层, 共得到 21 个不同分辨率特征图。然后,按照式(2)——(8)计算每个特征的谱残差并重构得到其显著性度量图。再者,进行特征显著图的叠加:首先将不同分辨率的特征分量在各自分辨率下相加,得到 3 个特征不同分辨率的显著图共 9 个,其次将每个特征第 3 层的结果线性插值后与第 2 层相加,然后再经过线性插值后与第 1 层结果相加得到最终的 3 个特征显著图。最后利用本文前面描述的特征选择方法选出最有效的特征显著图作为最终的显著图 $S(x)$ 。

为了进一步获得更明显的显著区域(白色表示显著区域,黑色表示其余区域),我们使用动态阈值的方法:

$$O(x) = \begin{cases} 1 & S(x) > \text{阈值} \\ 0 & \text{其他} \end{cases} \quad (16)$$

实验中动态阈值的设置方法如下:图 4 所示是 $S(x)$ 的直方图。其中 A_0 为直方图的期望, A_1 为期望和标准差的差, A_2 为期望和标准差的和, A_3 为直方图统计数据中第一个个数小于 10 的点。根据概率论和数理统计可知,直方图中大部分数据都分布在 A_1 和 A_2 之间。因此,阈值可以取 A_2 和 A_3 之间的某个值,它可以有效的消除显著图中的噪声点,达到检测显著目标的目的。本文选择期望与标准差之和的 1.3 倍作为动态阈值获得了较好的检测效果。

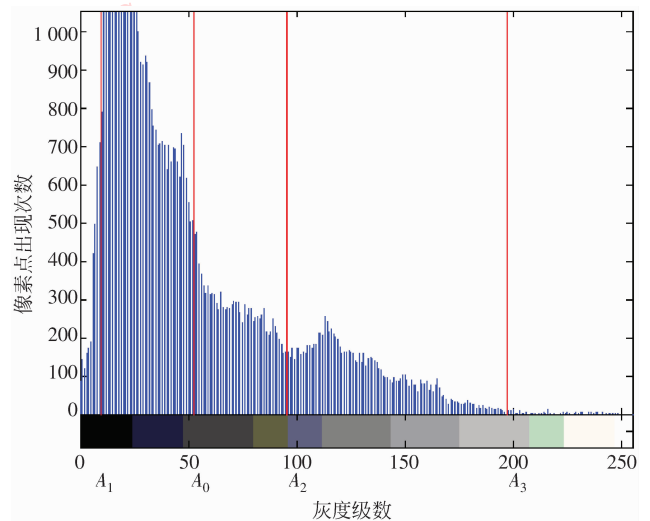


图4 显著图的直方图

Fig. 4 Histogram of Saliency map

3 实验结果与分析

为了验证显著性检测模型的有效性,从网站^[10]下载了 62 幅场景图片,并从 Corel 图像库中选择 300 幅图片对图像库进行了扩充,构成实验图像库。让 5 个观察者标记出每幅图片中的显著目标,设第 k 个观察者标记的显著目标图为 $O_k(x)$,图中 1 代表目标物体,0 代表背景,用击中率 HR(hit rate)和误警率 FAR(false alarm rate)来评价检测效果,计算公式为

$$HR = E\left(\prod_k O_k(x) \cdot S(x)\right) \tag{17}$$

$$FAR = E\left(\prod_k (1 - O_k(x)) \cdot S(x)\right) \tag{18}$$

这种评价标准说明一个最优的显著性检测系统其 HR 值应该在没有被标注为显著目标的区域具有最小值,在被大多数人标为显著目标的区域具有最大值。

实验中,为了更清楚地看到对比结果,通过如下两组实验,将本文模型、Itti 模型^[3]以及 Hou 模型^[6]进行了对比:1)以 Itti 模型为标准,将其余两种模型的 HR 值调整为和 Itti 模型相同,计算调整后的显著图对应的 FAR 值;2)以本文模型为标准,将其余两种模型的 FAR 值调整为和其相同,计算调整后的显著图对应的 HR 值。实验结果如表 1。

表 1 3 种方法的 HR 和 FAR 比较结果

Tab.1 HR and FAR Results for different methods

方法	HR	FAR	HR	FAR
Itti	0.506 2	0.268 4	0.202 3	0.107 2
Hou	0.506 2	0.142 4	0.381 0	0.107 2
本文方法	0.506 2	0.127 3	0.426 4	0.107 2

实验数据表明,用本文方法检测的准确性优于其他两种方法。从测试图像库中随机选取 10 幅复杂背景彩色图像作为实例进行展示,如图 5 所示。其中每组图片第 1 列为输入图像,第 2 列为 Itti 模型^[3]获得的显著图,第 3 列为 Hou 模型^[6]获得的显著图,第 4、5 列分别为本文模型获得的显著图及其阈值处理后的二值图。

由图可见,Itti 模型^[3]虽然能够检测出人所感兴趣的人或物体,但存在较高的误判率,一些简单的背景也被检测出来,如图 5 中第 4 个样本的背景在显著图中占据大部分面积,以至对飞机的检测失效;Hou 模型^[6]较 Itti 模型^[3]有较大改进,能够剔除一些简单的背景,获得相对满意的结果,但是在一些边界对比强烈的地方会出现误判,如图 5 第 10 个样本中天空和山的交界处,以及雪地里露出的小路,都在 Hou 模型^[6]的显著图中有较强的体现,而人物却淹没在这些误判的区域中。

图 5 中第 4 列所示为本文模型的显著图。和 Hou 模型^[6]的显著图对比可见,本文模型获得的显

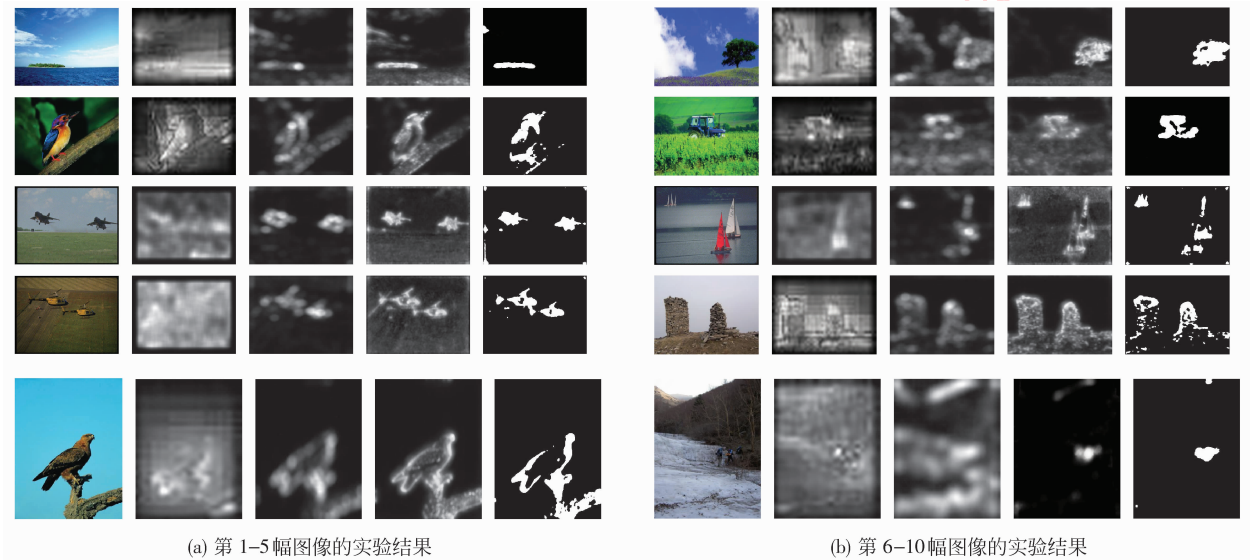


图 5 Itti 模型、Hou 模型以及本文模型的实验结果

Fig.5 The result of our method in comparison with Itti's and Hou's method

著图更加接近显著目标的真实外观,准确性更高,具有很好的抗干扰能力,没有误判、漏判的情况。本文模型的实验结果明显优于以上两者。

4 结 论

根据人类视觉的选择注意特性以及自然图像统计特性所具有的变换不变性,提出了融合谱残差和多分辨率分析的显著目标检测方法。该方法在不同分辨率下提取多种底层特征,进而实现了一种特征选择方法。将此方法应用于自然图像,在检测效果、抗噪能力等方面获得了较为满意的结果。

与 Hou 算法^[6]相比,本文算法具有以下几个突出特点:首先,算法实现了一种自动的特征选择方法,解决了亮度特征在检测某些图像显著性会失效的问题;其次,使用了多分辨率分析方法,这样选取的注视点和人观察图像时的实际注视点更加吻合;最后,采用动态阈值比固定阈值能更好的检测出显著目标区域,减少噪声点的干扰。

算法没有将检测出的显著区域应用到具体的图像分析任务中去,今后我们将结合目标识别、基于内容的图像检索和图像压缩等具体应用研究如何发挥显著目标检测的作用。

参考文献 (References)

[1] Kadir Timor, Brady Michael. Saliency, scale and image

description[J]. International Journal of Computer Vision, 2001, 45(2):83-105.

[2] Gesu Vito Di, Valenti Cesare, Strinati Laurent. Local operators to detect regions of interest[J]. Pattern Recognition Letters, 1997, 18(11-13): 1077-1081.

[3] Itti Laurent, Koch Christof, Niebur Ernst. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259.

[4] Barlow H B. Possible principles underlying the transformation of sensory messages[C]//Sensory Communication. Cambridge, MA: MIT Press, 1961:217-234.

[5] Daniel L Ruderman. The statistics of natural images[J]. Network: Computation in Neural Systems, 1994, 5(4):517-548.

[6] Hou Xiaodi, Zhang Liqing. Saliency detection: a spectral residual approach[C]//IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, Minnesota, USA: IEEE, 2007:1-8.

[7] van der Schaaf A, van Hateren J H. Modeling the power spectra of natural images: statistics and information[J]. Vision Research, 1996, 36(17):2759-2770.

[8] Wang Zheshen, Li Baoxin. A two-stage approach to saliency detection in images[C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Las Vegas, NV: IEEE, 2008:965-968.

[9] Walther Dirk, Koch Christof. Modeling attention to salient proto-objects[J]. Neural Networks, 2006, 19(9):1395-1407.

[10] Hou Xiaodi. Xiaodi's Homepage [EB/OL]. (2008-09-03) [2009-03-20]. <http://www.its.caltech.edu/~xhou/>.