

# 复杂环境下基于视觉显著性的人脸目标检测

陈 凡<sup>1</sup>, 童 莹<sup>2</sup>, 曹雪虹<sup>2</sup>

(1. 南京邮电大学 通信与信息工程学院, 江苏 南京 210003;  
2. 南京工程学院 通信工程学院, 江苏 南京 211167)

**摘 要:**当前复杂环境下人脸识别因受目标背景杂乱等因素影响,分类效果不理想。针对此问题,提出了基于视觉显著性的人脸目标检测方法,利用基于图论的视觉显著性算法(Graph-Based Visual Saliency, GBVS)提取复杂环境中的人脸目标的显著图,对显著图进行阈值分割和形态学操作得到二值人脸目标区域,以区域质心为中心,以区域边缘到质心的最小距离为边长,截取图像中的准确人脸区域,实现复杂环境下的人脸目标检测。在LFW数据库上的实验结果表明,所提算法能够准确地完成“抠图”的任务,具有较为理想的人脸检测效果,因算法实现过程无需人工干预,可有效摒除杂乱背景干扰,且提高了检测速度,实现了无监督的人脸检测,为智能化人脸识别提供了理论研究基础。

**关键词:**复杂环境;基于图论的视觉显著性算法;视觉显著性;人脸检测

**中图分类号:**TP273

**文献标识码:**A

**文章编号:**1673-629X(2017)01-0048-05

doi:10.3969/j.issn.1673-629X.2017.01.011

## Face Target Detection of Visual Saliency in Complex Environment

CHEN Fan<sup>1</sup>, TONG Ying<sup>2</sup>, CAO Xue-hong<sup>2</sup>

(1. College of Communications and Information Engineering, Nanjing University of Posts and  
Telecommunications, Nanjing 210003, China;

2. College of Communication Engineering, Nanjing Institute of Technology, Nanjing 211167, China)

**Abstract:** At present, the classification accuracy of face recognition in complex environment is not satisfied because of the background clutter and other factors. To solve this problem, a face detection algorithm based on visual saliency has been proposed, in which graph based visual salient algorithm is employed to extract salient maps of face region in complex environment and then threshold segmentation and morphological operations is run to get the binary face marks and the centroid is taken as center as well as the minimum distance of the region edge and centroid as side length to crop the accurate area of target face to achieve the goal of face detection in complex environment. Results of experiments on the LFW image database show that the proposed algorithm can accurately fulfill “matting” tasks and achieve good results in face detection, the process of which needs no artificial participation, and can effectively exclude the interference of pell-mell background besides having improved the detection rate, realizing the unsupervised face detection, providing a theoretical foundation for intelligent research of face recognition.

**Key words:** complex environment; GBVS; visual saliency; face detection

## 0 引言

随着图像处理和模式识别技术的发展,人脸识别技术已经成为现代模式识别和人工智能领域的研究热点之一。其中,人脸检测是人脸识别的预处理步骤,可以有效去除背景干扰,尤其对复杂环境下的人脸识别,进行人脸检测尤为重要。人脸检测在近二十年的时间内取得了长足的发展,各种各样的人脸检测算法层出

不穷,主要包括基于模板匹配的人脸检测算法和基于肤色模型的检测算法两类<sup>[1]</sup>。基于模板匹配的人脸检测方法<sup>[2-3]</sup>是根据先验数据归纳出一个统一的模板,然后根据一个能量函数确定被检测域中与模板匹配度最高的区域,即人脸区域。基于肤色模型的人脸检测算法<sup>[4]</sup>主要依据是在彩色图像中,人脸的肤色是区别与非脸的一个显著特征。两种方法各有利弊<sup>[5]</sup>。基于

收稿日期:2016-02-25

修回日期:2016-07-06

网络出版时间:2017-01-04

**基金项目:**国家自然科学基金资助项目(61471162);江苏省自然科学基金(BK20141389);南京工程学院科研基金(QKJA201304)

**作者简介:**陈 凡(1991-),女,硕士研究生,研究方向为图像处理与模式识别;童 莹,副教授,研究方向为图像处理与模式识别;曹雪虹,教授,研究方向为无线通信系统与信息理论。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170104.1023.038.html>

模板匹配的人脸检测算法原理简单,速度快,但由于不同人脸差异很大,很难设计出精确匹配的模板,再加上环境的复杂性,很难获得理想的人脸检测效果。基于肤色模型的算法是在彩色图像中检测人脸的一种很常用的算法,但由于人体其他部位可能具有和人脸相同的肤色,比如男性赤膊等,并且某些场景中的背景也可能和人脸皮肤有相同的模型分布,因此该方法鲁棒性差。

为了解决复杂环境下的人脸检测问题<sup>[6-7]</sup>,比如光照、多姿态、面部遮挡、复杂背景等因素的影响,提出了基于视觉显著性的人脸检测算法。视觉注意机制是人类视觉的一项重要心理调节机制,视觉显著性是利用人的视觉注意机制,在大量的视觉信息中迅速找到显著的或感兴趣的物体。当人眼看到一幅带有复杂背景的图片时,一般情况下,会把注意力放在图像中最显著的区域。因此,所提出的基于视觉显著性的人脸检测算法,可以利用视觉注意机制,从复杂背景中快速锁定人脸目标。相比传统的人脸检测算法,不仅可以不受光照、姿态、遮挡等因素的干扰,准确检测到人脸区域,而且无人工干预,实现过程更加智能化。为后续人脸特征提取、分类提供了有效信息输入,一方面降低了提取特征的计算量,缩短了特征提取的时间,另一方面,去除了干扰因素,使特征提取更准确,识别率也将大大提高。

当前,视觉显著性算法的研究已取得很多成果。早期 Itti 提出生物学启发计算模型<sup>[8]</sup>。之后, Harel 等改进了 Itti 模型,引入马尔可夫链,提出基于图的显著性 (Graph-Based Visual Saliency, GBVS) 模型<sup>[9]</sup>,拓宽了显著性方法计算的思路。接着,提出了基于频率调制的显著性 (Frequency-Tuned Saliency, FTS) 模型<sup>[10]</sup>,此算法在自然图像的显著性检测方面有不错的效果。为弥补之前显著图算法只能检测出独立物体或者主体物体的弊端,Goferman S 等提出基于内容感知的显著性检测 (Context-Aware Saliency Detection, CASD) 模型<sup>[11]</sup>,此方法可以检测出体现图像语义的区域。因为 GBVS 方法提取的显著图的灰度较平均,可以定位出大致目标人脸区域,而其他显著图方法或忽略了目标信息在显著图中的完整性<sup>[12]</sup>,或检测的目标区域细节纹理太过清晰,显然不利于后续对显著图进行阈值分割得到模板的工作。为此,提出了基于视觉显著性的人脸目标检测方法,该方法利用基于图论的显著性模型获取人脸区域的显著图。在输入具有复杂背景的图片后,应用 GBVS 获取人脸图像的视觉显著图,并对其进行阈值分割得到一幅二值模板。因该模板是比较粗糙的,包含一些小的信息干扰或丢失,因而引入形态学操作得到更加精确的二值模板。寻找精确的二值模板

中白色区域的质心,并以区域质心为中心,以区域边缘到质心的最小距离为边长,截取图像中的准确人脸区域,从而检测到目标人脸。

## 1 基于图论的视觉显著性

看到一幅图像,第一眼关注的是图像中最吸引眼球区域。这部分人眼感兴趣的区域因为和周围背景有很大的差异,所以更容易让人察觉,这就是人类的视觉注意机制。显著区域最能表现此幅图像的主要内容,包含原图的信息量最大,因此可以利用视觉注意机制进行复杂环境下的目标人脸检测。GBVS 算法在特征提取的过程中类似 Itti 算法去模拟视觉原理,但在显著图生成的过程引入马尔可夫链,用纯数学计算得到显著值。图 1 是视觉显著图的获取流程。

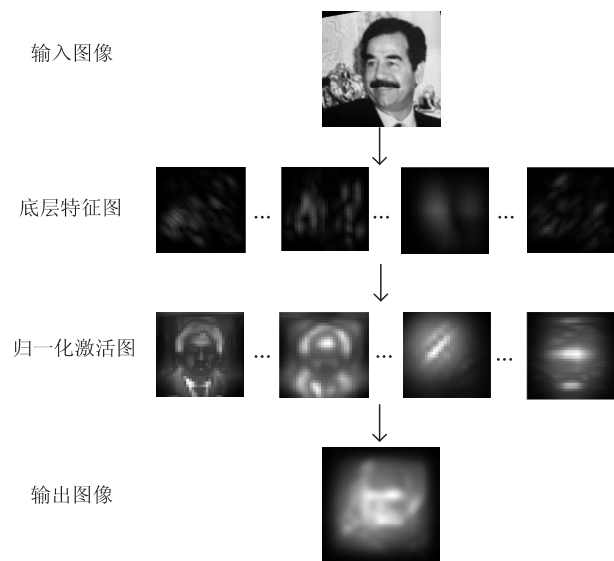


图 1 GBVS 显著图获取示意图

视觉显著图获取步骤如下:

(1) 输入一幅大小为  $250 \times 250$  的灰度图片,利用高斯核函数平滑图像,每次将分辨率降低为原来的  $1/2$ ,下采样 4 次,实验中只用第 2,3,4 次采样得到的图像。分别提取这 3 幅图像的亮度和方向特征,其中方向特征是提取的  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  方向的信息,最后可得到 15 幅底层特征图(图像尺寸  $32 \times 32$ )。

(2) 把这 15 幅特征图依次作为输入,计算每幅图的激活图。对每一幅特征图,以图中的每一个像素点为节点,根据像素点间的灰度值相似度和像素点位置间的距离(欧氏距离)作为连接权值,建立一个全连通的有向图  $G_A$ ,如图 2 所示。

从节点  $(i, j)$  到节点  $(p, q)$  的有向边会赋予一个权值  $W((i, j), (p, q))$ ,权值定义为:

$$W((i, j), (p, q)) = d((i, j) \parallel (p, q)) \cdot F(i - p, j - q) \quad (1)$$

其中,  $d((i, j) \parallel (p, q))$  表示节点  $(i, j)$  和节点

$(p,q)$  之间灰度值  $M(i,j)$  和  $M(p,q)$  的相似程度,计算公式为:

$$d((i,j) \parallel (p,q)) = \left| \log \frac{M(i,j)}{M(p,q)} \right| \tag{2}$$

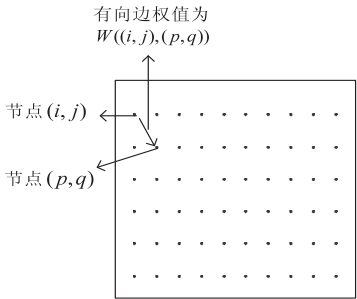


图 2 有向图构建示意图

$F(a,b)$  表示节点  $(i,j)$  与节点  $(p,q)$  位置间的欧氏距离,计算公式为:

$$F(a,b) = \exp\left(-\frac{a^2 + b^2}{2\sigma^2}\right) \tag{3}$$

(3)连接权值矩阵(1 024×1 024),并进行归一化,使矩阵每列之和为 1,形成马尔可夫状态转移矩阵。

(4)对马尔可夫转移矩阵进行多次迭代,直到马尔可夫链达到平稳分布。马尔可夫链的平稳分布反映了随机游走者到达每个节点/状态消耗的时间。节点视觉特征越相似,权值就越大,转移概率越大,在两点之间游走花费的时间就短;反之,则越长。视觉特征越不相似的点越显著。

(5)找到马尔可夫矩阵的主特征向量(1 024×1),主特征向量是主特征值对应的向量,矩阵的多个特征值中模最大的特征值叫主特征值,对应图像的显著节点。把主特征向量重新排列成 2 维(32×32)的形式,就得到了激活图,并进行归一化。

(6)按照上面的方法得到每个特征通道的特征图的激活图,再把各个特征通道内激活图相加,最后把亮度和方向特征通道激活图都叠加起来,就得到了视觉显著图。

2 基于视觉显著性的人脸检测算法

算法流程图如图 3 所示。

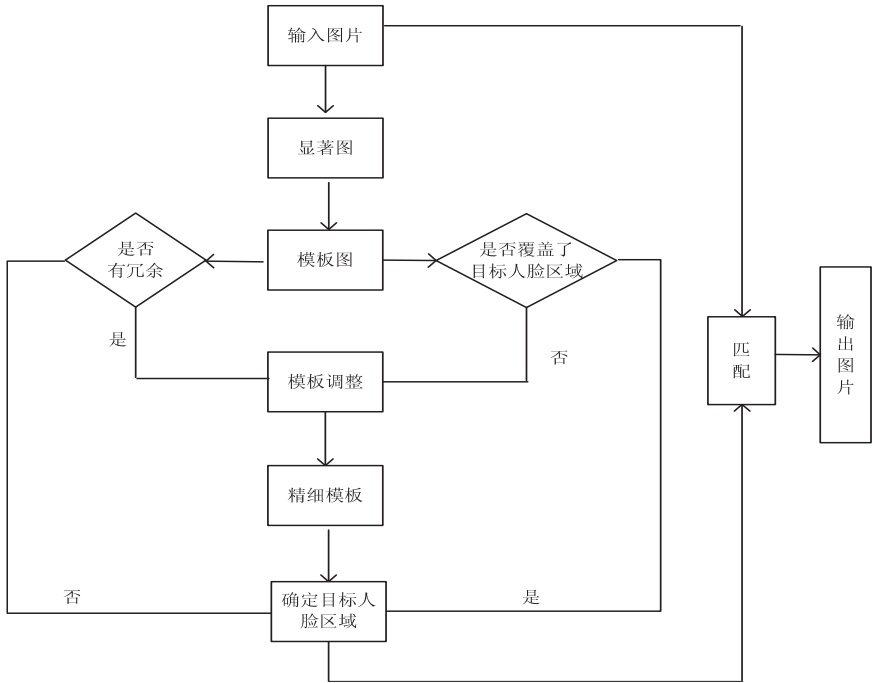


图 3 所提算法流程图

所提出的算法,其输入是一幅 250×250 的灰度图像,输出是包含目标人脸区域的图像。算法可分为两部分:第一部分是利用视觉注意机制获取目标人脸区域的显著图部分,具体算法已给出;第二部分是利用显著图提取人脸区域图像部分,具体实现步骤如下:

(1)输入图像是一幅 250×250 的灰度图片  $I(x,y)$ ,利用 GBVS 算法提取目标人脸区域显著图,记为  $S(x,y)$ 。

(2)采用双三次插值算法调整  $S(x,y)$  图像尺寸为 250×250,把  $S(x,y)$  中的每个像素按照灰度值强弱

划分成  $n$  个等级( $n$  一般取 99),返回第  $m$  个百分位值,此处  $m$  取 60。把这个值作为阈值,分割显著图  $S(x,y)$  得到模板  $M_1(x,y)$ 。

(3)判断模板  $M_1(x,y)$  是否有冗余,若是,则执行步骤(5);否则执行步骤(6)。

(4)判断模板  $M_1(x,y)$  是否完全覆盖了目标人脸区域,若否,则执行步骤(5);若是,则执行步骤(6)。

(5)调整模板  $M_1(x,y)$  得到精细模板  $M_2(x,y)$ 。如果模板  $M_1(x,y)$  余留了目标人脸以外的区域,就把  $M_1(x,y)$  中的冗余区域删除,只保留有用的区域;如果



$M_1(x,y)$  没有完全覆盖整个目标人脸区域,就对  $M_1(x,y)$  实施闭运算;如果  $M_1(x,y)$  丢失信息过多,就对  $M_1(x,y)$  进行孔洞填充。

(6) 确定目标人脸区域。初始化变量  $sum\_x$ 、 $sum\_y$ 、 $area$ ,用以记录精细模板  $M_2(x,y)$  中非零区域横纵坐标之和与非零区域数目。遍历  $M_2(x,y)$ ,如果其中某个像素点值为 1,就把  $area$  值加 1,  $sum\_x$  和  $sum\_y$  的值分别加上该点对应的横纵坐标值。遍历完  $M_2(x,y)$  之后,用  $area$  对  $sum\_x$  和  $sum\_y$  求平均就得到  $M_2(x,y)$  的质心  $p$ 。用  $edge$  函数处理  $M_2(x,y)$ ,得到  $M_2(x,y)$  区域边缘的二值图像,求出质心  $p$  到区域边缘的最小距离  $d$ 。以  $p$  为中心,以  $d$  为边长作正方形,这个正方形区域就是目标人脸所在的区域。

(7) 将输入图片  $I(x,y)$  与目标人脸区域作匹配就得到最终的人脸检测图。

3 实验仿真

对所提出的算法在 LFW (Labeled Faces in the Wild) 人脸数据库<sup>[13]</sup>上进行了实验。这个数据库包含 5 749 个不同人 13 233 张人脸图片,图片大小为 250×250,其中 1 680 个人有两张以上的图片,剩余的 4 069 个人只有一张图片,有些图片中含有不只一张人脸,但是目标人脸是位于图片中间,不是目标人脸的人脸均被视为背景,每张图片都分配了唯一的统一格式的名字标识。实验采用的计算机硬件配置是 Intel (R) Core(TM) i5-5200@ 2.20 GHz 2.20 GHz,4.00 GB 内存,其软件环境是 Windows 8 操作系统,程序使用 Matlab 2014a 语言进行编写。主要包括以下三个实验仿真。

3.1 所提出算法与 Itti、CASD 人脸检测效果比较

采用 GBVS、Itti、CASD 三种显著图算法提取同一幅图片的视觉显著图,再用阈值分割的方法得到二值模板,然后把二值模板与输入图片相乘得到人脸区域分割图,结果如图 4 所示。

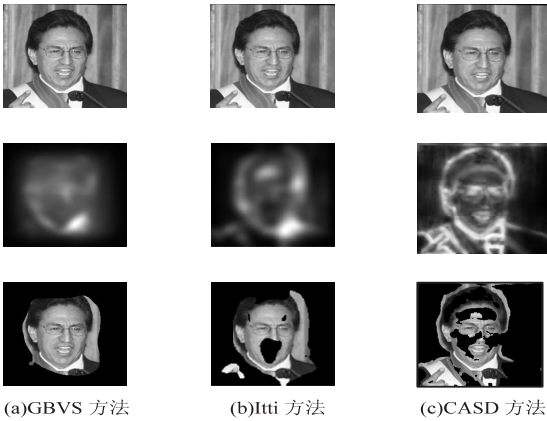


图 4 视觉显著图方法分割的人脸效果图

通过分析比较图 4 中三种显著图方法分割人脸区域的效果,可以看出,GBVS 显著图方法提取的目标显著区域轮廓模糊,阈值分割更容易区分与非脸区域,因此更完整地分割出了目标人脸,既无较多冗余信息,也没丢失面部信息;Itti 方法忽略了目标信息在图像中的完整性,效果次之,可以看到嘴巴区域信息完全丢失,此外,还残留了些许的衣服领口信息;CASD 方法提取的显著区域纹理过于详细,灰度值差异大,不利于二值化分割,不仅额头、脸颊、鼻梁有严重的信息丢失,而且绶带和衣领部分有较多冗余信息,分割效果最差。

3.2 所提出算法与基于模板的人脸检测效果比较

基于模板的人脸检测 (Template-Based Face Detection, TBFD) 方法是通过定位输入图片中人的两只眼睛,获取眼睛之间的线段距离  $D$ ,再分别以  $D$  的中点为端点,向上  $0.5D$  距离和向下  $1.5D$  距离为另一端点作线段,以  $D$  和  $2D$  确定的矩形区域为人脸模板截取输入图片,从而达到人脸检测的目的。对大多数图片而言,TBFD 方法和基于 GBVS 的人脸检测方法都能得到很好的效果,将其称为典型图片。但有的图片不然,这些图片用所提出的基于 GBVS 的人脸检测方法要比 TBFD 的人脸检测方法好得多,称之为非典型图片。图 5(a)、(b)各列出了几张典型和非典型的图片,以及采用所提出的算法和 TBFD 方法的人脸检测效果。



(a)典型图片 (b)非典型图片  
图 5 所提出算法与 TBFD 方法检测得到的典型与非典型图片

从图 5(a)、(b)中可以看出,在处理正面人脸图片时,所提出算法和 TBFD 方法得到的人脸检测效果相差无几,都准确提取出了图片中的目标人脸区域。但是如果图片中人脸存在姿态的变化,因为人的两眼之间的距离  $D$  随姿态发生改变,TBFD 方法中以  $D$  和  $2D$  确定的矩形区域不能再截取到完整的目标人脸,所以图 5(b)中 TBFD 方法检测到的人脸信息丢失严重,单从人脸检测结果判断已经不能辨识出这个人了,因此检测效果不如所提出的算法好。此外,还可以从算法消耗时间上比较两种算法的优劣。算法消耗时间从输入一张图片开始计时直到目标人脸被检测到停止计时。通过实验得到所提出算法和 TBFD 方法消耗的时间结果如表 1 所示。

表 1 所提出算法与 TBFD 方法检测时间比较

算法	检测时间/s
所提出算法	1.27
TBFD 方法	4.80

从表 1 中可以看出,所提出算法与 TBFD 方法检测目标人脸所消耗的时间是不同的,所提出算法消耗时间明显低于 TBFD 方法。所提出算法耗费的时间在显著图获取、模板获取、模板调整、目标人脸区域获取上,因为算法运行迅速,所以处理一张图片只需 1.27 s。而 TBFD 方法消耗的时间在定位人的两只眼睛和人的反应时间上,其中人的反应时间更长,所以处理一张图片平均消耗的时间是 4.80 s。此外,TBFD 方法还存在检测方法机械,检测出的目标人脸图像分辨率较低等问题。综合考虑两种算法在人脸检测效果与消耗时间两个方面的性能,前者算法更有效。

3.3 所提出算法在非控人脸识别中的应用

从 LFW 数据库中选出有 20 张以上(包括 20 张)图片的人作为实验数据,总共 62 类人,共 3 023 张图片。把这 3 023 张图片作为输入,利用所提出算法提取目标人脸区域得到另一个新的人脸数据库(记为 LFW-GBVS)。随机抽取数据库中每类人的 10 张图片作为训练样本,每类人剩下的图片留作测试,即每个人脸数据库共 620 个训练样本,2 403 个测试样本。

采用 HOG 算法<sup>[14]</sup>对两个人脸数据库进行特征提取,并用 SVM 分类器<sup>[15]</sup>进行分类,得到的实验结果如表 2 所示。

表 2 LFW 与 LFW-GBVS 人脸库的识别性能比较

数据库	识别率/%	识别时间/s (识别+分类)
LFW	31.79	211.7
LFW-GBVS	72.53	57.8

从表 2 可以看出,LFW-GBVS 人脸库的识别性能得到很大提升,识别率提高到 72.53%,是 LFW 人脸库识别率的 2 倍;识别时间为 57.8 s,降为原来的 1/4。因此,所提出算法在非控人脸识别方面是实用且有效的,这为今后研究非控环境下的人脸识别提供了技术支持。

4 结束语

为解决复杂环境下因受目标背景杂乱等因素影响人脸识别分类效果不理想的问题,提出了一种基于视觉显著性的复杂环境下的人脸检测方法。该方法根据 GBVS 算法获得人脸显著图,根据显著图选择阈值进行阈值分割得到二值模板,对模板进行调整直到最优,找到模板中的目标人脸所在区域,将目标人脸区域与

输入图片匹配检测到人脸目标。实验结果表明,所提出的方法能有效摒除杂乱背景干扰,显著提高非控环境下人脸识别的准确率,且与其他人脸检测方法相比,不仅准确度高,而且算法运行快速,适用于非控环境下的人脸检测问题。

参考文献:

[1] Yun J U, Lee H J, Paul A K, et al. Face detection for video summary using illumination-compensation and morphological processing[J]. Pattern Recognition Letters, 2009, 30(9): 856-860.

[2] Yilmaz S. Gray level based face detection using template face mask and L1 norm[J]. IJWA, 2010, 2(4): 243-249.

[3] Cheng J. A multi-template combination algorithm for protein comparative modeling [J]. BMC Structural Biology, 2008, 8(1): 18.

[4] Du C, Zhu H, Luo L M, et al. Face detection in video based on AdaBoost algorithm and skin model[J]. Journal of China Universities of Posts and Telecommunications, 2013, 20(13): 6-9.

[5] 姚 坤. 人脸检测技术综述[J]. 电子技术与软件工程, 2014(13): 122.

[6] 廖广军. 复杂条件下的人脸检测与识别应用研究[D]. 广州: 华南理工大学, 2014.

[7] 叶海燕. 复杂条件下的人脸检测与识别应用研究[J]. 新课程, 2015(1): 6-7.

[8] Itti L, Koch C. Computational modelling of visual attention [J]. Nature Reviews Neuroscience, 2001, 2(3): 194-203.

[9] Harel J, Koch C, Perona P. Graph-based visual saliency [C]//Advances in neural information processing systems. [s. l.]: [s. n.], 2006: 545-552.

[10] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[C]//Proc of IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2009: 1597-1604.

[11] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10): 1915-1926.

[12] 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.

[13] Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments [R]. Massachusetts: University of Massachusetts, Amherst, 2007.

[14] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Proc of IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2005: 886-893.

[15] Cortes C, Vapnik V. Support vector machine [J]. Machine Learning, 1995, 20(3): 273-297.