

基于生物视觉特征和视觉心理学的视频显著性检测算法

方志明 崔荣一 金璟璇

Video saliency detection algorithm based on biological visual feature and visual psychology theory

Fang Zhi-Ming Cui Rong-Yi Jin Jing-Xuan

引用信息 Citation: *Acta Physica Sinica*, 66, 109501 (2017) DOI: 10.7498/aps.66.109501

在线阅读 View online: <http://dx.doi.org/10.7498/aps.66.109501>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2017/V66/I10>

您可能感兴趣的其他文章

Articles you may be interested in

基于空间自适应非凸正则项全变差相干斑噪声抑制

[Spatially adapted total variation model with nonconvex regularizer based speckle reduction](#)

物理学报.2014, 63(18): 189501 <http://dx.doi.org/10.7498/aps.63.189501>

一种抑制合成孔径雷达图像相干斑的各向异性扩散滤波方法

[An anisotropic diffusion filtering method for speckle reduction of synthetic aperture radar images](#)

物理学报.2014, 63(17): 179502 <http://dx.doi.org/10.7498/aps.63.179502>

基于纹理显著性的微光图像目标检测

[Low light level image target detection based on texture saliency](#)

物理学报.2014, 63(6): 069501 <http://dx.doi.org/10.7498/aps.63.069501>

基于图像熵的各向异性扩散相干斑噪声抑制

[Speckle reduction by image entropy anisotropic diffusion](#)

物理学报.2013, 62(9): 099501 <http://dx.doi.org/10.7498/aps.62.099501>

一种改进的用于合成孔径雷达图像相干斑抑制的双边滤波参数配置方法

[A modified method to configure the parameters of the bilateral filtering for synthetic aperture radar image speckle reduction](#)

物理学报.2012, 61(18): 189501 <http://dx.doi.org/10.7498/aps.61.189501>

基于生物视觉特征和视觉心理学的视频显著性检测算法*

方志明 崔荣一 金璟璇†

(延边大学工学院计算机科学与技术系, 智能信息处理实验室, 延吉 133002)

(2016年11月18日收到; 2017年2月18日收到修改稿)

提出了一种空域和时域相结合的视频显著性检测算法. 对单帧图像, 受视觉皮层层次化感知特性和 Gestalt 视觉心理学的启发, 提出了一种层次化的静态显著图检测方法. 在底层, 通过符合生物视觉特性的特征图像(双对立颜色特征及亮度特征图像)的非线性简化模型来合成特征图像, 形成多个候选显著区域; 在中层, 根据矩阵的最小 Frobenius-范数(F-范数)性质选取竞争力最强的候选显著区域作为局部显著区域; 在高层, 利用 Gestalt 视觉心理学的核心理论, 对在中层得到的局部显著区域进行整合, 得到具有整体感知的空域显著图. 对序列帧图像, 基于运动目标在位置、运动幅度和运动方向一致性的假设, 对 Lucas-Kanade 算法检测出的光流点进行二分类, 排除噪声点的干扰, 并利用光流点的运动幅度来衡量运动目标运动显著性. 最后, 基于人类视觉对动态信息与静态信息敏感度的差异提出了一种空域和时域显著图融合的通用模型. 实验结果表明, 该方法能够抑制视频背景中的噪声并且解决了运动目标稀疏等问题, 能够较好地复杂场景中检测出视频中的显著区域.

关键词: 显著性检测, 非线性简化, Gestalt 视觉心理学, Lucas-Kanade 算法

PACS: 95.75.Mn, 42.30.Tz, 42.30.Va

DOI: 10.7498/aps.66.109501

1 引言

近几十年, 随着神经心理学和神经解剖学的发展, 视觉显著性逐步成为计算机视觉的热点. 视频的显著性区域检测可用于简化复杂视频场景, 过滤与任务相关性较弱的信息, 保留与当前任务相关度较高的信息^[1]. 自动完成视频的显著性区域检测成为视频内容感知^[2]、视频编码^[3,4]、无人驾驶^[5]和视频摘要等^[6-8]任务的重要基础任务.

显著性检测方法主要分为四类: 基于像素的检测模型、基于块的检测模型、基于频率的检测模型和基于低秩重建的检测模型. 基于像素的检测模型以像素为基本单位, 在不同特征下进行图像显著性的计算. Itti 等^[9]模拟视觉系统的神经机制, 提

出了一种基于多尺度图像特征融合的显著性检测方法. 在此基础上, Itti 和 Koch^[10]增加了运动特征, 从而将该模型扩展到视频中的显著性检测. 该类方法计算量较大, 对噪声敏感, 受图像复杂度的影响较大. 基于块的方法, Cheng 等^[11]将原图像过分割成多个区域, 然后提取颜色直方图和空间位置来计算区域的显著度, 从而提取显著性目标. 该类方法的检测结果取决于块分割的准确性. 此外, Liu 和 Wang^[12]基于中心-周围差异, 提出了一种结合局部块的变化对比和全局感知的方法, 该方法简单快速. 基于频率谱的方法, Guo 等^[13]通过实验发现图像的相位谱残差(PFT)的显著性检测方法优于幅度谱残差(SR)^[14]方法. 在此基础上, 将颜色、亮度和运动特征组成一个四元数组并提出了QPFT的方法, 同时可用于视频的显著性检测. 该

* 吉林省科技发展计划项目(批准号: 20140101186JC)资助的课题.

† 通信作者. E-mail: 1537161104@qq.com

类方法不仅受背景复杂程度干扰严重,且检测出的显著区域为一系列分散的点,不利于显著区域的完整分割与提取.此外,该方法运动显著性检测结果受时间间隔参数的影响较大且无法衡量运动显著性.基于低秩重建的方法,Zhu和Wang^[15]以及Tao等^[16]将图像表示为低秩部分(非显著性部分)和稀疏部分(显著性部分),通过低秩矩阵恢复得到显著图.Xue等^[17]通过在 $X-t$ 和 $Y-t$ 方向将低秩矩阵分解的方法来提取视频中的显著性目标.此类方法检测结果受参数选取及背景复杂度的影响大且只保证了检测结果是稀疏的,并不意味检测结果是显著的.

此外,从目标和背景的分割角度考虑显著性目标提取,马兆勉和陶纯堪^[18]以及金左轮等^[19]认为目标前景在纹理特征上相对于背景更加光滑.金左轮等利用纹理粗糙度来计算图像的显著性,由于缺少颜色特征,导致彩色目标漏检.纹理特征与颜色特征具有相关性,纹理粗糙程度在颜色空间分布上呈现出连贯性和集中性等特点.因此,在自然图像中,纹理特征的部分信息可以由颜色特征来体现.从信息论的角度考虑,人造目标的出现会引起自然场景的统计特性发生变化^[20].许元男等^[21]利用Wigner-Ville分布和Rényi熵来计算显著图.由于缺少空间分布和颜色分布等先验知识,该方法只能应用于灰度图像,且检测结果的完整性较差.

除了以上方法及以上方法的改进算法,近两年出现了大量利用深度学习^[22-24]的方法来做显著性检测的文献,其原理为通过构建、训练神经网络来生成显著图.此类算法需要庞大的数据集和手工标注数据集,计算量大且不宜用于视频显著性的检测.

以上文献中视频显著性检测都是在图像显著性检测算法中,将使用帧差法检测出的运动目标^[13]作为显著性运动目标,因此检测结果不理想.文献^[25]总结相关文献发现视频显著性检测大多是将视频显著值归结为先计算两种显著值即运动显著值和静态显著值,然后融合两者结果.与此同时,该文献采用结合滤波器的金字塔光流法进行动态显著性估计,该方法计算量较大,且金字塔模型不适用于分辨率较低的视频.目前,视频运动目标检测主要有帧差法^[26]、背景差分法^[27]、光流法^[28]

等三种方法.帧差法算法简单快速,但不能检测出运动幅度的大小.背景差分法过度依赖背景模型的准确性.光流法主要优点是能够用于计算各像素的运动幅度和方向,主要缺点是检测结果受噪声的干扰严重.

Elazary和Itti^[29]通过大量标注的图像数据测试,认为场景中的显著性区域受低层视觉属性的影响较大.针对上述文献存在的问题,本文提出了一种基于双对立颜色特征、Gestalt视觉心理学和光流法的空域和时域显著性融合的显著性检测算法.在空域,显著性检测基于生物视觉特性的双对立颜色特征和亮度特征,模型具有层次结构,自底向上,图像逐渐简化,利用Gestalt视觉心理学主要理论使得显著目标具有一定的整体性.在时域,基于运动目标所在位置、运动幅度和运动方向一致性假设,利用Lucas-Kanade算法(简称LK光流法)^[30]并通过二分类进行降噪处理来计算运动显著性,使其适用于低分辨率复杂视频显著性检测.最后在不同颜色空间中融合了空域和时域的显著性检测结果,该模型实现了视频的时空显著性检测,实验中对低分辨视频进行测试,取得了较好的实验结果.

2 视频显著性检测框架

视觉显著性检测的原理是通过模仿人类视觉注意机制的方法来获得显著性区域.视觉显著性描述了一个目标区域在一个场景中的独特性或吸引视觉注意的能力,这种能力来自生物视觉特性或由观察者受先验知识导致的.视频显著性检测和图像显著性检测的主要区别在于视频具有运动特征.

由于视频种类繁多,目前没有一种时空显著性融合方法能够应用于所有类型的视频.本文将人眼对彩色信息比灰度信息更为敏感和人眼对运动信息比静态信息更敏感的两大特性一一对应,将视频的单帧图像静态显著性和序列帧图像显著性检测结果分别用灰度颜色模型和孟塞尔色系模型^[31]表示,提出了一种通用各个场景、基于视觉敏感度的显著性可视化的表示方法.在复杂场景视频的单个画面中既能够同时显示两种显著性的结果,又能够不致使画面过于复杂、混乱.显著性检测框图如图1所示.

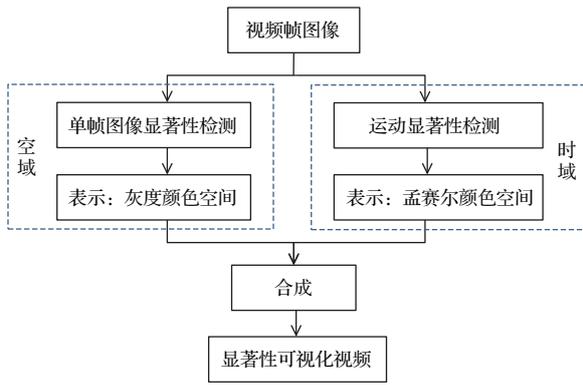


图1 视频显著性检测框架图

Fig. 1. Video saliency detection framework.

3 空域显著性检测

大脑皮层中主要负责处理视觉信息的部分是视觉皮层 (visual cortex). 人类的视觉皮层包括初级视皮层 (V1, 也称作纹状皮层 (striate cortex)) 和纹外皮层 (extrastriate cortex, 如 V2, V3 等). 作为第一个进行视觉处理的区域, V1 主要接收与外观感知有关的电信号, 响应结果进一步传导到 V2 等更高级视皮层区域进行处理.

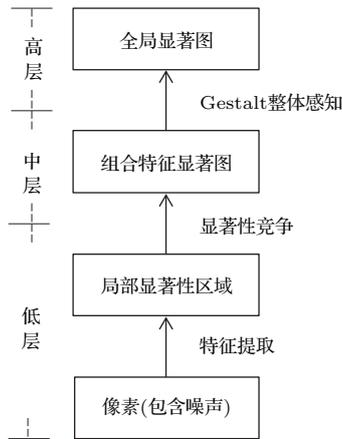


图2 图像显著性检测框图

Fig. 2. Image saliency detection block diagram.

受视皮层结构和 Gestalt 视觉心理学的启发 [32], 本文建立了具有三层结构的空域显著性检测模型, 逐层对图像进行简化, 并将各个简化结果加以组合, 使之成为一个知觉上易于处理的整体. 本文提出的图像显著性检测模型框图如图 2 所示, 图 3 为对应的模型示意图. 低层为视觉特征提取层, 用于提取底层特征, 低层图像中的各个像素, 包括了噪声点 (灰色点). 低层经过特征提取, 提取出各个特征对应的显著性区域, 从而降低图像的复杂度; 在中层, 多个特征显著性目标区域进行竞争,

得出单个特征对应的局部显著性区域; 高层为组织层, 利用 Gestalt 视觉心理学的主要理论对中层竞争胜出的局部显著区域进行整体感知, 进行目标区域的整合. 随着层次由低到高, 图像不断被简化, 形成整体. 本文所构造的模型首先将图像从像素点映射到多个区域作为候选显著性目标, 然后将单目标和邻近候选区域进行整体感知整合, 最终提取出符合视觉认知心理的显著性目标.

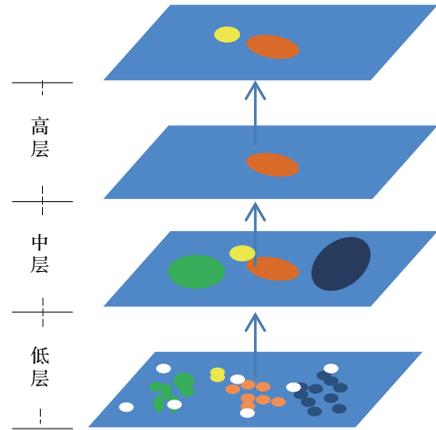


图3 (网刊彩色) 图像显著性检测示意图

Fig. 3. (color online) Image saliency detection schematic diagram.

3.1 低层-像素点到多区域

本文利用图像数据的分布特性, 自适应地合并具有相似特征的像素点, 形成与特征对应的区域群. 基于视觉对可见光颜色的敏感度差异, 本文采用双对立颜色对 [33] 及亮度特征属性对应的特征图像来共同体现图像的显著性. 双对立颜色特征图像矩阵 (RG, BY) 和亮度特征图像矩阵 (I) 定义同文献 [9], 分别由 (1), (2) 和 (3) 式定义. 其中, R, G, B, Y 分别为对立色中的红、绿、蓝、黄颜色; rc, gc, bc 分别表示 rgb 颜色模型中的原图像的红、绿、蓝的颜色通道图像矩阵.

$$RG = |R - G| = \left| \left(rc - \frac{gc + bc}{2} \right) - \left(gc - \frac{rc + bc}{2} \right) \right|, \quad (1)$$

$$BY = |B - Y| = \left| \left(bc - \frac{rc + gc}{2} \right) - \left(\frac{rc + gc}{2} - \frac{|rc - gc|}{2} - bc \right) \right|, \quad (2)$$

$$I = (rc + gc + bc)/3, \quad (3)$$

$$Sim_F = \underbrace{F * F \cdots * F}_S \quad (S = 1, 2, 3). \quad (4)$$

(4)式中的 Sim_F 表示特征图像矩阵 $F(F = RG, BY, I)$ 经过简化的结果, S 为简化参数($S = 1, 2, 3$), 运算符 $*$ 表示矩阵对应元素相乘. 该操作可以减少图像中对比度弱的区域. 若简化参数 S 取不同的等级值, 则图像中对比度弱的区域减少的程度就不同. 图5为图4所示的原始彩色图像的灰度图像对应的灰度分布直方图, 从左到右依次为灰度图的简化参数 S 从1增至3的直方图. 由直方图的分布情况可以看出, 随着简化等级的升高, 直方图由灰度分布较均衡的情况逐渐变为两端分布的情况. 由图6可以看出, 随着简化等级 $S(S = 1, 2, 3)$ 的提高, 亮度特征图像(第一列)中的较亮区域(警示牌等区域)进一步被突出, 红绿

特征图像(第二列)中的红绿差异被突显(红色三角形), 而蓝黄特征图像(第三列)中的偏黄区域进一步被凸显(孩子的衣服). 因此, 随着简化等级的提高, 特征图像中对应特征属性的区域逐步被凸显, 而其他特征属性对应的区域被合并和抑制.



图4 (网刊彩色) 原始图像
Fig. 4. (color online) Original image.

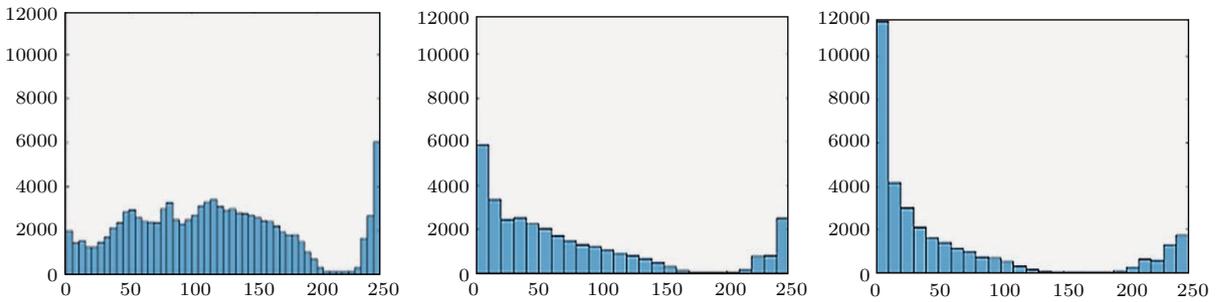


图5 简化等级-灰度分布变化图
Fig. 5. Simplified level gray distribution map.

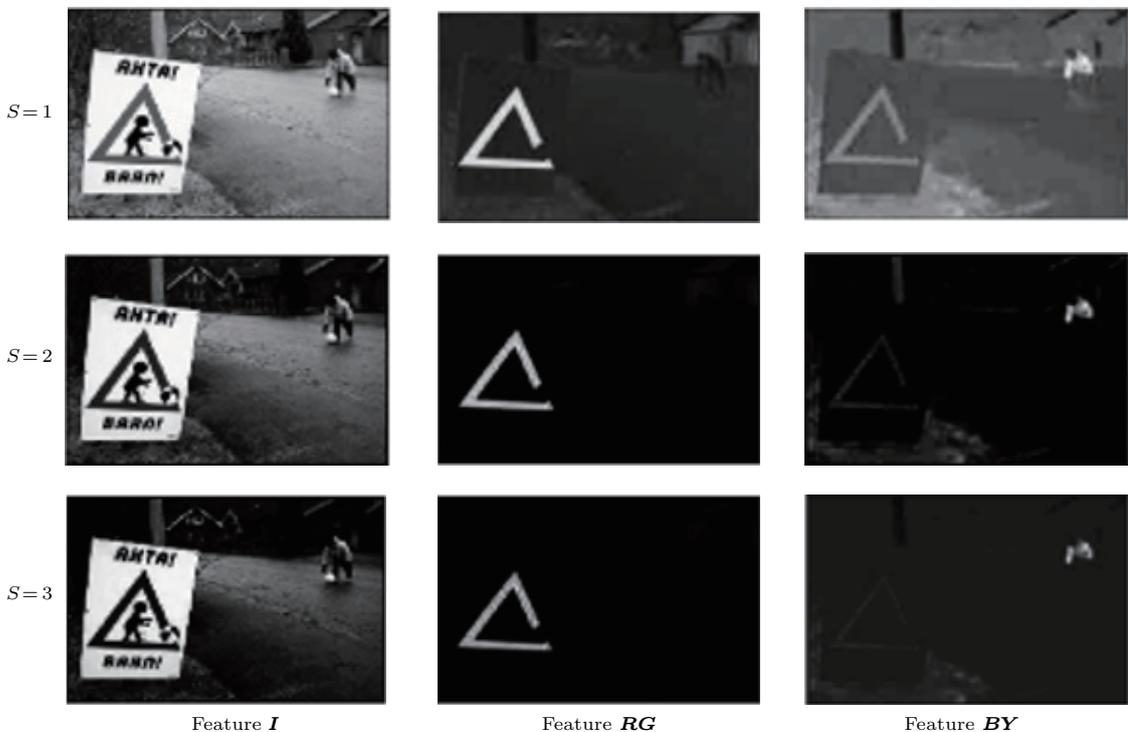


图6 特征-简化等级示意图
Fig. 6. Feature image with different simplified levels.

3.2 中层-组合特征竞争

中层为竞争层, 图像对应的三个特征对显著区域形成的贡献进行竞争. 传统算法对多特征显著图进行线性叠加, 而 Gestalt 理论的基本出发点是, 整体不能用部分之和去理解, 因此本文采用多特征图像非线性组合并利用最小 F -矩阵范数进行约束得到竞争力最强的局部显著性区域. 矩阵 A 的 F -范数由 (5) 式表示, (6) 式表示特征图像的非线性组合方式由参数 $\theta = (a, b, c, d, e, g)$ 确定. 其中, a, b, c 为简化参数, $a, b, c \in \{1, 2, 3\}$, 同 (4) 式中的 S ; d, e, g 为组合参数, $d, e, g \in \{-1, +1\}$. 组合参数取值为 -1 表示该属性下的特征区域对提取最佳显著区域起负作用, 取值为 $+1$ 表示该属性下的特征区域对提取最佳显著区域起正作用. 如 (7) 式所示, 利用

最小 F -范数来求得对应特征的显著性区域的非线性组合参数, 从而得到显著性最强的局部显著图, 并且保证了非线性组合特征对应的显著性区域足够稀疏.

$$\|A\|_F = \sqrt{\sum_i \sum_j a_{ij}^2}, \quad (5)$$

$$M_f = d * I^a + e * RG^b + g * BY^c, \quad (6)$$

$$\theta = \arg \min_{\theta} \|M_f(\theta)\|_F. \quad (7)$$

将不同简化组合参数对应的显著图按照其对应矩阵的 F -范数值从小到大排列, 排列结果如图 7 所示, 逐行从左到右, F -范数逐渐增大. 图 7 中的第一幅图像为最小 F -范数对应的特征图像的非线性组合. 最后一幅图像为最大 F -范数对应的特征图像的非线性组合.



图 7 范数-显著性区域变化示意图
Fig. 7. Norm-saliency regional variation schematic diagram.

3.3 高层-Gestalt 视觉整体感知

Gestalt 理论明确提出: 在眼和脑的作用下, 图像不断地进行组织、简化及统一. Gestalt 的组织过程是有选择地将一些元素统一在一起, 我们能感知到它是一个完整的单位. 本文主要应用以下 Gestalt 的主要理论作为约束条件进行显著性区域

整合:

- 1) 简化 从图像中排除不重要的部分, 只保留那些绝对必要的组成部分, 从而达到视觉的简化;
- 2) 主体和背景 场景的特点会影响视觉系统对场景中的主体和背景的解析, 当一个小物体(或色块)与更大的物体重叠时, 我们倾向于认为小的物体是主体而大的物体是背景;

3) 整体与局部 由知觉活动组织成的经验中的整体, 在性质上不等于部分的简单线性叠加;

4) 接近 指单独的视觉单元无限贴近, 以至于它们彼此黏连, 在视觉上就形成了一个较大、统一的整体;

5) 闭合 封闭的图形往往看成一个整体.

其中约束条件 1) 和约束条件 3) 已在低层和中层体现. 低层中的特征图像通过简化参数进行简化; 在中层, 通过对特征图像进行非线性组合来构造局部显著性区域.

在高层, 根据 Gestalt 约束条件 4) 和约束条件

5) 判断各个特征图像的最简化图像 ($S = 3$) 经过 mean shift 聚类的显著性区域块的外接矩形框 S_i 与最小 F -范数对应的局部显著区域块外接矩形框 S_{\min} 是否有交集, 并按就近原则合并. 合并条件为: 两个外接矩形框区域存在重叠部分 (满足理论第 4, 5 条), 如 (8) 式所示,

$$S_i \cap S_{\min} \neq \emptyset. \quad (8)$$

合并停止条件为: 合并满足重合条件的显著性区域面积大于背景面积或接触图像边界 (满足约束条件 2)). 静态图像的显著性检测流程示意图见图 8.

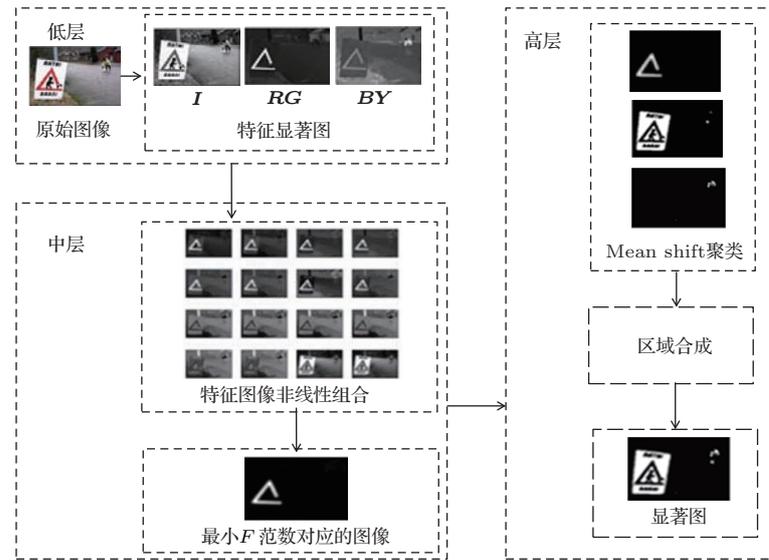


图 8 显著性检测流程示意图

Fig. 8. Saliency detection process schematic diagram.

4 时域显著性检测

Lucas-Kanade 即 LK 光流法应用于输入图像的一组特征点上时比较方便, 因此被广泛应用于稀疏光流场. 由于真实场景视频的背景复杂度高, 噪声点多, LK 光流法需要计算每一个像素的光流, 在实际应用中的检测结果不是很好. Bouquet [34] 采用图像金字塔的方法来实现运动的跟踪, 但由于视频分辨率低, 特征点精确度低, 导致光流点过稀疏情况发生, 运动目标缺失严重, 不适用于显著性运动目标的运动幅度可视化. 同多数稀疏光流法, 在运动目标与环境相似时, 利用特征点的光流法容易出现检测失败和光流点稀疏的问题. 为了解决光流点稀疏问题, 本文基于传统的 LK 光流法检测结果, 根据光流点的位置、运动方向和运动幅度值特征,

对光流点进行二分类, 将其分为噪声点和运动目标点. 算法步骤如图 9 所示.

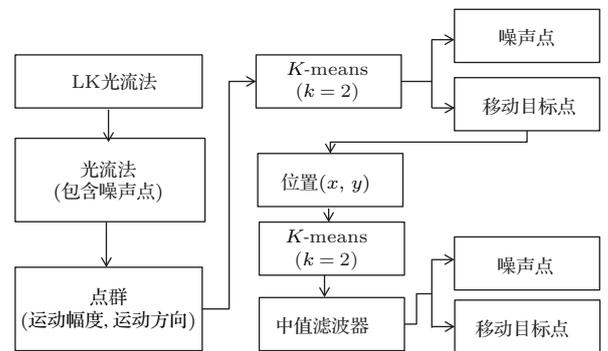


图 9 LK 改进算法框图

Fig. 9. Improved LK algorithm block diagram.

图 10 中的 (a1) 和 (c1) 为测试视频, 其画面分辨率较低, 背景复杂度较高. 其中, 图 10 (a1) 为场

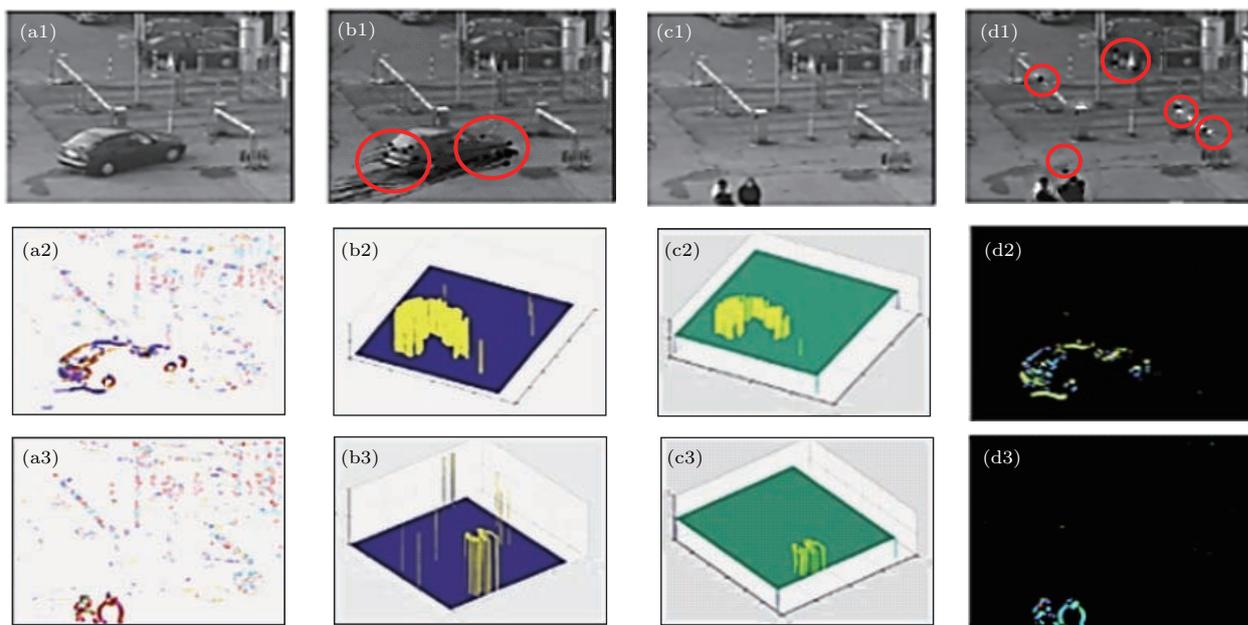


图 10 (网刊彩色) 基于 LK 光流法的改进方法与金字塔光流法得对比 (a1) 快速移动, (b1) 金字塔光流法, (c1) 缓慢移动, (d1) 金字塔光流法; (a2), (a3) LK 光流法, (b2), (b3) K -均值聚类, (c2), (c3) 中值滤波, (d2), (d3) 本文结果

Fig. 10. (color online) comparison between the improved method based on LK optical flow method and Pyramid optical flow method: (a1) Fast motion, (b1) pyramid optical flow, (c1) slow motion, (d1) pyramid optical flow; (a2), (a3) LK optical, (b2), (b3) K -means cluster, (c2), (c3) median filter, (d2), (d3) our result.

景中运动速度快的片段中的一帧, 图 10(c1) 为场景中运动速度较慢的片段中的一帧, 图 10(b1) 和图 10(d1) 分别为用金字塔光流法检测其与前一帧的光流实验结果. 可以看出, 检测到的运动目标过于稀疏, 无法检测出较完整的个体, 不适用于视频显著性表示.

图 10(a2) 和图 10(a3) 表示传统 LK 光流法分别对图 10(a1) 和图 10(c1) 的检测结果, 可以看出传统方法虽然能够检测出较为完整的运动目标, 但对噪声极其敏感. 为了得到较为完整的运动目标, 本文基于传统的 LK 光流法, 对检测出的光流点进行聚类, 将光流点分为噪声点和运动目标点. 图 10(b2) 和图 10(b3) 为运动目标聚类结果, 可以看出, 本文的方法能够较好地将噪声点去除. 在此结果上, 基于运动目标点具有局部集中的特性, 本文采用中值滤波消除孤立点, 图 10(c2) 和图 10(c3) 为孤立点去除的结果. 从图 10(d2) 和图 10(d3) 可以看出本文的方法既能够有效保留运动目标, 又能够抑制噪声.

5 实验及可视化分析

本文从两方面进行实验结果评价. 一方面是将

低分辨率、噪声点多的视频单帧图像和静态自然图像的显著性检测结果同其他经典方法的结果进行可视化对比, 并利用 MSRA1000 数据库 ECSSD 数据库^[35] 对本文方法与其他方法进行对比分析; 另一方面结合光流法对视频多帧图像显著性检测结果进行可视化评价. 实验软件为 Matlab2015a, 硬件条件为 4 GB RAM 的 Intel Xeon CPU E5-2603.

5.1 单帧/静止图像实验评价

图 11 为视频单帧图像由本文方法进行显著性检测的结果, 背景复杂程度从左到右依次增加. 图 11 中的子图 (a) 为交通视频单帧图像及对应的显著性检测结果, 图像中用于辅助驾驶的标志性目标均能较好地被检测出, 检测结果符合工程需求; 图 11 (b) 为室内视频单帧图像及其显著性检测结果, 其中颜色鲜艳的目标被成功检测出来且目标较为完整, 如桌面上的红色袋子, 枕头和手臂, 红色和黄色的魔方平面等. 为了进一步验证本文方法的准确性和鲁棒性, 将本文方法与频率调整 (FT), 直方图反差 (HC), 区域反差 (RC), LC 等方法进行了可视化对比.

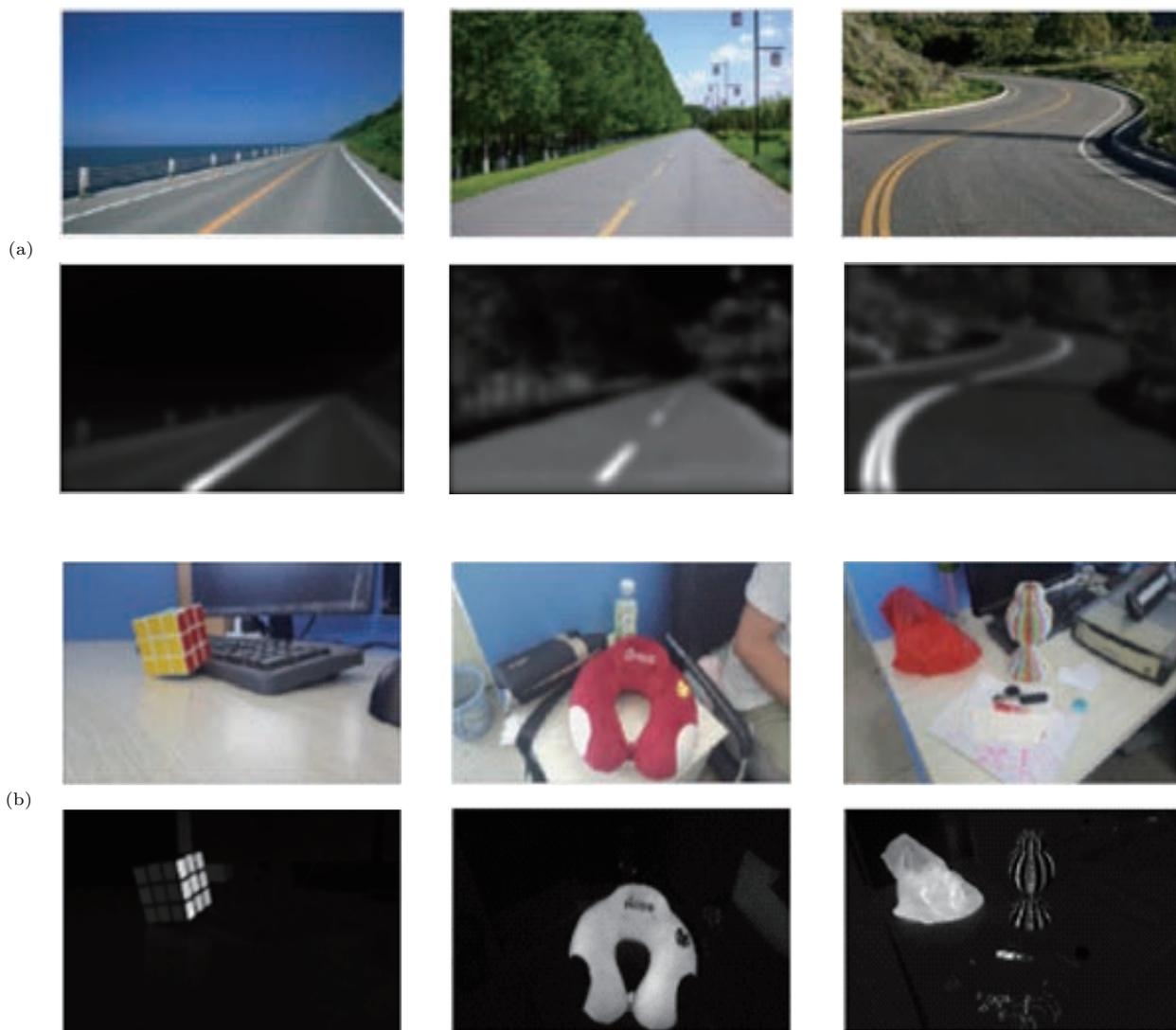


图 11 (网刊彩色) 室外室内场景视频单帧图像显著性检测结果示意图 (a) 交通视频单帧图像; (b) 室内视频单帧图像

Fig. 11. (color online) Single frame image saliency detection result diagram of outdoor and indoor scene video: (a) Results of single frame image detection in traffic scene video; (b) results of single frame image detection in indoor scene video.

图 12 是对 MSRA1000 数据库利用不同方法进行显著性检测的结果示例, 对于复杂背景的图像, 频率分析的方法 (如 SR) 检测结果受背景干扰严重. 由实验结果可以看出, 以上方法按照抗背景干扰能力的排序为: Ours>RC>HC>FT>LC>SR, 本文方法受复杂背景的影响最小. 相对于其他方法, 本文方法不仅成功地检测出目标且受背景干扰最小, 而且检测出的显著性区域更接近手工标注 (GT) 的区域且检测结果边界清晰, 易于显著性目标分割.

图 13 是对 ECSSD 数据库利用不同方法进行显著性检测的结果示例, 该数据库的显著性目标所在的背景相对于 MSRA1000 数据库更为复杂, 而且显著性目标的个体数量和颜色组成上更具有多样性, 给显著性检测带来极大的挑战. 第 5 行、第 7 行

和第 8 行的图像受背景和自身复杂的目标颜色和纹理构成的影响, 检测结果较差, 仅本文方法可以将显著性目标与背景分离. 经 ECSSD 数据库检测, 对背景复杂的图片, SR 方法检测结果最差, 本文方法抑制背景复杂程度干扰的能力最强, 检测出的目标较为完整且易于分割, 有利于进一步的视频显著性目标分割、视频压缩等工作.

本文方法与其他典型方法 (CH, GS-SP, LR, SF, CB, RC, HC, RA, FT 及 IT) 的实验结果^[36]及 SR, PQFT^[14]的实验结果进行量化对比, 选取不同阈值得到不同方法的 PR (precision-recall) 曲线. 图 14 中各个子图的纵轴和横轴分别代表查准率 P (precision) 和召回率 R (recall). 图 14 (a) 的两幅图为不同方法在 MSRA1000 数据库的测试

结果, 图 14(b) 为不同方法在 ECSSD 数据库的测试结果. 从图 14 可以看出, 本文方法的检测结

果优于其他文献方法的实验结果, 具有较高的准确率.



图 12 (网刊彩色) MSRA1000 数据库显著性检测示例

Fig. 12. (color online) Results of MSRA1000 database saliency detection.

5.2 视频多帧实验可视化分析

为了验证本文算法的效果, 本文算法对多帧视频进行了显著性实验, 如表 1 所列, 选取了 3 个室内

视频、2 个室外视频, 且视频中的运动均为非刚体运动, 所有视频都存在噪声和画面的局部抖动. 视频 1、视频 2 和视频 3 为室内视频, 其中视频 1 到视频 3 的背景复杂度依次增高.

表 1 实验视频简介

Table 1. Test video introduction.

序号	环境	分辨率	场景描述	复杂度
1	室内	352 × 288	人从移动到静止, 画面较为模糊	一般
2	室内	432 × 240	挥动手臂, 手上拿有颜色显著的红盒	中等
3	室内	352 × 288	人向前移动, 移动衣服目标颜色不显著, 桌子颜色显著	复杂
4	室外	360 × 240	移动的火车和前行的两个人 (其中一个衣着较为显著)	复杂
5	室外	320 × 240	树叶抖动, 两个人在行走, 其中一个人所处位置的背景复杂	复杂



图 13 (网刊彩色) ECSSD 数据库显著性检测示例

Fig. 13. (color online) Results of ECSSD database saliency detection.

表 1 中对应序号视频的静态显著性检测结果和运动显著性检测结果如图 15 所示. 图 15 的第一列表示视频序列中的一帧, 第二列表示该帧的静态显著性, 第三列表示该帧相对于前一帧的运动显著性检测结果. 实验表明, 本文的方法能够有效提取视频单帧的颜色显著性区域和运动显著性区域, 且检测结果受复杂背景的干扰较小, 提取显著区域的效果较好. 经过不同低分辨率视频和图像数据库的验证, 本文的静态显著性检测方法能够适用于多种类型视频的显著性检测. 静态显著性和动态显著性融

合结果示例如图 16 所示, 人由图 16(a) 的移动状态变为图 16(b) 的静止状态. 图 16(a) 中彩色部分为运动显著性区域, 灰色部分为静态显著性区域, 随着运动的停止, 图 16(b) 中彩色区域消失.

5.3 视频多帧实验定量分析

统计包括表 1 中的 5 个低分辨率视频的实验结果数据, 总计 10 个低分辨率视频, 每个视频随机选取 125 帧, 共 1250 帧的视频画面图像. 经 10 个人进行手工标注, 共同标注次数超过 5 次的区域记为有效

标注区域, 总计 4386 个静态显著性区域, 973 个动态显著性区域. 利用矩形框手工标注单帧图像中的静态显著性区域和运动显著性区域. 实验结果中的标记区域包括: 正确检测结果和错误检测结果. 手工标注区域但实验未标记区域为漏检区域. 标记区域矩形框与手工标注区域矩形框的重叠面积达到手工标注框的面积 60% 以上才记为正确标记, 否则为错误标记 (虚报区域). 表 2 的数据表明, 本文的算法有较高的准确率, 较低的漏检率和错误率.

表 2 检测结果数据统计
Table 2. Statistic data of the test result.

		类别	数量	百分比/%
静态显著性	标记	正确	3962	90.33
		错误	371	8.46
	未标记	漏检	424	9.67
动态显著性	标记	正确	887	91.16
		错误	35	3.6
	未标记	漏检	86	8.84

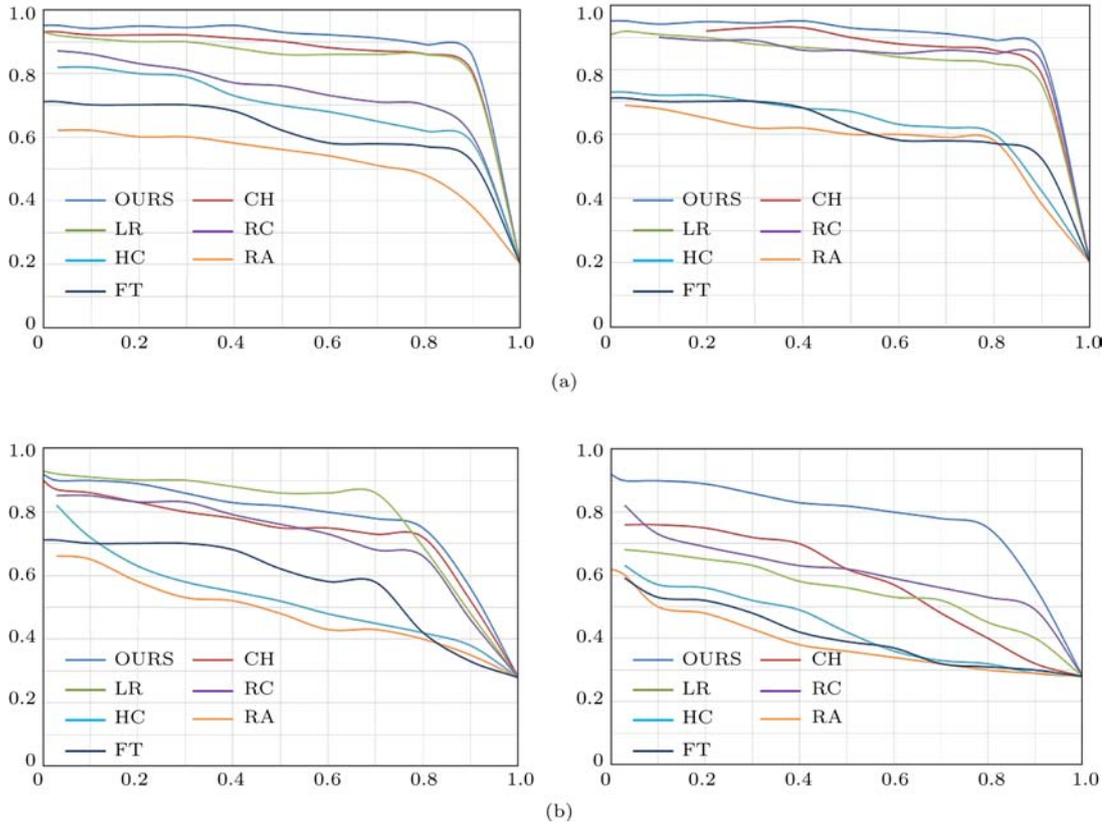


图 14 (网刊彩色) 不同方法在不同数据库的性能指标对比 (a) MSRA1000 数据库测试结果; (b) ECSSD 数据库测试结果
Fig. 14. (color online) Comparison of test results of different methods using different databases: (a) Test results of MSRA1000 database; (b) test results of ECSSD database.

6 结 论

本文利用双对立颜色特征和 Gestalt 视觉心理学理论作为约束条件, 基于层次结构的方法来检测图像显著性区域, 该方法降低了纹理特征带来的干扰, 检测出的显著性区域较为完整且符合视觉的基本规律. 在此之上, 结合改进的光流法, 对视频帧中的运动信息进行检测. 最后将图像的显著性和运动显著性利用不同的色系同时展现在一个画面中, 使融合后的画面更清晰. 与传统方法相比, 基于生物视觉特性的方法对不同特征得到的显著性

区域进行简单线性叠加, 未考虑视觉心理学相关的因素, 不符合普遍的视觉规律. 而基于数学计算方法 (如计算对比度和频率分析的方法) 未考虑生物视觉感知特性, 导致检测结果未能较好地与生物视觉特性感知结果相符. 本文的方法模仿生物视觉特性, 从底层提取相关特征的显著性区域, 同时利用高层先验, 即视觉心理学的相关成果, 对不同特征对应的显著性区域进行操作, 使通过生物特征计算得到的结果满足视觉心理学的相关理论. 实验结果表明, 本文方法的图像和视频显著性检测结果与手工标注区域基本一致.



图 15 (网刊彩色) 视频显著性检测结果 (第一列为视频单帧图像、第二列为静态显著性检测结果, 第三列为动态显著性检测结果)
 Fig. 15. (color online) Video saliency detection results (the first column for the video single frame image and second columns for the static saliency detection results, the third column for the dynamic saliency detection results).

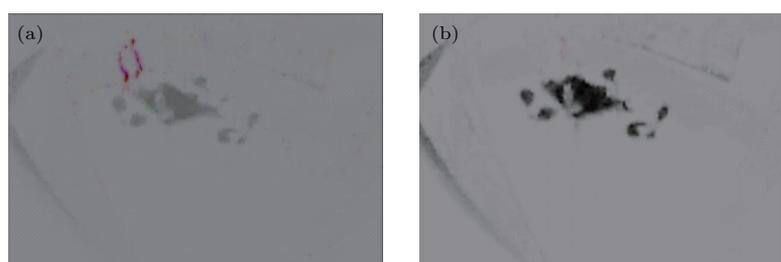


图 16 运动状态改变的视频显著性检测示例 (a) 运动状态; (b) 静止状态
 Fig. 16. Video saliency detection results of motion state changes: (a) Motion state; (b) static state.

人的视觉还受先验知识的影响, 针对不同场景, 静态显著性应当考虑更多的先验知识, 如交通场景中应当添加人脸识别、人的个体识别、交通警示牌识别等技术; 运动显著性应引入更多帧的运动情况, 以确定视频中运动显著性较强的时间段. 下一步工作将考虑以上因素, 对算法做进一步改进和

完善.

参考文献

[1] Borji A, Sihite D N, Itti L 2015 *IEEE Trans. Image Process.* **24** 5706

- [2] Cichy R M, Pantazis D, Oliva A 2016 *Cerebral Cortex* **26** 3563
- [3] Li Z C, Qin S Y, Itti L 2011 *Image Vision Comput.* **29** 1
- [4] Wu G L, Fu Y J, Huang S C, Chen S Y 2013 *IEEE Trans. Image Process.* **22** 2247
- [5] Franke U, Pfeiffer D, Rabe C, Knoeppel C, Enzweiler M, Stein F, Herrtwich R 2013 *Proceedings of IEEE Conference on Computer Vision* Sydney, Australia, December 1–8, 2013 p214
- [6] Ma Y F, Hua X S, Lu L, Zhang H J 2005 *IEEE Trans. Multimed.* **7** 907
- [7] Ejaz N, Mehmood I, Baik S W 2014 *Comput. Elec. Engr.* **40** 993
- [8] Evangelopoulos G, Zlatintsi A, Potamianos A, Maragos P 2013 *IEEE Trans. Multimed.* **15** 1553
- [9] Itti L, Koch C, Niebur E 1998 *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 1254
- [10] Itti L, Koch C 2001 *Nat. Rev. Neurosci.* **2** 194
- [11] Cheng M M, Zhang G X, Mitra N J, Huang X, Hu S M 2011 *Proceedings of Computer Vision and Pattern Recognition* Colorado Springs, November 15–18, 2011 p409
- [12] Liu J, Wang S 2015 *Neurocomputing* **147** 435
- [13] Guo C, Ma Q, Zhang L 2008 *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Anchorage, Alaska, January 17–18, 2008 p1
- [14] Hou X D, Zhang L Q 2007 *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Minneapolis, Minnesota, June 19–21, 2007 p18
- [15] Zhu Z, Wang M 2016 *J. Comput. Appl.* **36** 2560
- [16] Tao D, Cheng J, Song M 2016 *IEEE Trans. Neur. Netw. Lear. Syst.* **27** 1122
- [17] Xue Y W, Guo X J, Cao X C 2012 *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* Kyoto, Japan, March 25–30, 2012 p1485
- [18] Ma Z M, Tao C K 1999 *Acta Phys. Sin.* **48** 2202 (in Chinese) [马兆勉, 陶纯堪 1999 物理学报 **48** 2202]
- [19] Jin Z L, Han J, Zhang Y, Bo L F 2014 *Acta Phys. Sin.* **63** 069501 (in Chinese) [金左轮, 韩静, 张毅, 柏连发 2014 物理学报 **63** 069501]
- [20] Wu Y Q, Zhang J K 2010 *Acta Phys. Sin.* **59** 5487 (in Chinese) [吴一全, 张金矿 2010 物理学报 **59** 5487]
- [21] Xu Y N, Zhao Y, Liu L P, Zhang Y, Sun X D 2010 *Acta Phys. Sin.* **59** 980 (in Chinese) [许元男, 赵远, 刘丽萍, 张宇, 孙秀冬 2010 物理学报 **59** 980]
- [22] Wang X, Ma H, Chen X 2016 *Proceedings of International Conference on Image Processing* the Phoenix Convention Centre, Phoenix, Arizona, USA, September, 2016 p25
- [23] He S, Lau R W, Liu W 2015 *Int. J. Comput. Vision* **115** 330
- [24] Li H, Chen J, Lu H 2017 *Neurocomputing* **226** 212
- [25] Huang Y 2016 *M. S. Thesis* (Beijing: Institute of Optoelectronic Technology) [黄焯 2016 硕士学位论文 (北京: 中国科学院)]
- [26] Paragios N, Deriche R 2000 *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 266
- [27] Tsai D M, Lai S C 2009 *IEEE Trans. Image Process.* **18** 158
- [28] Barron J L, Fleet D, Beauchemin S S 1994 *Int. J. Comput. Vision* **12** 43
- [29] Elazary L, Itti L 2008 *J. Vision* **8**
- [30] Lucas B D, Kanade T 1981 *Proceedings of International Joint Conference on Artificial Intelligence* Vancouver, BC, Canada, August, 1981 285
- [31] Baker S, Scharstein D, Lewis J P, Roth S, Black M J, Szeliski R 2007 *Proceedings of IEEE International Conference on Computer Vision* Rio de Janeiro, Brazil, October 14–21, 2007 p92
- [32] Koffka K 1935 *Principles of Gestalt Psychology* (London: Lund Humphries)
- [33] Mullen K T 1985 *J. Phys.* **359** 381
- [34] Gary B, Adrian K 2008 *Learning OpenCV* (America: O'Reilly Media) pp356–370
- [35] Shi J, Yan Q, Xu L, Jia J 2016 *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 1
- [36] Li X, Li Y, Shen C H, Dick A, Hengel 2013 *Proceedings of Computer Vision* Sydney, NSW, Australia, December 8, 2013 p3328

Video saliency detection algorithm based on biological visual feature and visual psychology theory*

Fang Zhi-Ming Cui Rong-Yi Jin Jing-Xuan[†]

(Intelligent Information Processing Laboratory, Department of Computer Science and Technology, Faculty of Engineering,
Yanbian University, Yanji 133002, China)

(Received 18 November 2016; revised manuscript received 18 February 2017)

Abstract

In order to solve the problems of video saliency detection and poor fusion effect, a video saliency detection model and a fusion model are proposed. Video saliency detection is divided into spatial saliency detection and temporal saliency detection. In the spatial domain, inspired by the properties of visual cortex hierarchical perception and the Gestalt visual psychology, we propose a hierarchical saliency detection model with three-layer architecture for single frame image. The video single frame is simplified layer by layer, then the results are combined to form a whole consciousness vision object and become easier to deal with. At the bottom of the model, candidate saliency regions are formed by nonlinear simplification model of the characteristic image (dual color characteristic and luminance characteristic image), which is in accordance with the biological visual characteristic. In the middle of the model, the candidate regions with the strongest competitiveness are selected as the local salient regions according to the property of matrix minimum Fresenius-norm (F- norm). At the top level of the model, the local salient regions are integrated by the core theory of Gestalt visual psychology, and the spatial saliency map is obtained. In the time domain, based on the consistency assumption of a moving object in target location, motion range and direction, the optical flow points detected by Lucas-Kanade method are classified to eliminate the noise interference, then the motion saliency of moving object is measured by the motion amplitude. Finally, based on the difference between the visual sensitivity of dynamic and static information and the difference in visual sensitivity between color information and gray information, a general fusion model of time and spatial domain salient region is proposed. The saliency detection results of single frame image and video sequence frame image are represented by the gray color model and the Munsell color system respectively. Experimental results show that the proposed saliency detection method can suppress the background noise, solve the sparse pixels problem of a moving object, and can effectively detect the salient regions from the video. The proposed fusion model can display two kinds of saliency results simultaneously in a single picture of a complex scene. This model ensures that the detection results of images are so complicated that a chaotic situation will not appear.

Keywords: saliency detection, nonlinear simplification, Gestalt visual psychology, Lucas-Kanade method

PACS: 95.75.Mn, 42.30.Tz, 42.30.Va

DOI: 10.7498/aps.66.109501

* Project supported by the Science and Technology Development Plan Foundation of Jilin Province, China (Grant No. 20140101186JC).

[†] Corresponding author. E-mail: 1537161104@qq.com