虚拟现实全景图像显著性检测研究进展综述

丁 颖[1,2],刘延伟[1],刘金霞[3],刘科栋[1],王利明[1],徐 震[1]
(1. 中国科学院信息工程研究所,北京 100864;2. 中国科学院大学网络安全学院,北京 100049;
3. 浙江万里学院,浙江宁波 315100)

摘 要: 随着虚拟现实处理技术的发展,虚拟现实全景图像的显著性检测成为近年来学术界和工业界关注的研究热点. 本文分析虚拟现实全景图像的特性,综述虚拟现实全景图像显著性检测算法的研究进展. 将已有的虚拟现实全景图像显著性检测算法进行分类、分析以及对比,本文总结了当前虚拟现实全景图像显著性检测面临的挑战,并对其发展趋势进行展望.

关键词: 虚拟现实; VR 全景图像; 显著性检测
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2019)07-1575-09
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.07.024

An Overview of Research Progress on Saliency Detection of Panoramic VR Images

DING Ying[1,2], LIU Yan-wei[1], LIU Jin-xia[3], LIU Ke-dong[1,2], WANG Li-ming[1], XU Zhen[1]
(1. *Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100864, *China*;
2. *School of Cyber Security, University of Chinese Academy of Sciences, Beijing* 100049, *China*;
3. *Zhejiang Wanli University, Ningbo, Zhejiang* 315100, *China*)

Abstract: With the development of virtual reality (VR), the saliency detection for panoramic VR image has been a hot research topic in both academic and industry worlds. After analyzing the particular characteristics of panoramic VR image, this paper summarizes the research progress of the saliency detection algorithms of panoramic VR image. Existing VR image saliency detection algorithms are firstly classified, analyzed and compared. Then, the current research challenges and future directions of VR saliency detection are discussed.

Key words: virtual reality; VR panoramic image; saliency detection

1 引言

虚拟现实(Virtual Reality, VR)全景图像是一种以360度实景图像为基础构造虚拟环境的技术. 与传统图像相比,VR 全景图像可以捕捉到更多的场景信息[1]. 通过使用头戴式显示器实时渲染不同的视角图像,观察者可以自主观看任意方向上的场景,得到身临其境的视觉体验. 基于这些特征,VR 全景图像受到人们的重视并被广泛用于娱乐、医疗、教育、影视等领域.

VR 全景图像的分辨率是传统图像的几倍,这使得 VR 全景图像的存储和传输都变十分困难. 然而,人的视觉注意机制体现了一种具有选择性的注意能力,即面对一个场景时,人类能够自动地处理感兴趣区域,而选择性地忽略不感兴趣区域. 因此,有必要对 VR 全景图像中的信息进行显著性检测,以便合理地减少 VR 全景图像中的视觉冗余信息. 对 VR 全景图像进行显著性检测,不仅可以提高 VR 全景图像的压缩效率[2],减少传输带宽[3],而且对 VR 全景图像编辑[4]起着至关重要的作用.

近年来,针对传统图像的显著性检测技术已经相对成熟,出现了很多优秀的检测模型[5,6]. 这些模型同样可以应用于 VR 全景图像. 然而,VR 全景图像具有独

收稿日期:2018-08-22; 修回日期:2018-11-28; 责任编辑:孙瑶
基金项目: 国家自然科学基金(No. 61771469)

特的观看方式,如果将传统图像的显著性检测模型直接应用于 VR 全景图像,效果并不理想. 因此,需要研究符合 VR 全景图像特性的显著性检测技术. 在最近的文献中,已经出现了一些针对 VR 全景图像显著性检测的研究工作[7],并取得了一定的进展,但仍存在一些值得深入剖析的问题. 基于这一点,本文首先分析 VR 全景图像的独特观看特性及其对视觉注意机制的影响,然后基于平面图像到 VR 全景图像显著性检测的演进过程,对现有的 VR 全景图像显著性检测算法归纳总结,最后讨论了未来 VR 全景图像与视频显著性检测的研究趋势.

2 VR 全景图像的观看特性

与传统图像不同,VR 全景图像在显示、传输和存储时不以同一种数据格式表征,观看 VR 全景图像时需要进行一些投影变换. 如图 1 所示,球面(Sphere)表示 VR 全景图像,在某一时刻,全景图像投影至视角 ABCD 被用户观看到. 在存储时,VR 全景图像大多使用等矩形投影(Equirectangular Projection,ERP)格式. 在编码传输过程中,VR 全景图像可能会转换到其他投影域,例如,立方体投影(CubeMap Projection,CMP)[8]等. 在不同投影域中,VR 全景图像的采样频率以及空间结构会发生改变. 图像的采样频率和空间结构直接影响人类对图像的视觉关注度. 因此,投影变换也影响着 VR 全景图像的显著性检测性能.

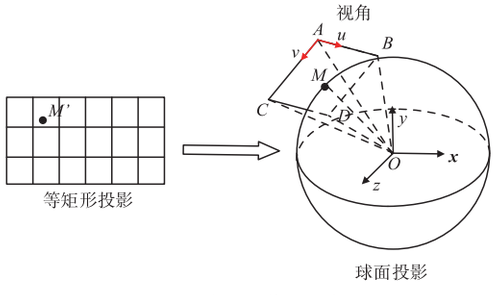
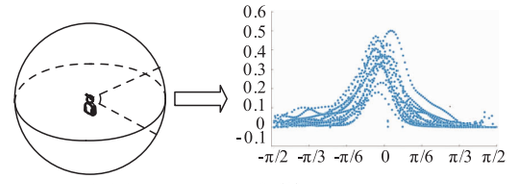


图1 VR全景图像的视角投影

VR 全景图像与传统图像的另一个显著的不同点在于用户的观看行为. 在观看 VR 全景图像时,用户与图像之间具有交互过程. 在观看 VR 全景图像过程中,用户除了眼球运动外,还存在头部运动. 如图 2(a)所示,用户处于视场中心,可以自由移动头部进而选择希望看到的场景. 对于用户而言,人的视角范围有限,在平视过程中,只能看到球面赤道附近的区域,看不到球面的两极区域. 然而,长时间低头或者仰头观看是不舒服的. 这导致赤道附近的内容更容易被用户关注. 本文将这种现象称为“赤道偏倚”. 图 2 给出了在 VR 全景图像数据集[7]上得到的用户关注程度(用户观看概率)与纬度之间的关系图. 从图中可以看到,在赤道附近(纬度 $-\frac{\pi}{6}$至$\frac{\pi}{6}$)集中了大量的关注度.



2 (a) 2 (b)

图2 赤道偏倚

3 VR 全景图像显著性检测算法

图像显著性检测的研究可以追溯至 1985 年,Koch 等人[9]首次提出显著性图的概念. 最初,Itti 等人[10]从生物学角度出发,提取亮度、颜色和方向特征,通过使用中心—邻域、归一化及合并操作得到显著性图. 后来,很多研究者对 Itti 模型进行了优化. 例如,Harel 等人[11]优化了 Itti 模型的特征融合方法,使用马尔科夫随机场计算图像显著图,提出基于图的显著性检测算法(Graph-Based Visual Saliency,GBVS).

之后,许多学者另辟蹊径,开发了一系列纯计算的显著性检测模型[12]. Hou 等人[13]从频域入手,在图像的幅度谱中减去先验知识的幅度谱,将幅度谱残差作为图像显著性图. Zhang 等人[14]提出了一种基于布尔图的显著性检测模型(Boolean Map Saliency,BMS),使用一系列的布尔图表示图像,通过分析布尔图的拓扑结构得到最终的显著性图.

随着深度学习技术的发展,基于深度学习的图像显著性检测算法应运而生. 2012 年,Krizhevsky 等人[15]提出 Alexnet 卷积神经网络模型,通过卷积神经网络(Convolutional Neural Network,CNN)提取特征. Mathieu 等人[16]改进了 Alexnet 模型,提出了使用同一个卷积网络完成图像分类、定位、检测三个任务的方法——OverFeat 模型.

传统的显著性检测算法都是针对平面图像的,随着 VR 全景图像的出现,VR 全景图像的显著性检测算法逐渐涌现. 受到传统图像显著性检测算法的启发,VR 全景图像的显著性检测算法可以分为(如表 1 所示):针对 VR 全景图像改进的显著性检测算法、基于注视位置数据映射的 VR 全景图像显著性检测算法、基于深度学习的 VR 全景图像显著性检测算法. 在表 1 中涉及到的传统显著性检测算法,除了前文已经提到的一些算法,还包括 EDN(Ensemble of Deep Networks)、ResNet 视觉注意模型(Saliency Attentive Model-Residual Network,SAM-ResNet)、ImgSig(Image Signature)、AWS(Adaptive Whitening Saliency)等.

表 1　VR 全景图像显著性检测算法总结

方法		操作域	是否考虑赤道偏倚	相关论文
分类	具体方法			
改进的传统方法	颜色词典、GBVS、SAM-ResNet	等矩形投影	是	[17][18]
			否	[19]
	Itti 模型、GBVS	视角	是	[18][20][21][22]
			否	
	Itti 模型	球面	是	
			否	[23]
	GBV、SAM-ResNet、BMS、ImgSig、AWS	立方体投影	是	
			否	[19][24]
基于数据映射的方法	数据映射 + 高斯滤波	等矩形投影	是	[25]
			否	[26]
基于深度学习的方法	卷积神经网络	视角	是	[27][28]
			否	
	卷积神经网络、深度强化学习	等矩形投影	是	
			否	[29][30]
	卷积神经网络	球面	是	
			否	[31]

3.1　针对 VR 全景图像改进的传统显著性检测算法

目前,很多 VR 全景图像显著性检测算法都是在传统显著性检测技术上改进后的算法. 改进主要涉及两个方面:投影变换和赤道偏倚. 根据 VR 全景图像具有多种投影方式的特性,VR 全景图像的显著性检测可以在不同的投影域中进行. 由于输入的 VR 全景图像以及最终的显著性图都是 ERP 格式的图像. 因此,在显著性检测时需要进行投影变换. 根据 VR 全景图像的投影域,基于传统算法改进的 VR 全景图像显著性检测算法框架总结如图 3 所示.

一些算法注意到 VR 全景图像中的赤道偏倚现象,将其应用于显著性检测中. 赤道偏倚反映了用户观看 VR 全景图像时的行为特性. 然而,目前刻画赤道偏倚现象的模型并不统一. 最简单的方法是将 VR 全景图像划分为几部分区域,每一部分区域赋予不同的显著性权值[20]. 这种方法简单方便,但是只在不同的图像区域操作,没有细化到像素级别,处理的粒度较粗,缺乏准确性. 因此出现了一些更准确的方法:使用高斯[21]、拉普拉斯[27]等模型模拟赤道偏倚.

3.1.1　等矩形投影域显著性检测算法

等矩形投影是 VR 全景图像中最常见的一种投影方式. 因此,大多数的 VR 全景图像显著性检测算法在等矩形投影域上进行[18~20]. 这类方法改进传统的显著性检测算法使其适用于 VR 全景图像,以得到基本的 VR 全景图像显著性图. 例如,基于颜色词典的 VR 全景图像显著性预测模型[17],基于 GBVS 和 SAM-ResNet 的 VR 全景图像显著性检算法[19],以及适应于 VR 全景图像的 BMS360 算法[18].

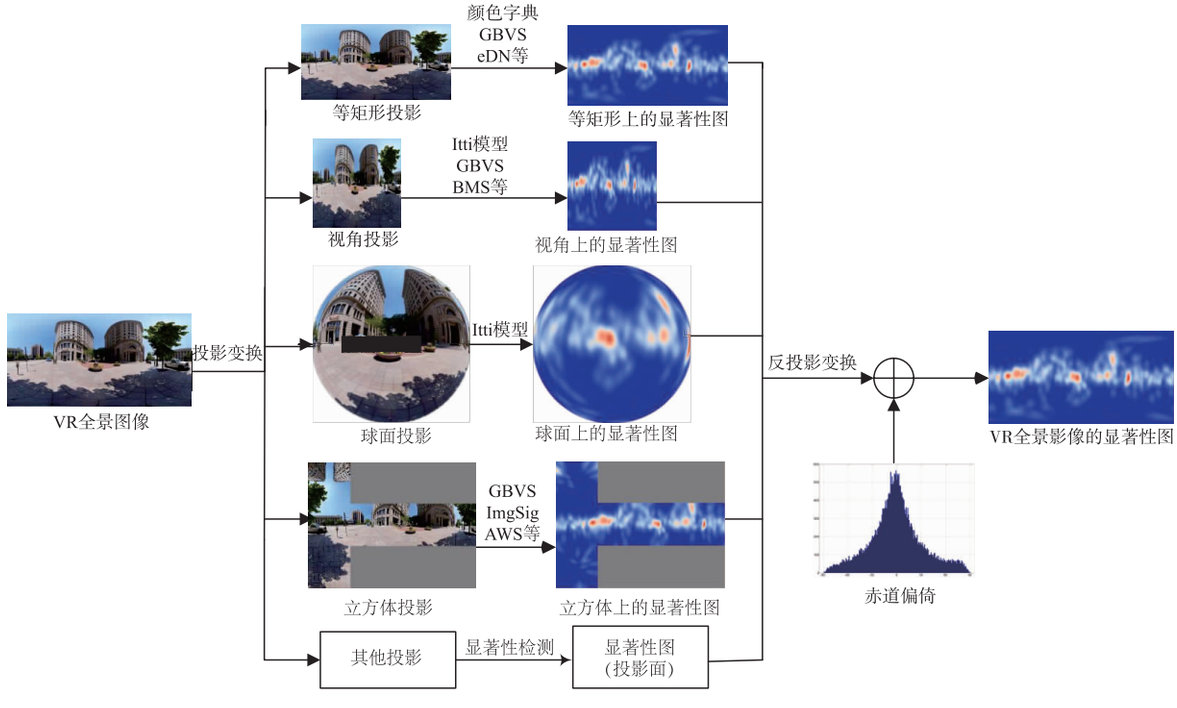


图3　基于传统算法的VR全景图像显著性检测技术框架

VR 全景图像转换至等矩形域时,图像中对应球面的两极点被转换至上下边界,因此,图像两极区域被过度表示. 此外,在 VR 全景图像等矩形域中,图像的左右边界并不是真正意义上的边界. 因此,VR 全景图像在等矩形域投影存在左右边界失真以及图像两极区域被过度表示的问题.

为了解决左右边界失真的问题. 文献[18,19]从不同投影域入手,融合 VR 全景图像几个投影平面中的显著性图. 文献[18]融合 VR 全景图像在以不同经度为边界的几个等矩形投影中的显著性图,通过将获得的几个显著性图平均的方式得到最终的显著性图. 文献[19]在 VR 全景图像的等矩形域、后向等矩形域分别进行显著性检测,并将两个显著性图融合. 为了解决图像两极区域被过度表示的问题,文献[19]从投影域入手,提出了扩展的立方体投影. 文献[18]则修改了传统的 BMS 算法,在布尔映射的归一化过程中进行编码映射,将二进制特征转换为具有连续值的特征.

在 VR 全景图像的等矩形投影域上进行显著性检测是比较简单方便的方法,适用于对算法的速度有很高要求的情况. VR 全景图像的等矩形投影与传统图像在形式上几乎无差别. 因此,传统的显著性检测算法一般可以直接用于 VR 全景图像的等矩形投影上. 此外,这类算法比其它算法少了很多投影变换,因此,这类算法的复杂性较低,执行速度较快. 然而,这类算法的精准率一般不高.

3.1.2 视角投影域显著性检测算法

考虑到用户所看到的 VR 全景图像以连续视角的形式显示,一些显著性检测算法[18,20~22]在视角域进行显著性检测. 由于输入的 VR 全景图像是 ERP 格式的,需要将图像从等矩形域转换到视角域. 在视角上应用 Itti 模型、GBVS 等方法提取特征,得到基本的 VR 全景图像显著性图. 最后将在视角域检测的显著性图反投影至等矩形域.

Zhu 等人[22]利用 Itti 模型,在视角上构建多层金字塔. 通过融合颜色、方向和空间频率等自下而上的特征以及车辆、人员轮廓和赤道偏倚等自上而下的特征,得到每个视角的显著性图. Battisti 等人[20]在视角上提取纹理、色调、饱和度作为低级特征,提取肤色、面部以及人数作为高级特征. 以赤道偏倚作为权重,将低级特征与高级特征融合,生成显著性图. 根据 GBVS 算法,Lebreton 等人[18]在视角上基于方向分析执行特征提取,而后将提取的特征反投影至等矩形域,完成后续步骤.

从频域出发,Ding 等人[21]提出在每个视角上使用高斯差分滤波,将视角中心点的 DoG 值映射到 VR 全景图像中,计算每个像素的 DoG 值与图像中所有像素的平均 DoG 值之间的欧几里德距离. 融合赤道偏倚,得到最终的显著性图.

显著性检测旨在检测出用户关注的部分,在视角域进行 VR 全景图像显著性检测更符合实际情况. 这类算法适用于以用户体验质量做为重要考量的情况. 然而,由于一幅 VR 全景图像可以投影至很多视角,这类算法的速度比等矩形投影域检测算法慢,复杂度也相对较高. 此外,使用这类算法进行 VR 全景图像显著性检测时需要注意不同视角之间的重叠影响.

3.1.3 球面投影域显著性检测算法

VR 全景图像的球面域是非变形域. 在球面上,VR 全景图像上的任何一处均不存在形变. VR 全景图像的球面域图像具有连续性,保留了位置信息,提供了完整的视野. 因此,一些 VR 全景图像的显著性检测算法在球面投影域上进行. Bogdanova 等人[23]基于 Itti 模型,提出了一种球面域 VR 全景图像显著性检测算法. 该算法构建了球形高斯金字塔,在每一层上提取特征. 通过中心—邻域、归一化及合并操作,获得球形显著图.

这类算法适用于对检测的精准度有很高要求的情况. VR 全景图像实质上以球面投影方式存在,球面域的 VR 全景图像保留了更多更准确的信息. 因此,在球面域进行显著性检测比在其他域进行更精确. 然而,这类方法在实现上比其它方法困难许多. 由于传统的显著性检测方法是基于平面图像的,而球面域的 VR 全景图像是立体空间中的图像. 这类方法需要将平面上的显著性检测算法扩展到立体空间.

3.1.4 立方体投影域显著性检测算法

一些传统的显著性检测算法,例如,GBVS、BMS、SAM-ResNet、ImgSig、AWS,通过合理的优化可以应用于 VR 全景图像的立方体投影域. 一般的立方体投影使图像失去了全局性,并引入了 24 个较小的边界(每个面 4 个),极大地恶化了显著性检测的质量. 为了保留图像中的所有原始图像信息和上下文信息,Startsev 等人[19]设计了一种扩展的立方体投影. 因此,文献[19]对立方体的每个投影面加入与其相邻的投影面,保留了图像投影面边界处的连续信息,消除了两极区域失真的问题. Maugey 等人[24]提出了一种新的 VR 全景图像投影方式——双立方体投影. 双立方体投影将 VR 全景图像映射至两个相差 45 度的立方体投影上,通过融合两个立方体投影上的显著性图,消除立方体投影中的边界区域不连贯问题.

3.1.5 其他投影域显著性检测算法

除了以上几种投影域之外,VR 全景图像还包含一些其他的投影域,比如八面体、二十面体、截断的方形金字塔投影[8]等. 由于这些投影域在实际应用中并不常见,目前没有在这些投影域中的显著性检测算法. 在这些投影域中,传统的显著性检测算法依然可以通过

改进用来进行显著性检测.

3.2　基于注视位置数据映射的显著性检测算法

一些 VR 全景图像显著性检测算法利用用户注视位置数据构建显著性图. 这类方法将有限个观看者的有限个注视点扩展到一般意义上的 *VR* 全景图像显著性图,可以看做从个体到整体的过程,其基本框架如图 4 所示.

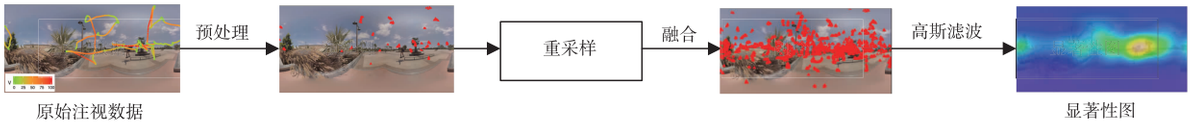


图4　基于注视位置数据映射的VR全景图像显著性检测框架[26]

这类算法主要分四个基本步骤:预处理、重采样、融合、滤波. 预处理,即对原始注视位置数据清洗,排除非用户注视的数据. 对注视点位置数据重采样,以确保符合球面上数据点的平均分布. 将不同观看者的注视位置融合以获得统计信息,得到 VR 全景图像的离散注视图. 在 VR 全景图像的离散注视图上执行滤波以产生连续的 VR 全景图像显著性图.

不同算法区分关注点与非关注点的方法不尽相同. 例如,Upenik 等人[26]提出了一种利用头部方位移动轨迹的 VR 全景图像的显著性检测模型,以头部角速度区分用户是否关注图像中某一位置. Abreu 等人[25]针对参与者的视角中心轨迹,依据注视时间长短区分用户的关注度.

基于注视位置数据映射的显著性检测算法需要一些观察者观看该图像的注视位置数据,其结果取决于采集到的数据的规模与质量. 如果采集到的数据不准确或不具有代表性,最后得到的显著性图将很糟糕. 这种方法需要耗费大量的人力和物力,一般只用于根据数据集产生基准显著性图.

3.3　基于深度学习的显著性检测算法

近几年,深度学习技术取得了快速的发展,并已经应用于各个领域,如语音识别和目标检测等. 基于深度学习的 VR 全景图像显著性检测算法也开始涌现. 这种类型的算法需要大量的已有显著性图训练显著性检测模型,通过训练好的显著性模型对 VR 全景图像进行显著性检测,得到最终的显著性图. 下面以 CNN 为例理解图 5 中基于深度学习的 VR 全景图像显著性检测框架.

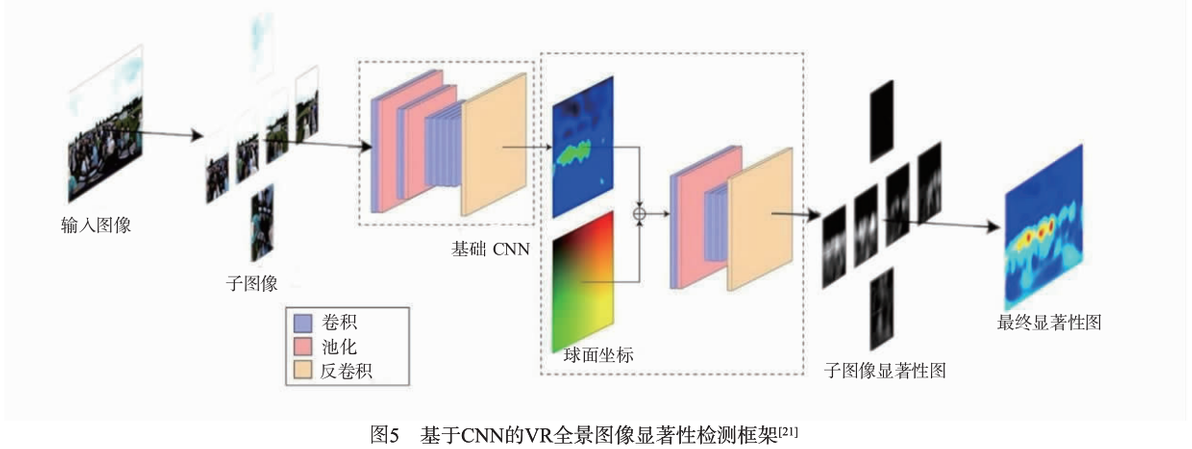


图5　基于CNN的VR全景图像显著性检测框架[21]

CNN 模型是最常见的应用于显著性检测的深度学习模型. 在这种模型中,一般需要对 VR 全景图像进行预处理,比如去均值和归一化等. 将预处理过后的 VR 全景图像数据作为 CNN 的输入数据. 使用卷积层对输入数据卷积,提取出图像的特征. 卷积后可以采集到数据中的很多特征. 为了减少计算量,需要将采集到的特征抽样或者聚合,这个过程称为池化. 根据显著性检测模型的需要,卷积层和池化层的组合可以出现很多次. 通过全连接层将前面经过多次卷积后高度抽象化的特征整合、归一化,并输出图像中像素的显著性概率,形成 VR 全景图像的显著性图.

如图 5 所示,Monroy 等人[28]以端到端的方式将 CNN 用于 VR 全景图像. VR 全景图像被映射成六个类似于视角的切片子图像,子图像分别作为 CNN 的输入. 经过卷积、池化和反卷积处理,输出子图像显著图. 此外,考虑到赤道偏倚影响,根据像素在球面的位置、决定突出其显著性或者降低显著性. 将六个切片子图像的显著性图映射至等矩形平面,即为 VR 全景图像的显著性图.

使用多个小卷积核作为 CNN 中的卷积层,Sitzmann

等人[27]将提取的特征映射到编码网络. 基于拉普拉斯模型拟合出赤道偏倚函数,将赤道偏倚函数与之前的显著性图融合,生成最终显著性图. Assens 等人[29]利用深度卷积神经网络,前三层使用多个小卷积核作为CNN 中卷积层,并在最后一层卷积层中使用单个滤波器,生成 VR 全景图像的显著性图.

鉴于强化学习具有强大的推理能力,Xu 等人[30]使用深度强化学习(Deep Reinforcement Learning,DRL)进行 VR 全景视频显著性检测,生成头部运动(Head Movement,HM)图. 算法以当前操作预测的和基准 HM 扫描路径之间的相似性作为奖励,通过奖励的值决定 DRL 的动作. 运行多个 DRL 工作流以确定每帧 VR 全景图像的潜在 HM 位置,形成最终的 HM 图.

利用球面投影上 VR 全景图像保真方面的优势,Zhang 等人[31]提出基于球面卷积神经网络的显著性检测模型(Spherical Convolutional Neural Networks,SCNN). 在 SCNN 中,卷积核在球顶位置,卷积可以看做卷积核沿球体的旋转. 考虑到 VR 全景图像通常以等矩形投影的形式存储,根据要卷积的块的位置拉伸和旋转卷积核实现 VR 全景图像的球面卷积.

基于深度学习进行 VR 全景图像显著性检测是一种以此推彼的过程. 这类算法一般具有较高的准确度,因为可以通过训练从数据中提取到 VR 全景图像的很多特征,其中可能包含一些还没有被研究人员发现的 VR 全景图像隐性特征. 然而,这类算法也有一定的局限性,例如,需要很多用于训练的 VR 全景图像和基准显著性图,检测结果受训练数据的影响,训练时间长,训练的特征里也可能会包含影响结果的错误信息等.

4 数据集及评价指标

为了评估显著性检测算法的性能,需要相关的数据集和评价指标. 数据集可以对显著性检测算法效果进行主观评价. 与评价指标结合,数据集可以对显著性检测算法提供客观评价基准. 此外,统一的数据集和评价指标也便于比较各种显著性检测算法的性能.

4.1 数据集

目前,基于 VR 全景图像和视频的数据集很多. 表 2 归纳了现有的一些数据集. 基于数据集提供的图像的多样性以及数据的可用性,本节详细叙述了一个优秀的标准数据集,即用于比赛—the Grand Challenge "Salient360!"的 Gutiérrez[7]等人收集的数据集.

数据集中包括一定量的 VR 全景图像以及观看者,同时记录了观看者在观看这些 VR 全景图像时的头部运动或者眼部运动等. 部分数据集还专门提供了基准显著性图,方便使用者评估显著性检测算法. 此外,数据集除了可以用于研究 VR 全景图像显著性检测之外,还可以用于研究用户行为以及 VR 全景图像的压缩、传输、编辑等.

表 2 VR 全景图像/视频数据集

数据集	场景数量(个)	场景内容信息	记录数据类型	有无基准显著性图	视频/图像
Gutiérrez[7]	85	室内、室外	头部 眼部运动	有	图像
Sitzmann[27]	22	室内、室外	头部运动	有	图像
Abreu[25]	21	室内、室外	头部运动	无	图像
Corbillon[32]	7	室外	头部运动	无	视频
Wu[33]	18	室内、室外	头部运动	无	视频
Lo[34]	10	室外	头部运动	有	视频

目前大多数的数据集只采集头部数据,而 Gutiérrez 等人[7]所介绍的数据集既采集了头部运动数据又采集了眼部运动数据. 该数据集的数据类型丰富,而且规模明显比其它数据集大. 该数据集包含了 85 个 VR 全景图像,考虑到了各种属性,例如拍摄环境、空间复杂度、前景物体的数量等.

Sitzmann 等人[27]提出的数据集记录了参与者使用头戴式显示器在站立、就坐以及在桌面监视器上观察单声道相同场景三种状态下的数据. Abreu 等人[25]收集了参与者分别观看 VR 全景图像 10 秒和 20 秒的视角中心轨迹并对比. Corbillon 等人[32]既提供了可用的数据集,又提供了用于收集数据集的开源软件. Wu 等人[33]进行了两次实验:有目的地观看和无目的地观看. Lo 等人[34]记录了 VR 全景图像的内容数据以及传感器数据,并使用原始日志文件中的时间戳增加额外的内容和传感器数据.

4.2 评价指标

4.2.1 主观评价

计算机视觉算法的性能评估近年来越来越受关注. 对显著图检测最简单的评价方法便是直接对比算法预测的显著性图与基准显著性图. 这种对比方法简单直观,但存在局限性:(1)难以比较使用了不同数据集的算法;(2)当两幅图像差距不大时,很难形成统一的结论.

4.2.2 客观评价

对于图像的显著性,简单地展示视觉效果不足以作为验证显著性检测算法鲁棒性的手段,使用性能指标能够客观评估显著性检测算法. 使用这些性能指标可以直接比较不同显著性算法或相同算法的不同配置的性能. 鉴于此,本节详细介绍一些常用的评价指标.

(1)线性相关系数

线性相关系数(Linear Correlation Coefficient,CC)用

于评价预测的显著图与基准显著图的吻合程度. 随着算法预测的显著图和基准显著图之间相关强度增加,CC 也增加,最高达到 1. CC 的计算方法如下:

$$\mathrm{CC} = \frac{\sum_{x,y}(S(x,y)-\mu_s)(G(x,y)-\mu_G)}{\sqrt{\sigma_s^2\sigma_G^2}} \tag{1}$$

其中 $S(x,y)$ 是算法的预测显著图,$G(x,y)$ 是基准显著图,μ_s 和 σ_s 是算法预测显著图的期望和方差,μ_G 和 σ_G 是基准显著图的期望和方差.

(2)KLD

KLD(KL-Divergence,KLD)用来衡量预测的显著图与基准显著图之间的距离. KLD 的值越小,表示算法预测的显著图与基准显著图之间越接近,算法的性能越好. KLD 的计算公式如下:

$$\mathrm{KLD} = \frac{1}{2}\sum_{k=0}^{255}\left(S_k\log\frac{S_k}{G_k} + G_k\log\frac{G_k}{S_k}\right) \tag{2}$$

其中 S_k 和 G_k 分别是归一化后的预测显著图与基准显著图在 k 处的值.

(3)标准化扫描路径显著性

标准化扫描路径显著性(Normalized Scanpath Saliency,NSS)用来计算归一化的预测显著性图在人眼注视点位置上的平均标准化显著性. 预测显著图首先需要被归一化为均值为 0,方差为 1. 在所有观察者的注视点提取归一化显著性值,将其平均即为 NSS. NSS 的计算如下:

$$\mathrm{NSS} = \frac{1}{n}\sum_{k=0}^{n}(S_k - G_s) \tag{3}$$

其中,n 是观察者的凝视点的个数.

(4)ROC 曲线和 ROC 曲线下面积

显著图可以被解释为像素点是否是注视点的二值分类器. 图像中的注视点被预测正确的概率记为真阳性率,图像中的背景点被预测错误的概率称为假阳性率. 在坐标系中画出真阳性率和假阳性率随阈值的变化曲线,即受试者工作特征曲线(Receiver Operating Characteristic Curve,ROC 曲线). ROC 曲线可以直观的表示出显著性检测模型在任意阈值下对于图像显著点的识别能力. ROC 曲线的精度是由 ROC 曲线下的面积(Area Under ROC Curve,AUC)来衡量的. AUC 的取值范围为 0. 5 到 1,AUC 的值越接近 1,表示算法的性能越好.

(5)F-measure

F-measure 为精度和召回值之间的加权调和平均值,用来评价预测显著性图的准确率和完整性. F-measure 越大表示显著性检测算法的效果越好,F-measure 的计算方式为:

$$F_\beta = \frac{(\beta^2+1)PR}{\beta^2 P + R} \tag{4}$$

其中,F_β 是 F-measure 值,β 是参数,P 是精确率(正确率),R 是召回率(查全率).

对于一个好的显著性检测算法,其 KLD 值应该很小,CC、NSS、AUC 和 F-measure 值应该很大. 但实际情况下,很难找到一个在所有指标上都有很好性能的算法.

4.2.3　算法性能对比

为了对各个算法以及评价标准有更清晰的认识,本小节对目前算法的性能对比分析,总结如表 3 所示.

表 3　算法性能对比

算法	KLD	CC	NSS	AUC	数据集
传统、等矩形域[17]	0. 477	0. 55	0. 939	0. 736	Gutiérrez[7]
传统、等矩形域[19]	0. 45	0. 58	0. 92	0. 75	Gutiérrez[7]
传统、视角域[20]	0. 81	0. 52	–	–	Gutiérrez[7]
传统、视角域[21]	0. 787	0. 658	–	–	Gutiérrez[7]
传统、视角域[18]	0. 698	0. 527	0. 851	0. 714	Gutiérrez[7]
传统、视角域[22]	0. 481	0. 532	0. 918	0. 734	Gutiérrez[7]
深度学习、视角域[27]	–	0. 49	–	–	Sitzmann[7]
深度学习、视角域[28]	0. 487	0. 536	0. 757	0. 702	Gutiérrez[7]
深度学习、等矩形域[29]	0. 1954	0. 8471	0. 7785	0. 6819	Gutiérrez[7]
深度学习、等矩形域[30]	–	0. 704	3. 275	–	Xu[30]
深度学习、球面域[31]	–	0. 6246	3. 5340	0. 8977	Zhang[31]

在这几种算法中,文献[27,30,31]所用的数据集与其它算法不同,无法通过主观对比将其与其它算法比较,只能通过客观评价指标对比.

如表 3 所示,文献[17 ~ 22]中所采用的方法都属于针对 VR 全景图像改进的传统显著性检测算法. 这类算法的性能差距不大,其中文献[21]算法表现较为突出. 原因在于,文献[21]算法考虑到了不同视角的相互影响. 同为等矩形域中的算法,文献[19]性能略优于文献[17],这是由于文献[19]算法使用了等矩形域、后向等矩形域、扩展的立方体投影域三种投影上显著性图的融合,解决了等矩形域投影造成的上下左右边界失真的问题.

文献[27 ~ 31]所采用的方法是基于深度学习的显著性检测算法,这类算法的性能差距很大. 其中文献[29,30]的算法性能明显高于其他算法,文献[28,31]的算法性能与针对 VR 全景图像改进的传统显著性检测算法基本持平,文献[27]的算法性能是所有算法里最差的. 这说明了针对 VR 全景图像改进的传统显著性检测算法是一种比较稳妥的方法,而在基于深度学习的显著性检测算法中,模型的选择及训练对结果的影响很大. 然而基于深度学习的显著性检测算法是一种很有潜力的模型,优秀的基于深度学习的显著性检测

算法的性能可以远远高于针对 VR 全景图像改进的传统显著性检测算法.

5 结论与展望

目前 VR 全景图像的显著性检测算法还不成熟,现有的 VR 全景图像显著性检测研究已经取得了一定的进展,但仍然存在一些挑战. 未来的 VR 全景图像显著性检测技术研究还有很大的提升空间.

(1)VR 全景图像的分辨率远远高于传统图像. 而且,在 VR 全景图像显著性检测过程中还可能存在多种投影变换. 因此,VR 全景图像的显著性检测算法的复杂度往往很高,如何降低复杂度是显著性检测能够被实时应用的关键.

(2)目前的视觉注意机制是面向传统平面图像的,但 VR 全景图像不同于传统平面图像,需要捕捉其特有的用户注意机制,了解各种视觉注意影响因素之间的相互作用. 目前,研究人员对 VR 全景图像注意机制的认识还不够清楚. 未来需要通过研究人类对 VR 虚拟场景认知的过程,总结其多维度注意认知结构,形成更深层次的认知模型.

(3)深度学习在传统图像、自然语言等领域中都表现出了极强的优势,可以将其应用于 VR 全景图像显著性检测. 基于深度学习的显著性检测算法[29]明显比其它显著性检测算法的结果好,这也就说明了 VR 全景图像还存在一些尚未发现的影响显著性检测的特性. 如何利用深度学习技术的进展挖掘这些特性并进一步提升 VR 全景图像显著性检测效率方面还存在着很大的优化空间.

(4)目前的显著性检测算法大多是针对 VR 全景图像的,针对 VR 全景视频的显著性检测很少. 在实际应用中,VR 全景视频的应用更为广泛. VR 视频的显著性检测,可以辅助提高 VR 视频的压缩与传输效率. 因此,将 VR 全景图像的显著性检测扩展到 VR 全景视频上更有意义.

参考文献

[1] 解凯,郭恒业,张田文. 图像 Mosaics 技术综述[J]. 电子学报,2004,32(4):630 – 634.
XIE Kai, GUO Heng-ye, ZHANG Tian-wen. A survey of image mosaics technology[J]. Acta Electronica Sinica, 2004,32(4):630 – 634. (in Chinese)

[2] LUZ G, ASCENSO J, BRITES C, et al. Saliency-driven omni-directional imaging adaptive coding: modeling and assessment[A]. Proceedings of the International Workshop on Multimedia Signal[C]. USA:IEEE,2017. 1 – 6.

[3] LIU K D,LIU Y W,LIU J X,et al. Joint source encoding and networking optimization for panoramic video streaming over LTE-A downlink[A]. Proceedings of the GLOBECOM[C]. USA:IEEE,2017. 1 – 7.

[4] SERRANO A, SITZMANN V, RUIZ-BORAU J, et al. Movie editing and cognitive event segmentation in virtual reality video[J]. ACM Transactions on Graphics,2017,36(4):1 – 7.

[5] BORJI A,CHENG M M,JIANG H,et al. Salient object detection:a survey[J]. ArXiv Preprint, 2014, 2(4):1411 – 5878.

[6] JUDD T, DURAMD F, TORRALBA A. A benchmark of computational models of saliency to predict human fixations[J/OL]. Technical Report, MIT, 2012, http://hdl.handle.net/1721.1/68590.

[7] GUTIERREZ J,DAVID E,RAI Y,et al. Toolbox and dataset for the development of saliency and scanpath models for omni-directional/360° still images[J]. Signal Processing Image Communication,2018,69:1 – 7.

[8] HE Y, VISHWANATH B, XIU X, et al. AHG8: interdigital's projection format conversion tool [S].

[9] KOCH C, ULLMAN S. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry [M]. Springer, Dordrecht: Matters of Intelligence, 1987. 115 – 141.

[10] ITTI L,KOCH C,NIEBUR E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998,V20(11):1254 – 1259.

[11] HAREL J,KOCH C,PERONA P. Graph-based visual saliency[A]. Proceedings of Advances in Neural Information Processing Systems [C]. USA: IEEE, 2007. 545 – 552.

[12] ACHANTA R,HEMAMI S,ESTRADA F,et al. Frequency-tuned salient region detection [A]. Proceedings of Computer Vision and Pattern Recognition [C]. USA: IEEE,2009. 1597 – 1604.

[13] HOU X,ZHANG L. Saliency detection:a spectral residual approach[A]. Proceedings of Computer Vision and Pattern Recognition[C]. USA:IEEE,2007. 1 – 8.

[14] ZHANG J,SCLAROFF S. Saliency detection:a boolean map approach[A]. Proceedings of International Conference on Computer Vision [C]. USA: IEEE, 2013. 153 – 160.

[15] KRIZHEVSHY A,SUTSKEVER I,HINTON G E. Imagenet classification with deep convolutional neural networks[A]. Proceedings of International Conference on Neural Information [C]. Qatar: Neural Information Processing,2012. 1097 – 1105.

[16] MATHIEU M,LeCun Y,FERGUS R,et al. OverFeat:integrated recognition,localization and detection using conv-

olutional networks[J]. ArXiv Preprint,2013,2(4):1 – 7.
[17] LING J,ZHANG K,ZHANG Y,et al. A saliency prediction model on 360 degree images using color dictionary based sparse representation[J]. Signal Processing Image Communication,2018,69:60 – 68.
[18] LEBRETON P,RAAKE A. GBVS360,BMS360,ProSal: extending existing saliency prediction models from 2D to omni-directional images[J]. Signal Processing Image Communication,2018,69:69 – 78.
[19] STARTSEV M,DORR M. 360-aware saliency estimation with conventional image saliency predictors[J]. Signal Processing Image Communication,2018,69:43 – 52.
[20] BATTISTI F,BALDONI S,BRIZZI M,et al. A feature-based approach for saliency estimation of omni-directional images[J]. Signal Processing Image Communication,2018,69:53 – 59.
[21] DING Y,LIU Y,LIU J,et al. Panoramic image saliency detection by fusing visual frequency feature and viewing behavior pattern[A]. Proceedings of the Pacific-Rim Conference on Multimedia[C]. China:MTAP,2018. 418 – 429.
[22] ZHU Y,ZHAI G,MIN X. The prediction of head and eye movement for 360 degree images[J]. Signal Processing Image Communication,2018,69:15 – 25.
[23] BOGDANOVA I,BUR A,HUGLI H. Visual attention on the sphere[J]. Computer Vision & Image Understanding,2010,114(1):100 – 110.
[24] MAUGEY T,MEUR O L,LIU Z. Saliency-based navigation in omni-directional image[A]. Proceedings of Multimedia Signal Processing[C]. USA:IEEE,2017. 1 – 6.
[25] ABREU A,OZCINAR C,SMOLIC A. Look around you: saliency maps for omni-directional images in VR applications[A]. Proceedings of Quality of Multimedia Experience[C]. USA:IEEE,2017. 1 – 6.
[26] UPENIK E,EBRAHIMI T. A simple method to obtain visual attention data in head mounted virtual reality[A]. Proceedings of Multimedia & Expo Workshops[C]. USA:IEEE,2017. 73 – 78.
[27] SITZMANN V,SERRANO A,PAVEL A,et al. Saliency in VR: how do people explore virtual environments? [A]. Proceedings of Transactions on Visualization and Computer Graphics[C]. USA:IEEE,2018. 1633 – 1642.
[28] MONROY R,LUTZ S,CHALASANI T,et al. SalNet360: saliency maps for omni-directional images with CNN[J]. Signal Processing Image Communication,2018,69:26 – 34.
[29] ASSENS M,MCGUINNESS K,GIRO-I-NIETO X,et al. SaltiNet:scan-path prediction on 360 degree images using saliency volumes[J]. International Conference on Computer Vision Workshops,2017,3:2331 – 2338.
[30] XU M,SONG Y,WANG J,et al. Modeling attention in panoramic video:a deep reinforcement learning approach[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018. https://arxiv.org/abs/1710.10755v1.
[31] ZHANG Z,XU Y,YU J,et al. Saliency detection in 360° videos[A]. Proceedings of the European Conference on Computer Vision[C]. Germany:Springer,2018. 488 – 503.
[32] CORBILLON X,DE Simone F,SIMON G. 360-Degree video head movement dataset[A]. Proceedings of the 8th ACM on Multimedia Systems Conference[C]. New York:ACM,2017. 199 – 204.
[33] WU C,TAN Z,WANG Z,et al. A dataset for exploring user behaviors in VR spherical video streaming[A]. Proceedings of the 8th ACM on Multimedia Systems Conference[C]. New York:ACM,2017. 193 – 198.
[34] LO W,FAN C,LEE J,et al. 360° video viewing dataset in head-mounted virtual reality[A]. Proceedings of the 8th ACM on Multimedia Systems Conference[C]. New York:ACM,2017. 211 – 216.

作者简介

丁　颖　女. 1995 年 5 月出生,河南安阳人. 2016 年毕业于郑州大学计算机科学与技术专业,2016 年进入中国科学院大学信息工程研究所,现为硕士研究生,从事全景图像显著性检测方面的有关研究.
E-mail:dingying@iie.ac.cn

刘延伟(通信作者)　男,1976 年出生,黑龙江人. 博士,中国科学院信息工程研究所副研究员,主要研究方向为沉浸式视频通信、多媒体信息处理与网络安全.
E-mail:liuyanwei@iie.ac.cn