

# 一种立体全景图像显著性检测模型

邱淼淼, 柴雄力, 邵枫\*

宁波大学信息科学与工程学院, 浙江 宁波 315211

**摘要** 三维全景给用户同时提供 360° 视角的同时给人以强烈的三维立体真实感, 虽然近年来研究人员开发了大量的算法来检测二维及三维图像中的显著区域, 但针对立体全景图像显著性检测的研究较少。考虑到全景图像的投影特点、立方体投影(CMP)图像有助于消除顶部与底部引起的扭曲及边框效果, 利用等矩形投影(ERP)图像中所有可用的上下文信息, 以 ERP 作为全局信息、CMP 作为局部信息, 融合了全局和局部的视觉显著图。提出的立体全景显著性检测模型由颜色相似度算法和区域对比度算法两部分组成。首先, 对图像进行多尺度线性迭代聚类超像素分割, 根据像素块的颜色差异得到颜色对比特征图; 然后依据空间分布紧凑性计算区域对比度; 根据颜色对比特征和区域对比度特征得到图像显著图。通过结合赤道偏移并且融入深度信息得到最终的立体全景显著图。最后, 将所得结果在公开的立体全景图像数据库 ODI 中进行了对比验证, 实验结果表明, 所提方法得到的显著结果具有较高的准确率、召回率和 F-measure 值, 其综合性能优于 6 种经典的显著预测算法。所提模型既能够充分利用图像信息, 又能有效地抑制复杂的背景区域, 可得到更加符合视觉感知的显著图。

**关键词** 图像处理; 显著性检测; 投影格式; 颜色相似度; 区域对比度

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202158.0810002

## Saliency Detection Model for Stereoscopic Panoramic Images

Qiu Miaomiao, Chai Xiongli, Shao Feng\*

Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315211, China

**Abstract** A three-dimensional (3D) panorama provides a 360° perspective for users, giving them a strong 3D sense of reality. Although researchers have developed a large number of algorithms to detect salient areas in two-dimensional and 3D images in recent years, there are few studies on the saliency detection of stereoscopic panoramic images. All available context information in an equirectangular projection (ERP) image which is used as the global information is used, the cubic projection (CMP) image is used as local information, and global and local visual saliency maps are integrated, taking into account the projection characteristics of the panoramic image and the CMP image which helps to eliminate the distortion and frame effect caused by the top and bottom. In this work, the proposed stereo panoramic saliency detection model is composed of two parts, i. e., color similarity and regional contrast methods. First, multi-scale linear iterative clustering superpixel segmentation is carried out on the images, and the color contrast feature map is obtained according to the color difference of pixel blocks. Then, the regional contrast is calculated according to the compactness of spatial distribution. The saliency map is obtained based on feature maps of color contrast and regional contrast. The final stereoscopic panoramic saliency map is obtained by combining equatorial migration and fusing depth information. Finally, the results obtained are compared and verified in the public stereo panoramic image database ODI. Experimental results show that the saliency maps obtained by the proposed method have high precision, recall rate, and F-measure value, and the comprehensive performance of the proposed method is better than that of the six classical saliency prediction algorithms. The proposed model not only makes full use of the image information, but also effectively suppresses the complex background area, so as to

收稿日期: 2020-07-22; 修回日期: 2020-08-15; 录用日期: 2020-09-09

基金项目: 国家自然科学基金(61622109)

\* E-mail: shaofeng@nbu.edu.cn

obtain the saliency map that is more consistent with the visual perception.

**Key words** image processing; saliency detection; projection format; color similarity; region contrast

**OCIS codes** 100.2000; 100.2960; 110.2970

## 1 引言

随着虚拟现实技术的发展,大量虚拟现实(VR)设备进入消费者市场,三维全景给用户的全方位视角的同时给人以强烈的三维立体真实感,传统三维建模技术在它面前便表现得相形见绌<sup>[1]</sup>。此外,全景图像/视频在商业、直播、医疗、监控等各行各业得到广泛的应用。相比传统平面图像,全景图像的视野范围更广、场景也更为复杂。准确预测出人感兴趣的区域将为图像的传输、编码及渲染提供基础。目前,全景图像视觉注意机制中的显著性检测方法成为了研究热点。

在纷繁复杂的外界场景中,人类视觉系统总是能够快速并无意地定位重要的目标区域并对其进行细致分析,而对其他区域仅进行粗略分析甚至忽略,人眼的这种主动选择性的视觉心理活动通常被称为视觉注意机制<sup>[2]</sup>。根据视觉注意力是否由大脑视皮层知识引导,现有的视觉显著性检测模型可分为两种类型:基于自底向上机制的模型和基于自顶向下机制的模型。自底向上机制是一种数据驱动的、与任务无关的感知过程,用于自然场景的自动显著区域选择<sup>[3-7]</sup>;而自顶向下的机制受执行的任务、目标的特征分布等因素的影响<sup>[8-13]</sup>。根据检测任务的不同,现有的视觉显著性模型又可分为两种类型:视觉关注点/区域检测模型<sup>[14]</sup>和视觉显著对象检测模型<sup>[15]</sup>。

在过去几十年中,许多显著性检测模型成功应用于各种视觉处理,如重定向<sup>[16]</sup>、质量评价<sup>[17]</sup>、编码<sup>[18]</sup>、图像分割<sup>[19]</sup>等。Li等<sup>[20]</sup>在考虑图像细节和基于区域预测的基础上,提出利用正则化的随机漫步排序来构造基于超像素的背景和前景显著估计的像素显著映射。Zhou等<sup>[21]</sup>提出一种自底向上融合紧密性和局部对比度信息的显著区域检测方法。Goferman等<sup>[22]</sup>考虑到图像特征,提出了一种结合上下文信息的显著预测方法。Wang等<sup>[23]</sup>针对立体信息以及立体对的两个视图之间的相关性,提出了一种立体显著性检测算法。Xia等<sup>[24]</sup>根据人类视觉系统对中心-环绕(C-S)对比非常敏感的特点,构建了一个深度C-S推理网络来训练和学习不同的显著特征。

最近,VR在工业和学术界越来越受欢迎,其

中,全景图像被广泛用于相关的VR应用中。Maughey等<sup>[25]</sup>将360°图像投影到双立方体中,然后通过特征提取模型的聚合来估计它们的显著性,所述特征提取模型包括基于图形的视觉显著性(GBVS)<sup>[26]</sup>、图像签名<sup>[27]</sup>、自适应美白显著性模型<sup>[28]</sup>、基于多尺度稀疏度的显著性检测模型<sup>[29]</sup>、布尔映射方法<sup>[30]</sup>和人脸检测器<sup>[31]</sup>。Lebreton等<sup>[32]</sup>将现有的二维图像视觉显著预测模型(GBVS<sup>[26]</sup>)进行扩展并应用于ERP格式的全景图像,另外考虑到观察者更倾向于观看球体的赤道区域,提出一种将自适应赤道偏置和GBVS相结合的框架。Abreu等<sup>[33]</sup>利用融合显著图后处理方法来减小当前显著性模型的中心先验限制,通过水平平移获得4组360°图像,对使用显著性模型预测得到的显著图进行平均处理。Xu等<sup>[34]</sup>提出了一种生成对抗性模拟学习(GAIL)算法用于模仿人类头部轨迹并进行训练学习。在多个深度强化学习(DRL)模型中,根据生成的头盔(HM)轨迹采样头部固定位置以预测显著信息。Chen等<sup>[35]</sup>提出了一种用于360°图像显著预测的局部-全局深度网络,通过两个深度子网络获得显著特征,最后利用线性组合生成最终的显著结果。

目前,三维全景能够为用户提供全方位视角,与二维图像的视觉注意力模型不同,立体图像的视觉注意力模型需要考虑图像的深度特征。近年来,研究人员开发了大量的算法来检测平面及立体图像的显著区域,但对立体全景图像显著性的研究较少。分析现有的显著预测模型<sup>[21,36]</sup>可以发现,颜色对比度较高的区域总是会具有视觉独特性。另外,考虑到背景区域在整个图像上具有较宽的分布,而显著对象则具有紧凑的空间分布。因此本文将经典的平面显著预测模型扩展到立体全景显著预测方法中。本文研究的创新点是:1)不同于平面显著提取方式,本文结合了等矩形投影(ERP)和立方体投影(CMP)两种投影格式分别提取全局和局部特征,解决了全景图像特有的采样失真问题;2)本文将平面显著预测模型和立体显著深度信息相结合,全面提取了立体全景显著特征,取得了较好的效果。将本文方法在ODI数据库上进行测试,得到的图像质量预测效果良好且准确,与其他方法相比,本文方法对显著性的预测性能更为出色。

## 2 投影格式转换

近几年研究人员针对传统 2D 图像的显著预测技术进行了广泛的研究,然而,这些模型不能直接应用于球面图像。全景图像都是球形的,为了计算全景图像的显著性,人们将球体映射到 2D 图像上,如图 1 所示,首先,在赤道上等距采样  $N_0$  个视点,并计算两个相邻视点的夹角  $\theta = \frac{360}{N_0}$ ,然后在南北纬均匀采样  $N_1 = N_0 \cos \theta$  个视点,值得注意的是,南北纬  $90^\circ$  只采样一次,即  $N_2 = 1$ 。目前的图像存储格

式大多都是 ERP,最后在渲染时将图像投射到球面。另一种常见的投影格式是 CMP,如图 1 所示,把球面的中心投影到外切球体的立方体的 6 个面上,每个立方体具有  $90^\circ$  视场。每个立方体面中的失真不像 ERP 图像中那样明显,但是由立方体面之间的不连续性引起的边界附近失真仍然存在。本文以 ERP 图像作为全局视觉信息,以 CMP 的 6 个面作为局部视觉信息,利用 ERP 图像中所有可用的上下文信息,并考虑 CMP 图像有助于消除顶部与底部引起的扭曲,结合两种投影格式得到更为有效的显著预测模型。

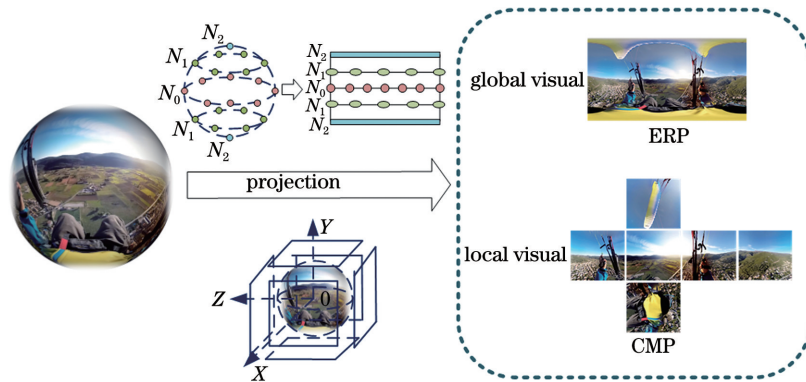


图 1 全景图像的投影格式示意图

Fig. 1 Illustration of projection format for panoramic image

## 3 显著性检测方法

本文介绍的模型可有效地预测全景图像的显著

性。整体框架图如图 2 所示,融合了模型预测的全局和局部显著图。所提出的显著性检测方法主要包括 4 个步骤:超像素分割、颜色相似度计算、区

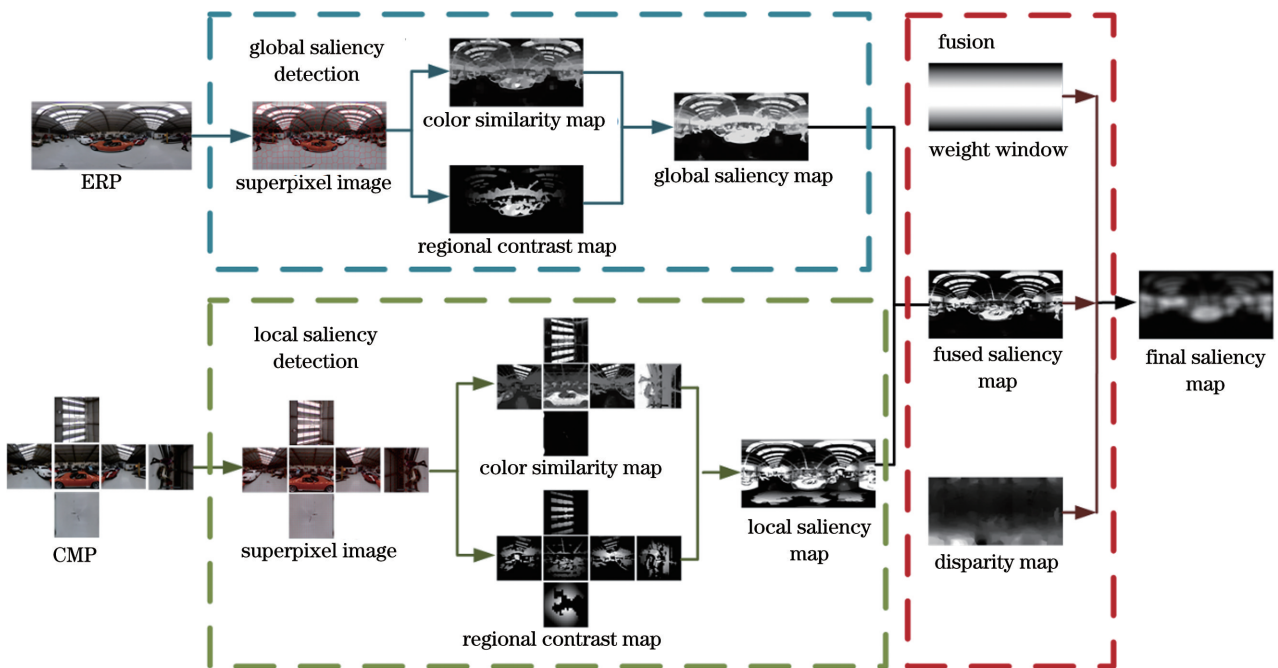


图 2 本文方法的整体框架

Fig. 2 Overall framework of proposed method

域对比度计算以及融合。图顶部所示部分将整个 ERP 格式作为输入,检测全局视觉注意力。第二部分,将 360°图像划分为立方体 6 个面。6 张图像作为输入,用于局部视觉显著性检测。然后,对 6 个面对应的输出进行逆投影得到 ERP 并将其集成到 360°显著图中。最后,通过权重窗口以及视差信息得到最终的立体全景显著图。

### 3.1 多尺度线性迭代聚类超像素分割

人类视觉系统在感知和理解一幅图像时偏向于关注整块区域而不是单独孤立的像素点。与传统的分块方法不同的是,简单线性迭代聚类(SLIC)算法<sup>[37]</sup>可以很好地保持对象的完整度。因此本文利用 SLIC 对图像进行预处理,并进行图像区域分割,以相对简单的格式表示图像并能减少图像的冗余。Cheng 等<sup>[38]</sup>所用的图像分割方法是基于图的方法,虽然基于图的分割算法也能快速产生超像素,但是超像素的数量和形状是无法控制的。鉴于此,本文采用基于 SLIC 的方法对图像进行分割,在 CIELAB 颜色空间利用 SLIC 算法提取第  $k$  个超像素块  $\mathbf{h}_k$  ( $k=1, \dots, i, \dots, j, \dots, K$ ),其中  $K$  为图像  $I$  的总超像素个数。根据所需的超像素个数  $K$  初始化种子并计算超像素之间的距离。其中,距离度量包括颜色差异和空间距离,两个距离的计算方法分别为

$$D(\mathbf{c}_i, \mathbf{c}_j) = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2}, \quad (1)$$

$$D(\mathbf{s}_i, \mathbf{s}_j) = \sqrt{(m_j - m_i)^2 + (n_j - n_i)^2}, \quad (2)$$

式中:  $\mathbf{c}_i$  和  $\mathbf{c}_j$  分别表示超像素  $\mathbf{h}_i$  和  $\mathbf{h}_j$  的颜色值;

$\mathbf{s}_i$  和  $\mathbf{s}_j$  分别表示超像素  $\mathbf{h}_i$  和  $\mathbf{h}_j$  的空间位置;  $l, a, b$  为 CIELAB 颜色空间中的 3 个通道变量;  $l_i, a_i, b_i$  表示超像素  $\mathbf{h}_i$  在 Lab 三个通道的颜色空间值;  $l_j, a_j, b_j$  表示超像素  $\mathbf{h}_j$  在 Lab 三个通道的颜色空间值;  $m_i, n_i$  分别表示超像素  $\mathbf{h}_i$  的横、纵坐标值;  $m_j$  和  $n_j$  分别表示超像素  $\mathbf{h}_j$  的横、纵坐标值;  $D(\mathbf{c}_i, \mathbf{c}_j)$  为  $\mathbf{h}_i$  和  $\mathbf{h}_j$  的颜色差异;  $D(\mathbf{s}_i, \mathbf{s}_j)$  为  $\mathbf{h}_i$  和  $\mathbf{h}_j$  的空间距离。

在本实验中,考虑到全景图像较平面图像的复杂度更高、分辨率也更大等特征,采用了多尺度超像素分割,即

$$\bar{\mathbf{S}} = \frac{1}{R} \sum_{K \in \{r, 2r, 4r\}} \mathbf{S}^K, \quad (3)$$

式中:  $\bar{\mathbf{S}}$  为多个尺度超像素下的均值;  $\mathbf{S}^K$  为超像素个数为  $K$  的超像素分割结果;  $R$  为设置的尺度数,本文设置  $R=3$ ;  $r$  为超像素初始值,设为 300; 超像素的个数  $K$  为 1200, 600, 300。

图 3 是不同原始图像以及 SLIC 在三个超像素尺度的分割结果,图 4 是以  $K=600$  为例,对 CMP 的 6 个面进行超像素分割的结果。超像素块的边界很大程度上贴近了原图像中的物体边界,每个像素块不仅包含底层的颜色、方向、亮度信息,而且还包含整个图像的结构信息。使用超像素块作为基本单位来进行显著性检测可以保证物体边界计算的准确性。从图 3 中可以看出,当  $K=300$  时,SLIC 对于熊的分割效果是三个尺度中最好的,但是对于细节信息的分割却需要更高的  $K$  值。因此,本文方法基于三个超像素尺度的均值得到显著结果,使得结构细节信息更为准确并且显著结果更加平滑。

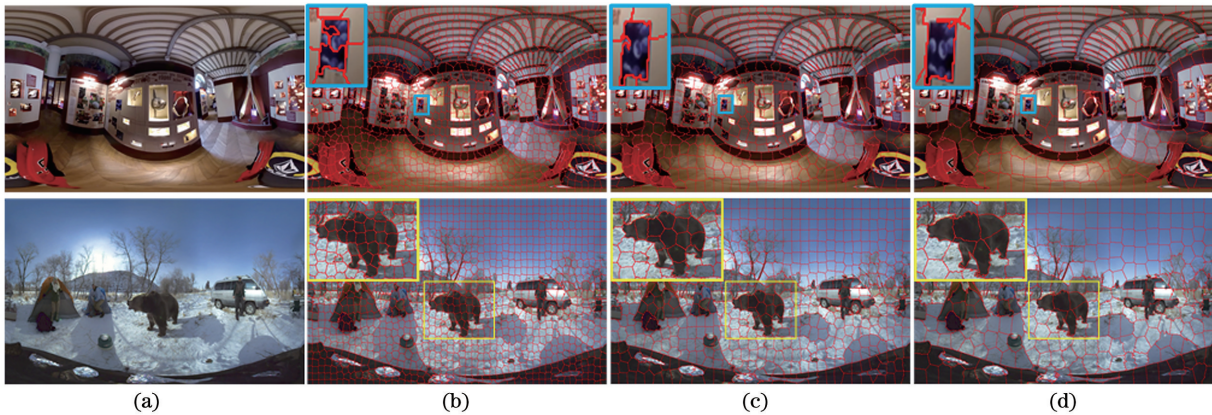


图 3 SLIC 对全局图像的超像素分割结果。(a)原始图像;(b) $K=1200$  的超像素分割结果;(c) $K=600$  的超像素分割结果;(d) $K=300$  的超像素分割结果

Fig. 3 Results of superpixel segmentation of SLIC for global images. (a) Original images; (b) results of superpixel segmentation for  $K=1200$ ; (c) results of superpixel segmentation for  $K=600$ ; (d) results of superpixel segmentation for  $K=300$

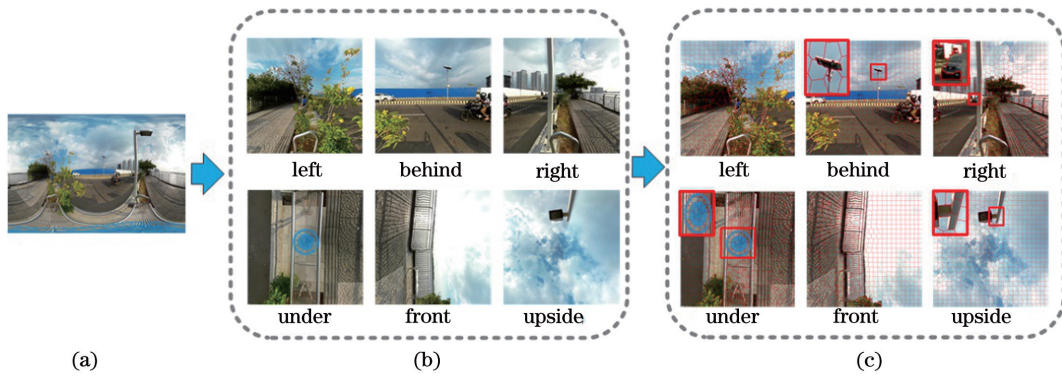


图 4 局部图像超像素分割结果。(a)原始图像;(b)CMP 6 个投影面;(c) $K=600$  的超像素分割结果图

Fig. 4 Results of superpixel segmentation of local images. (a) Original images; (b) six projection surfaces of CMP; (c) results of superpixel segmentation for  $K=600$

### 3.2 颜色相似度计算

针对全景图像的数据量更大、视野范围更广、场景也更为复杂等特点,为了减少需要考虑的颜色数目,采用了颜色量化的方法。另外,本文以超像素块作为基本元素来进行计算,以提高计算效率。如果定义超像素块  $h_i$  的颜色值为  $c_i$ ,  $c_i$  在 CIELAB 空间表示为  $c_i = (l_i, a_i, b_i)$ ,那么计算图像中各量化后颜色出现的频率就可以表示为  $f(c_i) = \frac{|c_i|}{M}$ ,其中,  $|c_i|$  为图像中像素颜色值量化为  $c_i$  的像素数量,  $M$  为图像中的像素总数。

观察现有的视觉显著模型可以发现,颜色对比度较高的区域总是会具有视觉独特性。如果一种颜色与图像中其他颜色的对比度较高,那么在视觉上,这种颜色通常就表现出较强的显著性。基于这一观察,对于任一幅彩色图像  $I$ ,可以定义某个像素块的显著性是该像素块与其他所有像素块之间的颜色对比度之和,即

$$S_{\text{Col}(h_i)} = D(h_i, h_1) + D(h_i, h_2) + \dots + D(h_i, h_K), \quad (4)$$

式中:  $D(h_i, h_K)$  为像素块  $h_i$  和  $h_K$  之间的颜色对比度。

此外,像素块  $h_i$  和  $h_j$  之间的颜色对比度又可定义为  $c_i$  和  $c_j$  的欧氏几何距离,即

$$D(h_i, h_j) = D(c_i, c_j). \quad (5)$$

对于图像  $I$  中任意两个相同颜色的像素块  $h_j$  和  $h_k$ ,有  $D(h_i, h_j) = D(h_i, h_k)$ ,显著性只与像素块  $h_i$  的颜色值  $c_i$  相关。因此,将(4)式右边相同颜色的像素进行整合,即

$$S_{\text{Col}(h_i)} = \sum_{j \neq i, j=1}^K f(c_i) D(c_i, c_j), \quad (6)$$

式中:  $f(c_i)$  是颜色值  $c_i$  在图像  $I$  中出现的频率。

本文选取若干图片进行实验,得到的结果如图 5 所示。图 5(a)是输入的彩色图像,图 5(b)是三种超像素尺度下获得的基于颜色相似度的均值图。从图 5(b)中可以看出,第一张图片的座位和第二张

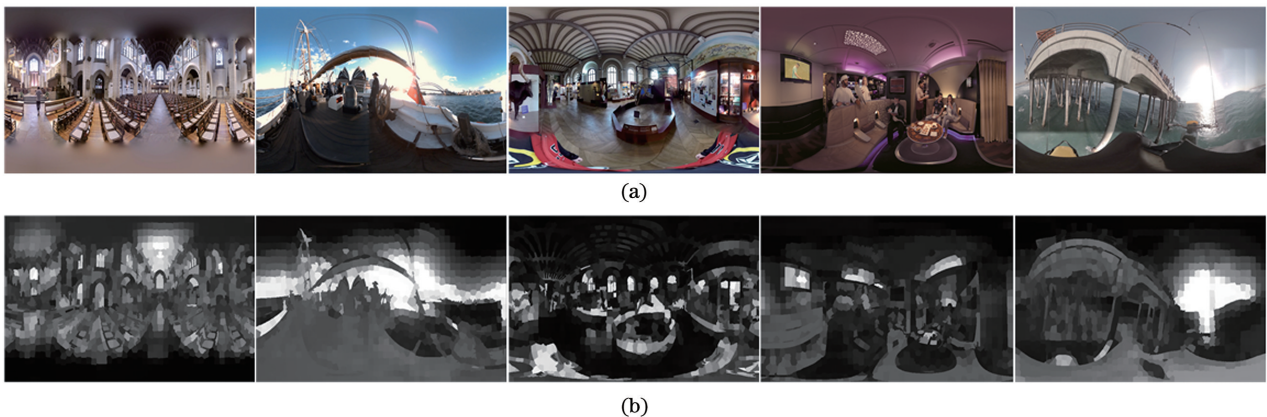


图 5 原始图像和基于颜色相似度得到的显著图。(a)原始图像;(b)基于颜色相似度得到的显著图

Fig. 5 Original images and saliency maps achieved based on color similarity. (a) Original images; (b) saliency maps achieved based on color similarity

图片中船上人物的信息,都表明了模型对于细节的处理有非常好的效果。但不足之处在于,单一地考虑整体颜色对比度的度量方法并不能很好地表征对人眼视觉注意机制具有重要影响的中心-周围机制<sup>[36]</sup>。也就是说,与当前超像素距离较近的区域对当前超像素的视觉显著的影响较大。

### 3.3 区域对比度计算

当前景区域与某些背景区域类似时,整体对比度方法可能会错误地抑制前景,而局部对比度方法可以基于前景区域与相邻区域的对比度突出显示前景。本文使用紧凑权重公式来抑制错误突出显示的背景区域,即

$$W_{ij} = \begin{cases} \exp[-D(s_i, s_j)/\sigma^2], & j \in N_i \\ 0, & j \notin N_i \end{cases}, \quad (7)$$

式中: $W_{ij}$  为像素间的权重信息; $\sigma$  为常量; $N_i$  为  $h_i$  的邻域集。

一幅图像中,突出对象通常被背景区域包围。在空间域中,与显著颜色相比,背景区域的颜色通常在整个图像上具有较大的分布。因此,在图像中显示较大颜色分布使显著的可能性更低,超像素  $h_i$  的空间方差为

$$V_{(h_i)} = \frac{\sum_{j=1}^N W_{ij} \times N_j \times |z_j - \mu_i|}{\sum_{j=1}^N W_{ij} \times N_j}, \quad (8)$$

式中: $N$  为像素点个数; $N_j$  是超像素块  $h_j$  的像素点个数; $z_j$  为  $h_j$  的质心(质心作为坐标轴的原点,记为 $[z_j^{(x_0)}, z_j^{(y_0)}]$ ),其中, $z_j = [z_j^{(x_0)}, z_j^{(y_0)}]$ ;  $\mu_i$  为空间均值,其中, $\mu_i = [\mu_i^{(x)}, \mu_i^{(y)}]$ , $[\mu_i^{(x)}, \mu_i^{(y)}]$  为像素块  $h_i$  于坐标 $[x, y]$ 区间的均值。

心理物理学研究表明,人们在拍照时通常会在图像中心附近拍摄感兴趣的物体<sup>[39]</sup>。显然,突出的对象更有可能在图像中心附近,并且背景通常分布在整个图像上。超像素与图像中心的空间距离为

$$D_{(h_i)} = \frac{\sum_{j=1}^N W_{ij} \times N_j \times |z_j - Z|}{\sum_{j=1}^N W_{ij} \times N_j}, \quad (9)$$

式中: $Z$  为图像质心的空间坐标向量, $Z = [Z^{(x_0)}, Z^{(y_0)}]$ 。

因此,基于区域对比度的显著结果的计算公式为

$$S_{Loc(h_i)} = 1 - \text{Norm}(V_{(h_i)} + D_{(h_i)}), \quad (10)$$

式中: $\text{Norm}(x)$  为将  $x$  规范化为 0 和 1 之间的函数。

图 6(a)为原始图像。图 6(b)展示的是基于区域对比度的显著图,通过三个超像素尺度的显著均值处理可得到更为平滑的显著结果。从图中可以看出,基于区域对比度计算的显著图可以较好地抑制未正确突出显示的背景区域。

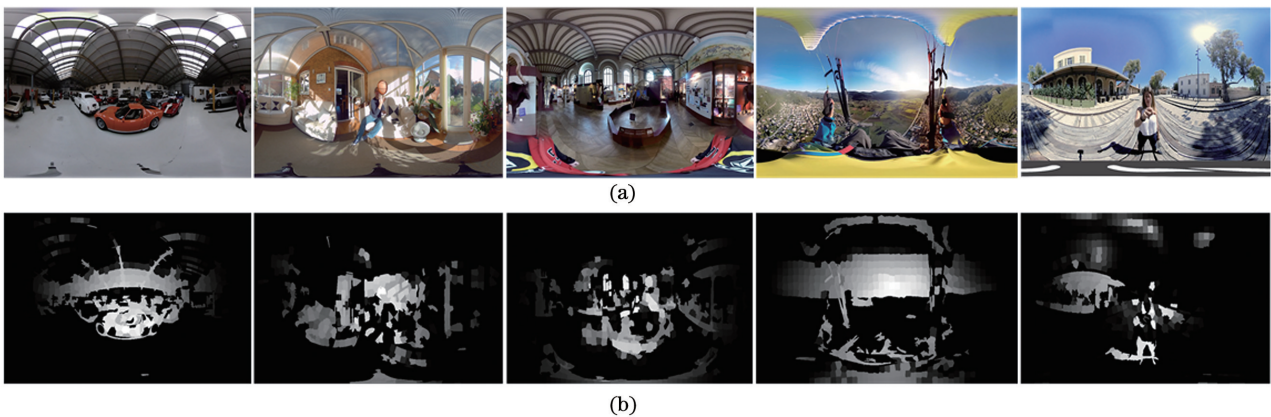


图 6 原始图像和基于区域对比度得到的显著图。(a)原始图像;(b)基于区域对比度得到的显著图  
Fig. 6 Original images and saliency maps achieved based on local contrast. (a) Original images; (b) saliency maps achieved based on local contrast

### 3.4 基于颜色相似度和区域对比度的显著性检测

若假设输入的彩色图像为  $I$ ,基于颜色对比度得到的显著图为  $S_{Col(h_i)}$ ,通过计算区域对比度得到的显著图为  $S_{Loc(h_i)}$ ,然后分配两幅显著图的权重和为 1,将两幅显著图相加得到的显著图  $S$  为

$$S = \beta S_{Col(h_i)} + (1 - \beta) S_{Loc(h_i)}, \quad (11)$$

式中: $\beta$  为权重因子,本文为了提高计算效率取  $\beta=0.5$ 。

图 7 是通过将两种算法相结合得到的显著图,可以看出,该算法对显著图有较好的改善,结合了两种

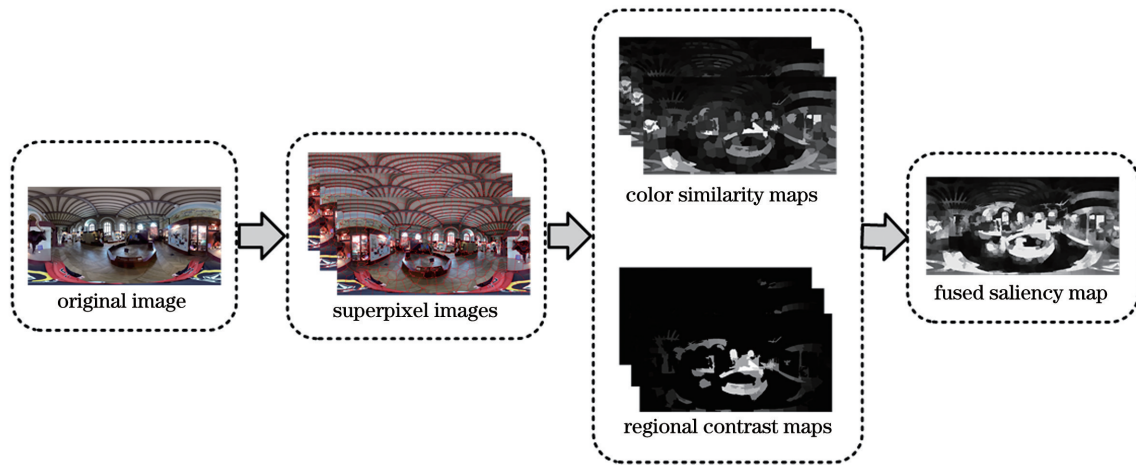


图 7 显著模型流程图

Fig. 7 Procedure of proposed saliency framework

算法的优点,得到的结果比较符合人眼的视觉效果。

### 3.5 全局视觉和局部视觉显著融合

所提出的融合方法首先需要将同一个立方体中的每 6 个局部显著图重新投影为 ERP 格式,利用简单乘积的融合方式将 ERP 图和 CMP 图进行融合。如图 8 所示, CMP\_ERP 即表示 CMP 逆投影到

ERP 格式的结果图。从图 8 可以看出,ERP 图像包含所有可用的上下文信息,而 CMP 消除了 ERP 图像顶部和底部的扭曲和边框,因此有机融合两种投影格式比单种投影的效果更好,从结果可以看出,所提出的方法不仅能够检测图像边缘显著目标,还能检测出复杂背景下的显著目标。

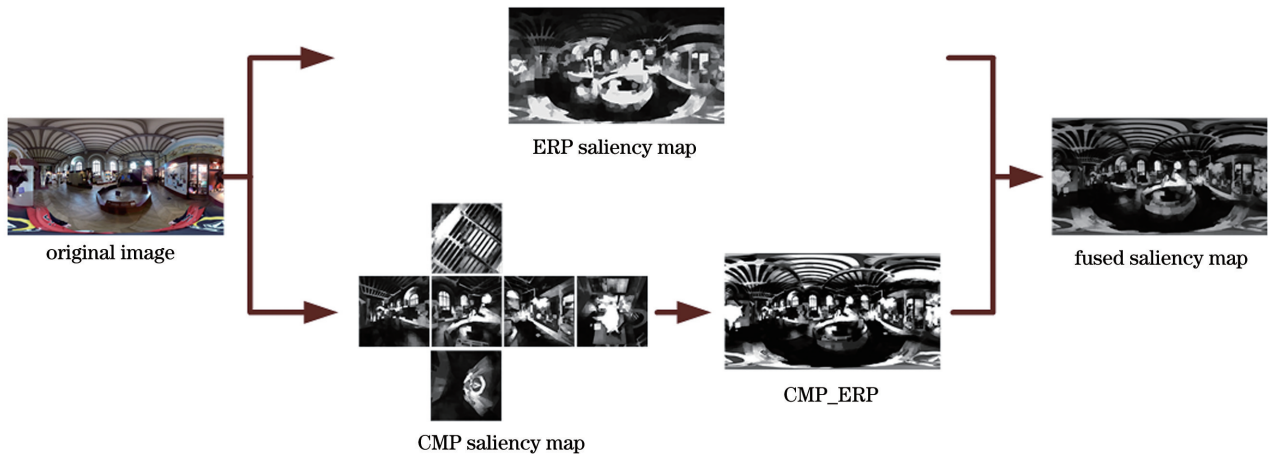


图 8 ERP 图和 CMP 图的融合显著图示意图

Fig. 8 Illustration of fused saliency map of ERP and CMP maps

对于全景图像,用户可以通过自由移动眼睛和头部进行观看,但目前存储的都是 ERP 格式,另外,考虑到图像都会存在的中心区域(即赤道区域)偏置的问题<sup>[40]</sup>,因此本文通过对等矩形平面上的像素赋予不同的权重来模拟球面的信息权重。根据 ERP 格式特征,加权窗口随着相对于赤道线的距离的减小而增加,权重大小只与纵坐标有关。对于  $M \times R$  (长度  $\times$  宽度)的等矩形投影格式的图像,加权窗口为

$$\omega(X, Y) = \cos \left[ \frac{\pi}{N} \times \left( Y + \frac{1}{2} - \frac{N}{2} \right) \right], \quad (12)$$

式中:  $(X, Y)$  代表图像的横纵坐标。

此外,研究表明人眼对场景中心区域和周围区域会表现出不同程度的视觉注意机制<sup>[41]</sup>,也就是说,那些与图像几何中心位置越近的像素点,一般具有越强的视觉显著性,随着像素点在图像中心偏离程度的增大,其视觉显著性也随之下降。本文采用高斯核函数模拟人眼视觉的中心-偏移特性:

$$CB(h_i) = \exp \left\{ - \left[ \frac{(z_i^{(x_0)} - Z^{(x_0)})^2}{2[\sigma^{(x)}]^2} \right] - \left[ \frac{(z_i^{(y_0)} - Z^{(y_0)})^2}{2[\sigma^{(y)}]^2} \right] \right\}, \quad (13)$$

式中:  $CB(h_i)$  为超像素  $h_i$  的高斯核函数;  $(z_i^{(x_0)}, z_i^{(y_0)})$  表示超像素块  $h_i$  的质心坐标值;  $(Z^{(x_0)}, Z^{(y_0)})$  为图像的质心坐标;  $[\sigma^{(X)}]^2$  和  $[\sigma^{(Y)}]^2$  分别为高斯核函数沿不同方向的方差。

## 4 结果与分析

### 4.1 数据库

为了验证本文方法的有效性和准确性,本文采用数据库<sup>[42]</sup>对算法的性能进行测试,数据库由 96 个立体 ODI 组成,包含了许多含有复杂背景和多种小物件的室内或室外全景图像。ODI 的分辨率大小在  $1920 \text{ pixel} \times 960 \text{ pixel}$  与  $4640 \text{ pixel} \times$

$2320 \text{ pixel}$  之间。

### 4.2 主观评估

与二维图像的视觉注意力模型不同,立体图像的视觉注意力模型需要考虑图像的深度特征,本文利用光流估计算法<sup>[43]</sup>得到视差结果。为了更准确地检测立体图像视觉显著区域,需要同时考虑各种低层次特征及深度感知特征对视觉注意分布的影响<sup>[44-46]</sup>。本文通过将二维图像的显著图和深度特征线性进行结合得到立体图像的视觉显著图。图 9 是立体显著图的部分实验结果,观察图 9(d)可以看出,通过加权窗口执行滤波可以获得更为平滑的显著图。



图 9 立体显著结果。(a)原始图像;(b)视差图;(c)立体显著图;(d)最终融合显著图

Fig. 9 3D saliency maps. (a) Original images; (b) disparity maps; (c) stereoscopic saliency maps; (d) final fused saliency maps

本文选取了 6 种经典的视觉显著性检测模型进行对比: GBVS 模型<sup>[26]</sup>、RR 模型<sup>[20]</sup>、RD 模型<sup>[21]</sup>、CA 模型<sup>[22]</sup>、TC 模型<sup>[23]</sup>以及 SAE 模型<sup>[24]</sup>。实验中,直接采用作者提供的源代码或者可执行程序对这些对比算法的性能进行测试。其中,GBVS、RR 及 RD 模型用于平面显著模型,TC 模型用于立体图像显著模型,SAE 模型中所提供的代码用于全景图像的显著性检测,因此,为了使测量结果具有可比性,对比模型分数都是通过融合左图像及深度信息得到的。图 10 给出了本文方法以及所有对比算法的显著图对比结果。第一列

是原始测试图像,第 2~9 列分别对应 6 种对比算法、本文方法和真实显著图。从图中可以看出,GBVS、RR 和 SAE 模型不能很好地抑制场景中的背景区域,前景显著对象不够清晰;RD、CA 和 TC 模型对于边缘显著目标的检测存在一定的局限性。从图 10 中图像 2 可以看出,对于复杂的地面背景,本文方法仍然能够准确地检测出握着降落伞的手部信息。相比于其他模型,本文方法的适用性更广,不仅能够检测图像边缘显著目标,还能检测出复杂背景下的显著目标,得到更加符合视觉感知的显著图。



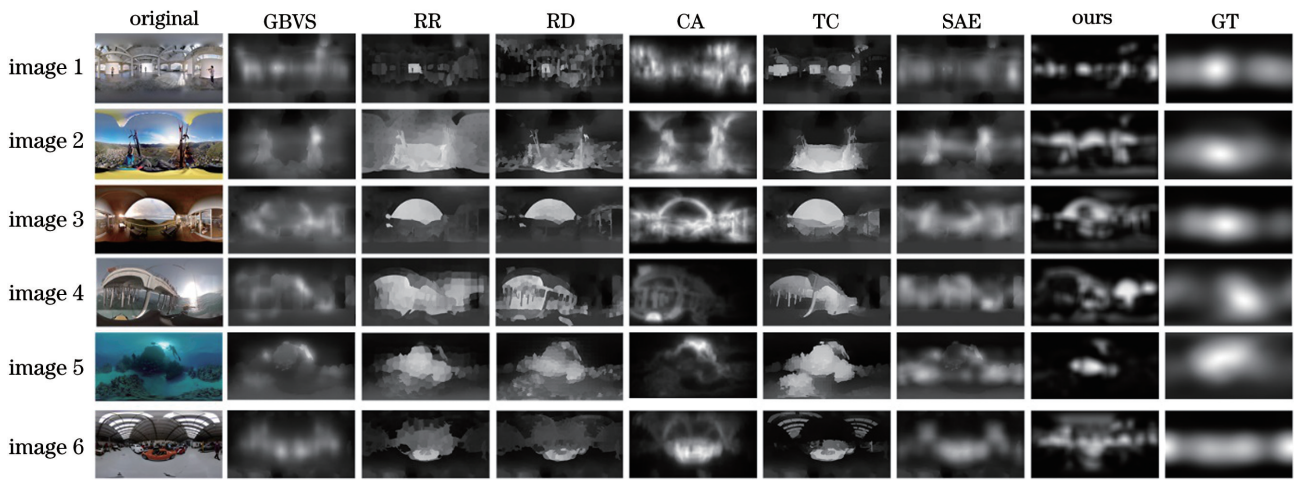


图 10 不同显著性检测模型的性能比较

Fig. 10 Comparison of performance of different saliency detection models

### 4.3 显著性检测客观评价结果

为了更加客观地评估本文方法的有效性,采用曲线面积(AUC)、相关系数(CC)、KL 散度(KLD)、准确率-召回率(PR)曲线以及综合指标(F-measure)值进行客观评价<sup>[35]</sup>。CC 为显著图和真实显著图之间的线性相关程度,KLD 用于测量这两个分布之间的差异。表 1 中的分数都是通过对数据库 ODI 中的图像求均值得到的,可以看出,本文方法的指标都高于其他显著性检测方法。另外,PR 曲线以及 F-measure 也被广泛用于文献中显著性检测模型的性能评估<sup>[21-22]</sup>。利用阈值可以将显著图

进行二值分割,显著图分为“正例”和“负例”,然后计算二值化的显著图和真实显著图之间的准确率 $[P = T_p / (T_p + F_p)]$ 和召回率 $[R = T_p / (T_p + F_n)]$ ,其中, $T_p$  为被正确分类的正例数, $F_p$  为本是负例却被错分为正例的数目, $F_n$  表示本是正例但被错分为负例的数目。准确率和召回率越高,说明模型的性能越好。图 11 给出不同显著模型在 ODI 数据库中的 PR 曲线,PR 曲线反映了分类器对正例的识别准确率和对正例的覆盖能力之间的权衡,明显可以看出在相同召回率的情况下,本文方法的准确率更高。

表 1 不同客观评价模型的性能评价指标比较

Table 1 Comparison of performance evaluation indexes of different objective evaluation models

Metric	GBVS	RR	RD	CA	TC	SAE	Ours
AUC	0.6767	0.6094	0.5810	0.6084	0.6416	0.6843	<b>0.8158</b>
CC	0.6535	0.4292	0.3506	0.4652	0.4778	0.7065	<b>0.7446</b>
KLD	0.2988	0.4476	0.5000	0.3734	0.4764	0.2488	<b>0.2321</b>

此外,本文还采用自适应阈值对显著结果进行二值化分割,自适应阈值  $T_a$  为

$$T_a = \frac{2}{M \times R} \sum_{i=1}^M \sum_{j=1}^R S_{sal}(i, j), \quad (14)$$

式中: $S_{sal}(i, j)$  为像素  $(i, j)$  的显著值大小。基于自适应阈值,显著图被分为“正例”和“负例”。然后,根据准确率和召回率计算评价模型好坏的统计量  $F$  值:

$$F = \frac{(1 + \beta^2) \cdot P \cdot R}{\beta^2 \cdot P + R}, \quad (15)$$

式中: $F$  为用于衡量分类模型好坏的指标; $\beta^2 = 0.3$ 。

从表 1 可以看出,本文提出的方法的预测指标

高于其他 6 种显著性检测方法。另外,为了比较投影格式对结果的影响,分别分析了 ERP 和 CMP 两种不同投影格式的 AUC、CC 及 KLD 结果,从表 2 可以看出,CMP 比 ERP 的预测效果更好一些。结

表 2 ERP 和 CMP 投影格式的性能指标比较

Table 2 Comparison of evaluation performance of ERP and CMP projection formats

Metric	ERP	CMP
AUC	0.7665	0.7750
CC	0.6434	0.6863
KLD	0.2680	0.2530

合表 1 和表 2,可以看出融合 ERP 的全局信息和 CMP 局部信息的方法的显著性能比基于单种投影格式的方法的预测效果更好。

图 11 和图 12 分别给出不同显著图模型在 ODI 数据库上的 PR 曲线及准确率、召回率、F-measure 柱形图。明显可以看出在相同召回率的情况下,本文方法的准确率更高。并且,从图 12 可以看出,GBVS 和 SAE 两种方法的准确率与本文方法相差不多,但是召回率较低。而本文方法在较高准确率的情况下,也能保证较高的召回率。该实验表明,本文方法的总体效果最好,可准确地检测图像中的显著目标。

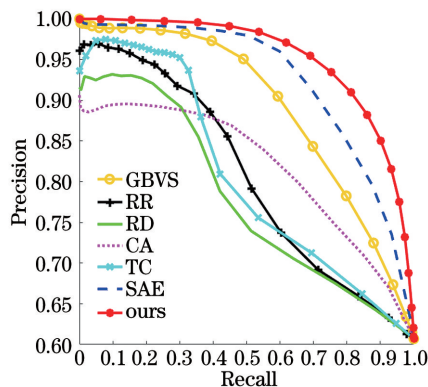


图 11 不同显著模型的准确率-召回率曲线对比

Fig. 11 Precision-recall curves of different saliency models

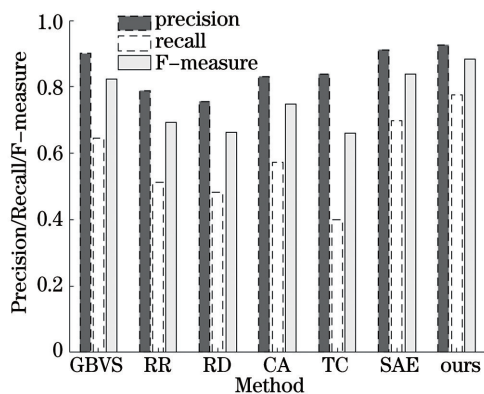


图 12 不同显著模型的 precision、recall、F-measure 值比较

Fig. 12 Precision, recall, and F-measure of different saliency models

但是,本文方法也存在一定的局限性。由于目前在立体全景领域比较成熟的显著预测模型还比较少,实验中的 6 种对比方法是分别基于平面、立体或全景图像模型并结合深度信息获得的。另外,针对模型而言,倘若场景中的显著区域与背景区域的颜色特征差异较小或者对于一些特殊的背景比较复杂的图像,本文方法的实验效果并不十分理想。因此,接下来的工作不仅需要考虑低层次特征还应该融合

高层次语义特征以提取显著特征。

## 5 结 论

提出了一种立体全景显著性的预测模型,该模型结合颜色相似度及区域对比度这两种底层特征进行显著性检测,通过融合全局视觉注意力和局部视觉显著性得到二维图像的显著性结果。最后,针对全景投影的特点进行加权滤波并且结合深度感知信息得到最终的立体全景显著图。实验结果表明,该模型在 ODI 数据库中具有较好的性能。在接下来的工作中,考虑基于全景投影多视口来探究立体全景显著性检测。另外,人眼视觉舒适度也是需要考虑的问题,深入探索引起视觉不舒适的因素并设计相关的评价模型有助于立体全景图像/视频技术从采集到渲染等各环节的优化,从而对研究更高视觉舒适度的立体全景图像/视频系统提供参考。

## 参 考 文 献

- [1] Papadakis P, Pratikakis I, Theoharis T, et al. PANORAMA: a 3D shape descriptor based on panoramic views for unsupervised 3D object retrieval [J]. International Journal of Computer Vision, 2010, 89(2/3): 177-192.
- [2] Peters R J, Iyer A, Itti L, et al. Components of bottom-up gaze allocation in natural images [J]. Vision Research, 2005, 45(18): 2397-2416.
- [3] Palmer S E. Vision science: photons to phenomenology [J]. The Quarterly Review of Biology, 2001, 76(4): 523-524.
- [4] Chang K Y, Liu T L, Chen H T, et al. Fusing generic objectness and visual saliency for salient object detection[C]//2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 914-921.
- [5] Fang Y M, Chi Z, Li J, et al. Visual attention modeling for stereoscopic video [C] // 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), July 11-15, 2016, Seattle, WA, USA. New York: IEEE Press, 2016: 1-6.
- [6] Ji C, Huang X B, Cao W, et al. Salient region detection of images based on deep learning [J]. Laser & Optoelectronics Progress, 2019, 56(9): 091007.
- [7] 纪超, 黄新波, 曹雯, 等. 基于深度学习的图像显著区域检测[J]. 激光与光电子学进展, 2019, 56(9): 091007.
- [7] Li S Q, Zeng C, Fu Y, et al. Optimizing multi-graph learning based salient object detection [J]. Signal

- Processing: Image Communication, 2017, 55: 93-105.
- [8] Li Y D, Xu X P. Video saliency detection method based on spatiotemporal features of superpixels [J]. Acta Optica Sinica, 2019, 39(1): 0110001.  
李艳荻, 徐熙平. 基于超像素时空特征的视频显著性检测方法 [J]. 光学学报, 2019, 39(1): 0110001.
- [9] Cui L Q, Chen J J, Qi B H, et al. Saliency detection based on background suppressing and foreground updating [J]. Laser & Optoelectronics Progress, 2020, 57(2): 021002.  
崔丽群, 陈晶晶, 齐博华, 等. 基于背景抑制和前景更新的显著性检测 [J]. 激光与光电子学进展, 2020, 57(2): 021002.
- [10] Liu T, Yuan Z J, Sun J, et al. Learning to detect a salient object [C] // IEEE Transactions on Pattern Analysis and Machine Intelligence. New York: IEEE Press: 353-367.
- [11] Yang J M, Yang M H. Top-down visual saliency via joint CRF and dictionary learning [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 2296-2303.
- [12] Torralba A, Oliva A, Castelhano M S, et al. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search [J]. Psychological Review, 2006, 113 (4): 766-786.
- [13] Zhao H, An W S. Image salient object detection combined with deep learning [J]. Laser & Optoelectronics Progress, 2018, 55(12): 121003.  
赵恒, 安维胜. 结合深度学习的图像显著目标检测 [J]. 激光与光电子学进展, 2018, 55(12): 121003.
- [14] Qu H C, Tian X R, Liu L M, et al. Image compression method based on multi-scale saliency region detection [J]. Journal of Image and Graphics, 2020, 25(1): 31-42.  
曲海成, 田小容, 刘腊梅, 等. 多尺度显著区域检测图像压缩 [J]. 中国图象图形学报, 2020, 25(1): 31-42.
- [15] Zhang Y Y, Ge H W. Image saliency detection based on graph and multi-feature diffusion [J]. Laser & Optoelectronics Progress, 2020, 57(4): 041020.  
张莹莹, 葛洪伟. 基于图和多特征传播的图像显著性检测 [J]. 激光与光电子学进展, 2020, 57(4): 041020.
- [16] Fang Y M, Chen Z Z, Lin W S, et al. Saliency detection in the compressed domain for adaptive image retargeting [J]. IEEE Transactions on Image Processing, 2012, 21(9): 3888-3901.
- [17] Yang S, Jiang Q P, Lin W S, et al. SGDNet: an end-to-end saliency-guided deep neural network for No-reference image quality assessment [C] // MM '19: Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 1383-1391.
- [18] Guo C L, Zhang L M. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression [J]. IEEE Transactions on Image Processing, 2010, 19 (1): 185-198.
- [19] Lei J J, Zhang H L, You L, et al. Evaluation and modeling of depth feature incorporated visual attention for salient object segmentation [J]. Neurocomputing, 2013, 120: 24-33.
- [20] Li C Y, Yuan Y C, Cai W D, et al. Robust saliency detection via regularized random walks ranking [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 2710-2717.
- [21] Zhou L, Yang Z H, Yuan Q, et al. Salient region detection via integrating diffusion-based compactness and local contrast [J]. IEEE Transactions on Image Processing, 2015, 24(11): 3308-3320.
- [22] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection [C] // IEEE Transactions on Pattern Analysis and Machine Intelligence. New York: IEEE Press: 1915-1926.
- [23] Wang W G, Shen J B, Yu Y Z, et al. Stereoscopic thumbnail creation via efficient stereo saliency detection [J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(8): 2014-2027.
- [24] Xia C, Qi F, Shi G M. Bottom-up visual saliency estimation with deep autoencoder-based sparse reconstruction [J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27 (6): 1227-1240.
- [25] Maugey T, Le Meur O, Liu Z. Saliency-based navigation in omnidirectional image [C] // 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSp), October 16-18, 2017, Luton, UK. New York: IEEE Press, 2017: 1-6.
- [26] Harel J, Koch C, Perona P. Graph-based visual saliency [C] // Advances in Neural Information Processing Systems 19, Proceedings of the 2006 Conference. Cambridge: MIT Press, 2017: 545-552.
- [27] Hou X D, Harel J, Koch C. Image signature: highlighting sparse salient regions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 194-201.
- [28] Garcia-Diaz A, Fdez-Vidal X R, Pardo X M, et al.

- Saliency from hierarchical adaptation through decorrelation and variance normalization [J]. *Image and Vision Computing*, 2012, 30(1): 51-64.
- [29] Riche N, Mancas M, Duvinage M, et al. RARE2012: a multi-scale rarity-based saliency detection with its comparative statistical analysis [J]. *Signal Processing: Image Communication*, 2013, 28(6): 642-658.
- [30] Zhang J M, Sclaroff S. Saliency detection: a Boolean map approach [C] // 2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 153-160.
- [31] Zhu X X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 2879-2886.
- [32] Lebreton P, Raake A. GBVS360, BMS360, ProSal: extending existing saliency prediction models from 2D to omnidirectional images [J]. *Signal Processing: Image Communication*, 2018, 69: 69-78.
- [33] de Abreu A, Ozcinar C, Smolic A. Look around you: saliency maps for omnidirectional images in VR applications [C] // 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), May 31-June 2, 2017, Erfurt, Germany. New York: IEEE Press, 2017: 1-6.
- [34] Xu M, Yang L, Tao X M, et al. Saliency prediction on omnidirectional images with generative adversarial imitation learning [EB/OL]. [2021-07-22]. [https://www.researchgate.net/publication/332439333\\_Saliency\\_Prediction\\_on\\_Omnidirectional\\_Images\\_with\\_Generative\\_Adversarial\\_Imitation\\_Learning](https://www.researchgate.net/publication/332439333_Saliency_Prediction_on_Omnidirectional_Images_with_Generative_Adversarial_Imitation_Learning).
- [35] Chen D W, Qing C M, Xu X M, et al. SalBiNet360: saliency prediction on 360° images with local-global bifurcated deep network [C] // 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), March 22-26, 2020, Atlanta, GA, USA. New York: IEEE Press, 2020: 92-100.
- [36] Shao F, Jiang Q P, Jiang G Y, et al. Prediction of visual discomfort of stereoscopic images based on saliency analysis [J]. *Optics and Precision Engineering*, 2014, 22(6): 1631-1638.  
邵枫, 姜求平, 蒋刚毅, 等. 基于显著性分析的立体图像视觉舒适度预测 [J]. *光学精密工程*, 2014, 22(6): 1631-1638.
- [37] Achanta R, Shaji A, Smith K, et al. SLIC superpixel compared to state-of-the-art superpixel methods [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282.
- [38] Cheng M, Zhang G, Mitra N, et al. Global contrast based salient region detection [C] // In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2011: 409-416.
- [39] Battisti F, Baldoni S, Brizzi M, et al. A feature-based approach for saliency estimation of omnidirectional images [J]. *Signal Processing: Image Communication*, 2018, 69: 53-59.
- [40] Safavi S M, Sundaram S M, Gorji A H, et al. Application of infrared scanning of the neck muscles to control a cursor in Human-Computer Interface [J]. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, July 11-15, 2017, Jeju, Korea (South), South Korea. New York: IEEE Press, 2017: 787-790.
- [41] Cui L Q, Yang Z Z, Duan T L, et al. Saliency object detection method based on complex prior knowledge [J]. *Laser & Optoelectronics Progress*, 2020, 57(10): 101019.  
崔丽群, 杨振忠, 段天龙, 等. 基于复合先验知识的显著性目标检测方法 [J]. *激光与光电子学进展*, 2020, 57(10): 101019.
- [42] Croci S, Knorr S, Smolic A. Sharpness mismatch detection in stereoscopic content with 360-degree capability [C] // 2018 25th IEEE International Conference on Image Processing (ICIP), October 7-10, 2018, Athens, Greece. New York: IEEE Press, 2018: 1423-1427.
- [43] Sun D Q, Roth S, Black M J. Secrets of optical flow estimation and their principles [C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 2010: 2432-2439.
- [44] Wang A H, Yu M, Peng Z J, et al. A quality assessment method of stereoscopic images based on human visual system [J]. *Opto-Electronic Engineering*, 2011, 38(1): 134-141.  
王阿红, 郁梅, 彭宗举, 等. 一种基于人眼视觉特性的立体图像质量客观评价方法 [J]. *光电工程*, 2011, 38(1): 134-141.
- [45] Fang Y M, Wang J L, Narwaria M, et al. Saliency detection for stereoscopic images [C] // IEEE Transactions on Image Processing. New York: IEEE Press, 2014: 2625-2636.
- [46] Gutiérrez J, David E, Rai Y, et al. Toolbox and dataset for the development of saliency and scanpath models for omnidirectional/360° still images [J]. *Signal Processing: Image Communication*, 2018, 69: 35-42.