

# 基于多角度分割的 360 全景图的显著性检测

苏 群<sup>1,2</sup>, 林春雨<sup>1,2</sup>, 赵 耀<sup>1,2</sup>, 李雅茹<sup>1,2</sup>, 刘美琴<sup>1,2</sup>

(1. 北京交通大学计算机与信息技术学院信息所, 北京 100044;

2. 现代信息科学与网络技术北京市重点实验室, 北京 100044)

**摘 要:** 与常规 2D 图像不同, 360 全景图包含当前空间的全部视觉信息, 因而在视频监控和虚拟现实等领域有着广泛的应用, 然而用户在某一时刻只能观看到一定的视角, 因此, 360 全景图的显著性区域检测对于视角预测至关重要。为此, 提出了多角度分割的 360 全景图的显著性检测。首先将全景图进行多角度分割, 将分割结果分别投影到立方体上去去除一定畸变; 然后对每个立方体面通过稠密稀疏重建进行显著性计算; 最后再将每个面的显著图投影到经纬映射方式的矩形上, 进行多角度融合以获得最终的显著图。通过人工标注的全景图显著区域进行实验对比, 结果表明该算法可以准确检测出 360 全景图的显著区域, 并优于当前先进算法。

**关 键 词:** 360 全景图; 显著性检测; 多角度分割; 稠密重构误差; 稀疏重构误差

中图分类号: TN919.81

DOI: 10.11996/JGj.2095-302X.2018061055

文献标识码: A

文章编号: 2095-302X(2018)06-1055-07

## Salient Detection of 360 Panorama Based on Multi - Angle Segmentation

SU Qun<sup>1,2</sup>, LIN Chunyu<sup>1,2</sup>, ZHAO Yao<sup>1,2</sup>, LI Yaru<sup>1,2</sup>, LIU Meiqin<sup>1,2</sup>

(1. Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China;

2. Beijing Key Laboratory of Advanced Information Science and Network, Beijing 100044, China)

**Abstract:** Unlike conventional 2D images, 360 panorama contains all the visual information of the current space, so it has a wide range of applications in video surveillance and virtual reality. However, a certain angle is available at a certain time. Therefore, the significant region detection of the 360 panorama is very important to visual angle prediction. To solve this problem, we propose a multi-angle segmentation based 360 panoramic image saliency detection. Firstly, the panoramic images are cut at multiple angles, and the segmentation results are projected to the cube to remove certain distortion. Then, the salient calculation is conducted for each cube surface through dense and sparse reconstruction. Finally, the saliency images of each surface are projected to the rectangular of the warp and weft mapping, and multi-angle fusion is made to obtain the final salient figure. The results of the 360 panorama test by manual annotation show that the algorithm can accurately detect the saliency and is better than the other methods for the saliency detection of the 360 panorama.

**Keywords:** 360 panorama; saliency detection; multi-angle segmentation; dense reconstruction error; sparse reconstruction error

视觉显著性与生物系统独特的感知机制有关, 使场景中的某个区域脱颖而出, 引起人们的注意。

而人眼所固有的强大能力是迅速捕捉场景中最明显的区域, 并将其传递到高层次的视觉皮层, 这种

收稿日期: 2018-04-16; 定稿日期: 2018-07-24

基金项目: 国家自然科学基金项目(61772066); 中央高校基本科研业务费专项资金项目(2018JBM011)

第一作者: 苏 群(1993-), 女, 山西运城人, 硕士研究生。主要研究方向为多媒体处理。E-mail: 16125146@bjtu.edu.cn

通信作者: 林春雨(1979-), 男, 辽宁绥中人, 副教授, 博士。主要研究方向为多媒体信息处理。E-mail: 16125146@bjtu.edu.cn

注意选择机制大大降低了视觉分析的复杂性,从而使人类视觉系统在复杂场景中识别物体的效率显著提高。许多生物学上的模型已经用来解释人类和动物的认知过程。高效显著性检测在许多计算机视觉任务中起着重要的预处理作用,包括图像分割、图像分类、目标识别与检测、图像压缩等。

以往关于显著性检测的研究主要集中在传统图像上。现有的显著性大多采用对比机制,主要分为全局对比和局部对比。局部对比度的算法有 ITTI 等<sup>[1]</sup>模拟和预测人类对图像的视觉定位研究; RAHTU 等<sup>[2]</sup>通过测量整个图像上的滑动窗口的中心-周围对比度进行显著性检测。全局对比的算法有 CHENG 等<sup>[3]</sup>提出的基于空间加权特征差异的全局对比度显著性算法; BORJI 和 ITTI<sup>[4]</sup>提出全局显著性计算与整个场景中出现的区域概率成反比。ACHANATA 等<sup>[5]</sup>的频谱残差法基于全局对比度,通过分析频域中显著区域的特征建立一幅显著图。文献[6]则用稠密稀疏重构(dense and sparse reconstruction, DSR)进行显著性检测,首先利用简单线性迭代聚类(simple linear iterative clustering, SLIC)算法对彩色图像进行多尺度分割,并取图像边界的超像素作为背景集;接着对每个超像素计算稠密、稀疏重构误差;然后,对上述两个重构误差利用  $K$  均值聚类的方法进行了基于上下文的平滑处理;最后,用贝叶斯准则和高斯模型进行多尺度融合,得到了一个全分辨率的显著图。

上述显著性算法主要针对传统图像,但随着 VR 技术的快速发展,360 全景图广泛应用在各个领域。比如,旅游景区的 360 全景展示,用户足不出户,带上头戴设备 HMD (head mount display)可观看整个场景空间的所有图像信息,身临其境。但通过 HMD 观看 360 全景图时,如何能快速定位到感兴趣区域呢?因此,获取全景图的视觉注意力机制变得越来越重要。与传统图像相比,全景图的分辨率更高,宽度是高度的 2 倍;一个典型的全景图像包括天空、山、水、地面等几个均匀的背景区域且分布水平细长,以及多个不同特征和尺寸的前景物体,并任意分布在图像中<sup>[7]</sup>。目前大多数全景图是以 2:1 的平面矩形图<sup>[8]</sup>(equi-rectangular projection, ERP)存储和展现的(图 1),其本质是将球面图的上下两极拉伸成圆柱体,然后展开成平面矩形,但这种投影方式使全景图上下两边发生畸变,用普通的显著性算法并不能准确的检测出 360 全景图的显著性。另一种常见的立方体投影(cube)方式,是通过

把球面上的点径向投影到外切球体的立方体的 6 个面上(图 2),然后展开 6 个面而获得的,可有效地去除上下两边畸变的影响<sup>[9]</sup>。综上,本文提出了一种基于多角度分割的 360 全景图的显著性算法。考虑到 360 全景图没有边界,先将 360 全景图进行多角度切割;并改用立方体映射的方法将全景图映射到 6 个面上,分别对每个面处理;DSR 采用边界上的区域作为背景,但是在全景图中没有边界,本文采支持向量机(support vector machine, SVM)分类器训练模型来获取背景集;然后用稠密稀疏重建误差来计算显著性;最后将 6 个面的显著图重新映射到矩形图上并进行多角度融合获得最终的显著图。

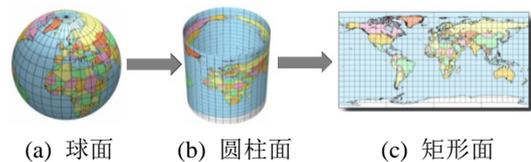


图 1 矩形平面投影示意图

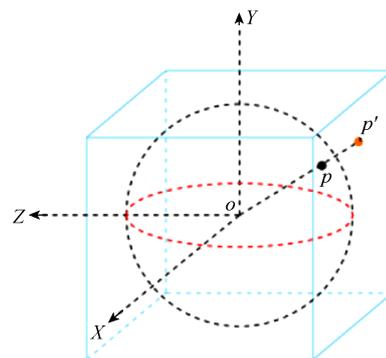


图 2 立方体投影示意图

## 1 方法

本文由以下部分组成:

- (1) 360 全景图的多角度分割及立方体映射;
- (2) 用 SVM 分类器训练模型获取背景集;
- (3) 用稠密稀疏重建误差计算显著图,并多角度融合得到最终的显著图;

(4) 为了验证显著性的准确性,人工标注了 20 个真值图供对比参考。

### 1.1 多角度分割及立方体映射

与普通图像不同,360 全景图通常包括多个不同特征、不同尺寸的前景目标,且分布在图像的不同区域。ERP 格式的全景图,其上下边界有严重畸变,通过立方体映射可以去除畸变的影响;使用多角度切割是因为全景图没有边界,对于一幅 ERP 格式的全景图,沿着经线切开后,左边图像移到右

边,重新拼接形成同一幅全景图。图3显示了多角度分割以及立方体映射的过程。本文将图3(a)中ERP格式的全景图从不同角度(如:  $-90^\circ$ ,  $90^\circ$ ,

$180^\circ$ ,  $0^\circ$ )切开重组成为4种同一场景下的全景图(图3(b)),再将其分别映射到立方体6个面上。以从 $180^\circ$ 切开为例(图3(c)),显示了立方体的6个面。

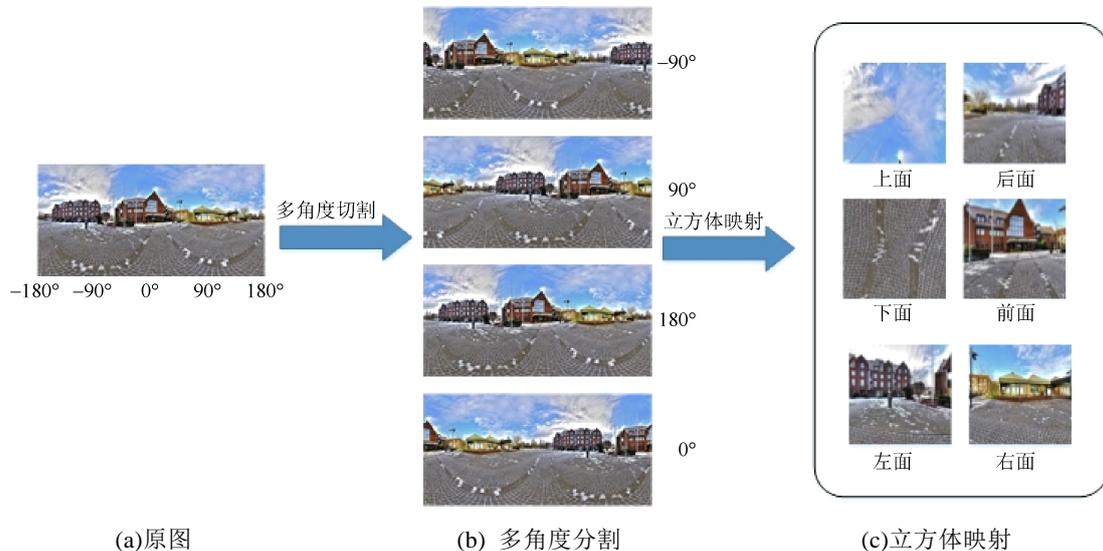


图3 多角度切割及立方体映射

## 1.2 背景模型

为了减少ERP映射方式带来的两极失真严重问题,先将ERP格式的全景图投影到立方体上,然后对每个面进行显著性测量。为了更准确获得图片的结构信息,先用SLIC算法将原彩色图分割成多个均匀紧凑的区域,即多尺度超像素分割,超像素的个数分别为: $K=50, 100, 150, 200, 250, 300, 350, 400$ ,对应的紧凑性分别为: $t=10, 10, 20, 20, 25, 25, 30, 30$ 。本文选用输入图像的颜色统计特征计算显著性,文献[4]指出,两种彩色空间Lab和RGB对显著性检测同样重要,为了有一个更好的全局对比度,本文采用Lab和RGB平均颜色特征以及最大、最小、平均灰度值表示每个超像素,即 $x=\{L, a, b, R, G, B, max, min, avg\}$ ,则整幅图像可表示为 $X=\{x_1, x_2, x_3, \dots, x_N\} \in R^{D \times N}$ , $N$ 为分块个数; $D$ 为特征维度。然后用SVM分类器对已公布的数据集MSRA10K进行训练,得到一个模型,将前景与背景分开。将背景集包含的每个超像素分块的 $D$ 维特征提取为 $b$ ,构造背景模板集合 $B=[b_1, b_2, b_3, \dots, b_M]$ ,其中 $M$ 为图像背景分块的个数。图4(a)展示了通过SVM分类器得到的背景模型。为了能有一个更好的全局对比度,本文在原来DSR的基础上,在表示每个超像素的时候加入了新的颜色特征<sup>[10-11]</sup>,即最大、最小、平均灰度值,并用SVM分类器建立背景集。

## 1.3 重构误差的显著性度量

本文的显著性检测是基于上述背景模型的稠密稀疏重构误差估计的过程。稀疏表示在过去近20年来主要应用于信号处理方面,其本质是构建一个超完备字典,用字典中的基函数来线性表示信号样本,目的是尽可能用较少的元素来获取信号中包含的信息。近年来,其在计算机视觉中也有广泛应用,在图像处理中,可以理解为任何图像都能被有限个自然图像块线性表示。但稀疏表示在计算显著性时存在一些不足<sup>[3]</sup>。例如,部分前景会被分割到背景集中,是因为前景目标出现在了图像的边界处,而此时,若用该背景集对前景目标进行稀疏表示,会由于已经包含了部分前景,从而使得重构误差变小,显著性接近于0,将该前景误检测成背景;同样,其他部分的显著精度也会降低;如图4(b)的第二行第一个图,部分前景目标未能检测出来;而稠密表示是在特征空间中对背景集建模,从而计算重构误差,但由于背景集有限,在进行显著性计算时会产生较大的误差,如图4(b)第一行第一个图,受背景影响较大,但能很好地处理位于图像边界处的前景目标。总的来说,稀疏重构误差对于处理复杂的背景更加鲁棒,而稠密重建误差能更精确地处理图像边界处的分块,稠密稀疏重建误差在测量显著性方面是相辅相成的。因此,本文用稠密稀疏表示来重建误差。

首先根据上述的背景集构造背景字典,基于背

景字典重建所有图像区域, 分别用稠密和稀疏表示, 将各个图像分块投影在背景字典上, 并获得重建系数, 然后计算该系数下的重构误差, 最后将重建误差归一化到 $[0, 1]$ 的范围。

对于稠密重构误差的计算, 基于背景模型的重建误差较大的分块更有可能为前景。本文用主成份分析(principal component analysis, PCA)对背景模型  $B = [b_1, b_2, b_3, \dots, b_M]$ ,  $B \in R^{D \times M}$  进行降维提取主要特征, 计算其特征向量  $U_B = [U_1, U_2, U_3, \dots, U_D]$ , 稠密重构误差为

$$\beta_i = U_B^T (X_i - \bar{X}) \quad (1)$$

$$\varepsilon_i^d = \|X_i - (U_B \beta_i + \bar{X})\|_2^2 \quad (2)$$

其中,  $i \in [1, N]$ ,  $\beta_i$  为第  $i$  个分块的重建系数;  $\bar{X}$  为  $X$  的均值;  $\varepsilon_i^d$  为分块  $i$  的稠密重构误差。

稀疏重构误差, 则利用基于稀疏表达的分类 (sparse representation based classifier, SRC) 的基本原理, 以背景样本集为字典对超像素进行稀疏重构, 其重构公式为

$$\alpha_i = \arg \min_{\alpha_i} \|X_i - B\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \quad (3)$$

$$\varepsilon_i^s = \|X_i - B\alpha_i\|_2^2 \quad (4)$$

在已获得 2 个重构误差的基础上, 采取第一步的处理, 如图 4(b) 所示, 用  $K$  均值聚类算法对所得的显著图进行基于上下文的误差传播机制处理, 然后对特征空间中的相邻区域加权均值处理, 采用多尺度分割超像素及重建误差, 通过高斯模型来获得像素级的显著性, 最后得到 6 个面的全分辨率的显著图, 如图 4(c) 所示。接下来第二步处理, 由于用

户观看全景视频时, 每个面被看到的概率不一样, 因此, 本文先对每个面的显著图通过显著部分在该面上的面积占比与该显著图的最大最小值的差的归一化加权求权重。具体如下:

(1) 根据上述步骤得到的每个面的显著图计算出最大值与最小值的差并归一化到 $[0, 1]$ , 归一化后的值为  $d_i$ , 其中  $i \in [1, 6]$ ; 假设通过立方体映射后, 每个面的面积为  $S$ , 设定一个二值显著性阈值  $T$ ,  $T \in [0, 255]$ , 采用自适应阈值法分别对以上求出来的 6 个面的显著图进行二值化, 然后求出每个面上白色区域的面积为  $S_i$ , 其中,  $i \in [1, 6]$ , 则第  $i$  个面的显著图的权重  $W_i$  的定义为

$$W_i = \frac{S_i}{S} + d_i \quad (5)$$

(2) 得到每个面的权重后, 把每幅图的 6 个面按权重比例重新映射成 ERP 格式的全景图, 最后, 将同一场景的 4 种显著图重组成从  $0^\circ$  切开的图像显示, 并求平均得到最终的显著图(图 5)。观看全景图时, 图片内容一定程度上决定了用户的观看视角, 即每个区域被看到的概率不相等, 在不同分割情况下的观看情况也不一样, 若有历史的头部运动数据, 则可根据历史数据建立模型得出每种分割情况的观看权重, 然后根据各自的权重去融合, 本文方法和数据集中没有历史头部数据问题, 此时平均法是最简单有效的方案。现有的显著性算法大多认为前景目标位于图像中间, 用传统的显著性算法处理全景图的 6 个面, 通过立方体投影到全景图后会产生明显的边界效应, 而通过多角度融合, 可以较好地平滑边界处的灰度值。

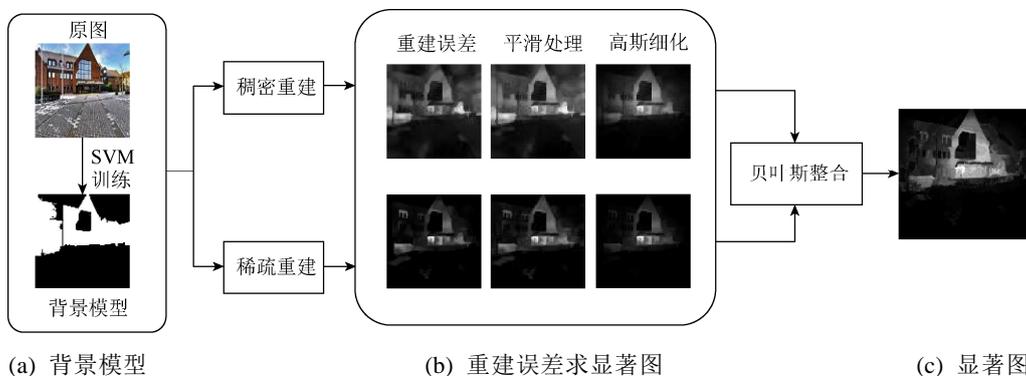


图 4 显著性检测过程图

## 2 实验结果

### 2.1 实验细节

为了寻求一个最优的多角度分割, 本文对 20

张全景图在不同分割次数下进行实验, 其中, 分割次数分别为  $M=1, 2, 4, 6, 8, 10$ , 得到的显著图为  $Sal_m$  ( $m=1, 2, 4, 6, 8, 10$ ), 然后计算出人工标注的真实值与实验结果的差值并求出每个像素

的平均差值。差值越小, 说明显著图与人工标注越接近, 则该分割次数下得到的结果更精确。统计结果如图 6 所示, 分割次数为 4 的时候差值最小, 约

为 26.2, 可以计算出平均每个像素的准确率为 0.897 ( $1-26.2/255=0.897$ ), 因此, 本文将全景图从 4 种不同角度分割。

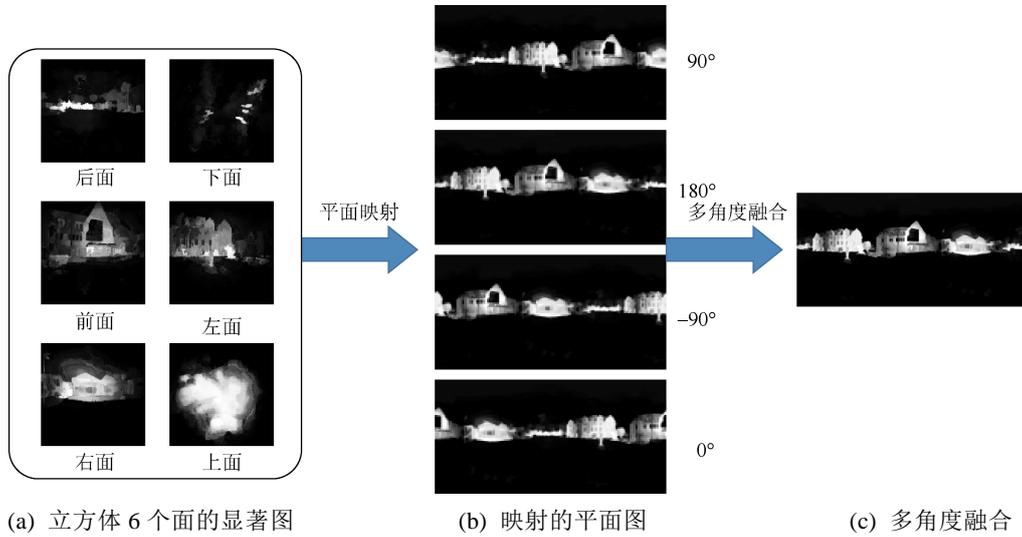


图 5 合成显著图

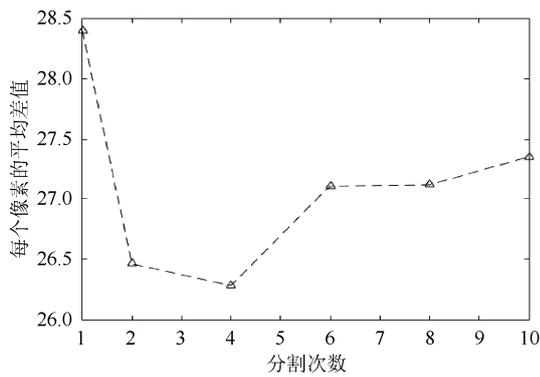


图 6 分割尺度统计结果图

将立方体映射后的 6 个面对应到 2:1 的平面图上后, 大概位置如图 7 所示。图中的 U(up)、D(down)、L(left)、F(front)、R(right)、B(back)分别对应立方体的上、下、左、前、右和后边。蓝色边界及蓝色字体表示分割前; 红色边界及红色字体表示分割重组后。可以观察到, 当进行 4 种不同角度分割时, 沿着第一条虚线切开后, 重新拼到图像的右边, 位于原来边界上的目标 A 经过分割重组后位于 R 面的中央, 因此, 通过显著性检测后的多角度融合可以有效去除边界的影响。若分割次数过多, 会造成过多

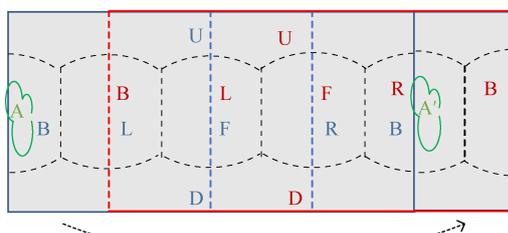


图 7 分割前后对比图

的冗余, 而且计算成本大。综合来看, 从 4 种不同角度分割最适宜。

### 2.2 评价指标

显著性的实验结果评估, 包括正确率 (Precision)-召回率 (Recall) 以及  $F$  值 ( $F$ -measure)。其中, Precision 的值表示检测出的正确显著像素占提取出的所有区域的像素的比例, Recall 的值表示检测出的正确显著像素占标准显著图中所有应被检测出的像素的比例, PR 曲线表明, 对于不同的二值显著性阈值 (0-255), 该实际显著图与标注数据 (Ground Truth) 的吻合程度如何。正确率 (Precision)-召回率 (Recall) 定义为

$$Precision = \frac{\sum S \cap G}{\sum S} \quad (6)$$

$$Recall = \frac{\sum S \cap G}{\sum G} \quad (7)$$

其中,  $S$  为不同阈值的二值显著像素;  $G$  为真实值;  $\sum$  为所有像素之和。  $F$ -measure 是由 Precision 和 Recall 的加权调和平均值计算出的整体性能评价指标, 定义为

$$F_{\beta} = \frac{(1 + \beta^2) Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (8)$$

本文用  $\beta^2=0.3$  来分配正确率与召回率的权重。

### 2.3 比较

本文从网上搜集了 20 张 360 全景图, 并进行了人工标注。为了验证本文算法的有效性, 把本

文的方法与 FT<sup>[5]</sup>、SR<sup>[12]</sup>、SIM<sup>[13]</sup>、CA<sup>[14]</sup>、HC<sup>[3]</sup>、GC<sup>[15]</sup>、SDSP<sup>[16]</sup>和 DSR<sup>[6]</sup>等算法做了比较。图 8 是 PR 曲线图, 最上边红色的曲线是本文的方法, 可以看出, 本文的方法对 360 全景图显著性检测的优越性, 比现有的显著性检测方法具有更高的正确率

和召回率。图 9 显示了以上方法的 *F-measure* 对比, 可以看出, 本文提出的方法比其他方法有更高的 *F-measure* 值, 说明本文算法的精度最高。图 10 是主观的展示了以上方法的视觉对比效果, 可以直观的看出本文的显著性算法更有效。

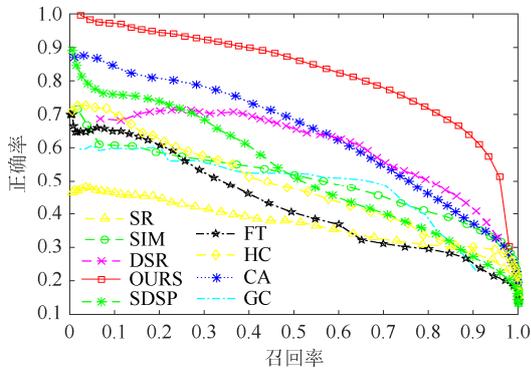


图 8 PR 曲线

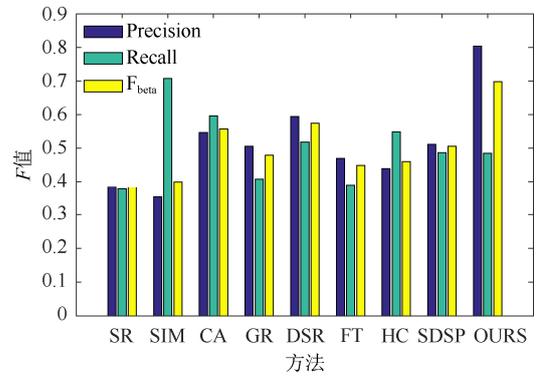


图 9 F-measure 对比图

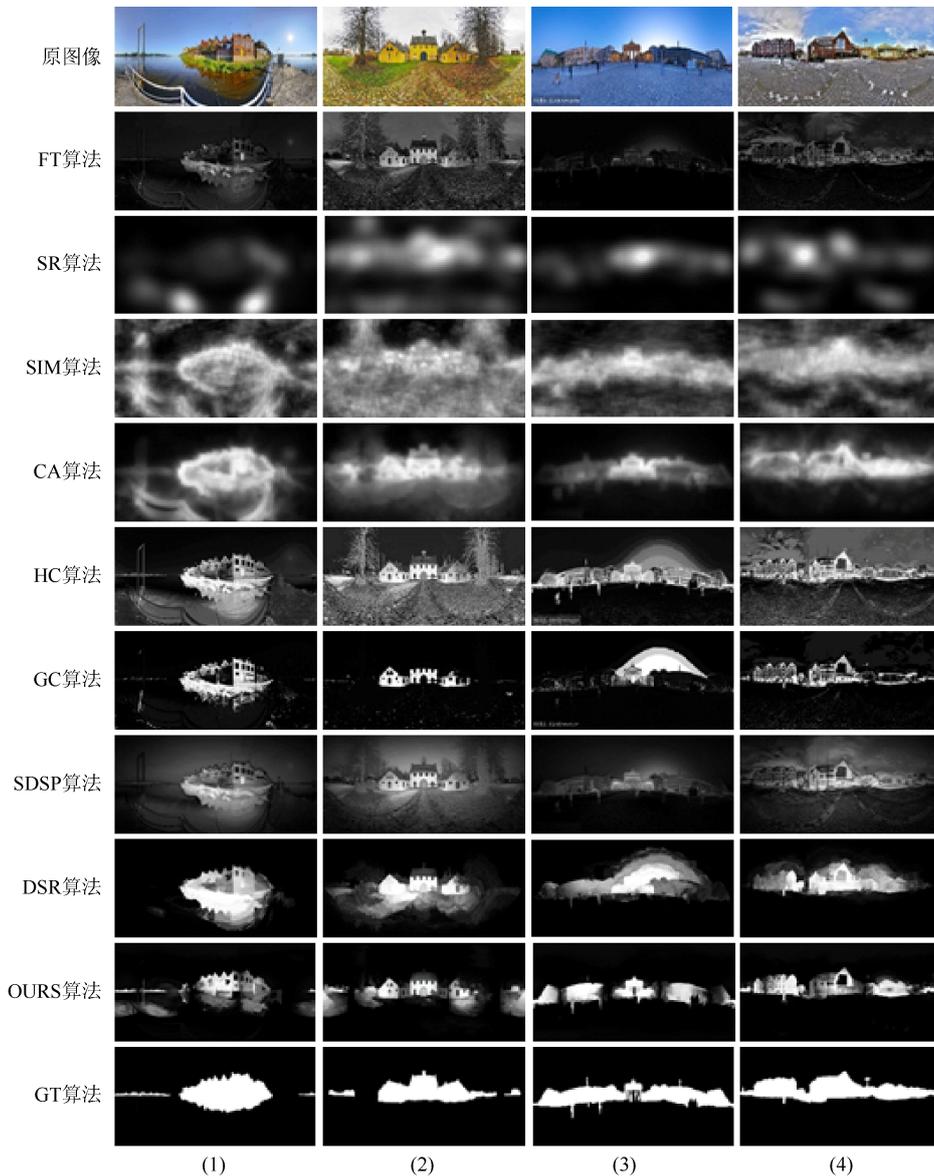


图 10 显著性图的视觉对比效果

### 3 结 论

本文主要贡献有: ①引入多角度分割和立方体映射; ②利用 SVM 分类器建立背景集, 同时引入新的颜色特征, 采用 Lab 和 RGB 的平均颜色特征以及最大、最小、平均灰度值表示每个超像素, 即  $x = \{L, a, b, R, G, B, max, min, avg\}$ ; ③进行立方体逆映射及多角度融合。

本文提出了一种基于多角度分割的 360 全景图的显著性检测算法。首先将 360 全景图进行多角度切割, 将其分别投影到立方体上, 去除畸变; 再对每个面通过稠密稀疏重建计算显著性, 最后再将每个面的图片投影到矩形上, 进行多角度融合。实验结果表明, 本文提出的显著性检测算法更适用于 360 全景图, 与现有的显著性方法相比较, 本文算法有更高的正确率、召回率以及 F-measure 值。

#### 参 考 文 献

- [1] ITTI L, KOCH C, NIEBYR E. 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.
- [2] RAHTU E, KANNAL J, SALO M, et al. Segmenting salient objects from images and videos [C]//*European Conference on Computer Vision*. Berlin: Springer, 2010: 366-379.
- [3] CHENG M M, ZHANG G X, MITRA N J, et al. Global contrast based salient region detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 569-582.
- [4] BORJI A, ITTI L. Exploiting local and global patch rarities for saliency detection [C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE Press, 2012: 478-485.
- [5] ACHANATA R, HEMAMI S, ESTRADA F, et al. Frequency-turned salient region detection [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2009: 1597-1604.
- [6] LI X H, LU H C, ZHANG L H, et al. Saliency detection via dense and sparse reconstruction [C]//*IEEE International Conference on Computer Vision (ICCV)*. New York: IEEE Press, 2013: 2976-2983.
- [7] ZHU C B, HUANG K, LI T H. Automatic salient object detection for panoramic images using region growing and fixation prediction model [EP/OL]. [2018-02-10]. <http://cn.arxiv.org/abs/1710.04071>.
- [8] VISHWANATH B, NANJUNDASWANY T, ROSE K. Rotational motion model for temporal prediction in 360 video coding [C]//*IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*. New York: IEEE Press, 2017: 1-6.
- [9] 董振江, 张东卓, 黄成, 等. 虚拟现实视频处理与传输技术[J]. *电信科学*, 2017, 33(8): 45-52.
- [10] LI Z H, TANG J H. Weakly supervised deep matrix factorization for social image understanding [J]. *IEEE Transactions on Image Processing*, 2017, 26(1): 276-288.
- [11] LI Z C, TANG J H. Unsupervised feature selection via nonnegative spectral analysis and redundancy control [J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 5343-5355.
- [12] HOU X D, ZHANG L Q. Saliency detection: a spectral residual approach [C]//*IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2007: 1-8.
- [13] RUBINSTEIN M, JOULIN A, KOPF J, et al. Unsupervised joint discovery and segmentation in internet images [C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE Press, 2013: 1939-1946.
- [14] GOFERMAN S, ZELNIK-MANOR L, TAIL A. Context-aware saliency detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(10): 1915-1926.
- [15] CHENG M M, WARRELLJ, LIN W Y, et al. Efficient salient region detection with soft image abstraction [C]//*IEEE International Conference on Computer Vision (ICCV)*. New York: IEEE Press, 2013: 1529-1536.
- [16] SEO H J, MILANFAR P. Static and space-time visual saliency detection by self-resemblance [J]. *Journal of Vision*, 2009, 9(12): 15-15.