**第3章 基于前景背景先验的室内场景多模态区域一致性显著性检测研究**

视觉显著性是复杂场景中对象检测的一种非常重要的特征。然而，基于图像显著性受到室内场景中杂乱背景和相似对象的影响，并且基于像素的显著性不能给整个对象提供一致的显著性。因此，在本文中，我们提出一种基于室内场景获得的多模态数据去计算视觉显著性图并且同时保持区域一致性的方法。来自一个场景的多模态数据首先是通过RGBD相机获取得到的。然后用自适应方法对该场景进行过分割，用来组合场景的颜色图和深度图。基于这些过分割结果，我们提出两种先验作为提高最终显著性图的领域知识，包括从彩色图像获得的焦点区域，以及从点云数据获得的平面背景结构。因此，通过融合场景中颜色数据、深度数据和点云数据产生新的显著性图。实验中，我们将所提出的方法与最前沿的方法进行了深入的比较，并且我们还将所提出的方法应用于真实的机器人系统以检测感兴趣的对象。实验结果表明，本文提出的方法在精度和召回率方面优于其他方法。

**3.1引言**

视觉显著性，作为一种重要的视觉特征，最近几年已经得到越来越多的关注。[1-4, 26, 27]视觉显著性表明子区域是否与周围环境有明显差异兵器能快速引起观察用户的注意。自从Itti[7]提出并应用计算机视觉的视觉注意力理论以来，所有后来提出的方法都是遵循文献[3]中总结的人类视觉注意四个基本原则中的一个或者几个，这些原则考虑了局部底层因素，全局因素，视觉组织规则以及高级因素。视觉显著性可以被用到许多计算机视觉任务[28]或者图像处理[29]中。许多机器人系统也利用视觉显著性去进行对象识别与检测[5, 6]。

当采用视觉显著性进行对象检测时，由于人类总是把对象看作一个整体，因此很重要的一点是一个对象中的所有子区域具有相同或相似的显著性值。然而，大多数已有的方法使用基于像素的方法去计算显著性，这意味着显著性值能在像素间变化。此外，大部分已存在的计算显著性值的方法仅仅是基于二维图像的，其中显著对象和背景之间的相似颜色和纹理会造成显著性的提取正确率降低。

为了解决上述提到的问题，通过采用多模态数据，我们提出一种新方法来计算基于区域的显著性。提出的方法首先精确地分割对象，然后将唯一的显著性分配给属于一个对象的区域中的所有像素点。基于多模态数据，本文提出的方法利用不同模态之间的互补性，因此减少其他对象和背景造成的影响。特别地，本文首先根据一种自适应方法，通过使用场景的颜色图和深度图，对场景进行过分割。接下来，通过扩展文献[10]中的方法，对颜色图和点云数据中的每个过分割去计算局部区域对比的显著性。然后，考虑基于两种先验（cues=先验？）的显著性，如计算颜色图焦点区域的先验，以及计算点云的平面背景的先验。最终，我们结合这两种先验信息，得到在区域中数值一致的多模态显著性图。

在实验中，我们首先用过分割精确度去评估颜色图和深度图的互补效果。然后，将我们的方法和目前提出的最好的方法进行比较，并且将我们提出的方法应用到真实机器人系统上去检测感兴趣对象。

本文的剩余部分结构安排如下。首先在下一节回顾显著性的相关工作。在第三节详细描述我们提出的方法，接下来是实验结果与讨论。最后一小节，我们得出结论。

3.2 相关工作

显著性是根据人类视觉注意机制，人类借助显著性就能够迅速注意到跟周围环境的差异性最大的区域。人类的视觉注意力的四个基本原则[3]已经被大多数显著性方法所采用（读者可以阅读详细的综述文章[8]）。本文主要关注的是自下而上的显著性检测相关文献，包括在[3]中的三个基本原则。这些方法可以被简单的分成两种原子计算单元的组别：基于像素的显著性检测方法[2, 3, 12-17]，以及基于区域的显著性检测方法[9-11]。

在第一组中，基于像素级别计算得到的显著性，其中每个像素有唯一的显著性值。在文献[2]中，作者提出采用高维高斯滤波去计算完整的对比度和显著性，得到像素精确的显著性图。文献[3]中，考虑多尺度的局部对比度和全局对比度，以及像素的直接上下文，来计算像素级别显著性图。文献[12][13]中，作者提出类似过滤器的方法，根据颜色对比度和光照对比度来计算显著性图。文献[14]中，作者提出采用马尔科夫方法，通过定义两个点的不相似性为其中一个点与两个量的比率之间的距离，来计算像素级别的显著性图。文献[15][16][17]中，通过采用增量编码长度或者有损编码长度来提出基于光谱信息的方法，以测量每个特征的透视熵增益。所有这些方法都能获得可以突出感兴趣对象的比较好的结果，然而，基于像素的方法因为环境干扰也可以给背景过高的显著性。比如图1，由于第一行桌面上的镜面反射引起的两个高亮区域具有很高的显著性。并且在第二行中的两个垃圾桶跟背景颜色相似，这两个垃圾桶的显著性值是很低的。

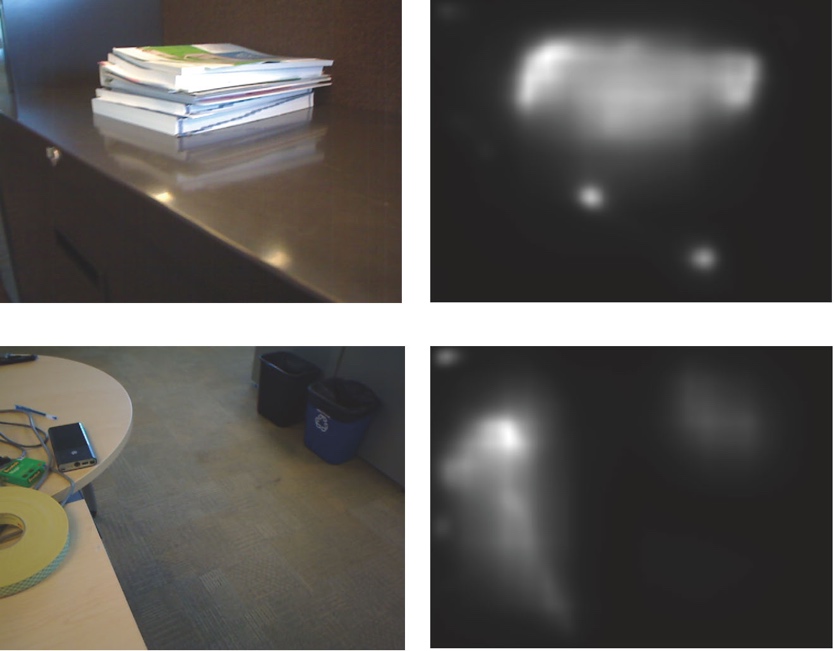


图3.1 被背景干扰的显著性图的例子

最近，越来越多的基于区域的显著性检测方法被研究人员提出来。这是因为当旨在解决显著对象检测问题时，区域一致显著性能获得更好的对象检测结果。文献[9]中，作者提出对图像进行过分割，然后通过图像的其他部分计算由一个或多个过分割区域组成的窗口的显著性值。Cheng等人[10]提出基于区域的显著性，其中一个过分割区域中的所有像素具有相同的显著性值，并且文献[30]在这个基础上结合颜色信息扩展了该方法。这些方法结合了超像素和显著性的优势，并且因此能检测具有高显著性值的显著对象。然而，这些方法只是针对二维图像，并且因此显著性检测的性能也会因为图像中存在两个相似的对象或者混乱的背景而降低。最近，Niu等人[11]扩展了文献[10]中的方法，其中提出结合二维图像到过分割图像中得到视差图，并且利用视差对比和领域知识去计算区域显著性值。然而，Niu等人提出的方法丢弃了图像信息并且，领域知识不适合深度图和点云数据。

所有上述提到的方法没有考虑背景信息。但是，背景也是显著性检测的一个重要的信息。最近，一些学者利用背景先验去计算显著性值并且获得了最好的性能[23-25]。在文献[23]中，作者提出一种鲁棒的软背景连通性去增强显著性的计算。文献[24]中，作者采用了边界先验和背景连通性先验。Yang等人也根据超像素的排序，利用背景信息去计算显著性值。这些方法给我们提供了显著性的考虑，然而，这些方法都是假设跟图像边界相连的区域属于背景，但这并不总是正确的，特别当我们在混乱场景中检测到显著的对象的时候。在本文的研究中，我们能从背景检测的多模态数据中获益。我们首先通过点云数据检测平面结构，然后将场景边缘相连的平面区域作为背景，这样操作很明显更加合理。

跟已有方法不同的是，我们提出的方法利用了多模态信息的优势，比如，从二维图像中提取前景先验，从三维点云数据中计算背景先验信息，以及利用二维图像和三维信息的互补效果产生更加精确的过分割区域。

3.2 算法介绍

我们提出的多模态区域一致的显著性由三个步骤组成。首先，我们采用颜色图像和深度图对场景进行过分割，然后，我们计算颜色图中的焦点区域作为前景先验信息，并且检测出点云中的平面区域作为背景先验信息。最终，我们通过结合全局对比度和先验计算最终的显著性图。如图3.2所示，结合标记为红色框感兴趣区域的彩色图以及第一行中的深度图，得到多模态过分割结果（第二行左图），并且产生针对平面分割的点云（第二行右上图）。平面分割结果映射到背景先验的深度图。第三行是最终的显著图。每个平面被指定了不同的颜色。

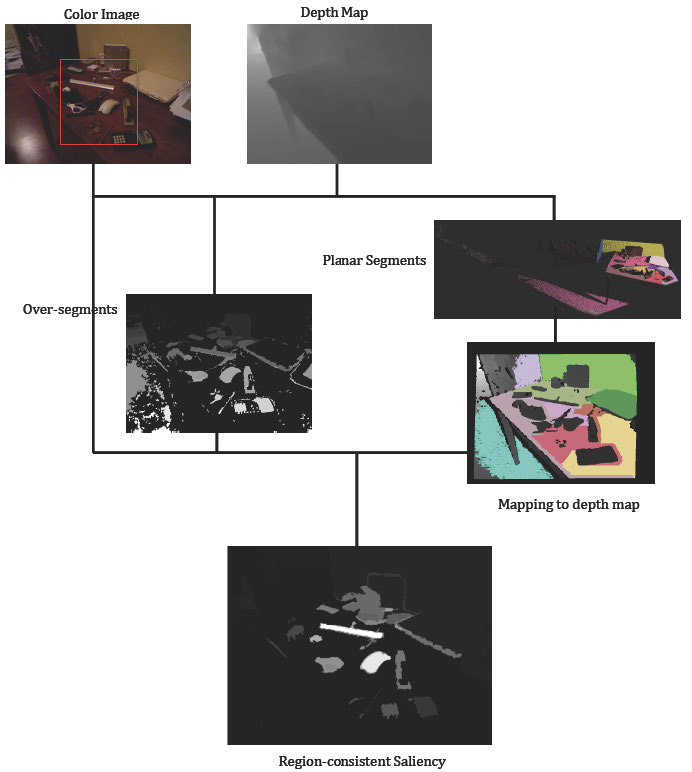


图3.2 提出的区域一致显著性方法。

**3.2.1多模态数据的过分割**

在本文的研究中，我们采用文献[18]中的方法去获取图像以及对应深度图的过分割。这个算法根据图像区域特征之间的差异来分割图像。图像间特征的差异定义为：

其中，和分别是组件和的两个节点，E表示邻接关系，定义了边界权重函数。在文献[18]中，边缘权重函数是用来计算两个相邻像素之间的绝对强度差，，其中，和表示两个像素的强度。在本文的研究中，边缘权重函数同时度量颜色空间和深度空间之间的差异。边缘权重函数的正式定义如下：

其中参数控制颜色和深度通道的权重。

尽管这个公式只是简单的改进了文献[18]中的算法，然而，如何确定参数是获得满意的分割结果的关键，需要谨慎的去选择。此外，所有图像不可能设置相同的参数值，因为不同的场景存在多种图像差异和深度差异。并且因此，对于不同场景的过分割，图像和对应的深度图有不同贡献。如图3.3所示，不同的参数值将会导致不同的分割结果。当使用深度图的时候，只有少数分割部分是在深度图很大的地方得到的。因此结果的显著性图对于区别不同的对象是不精确的。当增加颜色通道的权值，分割将会变的越来越多。如果忽略深度通道并只采用颜色通道，就会有过多的分割。在这种情况下，太多的区域去计算显著性将占据过多的计算资源。并且将会得到过多的不同区域的显著性，这会减少显著性的精确度。

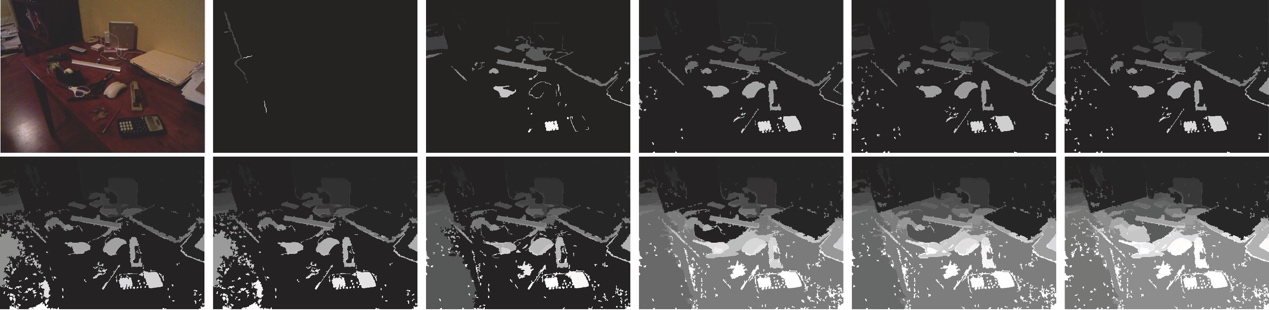


图3.3 给定不同权重参数的多模态过分割结果。

我们提出一种自适应方法确定权重。首先采用文献[18]中描述的相同参数去分别分割颜色图和对应的深度图。因此可以获得两个过分割。然后通过下面的公式计算出：

其中和分别表示深度图和颜色图过分割的数目。

尽管上面的自适应方法是简单的，但是对于我们的实验是有效的。根据上述公式，我们得到深度图中的过分割数目越大，使用的值就越小。这是合理的，因为在深度图中更多的过分割意味着可以提供更多的细节信息，并且权重会变大去弥补颜色信息。此外，深度图的权重减小，颜色图的权重应该增大，来得到足够多的用于显著性计算的过分割。实验中，我们发现的值的范围是5到50，的值的范围是200到500。因此参数涌来调整颜色图和深度图之间的过分割的占的比重。实验结果上，我们发现当的值在5到15之间的时候，过分割结果是稳定的。在实验中，我们设置为10。另一个参数设为0.8确保生成过分割的时候颜色图至少占20%的贡献。因此，通过这个公式深度图和颜色图能够自动互补。

3.2.2 前景先验

我们观察到当一个人拍照的时候，最重要的物体将会放在照片的最焦点的位置。因此首先通过主体区域抽取算法（Subject Region Extraction, SRE）[19]提取出焦点区域(Focus Region, FR)。采用一系列不同尺寸的高斯核去卷积图像，并且计算所有卷积图像中每个像素的水平梯度和垂直梯度。然后，每幅卷积图像像素的分布通过直方图近似计算，通过下面的梯度对数似然函数来度量一个像素块区域如何较好的用一个k乘k大小的高斯核来表达：

其中和分别表示卷积图像的水平梯度分布和垂直梯度分布。N(x)指的是像素x的邻接窗口，表示邻接窗口内的一个像素。找到不同高斯核大小中最大的，该高斯核定义为，用于最佳的解释图像中一个窗口区域。高斯核的尺寸说明了该区域块的清晰程度。因此的值越小，中心为x的窗口区域块越清晰。对于每个像素，存在对应的高斯核，通过定义一个二值图来确定清楚区域还是模糊区域：

将二值图U投影到水平轴和垂直轴，统计二值图中水平方向和垂直方向上白色像素点的数目，得到两对坐标和，在这两个范围内白色像素的个数等于二值图中所有白色像素的。当我们去检测用专业摄影师拍摄的照片中的焦点区域时，其中焦点区域的清晰度有明显的差异，检测到的焦点区域是准确的。然而，我们实验的图像是通过移动机器人拍摄的，没有专业的技巧去拍着这类照片。因此，如果被设为0.9，检测到的区域太大以至于不能准确的确定出焦点区域，如图3.4所示。在我们的实验中，我们发现较小的值是适合我们的测试数据的。图3.4显示了实验中的一个例子。很明显，当是0.6时，检测到焦点区域时最准确的。



图3.4 焦点区域检测的例子

检测到焦点区域后，我们定义第一个显著性先验信息。离焦点区域越近的区域，该区域的显著性值越高。因此我们正式得出下面的公式：

其中表示过分割块和焦点区域中心点之间的欧氏距离。表示所有过分割和焦点区域之间的最大距离。

一些其他的算法[2,3,14]强调了图像中心位置的显著性值，因为这些算法明确地或隐含地使用了图像中心店作为前景先验。但是，在一些情况下，关键的物体可能不在中心区域并且中心区域的显著性值也可能是不准确的。当采用焦点区域作为前景先验的时候，由于焦点区域不会被限定在图像的中心位置，上面这个缺陷可以被解决。

3.2.3 背景先验