**Multimodal region-consistent saliency based on foreground and background priors**

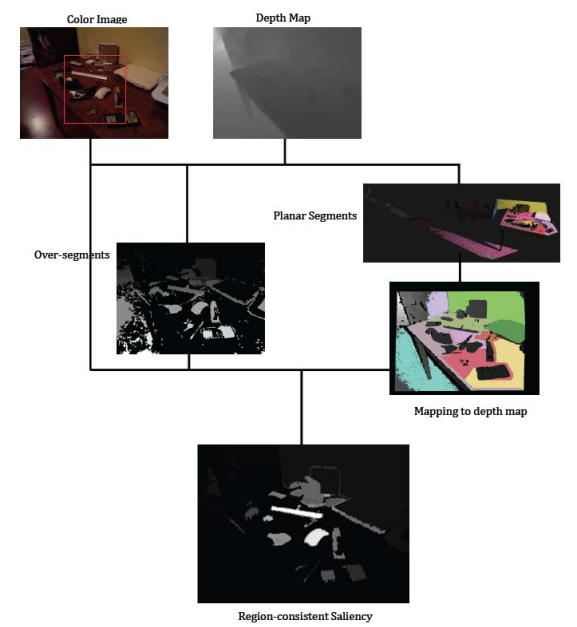
复杂环境中，视觉显著性是对象检测中一种非常重要的特征。但是，基于图像的显著性受到相同颜色和纹理的影响，而基于像素的显著性对于整个对象不能提供稳定的显著性。因此，本文提出一种新的方法—从多模态数据中计算出视觉显著性图同时保持区域一致性。场景中的多模态数据首先由RGB-D相机获取。 然后采用新的自适应方法结合颜色图和深度图，过分割场景。根据过分割结果，我们提出两种新的线索作为提高最终显著性图的邻域知识，包括从颜色图中得到的焦点区域以及从点云数据中得到的平面背景结构。因此，通过融合场景中颜色数据、深度数据和点云数据产生新的显著性图。实验中，我们对提出的方法和已有的前沿方法进行了深入比较，并且我们也将提出的方法应用到一个真实的机器人系统中，去检测感兴趣的对象。 实验结果显示，在准确率和召唤率上，我们提出的方法要优于其他方法。

1. Introduction

作为一个重要的视觉特征，最近几年视觉显著性已经得到越来越多的关注。视觉显著性指的是在环境中异于邻近区域的子区域，并且能够快速吸引观察者的注意力。自从由Itti提出并应用的针对计算机视觉的视觉注意力理论后，所有之后提出的方法都是基于文献[3]中提出的人类视觉注意力理论中的四项基本原则中一个或者几个，这些原则是局部底层原则，全局原则，视觉组织原则以及高层因子。视觉显著性的最重要的应用之一是从复杂场景中快速搜索显著性对象。许多机器人系统也用视觉显著性来进行对象识别和检测。

通过视觉显著性进行对象检测，希望包含对象的子区域维持相同或相似的显著性值，因为人类总是把对象看作一个整体。然而，大部分已经存在的方法使用基于像素的方法去计算显著性，得到显著性值随像素的不同而不同。此外，已存在的大部分方法只是基于二维图像，其中显著性对象和背景中相同的颜色纹理造成显著性的提取正确率降低。

为了去解决前面提到的问题，通过使用多模态数据，我们提出一种新的方法，计算基于区域的显著性。提出的方法首先精确分割对象，然后将不同的显著性分配到属于一个对象的区域的所有点。通过采用多模态数据，方法利用不同模态之间的互补性减少其他对象和背景的影响。特别地，首先基于新的自适应方法，将场景分割成采用颜色图和深度图的过分割。接下来，根据文献[10]，从颜色图模态中的每个过分割计算全局区域对比的显著性。然后，基于两个新线索的显著性，如从颜色图计算得到的焦点区域先验，以及通过点云计算得到的平面背景先验，同样被考虑进来。最终，我们将二者集成进来，得到一个多模态显著性图，使得在区域中具有相同的显著性值。图1阐释出上面提出的方法的提纲。



在实验中，首先通过计算过分割的精确度评估颜色图和深度图的互补性的影响。其次，将我们的方法和目前提出的最好的方法进行比较，最后在真实机器人系统上应用我们提出的方法去检测感兴趣对象。实验结果显示提出的方法能大大的提升感兴趣对象的显著性并且减少背景的影响。此外，一个对象绝大部分有相同的显著性，它简化了对象检测并提升了准确度。

本文的剩余部分安排如下：

1. **Related work**

显著性根据人类视觉注意力机制，人类能快速注意到区别性最大的区域。人类视觉注意力的四个基本原则被大多数显著性方法使用（读者可以参考最近一个比较详细的综述文献[8]）。本文，相关的文献大多关心的是自底向上的显著性检测，包含了文献[3]中前三个基本原则。根据原子计算单位，这些方法能够粗略被分成两大类：基于像素的方法[2][3][12][13][14][15][16][17]和基于区域的方法[9][10][11]。

在第一组中，显著性基于像素级别计算得到，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计算完整的对比度和显著性，得到一个像素精确的显著性图。文献[3]中，考虑多尺度的局部和全局对照，计算出像素级别的显著性图和像素的直接上下文环境。文献[12][13]中，作者提出类似过滤器的方法，根据颜色和光照的对比度来计算显著性图。文献[14]中采用马尔科夫方法，通过将两个点之间的不相似性当作1和两个数值的比率之间的距离，计算基于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耗的编码长度提出基于光谱信息的方法，计算出每个特征的透视熵。所有这些方法能得到高亮感兴趣对象的良好结果，但是基于像素的方法可能因为环境的干扰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。

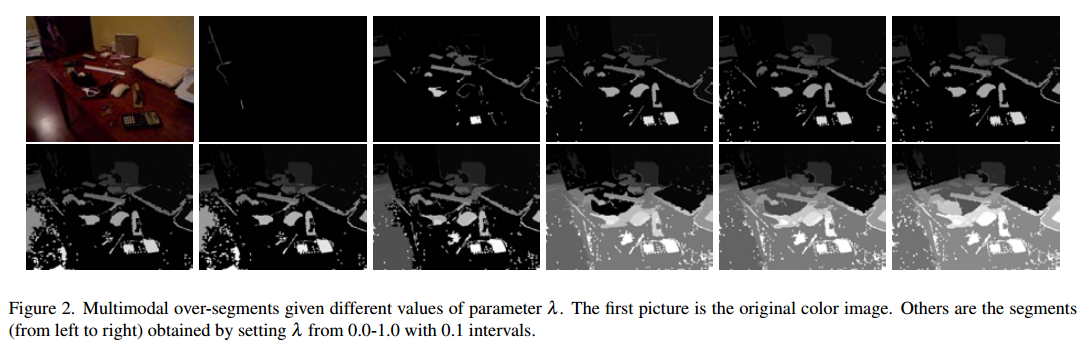
1. **Region-Consistent multimodal saliency**

如图一所示，提出的多模态的区域一致性显著性检测方法包含三个步骤。

首先，采用颜色图和深度图将场景分割成过分割。

接下来，在颜色图中计算焦点区域作为前景先验，并且在点云中，检测平面区域作为背景先验。

最终，通过结合全局的对比度和先验，计算出最终的显著性图。



**3.1 多模态过分割**

本文研究中，**我们采用文献18中的方法，得到来自图像和对应深度图的过分割。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，v\_i 和 v\_j是组件C\_1 和 C\_2各自的节点，E表示邻域系统，w(v\_i，v\_j)表示边权函数。在文献18中，边权函数**计算为两个相邻像素间的强度差的绝对值**，其中I(P\_i)和I(P\_j)表示两个像素的强度。 对于三通道颜色图，文献18中作者建议**在各自三个通道中分别进行分割**，然后**集成三个分割集合**。**研究中，我们把深度图作为第四通道。然后应用一个可替换的方法，同时边权函数计算颜色图和深度图中两个像素间的差值。我们采用这个方法是为了节省计算资源**。正式形式如下：

C:\Users\admin\AppData\Local\YNote\Data\qqE340EAF23D655E8767E1F26AD5AE1180\ed4bd91206a64ea2b8b1b889e0ac7fdb\3a7b713103c846ec84388260fd239857.jpg

其中参数lambda控制颜色和深度通道的权值。

尽管这个公式简单的扩展了文献18中的算法，然而，如何**确定参数lambda的值是解决满意分割的关键，需要小心仔细的选择**。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2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。

我们提出一种自适应的方法去确定权值λ。我们首先采用文献18中描述的方法，采用相同的参数去独立分割颜色图和对应深度图。因此会得到两倍的过分割数。因此通过下述公式去计算参数lambda：

C:\Users\admin\AppData\Local\YNote\Data\qqE340EAF23D655E8767E1F26AD5AE1180\1b987efa7ba94f86b020a8487aeec360\e18a8f79cbeb445b8610b12c37159227.jpg

其中 n\_dp 和 n\_cl是深度图和颜色图的过分割数目。

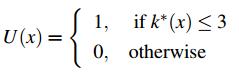
尽管这个自适应方法是简单的，但在我们的实验中是有效的。根据这个公式，我们得到的深度图中过分割的数值越大，参数lambda的值就越小。因为深度图中更多的过分割意味着能提供更多的细节，这是合理的，并且权值增加可以补充颜色信息。此外，深度图的权值应该减小，颜色图的权值应该增加，针对显著性计算得到足够多的过分割数。在实验中，我们发现n\_dp的值正常范围是5到50，并且n\_cl正常范围是200到500。因此，参数β用于调整颜色图和深度图之间过分割的分配比例。**实验中，我们发现当参数β的值在5到15之间，过分割的结果是稳定的**。在实验中，参数β值设为10。另一个参数α也被引入去保证颜色图的最小分配。**在实验中，当产生过分割时，参数α设为0.8确保至少20%是分配给颜色图。因此，根据这个公式，深度图和颜色图能相互自动补充。**

**3.2 前景先验**

我们发现当一个人拿着照片，一幅图的焦点点处的最重要的对象将被替换。因此**焦点区域(FR)**首先通过文献19中的**“对象区域检测（SRE）”算法**来检测。文献中**使用一系列不同尺寸的高斯核去对图像进行卷积操作**，并且**计算所有卷积图像中所有像素的水平和垂直梯度**。**然后，对于每个卷积图像中像素分布是通过直方图来近似得到的**，它可以**用于度量一批像素是如何解释通过一个k x k的核以及下面这个梯度的对数似然函数来得到**：

C:\Users\admin\AppData\Local\YNote\Data\qqE340EAF23D655E8767E1F26AD5AE1180\957551f50fdf451f97860fcae5a7b5de\clipboard.png

其中p^h\_k和p^v\_k 表示水平和垂直梯度的分布，从卷积图中计算得到，N(x)是像素x的领域窗口。通过找到所有核大小的l\_k\_(x)的最大值，我们能找到一个核，用k\*(x)表示，这可以被看做用于最好的解释图中的一批窗口。核的大小能反映碎片的干净度。因此，k\*(x)的值越小，在像素x中心处的碎片就越干净。对于每个像素，存在一个对应的k\*(x)，并且它能用于表示干净或者模糊的区域，通过定义一个二值图：



将U映射到水平和垂直轴上后，我们**统计二值图中从水平到垂直方向的白色像素的个数**去找到两队坐标，[h\_1, h\_2],[v1,v2]，其中在这两个范围中，白色像素的个数等于U中白色像素的总个数的(1-anta)/2。当专业摄影师通过捕捉相片中对象区域时，其中FR在清晰度上有明显的差异，检测到的FR值是非常精确的。然而，我们的图像是通过移动机器人捕获的，没有专业的捕获图像的能力。因此，如果参数anta设为0.9，检测到的区域太大以至于不能精确的表示出焦点区域，如图3b所示。在我们的实验中，我们**发现一个更小的anta值0.6是比较适合我们的测试数据**。图3c是其中一个例子。

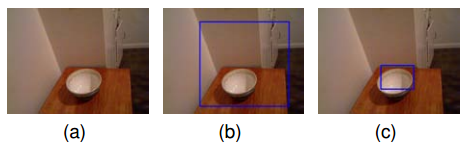
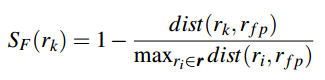


图3. 焦点区域检测例子。 (a)是原图。(b)中的蓝色矩形框通过设置参数anta为0.9检测得到。(c)中的蓝色矩形框通过设置参数anta为0.6。很明显(c)中检测到的焦点区域更加准确的集中在真正的焦点区域上。

在**检测焦点区域**时，我们能定义出**第一个线索显著性**。**距离焦点区域更近，对于这个区域的显著性更高**。因此我们有下面的定义：



其中dist()表示过分割r\_k的中心到焦点区域的中心的距离，并且表示焦点区域内所有过分割之间的最大距离。



**3.3 背景先验**

背景作为一个鲁棒性先验已经在文献[23][24]中使用，去构建精确的显著性图。然而，这些方法把大部分连接图像边界的区域看作是背景，这不总是正确的，因为在一些情况中，在图像边界处对象可能被替换掉，如图4所示。

为了克服这个缺点，我们**采用平面分割去检测背景**。在室内场景，存在许多平面结构，比如桌面，地面以及墙壁等。如果这些平面结构到了场景边界，那么一般被看作是背景。因此，我们**首先在从图像和对应的深度图像中产生的点云数据中检测平面区域**，并且**计算连接场景边界的边界的长度**。那么我们有下面的定义作为平面分割的背景连通度：



其中Len(p\_i)是边界的长度，它跟场景边界相邻，并且|p\_i|是分割平面p\_i的面积。

基于背景连通度，我们简单的定义一个过分割r\_k的背景权值：



其中 P 是所有检测到平面的集合。

因此，过分割r\_k和所有平面分割之间的重叠区域越多，Bw(r\_k)值越大。意味着r\_k有更大的可能性是属于背景，并且有很小的显著性。明显地，Bw(r\_k)值的范围是从0到1，并且因此我们能因此第二个DK显著性，基于背景权值



**3.4 多模态区域显著性**

最终，我们的多模态区域一致性显著性是从全局区域对比插入前景和背景先验计算得到的。全局区域差异(GRC)是通过扩展由文献[10]提出的算法来得到，它定义为



其中 D\_r(r\_k, r\_i) 和 D\_s(r\_k, r\_i)表示区域r\_k和r\_i之间各自颜色和空间距离，确定空间权值的长度，w(r\_i)表示区域r\_i的权值，它计算出来作为区域r\_i中的像素个数。

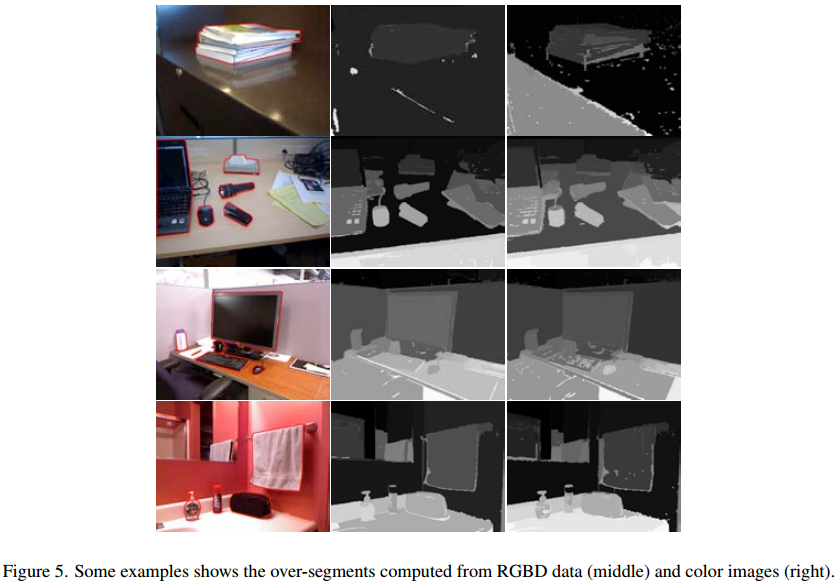
然后，最终的显著性图能通过下面的公式得到：



1. **实验**

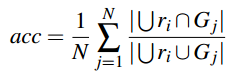
我们从多个方面来评估提出的方法。我们首先估计了从多模态数据中计算得到的过分割的精确度。然后，基于两种策略的感兴趣的分割对象召回率是通过跟目前最好的方法进行对比得到的。最终，我们在一个真实的机器人系统上实现了我们提出的方法，并且显示了一些检测感兴趣对象的结果。

前两个实验是基于文献[20][21]中两个公共RDB-D数据集进行的。对于第三个实验，机器人检测在我们实验室中感兴趣的对象。在我们提出的方法中，只有文献[18]中过分割方法使用的参数需要进行设置。我们设置所有实验的参数相同，其中，并且最小的成分尺寸是设为20。同样对于文献[10]中的方法，我们使用同样参数。



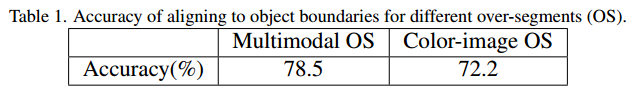
4.1 多模态过分割的评估

我们评估多模态过分割和颜色图像过分割的对象边界对齐的精确度。1000幅图和对应的深度图从两个数据集中被选为测试数据。因为在这两个数据集中只用边界框表示感兴趣对象，我们手动给所有对象添加对象轮廓。如图5第一列所示的例子。然后，我们计算出精确度，利用过分割和标定好的真实数据之间的重叠度。正式地，精确度定义如下：



其中N是对象个数，G\_j是第j个对象标定好的真实数据，Ur\_i是所有和G\_j相交的过分割的并集。

图1中列出来的精确度。多模态过分割能比颜色图像过分割获得更高的精确度，因为多模态过分割能避免相同颜色和纹理的影响。对于背景噪声，比如高亮反射，多模态过分割也能避免受到它的影响并提供更好的结果。图5所示的是一些例子。在第一行中，桌面上两个高亮反射区域被分割成在中间列的带有桌子的相同区域，但在右边列，这两个区域被分割成桌子的不同区域。在第三行，键盘和显示器之间的马克杯被正确的分割成中间列的独立分割部分，但是在右边列，分割是错误的。在第二和第四行中，我们的多模态过分割也获得了一个较好的调整到对象边界处。



4.2 多模态区域显著性的评估

在这个实验中，我们首先评估提出的方法的每一个部分对于最终显著性图的贡献，那么我们通过对比其他经典的和目前最好的方法去评估多模态区域显著性检测的准确度。为了更全面的评估提出的方法，我们采用文献[10]中用到的两个策略去分割显著性对象并计算他们的准确度。对于这两种策略，我们计算所有方法的召回率。

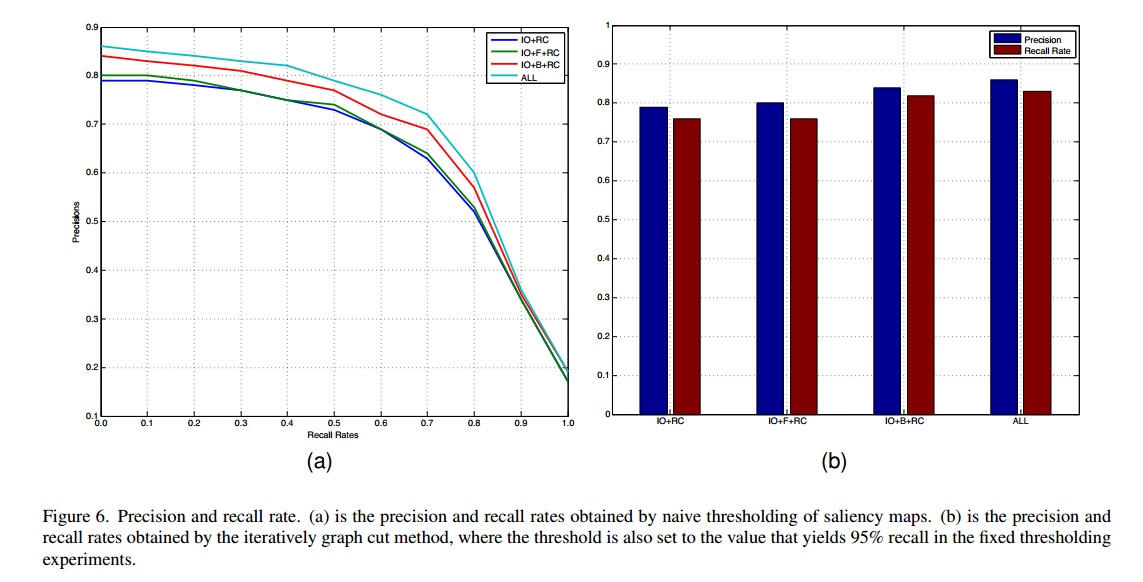
在我们的方法中，我们通过适当的调节颜色图和深度图对于最终的过分割贡献比例，然后结合前景和背景先验到全局区域对照显著性中。因此，我们采用四种不同的策略去计算显著性，比如，使用改进的过分割和全局区域对照（IO+RC），使用改进的过分割，前景先验和全局区域对照（IO+F+RC），使用改进的过分割，背景先验以及全局区域对照(IO+B+RC)，以及全部都融合进来的情况。召回率如图6所示。

从图中我们可以看到背景先验大大提升了最终的显著性检测。如图6a所示，”IO+RC”和”IO+F+RC”的召回率曲线有一些很小的不同，表示从“IO+RC”到“IO+F+RC”性能有轻微的提升。从图6b可以看到，我们也可以同样观察到召回率都提升了0.01。然而，“IO+RC”到“IO+B+RC”性能有很大的提升。图6b看到召回率分别提升了5%和6%。这暗示在室内场景，平面信息对于计算对象显著性很重要。我们在图7中举了一些例子。如第4,5列所示，我们观察到背景的显著性能被极度的压缩如果显著性是通过集成背景先验来计算得到的。

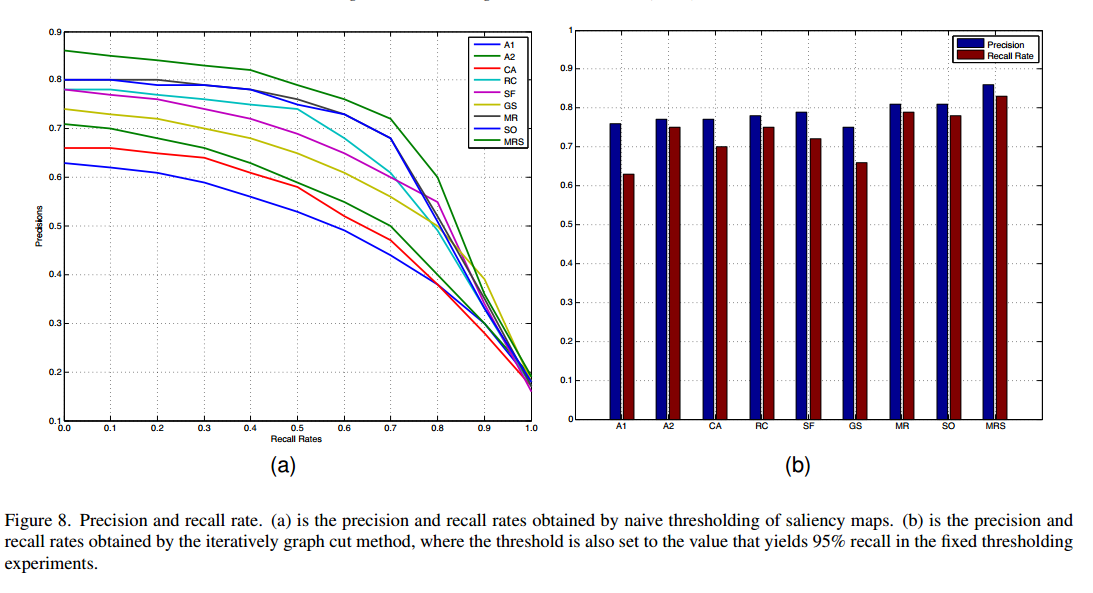
现在，我们通过比较三种基于像素的方法来评估我们提出的方法，比如。Achana的两种方法(A1)和(A2)，Goferman的方法(CA)，以及五种基于区域的方法等等。Cheng的区域对照显著性(RC)，显著性过滤(SF)，地表显著性（GS），基于图的流形排序(MR，Manifold Ranking)以及鲁棒性背景检测的显著性优化(SO)。

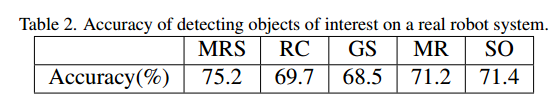
图8显示了定量对比结果。在图8a中，很明显我们提出的方法有很好的召回率。跟其他基于像素的方法进行对比(比如 A1 A2 和CA)，大多数基于区域的方法有更好的结果，因为它们能获得区域一致性显著性。提出的MRS方法优于其他室内场景中的方法，因为我们为这些场景特别设计了背景先验。相同的结果能在图8b中找到，其中我们的方法有最好的性能。

注意所有方法的性能是低于提出这些方法的文献中的性能。这是因为他们的性能是通过在一些经典的数据集上测试得到的。然而，我们的测试数据有一个更复杂的背景，并且在大多数图像中，存在至少一个感兴趣对象。因此，在真实环境中获得精确的显著性区域要更加困难。







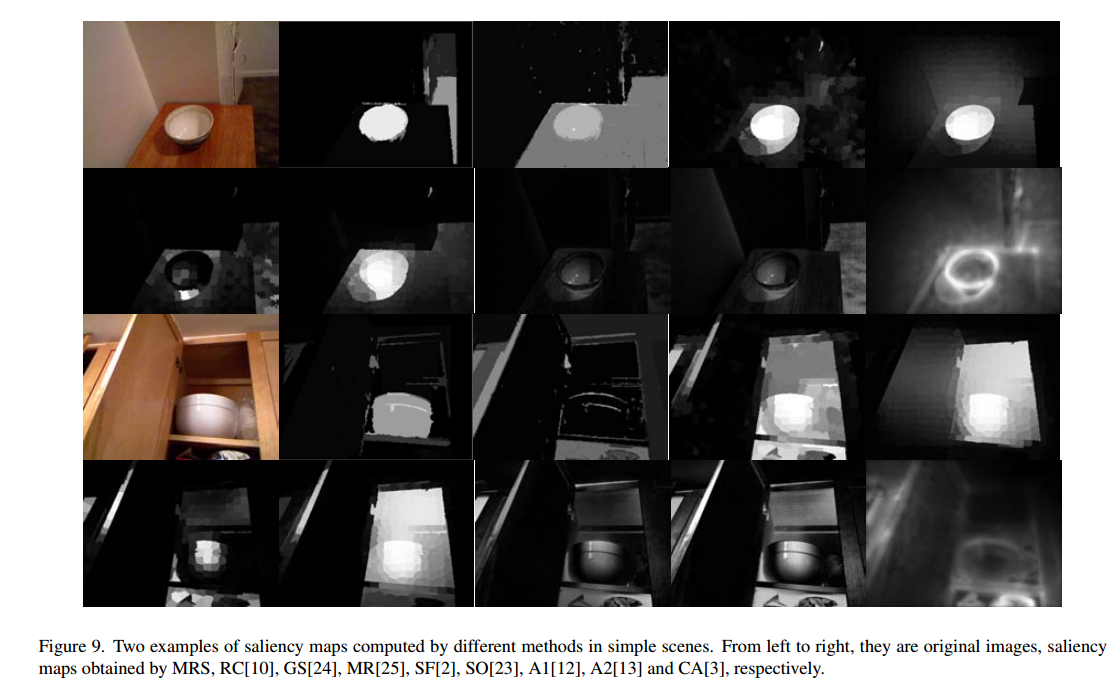


我们通过这些方法和我们提出的方法在图9简单的场景中以及图10复杂的场景中计算显著性图，其中我们的方法标识为MRS。很明显我们的方法是唯一能正确计算显著性对象的显著性的，比如在这两种场景中的两个碗。对于三种基于像素的方法，不可能去用显著性图去精确的检测出完整的对象。对于其他的基于区域的方法，通过采用简单的算法来检测完整的对象同样困难，因为这些方法没有利用三维信息，比如深度图和点云数据。对于图10中的复杂场景，我们的方法能很明显的避免掉背景的影响（白色墙壁），并且获得清楚的显著性图。对于这些方法，背景抑制了前景对象的显著性，比如通过SF，A1和CA方法得到的结果或者产生高显著性去干扰前景对象的显著性，比如通过RC，GS，MR，SO和A2方法得到的结果。

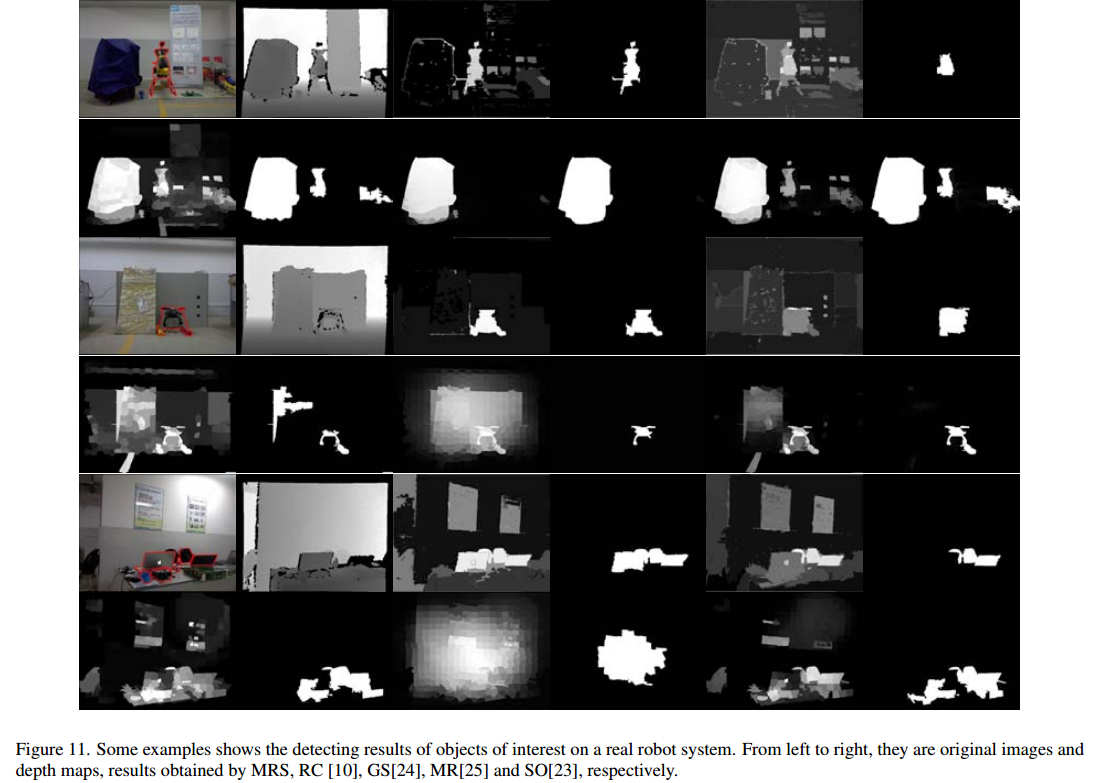
4.3. 在真实机器人系统上的评估

我们在真实机器人系统上比较RC，GS，MR和SO方法。这些方法是跟我们的方法最相关的。所有这些方法都是基于区域的方法。我们的方法使用了RC方法同样用到的计算区域对照的相同方法。GS，MR和SO方法利用了背景先验信息。

在实验中，我们**在我们的AS-R机器人上部署了我们的方法和其他的方法**，通过文献[10]中描述的**显著性割的方法去检测显著性对象**。在我们的实验环境中。**我们使用一个Kinect相机获取颜色图和深度图**。对于RC，GS，MR和SO方法来说，只用到了颜色图。对于我们提出的MRS方法，颜色图和深度图都用到了。因为我们的方法利用了将场景扩展为背景先验的平面结构，它能消除这些背景的影响，并因此得到更精确的对象检测结果。表2中列出的就是这些方法的精确度，图11显示的就是一些实验结果。很明显，我们的方法能更有效的检测感兴趣的对象。





 **5. 结论**

本文中，我们提出了一种新的方法计算**基于多模态数据的区域一致性显著性图**。提出的方法首先提出一个**简单但是有效的自适应多模态过分割方法去更有效的分割场景**。**基于这些精确的过分割**，**然后结合全局对照和新的前景背景线索去计算显著性**。当对比目前最好的方法它能得到一个更好的性能。我们发现室内的**平面结构能被使用作为背景先验用于提升显著性计算的精确度**，因为我们能**消除在聚类场景中的背景的影响**。此外，通过**检测焦点区域得到的前景先验对计算显著性同样能有积极的影响**。在带有大量颜色图和深度图的公共可使用数据集上，我们评估我们提出的方法。实验结果显示我们提出的方法在召回率上要优于其他的方法。