SHREC’16 Track：基于多模态视图的三维模型检索

**摘要**

本文展示了SHREC‘16 Track的结果：基于多模态视图的三维模型检索，它的目标是评估检索算法的性能，当多种视图被用于三维模型的表示。在本次工作中，生成了包含605个对象的集合，并为每一个图像提供了颜色图像和深度图像。选择200个模型用于查询，其中包含100个3D打印模型和100个真实三维物体，同时，其余的405个对象被选中用作测试，平均的检索性能被测量。本次Track吸引了7个参与者，提交了9个作品。相对于去年的结果，3D打印模型明显引入了一个坏的影响。今年的表现比去年的更差。这种情况也正好显示了基于多模式视图的三维模型检索方法具有一个很好的前景，并且在处理多模态数据上，揭示了一个有趣的现象。

分类和主题描述（根据ACM CCS）：I.3.3 [信息存储和检索]：内容分析和索引—提取方法。

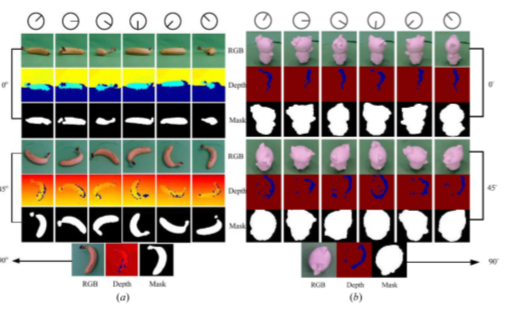
1. **简要介绍**

基于视图的三维对象检索旨在检索由多个视图组表示的三维对象。现有的方法大多是从三维模型信息出发，然而在实际应用中很难获得模型信息。在没有3D模型的情况下，一个3D模型的构造需要通过基于模型的方法收集图像来生成虚拟模型。我们注意到，三维模型的重构是想当昂贵的，同时，其性能也受到抽象图像的严格限制，严重限制了基于模型方法在实际中的应用。

随之颜色和深度视觉信息获取设备的广泛应用，例如Kinect和带有摄像机的移动设备，记录真实对象的颜色和/或深度视觉信息变得可行。这样，三维对象检索的应用可以进一步扩展到世界上的真实对象。现在的3D打印是如此受欢迎，以至于被广泛应用于许多领域。所以使用一些3D打印模型进一步丰富我们的数据集并进一步测试我们的方法是重要的。在2003年，从照明字段描述符[ CTSO03a ] ，近年来，大量的研究关注重点在基于视图的方法。Ankerst等人，[ AKKS99 ] 提出的二维视图从3D模型的最优选择，它关注于从三维模型的代表性特征中获得的数字特征。Shilh等人[SLW07]提出了3D模型的平移描述符特征，它不受3D模型的平移和缩放影响。然而，它不适合于由一组2D图像组成的3D模型。Tarik等人[ADV07] 提出了一种贝叶斯3D对象搜索方法，利用X均值[CTSO03B]来选择特征视图，并应用贝叶斯模型计算不同模型之间的相似度。Gao等人[GTH\*12] 提出了一种不依赖于相机排列约束的三维物体检索的通用框架。值得注意的是，通过视图检索模型仍然是一项艰巨的任务。挑战在于视图提取、视觉特征提取和模型距离测量。在用多模态视图检索3D模型的研究中，我们的目标是集中精力研究这个有趣的话题。这个研究的目标是通过多模态视图检索来检索3D打印模型和3D真实物体，多模态视图是每个3D模型的彩色图像和深度图像。我们的集合由605个对象组成，其中200个对象（包括100个3D打印对象和100个3D真实对象）被用作查询，而其他对象被用作测试对象。有7组参加了本次研究，一共提交了9个结果。评估结果展示了一个关于基于多模态视图的3D模型检索方法的非常有前景的方案，并且揭示了在处理多模态数据上非常有趣的现象。

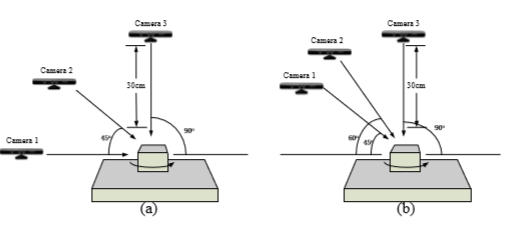
1. **数据集和查询**

一个包含实际物体和打印模型的3D模型数据集，包含他们的多模态视图，多视角RGB-D模型数据集（MV-RED）被收集，用于本次比赛。MV-RED数据集包括605个模型，包括100个3D打印模型和505个3D真实物体，可分为60个类别，如苹果、CAP、围巾、杯子、蘑菇和玩具。对于每个模型，RGB和深度信息都由3个Microsoft Kinect传感器从3个方向同时记录。也就是说，对于每个模型有两种图像数据——RGB和深度图像数据。示例视图可以从图1中看到。数据集是用3个Kinect传感器记录的（第一代），但是是在两种不同的相机设置下分开记录的，如图2（a）和图2（b）所示。202个模型是使用第一种相机排列方式记录的，303个模型是用另一种相机排列方式记录的。100个3D打印模型也使用第二种方式记录。相机1和相机2分别的捕获36个RGB和深度图像，通过一个由步进电机控制的均匀旋转的桌子。相机3仅仅捕获一个RGB图像和深度图像，在从上到下的视角。采用这种设置，一个模型可以获得73个RGB图像和73个深度图像。对于每个RGB和深度图像，图片的分辨率是640X480。为一些RGB图像提供了一些分割结果。这605个模型分别属于60类物体。这里，选择100个3D打印模型和100个3D真实物体用作训练，而其他405个对象用于测试。



图一：在MV-RED数据集中的例子。

（a）数据集中真实物体的图像示例；(b)数据集中的3D打印模型的图像示例



图二：每个物体的记录场景

1. **评估**

为了评估所有参与的方法的性能，采用以下评价标准[GJC\*14 ]。

1. Precision-Recall Curve（精确回忆曲线）综合地展示了检索性能，在平均召回率和平均准确度方面进行了评估，并被广泛应用于多媒体应用中。
2. NN（神经网络）评估第一次返回结果的检索准确率。
3. FT用于评估前τ个搜索结果准确度，其中τ是查询的相关模型的数量。
4. ST用于评估前2τ个搜索结果。
5. F-Meansure（F）联合评估了准确度和排名靠前的返回结果。在我们的实验中，前20个检索结果被用于F1计算。