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。

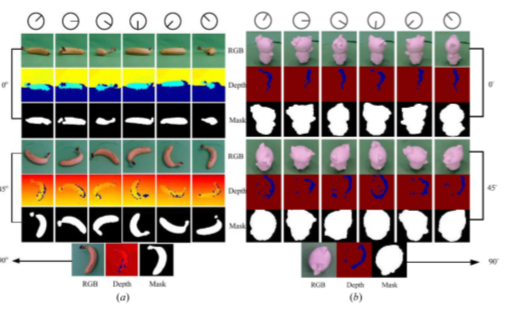
1. **简要介绍**

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。

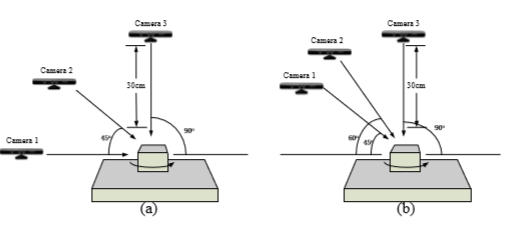
1. **数据集和查询**

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。
6. 归一化折现累积增益（NDCG）是在假设用户不太可能考虑较低结果的情况下，在具有较高权重的排名位置上分配相关结果的统计量。
7. 平均归一化修正检索秩（ANMRR）是一种基于秩的测度，它考虑在被检索出模型当中的相关模型的排序信息。较低的ANMRR值指示更好的性能，即相关模型排名靠前。
8. **参与者**

7个小组参加了这项研究，9个结果被提交。参与者的详细资料和相应的投稿人如下所示。

1. CGM-Zernike 和CGM-HoG方案由Lili Pan, Yali Zheng 和Yu Qiu提交，他们来自于中国电子科学技术大学。
2. NN-CNN方案由Hongjiang Wei 和 Yuyao Zhang提交，他们分别来自于加利福尼亚大学伯克利分校和杜克大学。
3. CVEM方案由Zhao Wang, Le An, Jun Zhang 和 Mingxia Liu提交，他们分别来自于麻省理工学院与北卡罗来纳大学。
4. Dense-BoW方案由Rongrong Ji, Liujuan Cao, Yan Zhang 和 Fuhai Chen提交，他们来自于中国厦门大学。
5. Hypergraph-Zernike 方案由Yue Gao 和 Qionghai Dai 提交，他们来自于中国清华大学。
6. RBoW/RBoW-He方案由Jiayun Hao, Liwei Wei 和Shuilong Dong提交，他们来自于中国北京工商大学。
7. Balancing Distance Learning (BDL)方案由Yu De 和 Zan Gao提交，他们来自于中国天津工业大学。

表一中提供了这些方法的简要总结。

表一：报名小组的清单

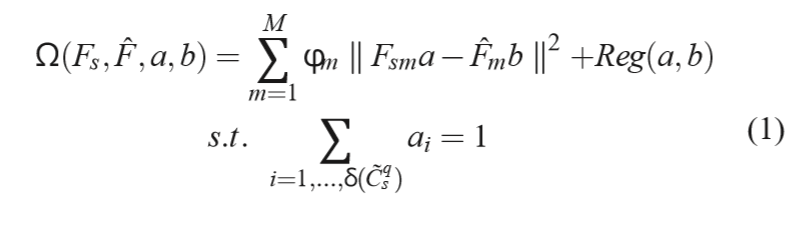
|  |  |  |
| --- | --- | --- |
| 参与者 | 方案名称 | 技术 |
| 中国电子科技大学 | CGM-Zernike  CGM-HoG | 图片配 |
| 加利福尼亚大学伯克利分校，杜克大学 | NN-CNN | 深度学习 |
| 麻省理工学院北卡罗来那大学 | CVEM | 图片配 |
| 厦门大学 | Dense-BoW | BoW |
| 清华大学 | Hypergraph-Zernike | 超图匹配 |
| 北京工商大学 | RBoW  RBoW-He | BoW |
| 天津工业大学 | Balancing Distance Learning (BDL) | 距离测量 |

1. **方法**

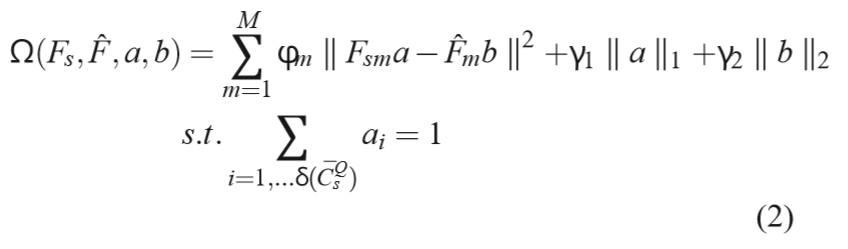
**5.1．基于CGM的三维模型检索，中国电子科学技术大学（CGM-Zernike/CGM-HoG）**

每一个三维模型由一组多视角的二维图像代表，它可以用一个图模型表示。一个经典的图G=（V，ε）包含节点集合=和边的集合.三维模型检索的任务需要计算查询模型和独立的个体候选模型之间的相似性得分。聂等人[NLGS15]的图匹配方法（CGM）用于计算查询模型和候选模型之间的相似性。CGM提出的团图进一步通过保持全局和局部结构，提出了一种团图匹配方法，它能有效处理多视图匹配问题。这里，每个图模型由两种元素组成：团集和与个体团相关联的属性集。

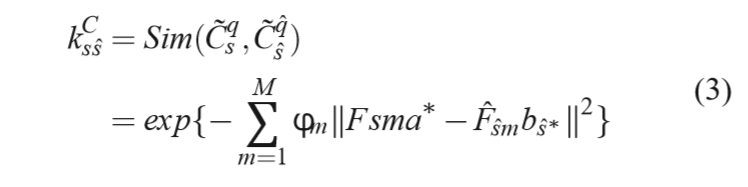
一个查询模型被表示为,一个候选模型被表示为。在一个查询模型中用特征集考虑一个团体，在一个候选模型中用多个特征集来考虑多个团体。团体间的相似性按照如下公式计算：



是在a中的系数，是在公式中是必要的，并且可以避免无效的结果;是用于重建的系数并且可以被分解为,其中是候选模型中与团相关联的系数的子向量。表示第模态的权重。这样，所提出的凸目标函数可以表述如下：



其中和是稀疏系数。通过对上述方程的最小化，可以得到最优的系数向量和可以被重写为。团集的相似性可以被重新定义如下：



最后，将团图匹配成功地引入到传统的图匹配方法中。经典的IQP框架可以用来处理这个匹配问题。在我们的结果中，提交了两组应用Zernike矩特征和HoG特征，即CGM-Zernike和CGM-HoG，的实验结果。

**5.2．基于卷积神经网络特征的最近邻域方案，美国加利福尼亚大学伯克利分校，杜克大学（NN-CNN）**

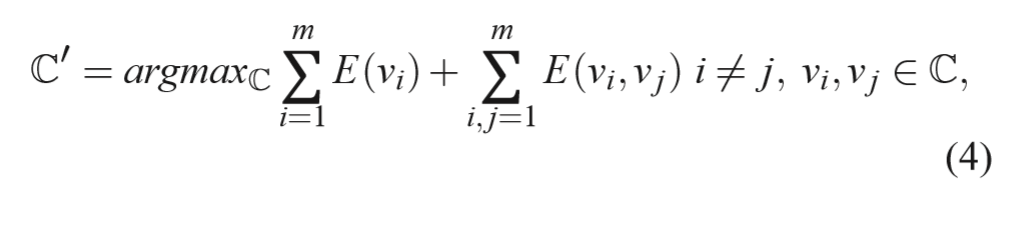
在本次竞赛中，卷积神经网络（CNN）模型被用来从多模态3D模型中提取特征。在本次研究中，CNN特征提取的整个流程包含两个步骤：第一步是以监督的方式训练CNN模型，然后从RGM图像和深度图像中提取深层次特征。最后应用最近邻域来计算不同模型之间的相似性。图3显示了框架的概述。

这次实验中，使用了一个19层的深度卷积神经网络模型，它在ILSVRC’12上进行了预训练，将每一个图像分类到1000个种类中，从RGB图像中提取CCN特征。同时，组织者为每个RGB图像提供深度信息。深度图像是描述三维物体，尤其是形状和距离信息的重要信息。CNN对RGB图像进行预训练可以适应产生深度图像的CNN特征。这种深度CNN特征涉及丰富的形状和结构信息。

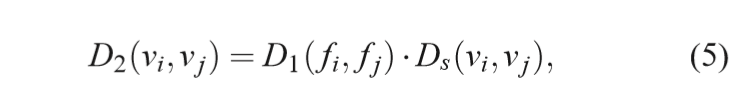
现在，颜色和深度特征已经从一个3D模型的每个视图中提取出来了，将每个特征投影到相似度量空间中，并且可以获得每个视图的相似度得分。然后将这些互补的多视角CNN评分通过加权融合方案相结合，以获得更准确的检索结果。

**5.2．基于图的特征视图集提取与匹配，麻省理工学院和北卡罗来纳大学(CVEM-HoG)**

对于每个基于视图的3D模型，在多个视图中有太多冗余信息，特别是在每个对象的73个视图中，每个对象的原始2D图像需要通过利用视觉和空间信息来去除冗余。图像聚类的规则是最大限度地提高类间相关性，同时最小化类间相关性。因此，视图约束聚类方法可以被表述为能量最小化问题。目标函数由两部分组成，它们可以被定义为：

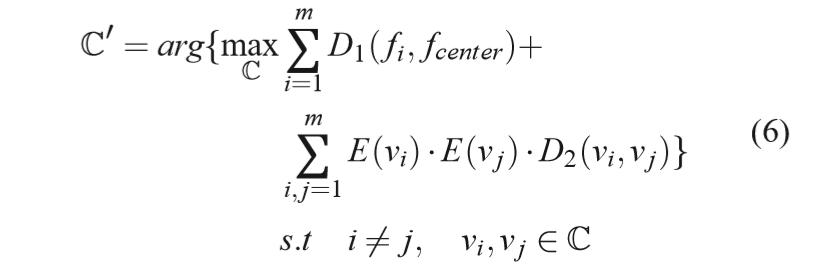


其中表示视图的能量，它表示了这个视图对该聚类C的贡献；表示不同视图之间的相关性。如果两个不同的视图和属于C，应该有更高的值。和的和代表了一个特殊的C聚类策略的全部能量。代表了C中的中心点的特征；代表了的特征；代表了和之间的相似性，它是按照欧几里得距离计算的。影响和的相关性。它可以用表示，它代表了和的相似性，计算方式如下：



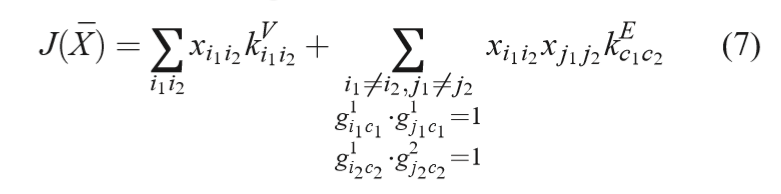
是根据欧几里得距离计算的，表示不同的两个视图之间的空间相似性，这是由和之间的球面距离计算的。球体的中心是这个3D模型的中心。

最后，Eq.4可以转换为：



在上述过程后，原始聚类问题已经成功地转化为一个能量最大化问题。在这项研究中，应用图形切割[TMN14]获得一组子集群。这里，我们从这个子簇中提取每个图像作为代表性的视图。

到目前为止，模型匹配问题可以表述为图匹配；其目的是确定下面函数最大值的图1和图2的节点之间的对应关系：



如果图形1的第个节点对应于图形2的第个节点，其中矩阵代表了节点对应关系，即，。

通过优化Eq.7，我们可以得到图1和图2之间的匹配得分，以处理检索问题。

**5.4. Dense-BoW ，厦门大学**

不同于直接利用RGB或深度特征来表示的其他方法，该方法主要是利用流行的bag-of-words（BoWs）表示和其他基于形状的特征来完成这项任务。在这种方法中，对于每张图片，两种类型的特征被提取出来，包括Zernike 和 Dense-BoW。

每个模型用一组视图集合，HoG特征是在密集采样点上提取的。使用的词汇的数量是。然后每个视图可以由一个维向量表示。为了获取形状信息，分别从每个图像中提取Zemike力矩，从而得到一个1×49矩阵MHoG。