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。

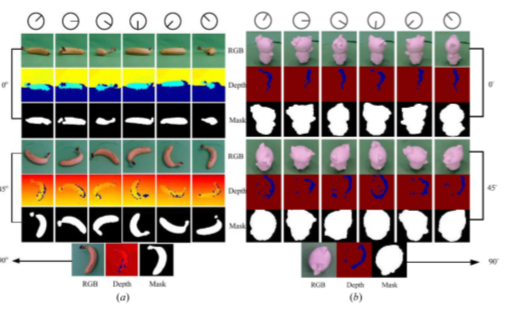
1. **简要介绍**

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。

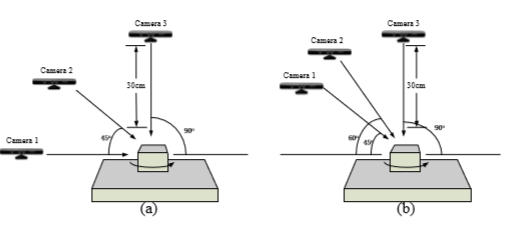
1. **数据集和查询**

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

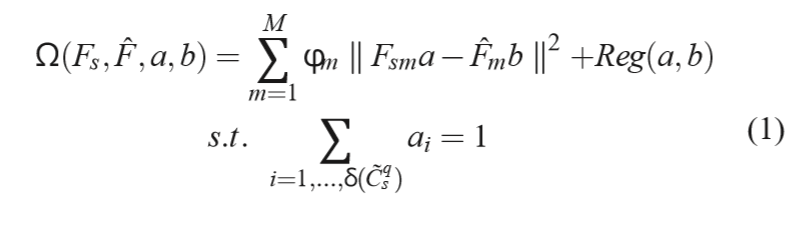
|  |  |  |
| --- | --- | --- |
| 参与者 | 方案名称 | 技术 |
| 中国电子科技大学 | CGM-Zernike  CGM-HoG | 图片配 |
| 加利福尼亚大学伯克利分校，杜克大学 | NN-CNN | 深度学习 |
| 麻省理工学院北卡罗来那大学 | CVEM | 图片配 |
| 厦门大学 | Dense-BoW | BoW |
| 清华大学 | Hypergraph-Zernike | 超图匹配 |
| 北京工商大学 | RBoW  RBoW-He | BoW |
| 天津工业大学 | Balancing Distance Learning (BDL)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

1. **方法**

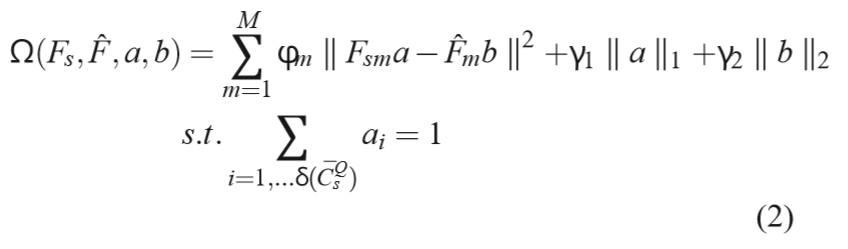
**5.1．基于CGM的三维模型检索，中国电子科学技术大学（CGM-Zernike/CGM-HoG）**

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。

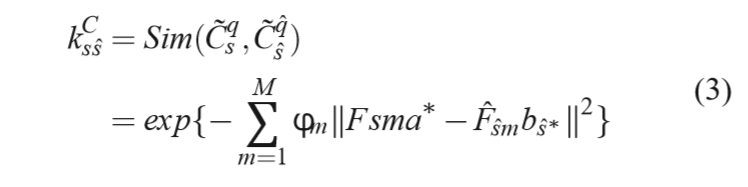
一个查询模型被表示为,一个候选模型被表示为。在一个查询模型中用特征集考虑一个团体，在一个候选模型中用多个特征集来考虑多个团体。团体间的相似性按照如下公式计算：



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。

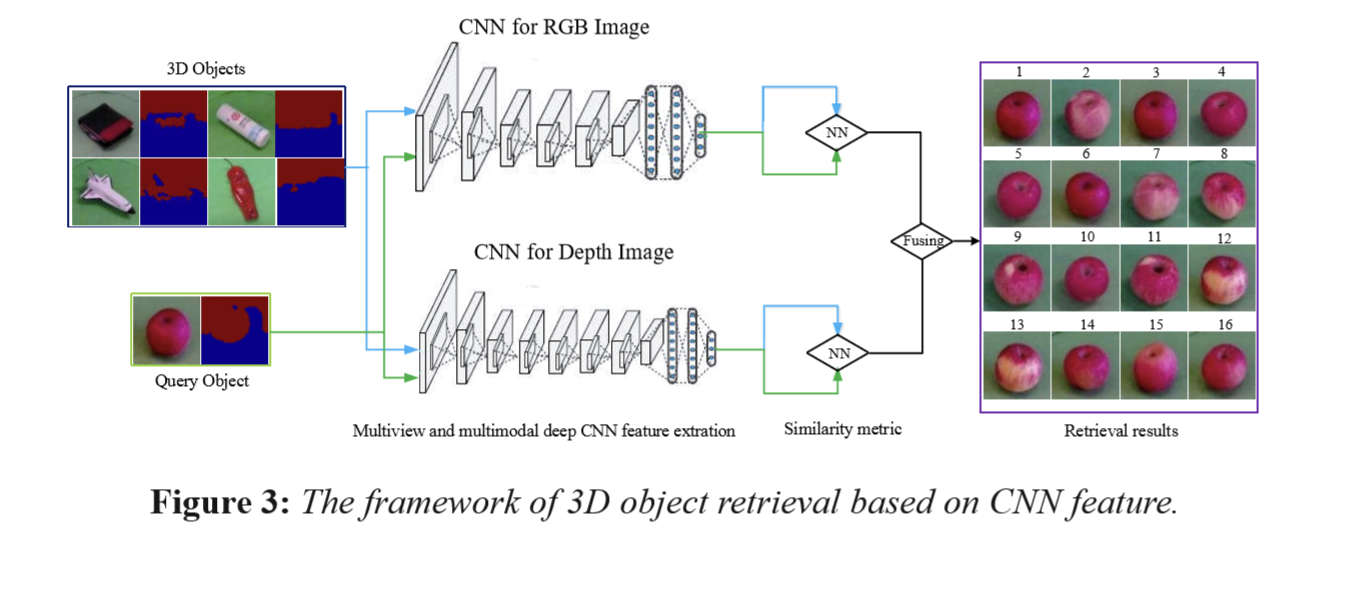
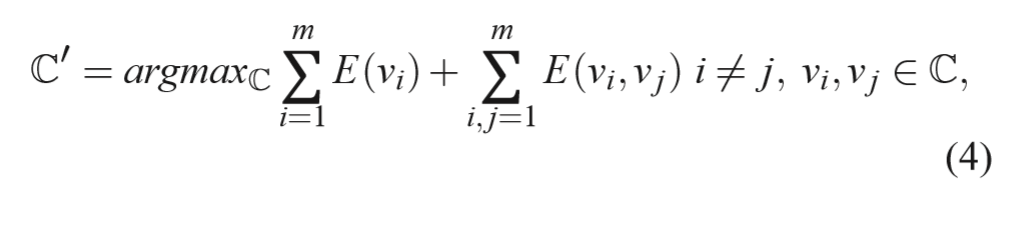


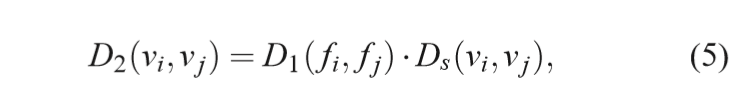
图3：基于CNN特征的3D模型检索框架

**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：

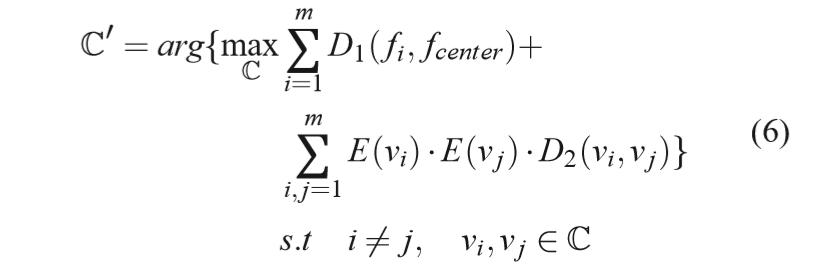


其中表示视图的能量，它表示了这个视图对该聚类C的贡献；表示不同视图之间的相关性。如果两个不同的视图和属于C，应该有更高的值。和的和代表了一个特殊的C聚类策略的全部能量。代表了C中的中心点的特征；代表了的特征；代表了和之间的相似性，它是按照欧几里得距离计算的。影响和的相关性。它可以用表示，它代表了和的相似性，计算方式如下：



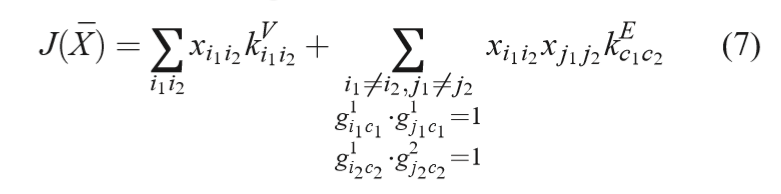
是根据欧几里得距离计算的，表示不同的两个视图之间的空间相似性，这是由和之间的球面距离计算的。球体的中心是这个3D模型的中心。

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，。

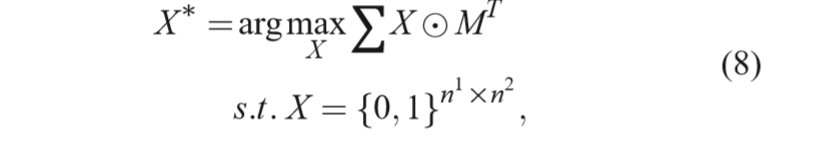
通过优化Eq.7，我们可以得到图1和图2之间的匹配得分，以处理检索问题。

**5.4. Dense-BoW ，厦门大学**

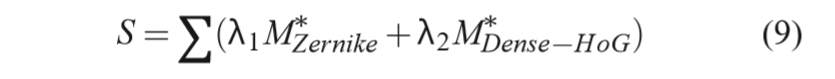
不同于直接利用RGB或深度特征来表示的其他方法，该方法主要是利用流行的bag-of-words（BoWs）表示和其他基于形状的特征来完成这项任务。在这种方法中，对于每张图片，两种类型的特征被提取出来，包括Zernike 和 Dense-BoW。

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。

比较两个3D模型和，对应的特征矩阵，和，可以被首先生成，其中代表了每一个视图的Dense-BoW特征。欧几里得距离被用来测量和之间的距离。然后一个的矩阵可以被用来表示和之间的关系。Eq.8被用来计算和之间的在不同特征空间中视图匹配结果。



利用贪心算法[EÖ98]处理该优化问题，得到最佳匹配结果X。根据不同特征空间中不同的匹配结果，使用Eq.9生成最终匹配得分。

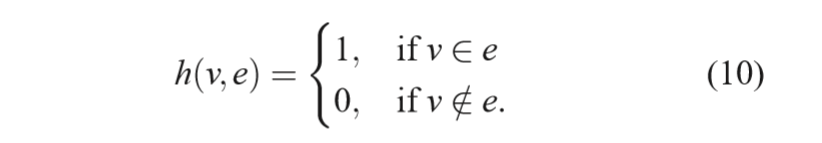


其中和是不同特征矩阵的权重，S是最终的匹配分数，它被用来表示和之间的相似性。3D模型检索是基于查询模型和数据库中模型之间的匹配得分S。

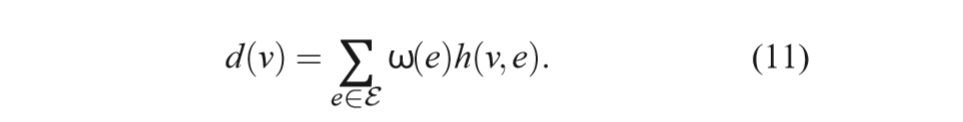
**5.5. 基于超图分析的三维物体检索与识别，清华大学（(Hypergraph-Zernike）**

在这项研究中，我们利用超图结构来表示一个基于视图的3D模型。我们将所有对象的视图分组成集群。然后将每个集群视为连接该集群中具有视图的对象的边缘（注意边缘可以连接超图中的多个顶点）。构建一个超图，它里面的顶点代表了数据库中的模型。我们根据集群中任意两个视图之间的视觉相似性来确定边缘的权重。通过改变集群的数目，许多超图可以被生成。超图实际采用不同的粒度编码了模型之间的关系。通过对这些超图进行检索和识别，我们可以避免模型距离估计问题，因为超图已经全面地描述了模型之间的关系。为了检索，我们使用相同的权值来融合超图。然而，我们利用训练数据来识别多个超图的最佳组合系数。

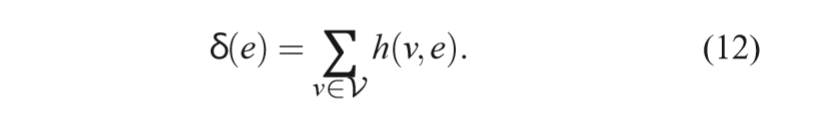
首先，我们介绍超图的结构。一个超图是由一个顶点集V和边的权重构成的。每一个边e对应一个权重。超图G可以用一个距离矩阵H来表示，其中每一个条目被定义为：



对于一个顶点v∈V，它的分数被定义为：

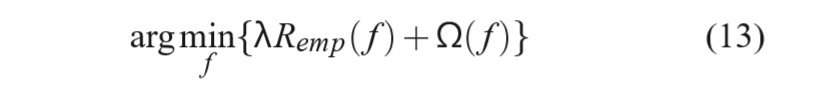


对于一条边e∈ε，它的分数被定义为：

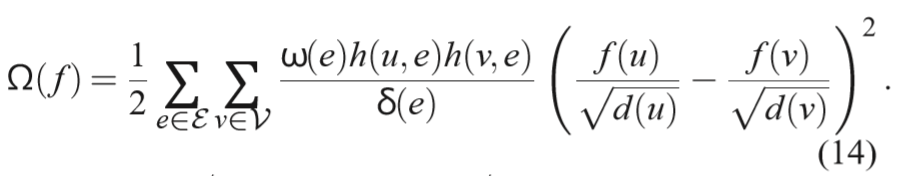


我们让和分别表示顶点度和边缘度的对角矩阵。用W表示边权的对角矩阵。

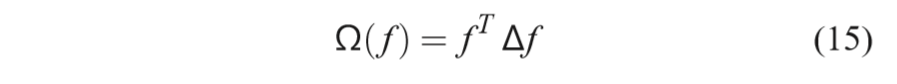
我们希望把检索任务看作是一类分类问题。可以在超图上执行不同的机器学习任务，如分类、聚类、排序和嵌入。在这里，我们使用了二元的分类框架[ZHS06]。



其中是分类函数，Ω()是一个正则化的超图，是一个经验损失，λ>0是折中的参数。超图的正则化被如下定义：

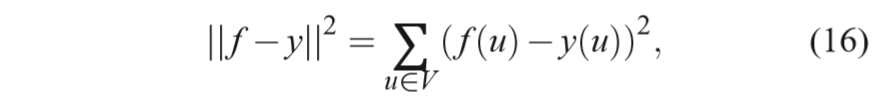


设，。归一化成本函数可以写成：

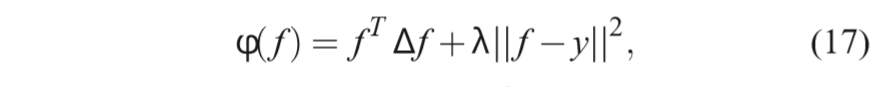


其中，是一个正半离散矩阵，通常称为超图拉普拉斯算子。

在这项研究中，所有的超图分享相同的符号V。因此，对于所有的.我们让。损失函数项被定义为：



其中y是标签向量。N代表数据库中模型的数量，并且假设i个模型被选做查询模型。让y表示一个n\*1维的向量，其中y中所有的元素都是0，除了i的值是1。3D模型检索的学习任务是最小化两个式子的和：



最终，我们通过求的微分，获得了,用于处理检索问题。

**5.6. BoW 和 BoW-He 方案，北京工商大学（BoW/BoW-He）**

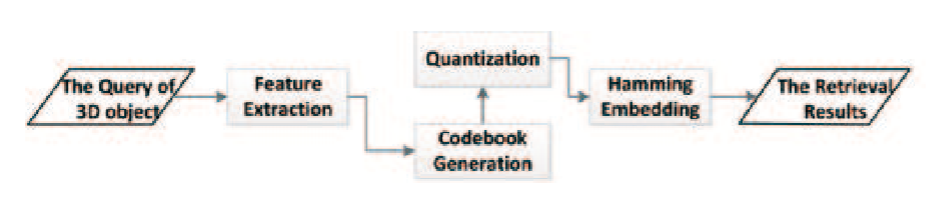
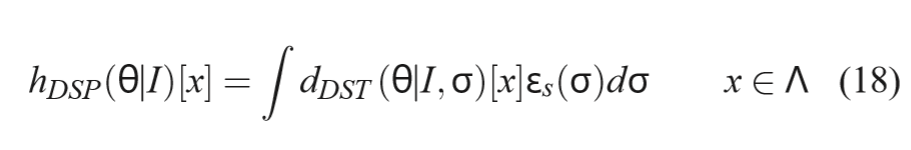


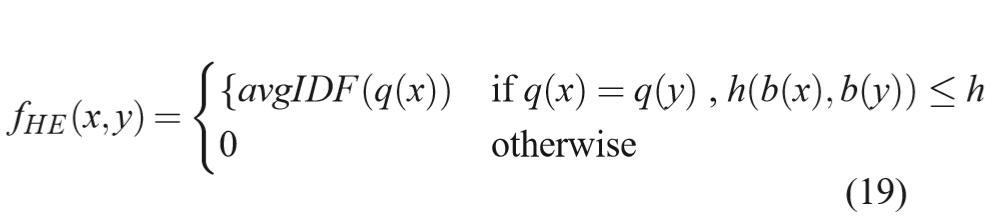
图4：BoW-He方案的框架

这个方案采用了Bag-of-words（Bow）模型。正如图4中展示的那样，我们方案的主要步骤如下：DSP-SIFT[DS14]是在这个步骤中，从3D对象的所有RGB图像中提取的。DSP-SIFT按照如下方式给出：

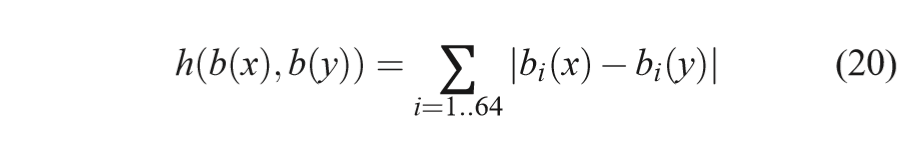


其中s是大小池的规模，ε是指数或其他单边密度函数。提取后，DSP-SIFT使用rootSIFT变换。利用近似K均值生成rootSIIFT[Zis12]码本。之后，使用近似近邻法（ANN）将每个描述符量化为码本中的近质心。利用多个赋值进行更好的回忆，其中描述符被分配给3个视觉单词，就像[ZWLT14]。

在这一步中，两种方法被分别使用。1）Refined BoW(RBoW)。avgIDF[LZ13]被用来计算相似距离。2）Reﬁned BoW with Hamming Embedding（RBoW-HE）。我们使用Hanmming Embedding(HE)[JDS08]来计算两个模型之间的相似性。如果描述符X被量化为其二进制签名，则两个描述符x和y之间的匹配函数可以被定义为



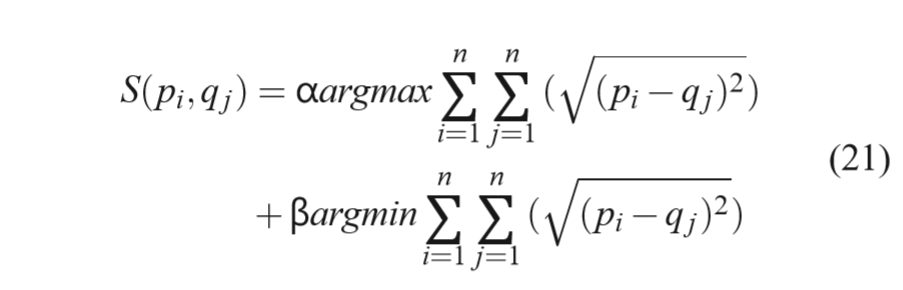
其中是一个修正的汉明阈值，就像,avgIDF在[4]中被定义。是汉明距离，被定义为



**5.7.基于平衡距离学习的三维模型检索，天津工业大学（BDL）**。

基于最近邻域算法和Hausdorff距离学习算法，提出了基于平衡距离学习的三维模型检索算法。对于最近邻域算法，它只考虑模型的相似性，而不考虑相似模型的偶然性和不同模型的差异。对于Hausdorff算法，它通过比较最远的距离来考虑相似物体的偶然性，但不考虑两个物体之间的一般因素。因此，我们讨论并开发了一种平衡偶然性、差异性和一般性的新算法。

对于同一类模型，它们的特征一般相似，但不能排除一些图像特征有很大差异的可能性。因此，在特征的比较上偶尔存在差异。对于不同的类模型，它们的特征通常是不同的。即使在两个不同的类模型之间最相似的图像特征，它们的特征距离也很大。因此，为了平衡影响因素之间的偶然性、差异性和一般性，我们增加了两个参数来控制这些因素？权重。我们的核心算法按照如下方法定义：



和表示来自两个模型的图像特征的每个维度，n是样本的数量。α和β是学习算法自动计算的影响因素。Eq.21可以计算两个模型的相似性。符号的值越大，这两个模型越不相似，否则，它们是相似的。该算法较好地平衡了两个不同模型之间的一般性和特殊性的关系，避免了最近邻域算法和Hausdorff算法的不足之处。

**6. 结果**

在这一节中，我们展示了七个小组为这个任务提交的9个的结果。图5展示了所有查询的定量评价结果。图6展示了100个真实查询模型的定量评价结果。图9展示了100个3D打印模型的查询的定量评价结果。图7显示了所有查询的精确回忆曲线。图8显示了100个真实3D查询的精确回忆曲线。

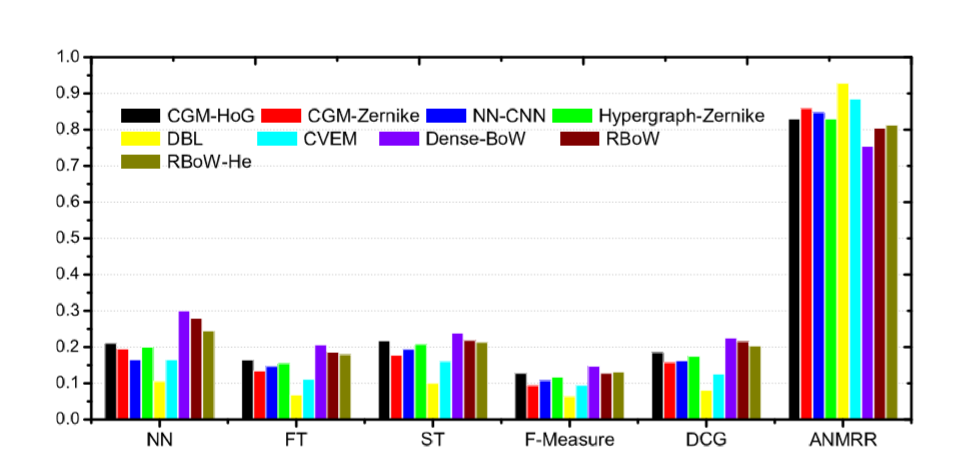


图5：基于每个模型的不同方法的评价得分。

从所有的参与者中看出，结果已经表明了使用多模态的视图的3D模型检索的性能。从结果来看，我们可以得到如下的观察结果。

1. 基于BoW的方法，例如，Dense-BoW，RBoW和RBoW-He方法相比于其他方法表现更好。这表明，即使在这样一个具有挑战性的任务下，基于BoW的学习能够探索对于3D对象的区分性特征。另外，在所有指标中密集BoW获得最佳检索结果；

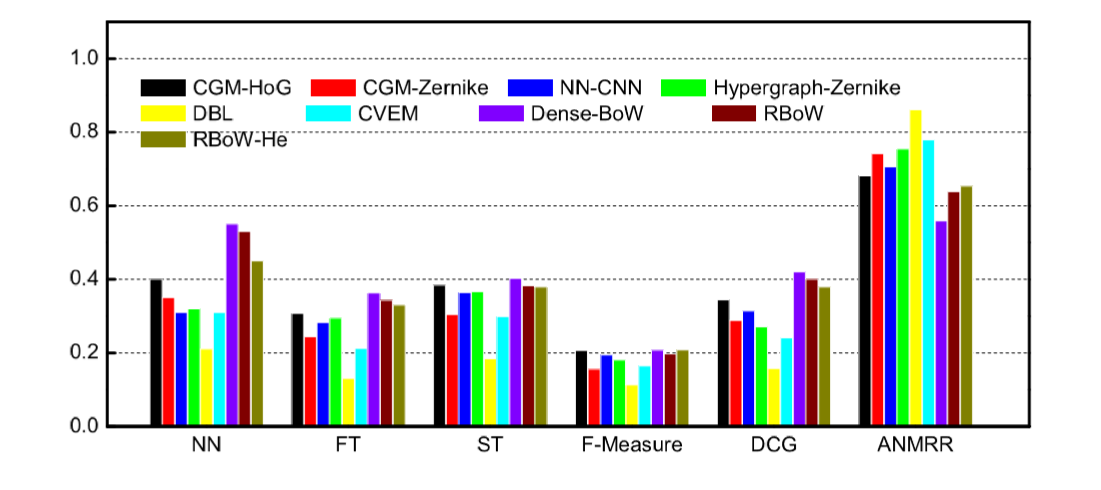


图6：基于每个3D真实物体的不同方法的评价得分

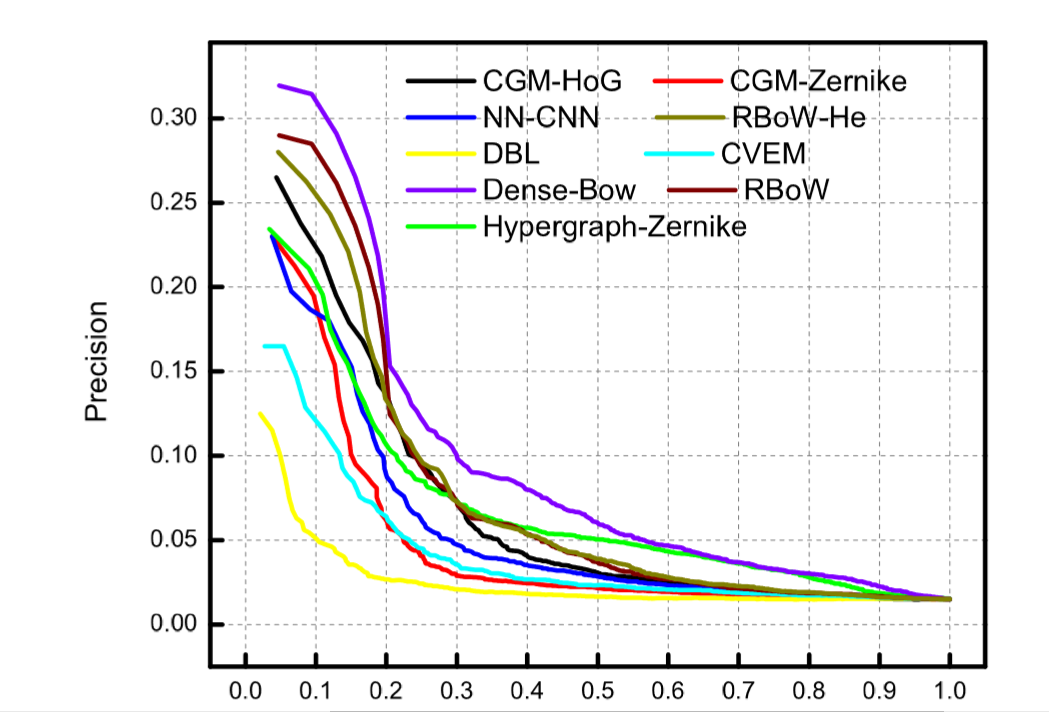


图7：基于每个模型的不同方法的精确回忆曲线

2. 在每个图中使用边缘的方法比使用每个图中的节点的方法更好。CGMZernike和CGM-HoG是利用边缘和节点之间的关系的两种方法。我们可以发现CGM-HoG比CGM-Zernike获得了更好的性能。另一个例子是CGM-Zernike和NN-CNN之间的比较。结果表明，深度数据可以传达更多的3D结构信息，并且可以比RGB数据更具区分性。

3. 从图9中，我们可以发现对于几乎所有的方案使用3D打印模型作为查询的结果没有比使用3D真实物体作为查询结果有更显著的进步。3D打印模型尺寸太小，无法保存足够深度信息或形状信息的原因。同时，3D打印模型的单色信息也限制了从RGB图像中提取的信息。在3D打印模型上的检索性能也表明民用3D打印机的技术还不足以满足所有人的需要。3D打印模型中不完整的形状信息也很难实现3D模型设计的大规模生产。这些问题将是今后非常具有挑战性的工作。

**7.结论**

综上所述，这项研究已经吸引了人们对使用多模态视图进行三维物体检索的研究。这是一项具有挑战性的任务，并且测试数据集中的所有数据都来自真实对象。我们有七个小组，他们成功地参加了本项研究，贡献了9个结论。本次研究作为一个平台，征求已有的基于视图的3D模型检索方法。在本文中，我们介绍了数据集，所有参与的方法和相应的性能。从分析结果来看，基于BoW的方法比其他方法更好。我们还观察到，深度特征可以比RGB特征更有效。对边缘相关的应用也进行了评价，得到了满意的结果。

虽然所有参与本次研究的方法已经达到了改进的性能，但任务仍然是挑战性的，并且结果远远不能令人满意和满足实际应用。基于视图的三维模型检索仍有很长的路要走。

**参考文献**

**[ADV07]** ANSARY T. F., DAOUDI M., VANDEBORRE J.-P.: A bayesian 3-d search engine using adaptive views clustering. TMM 9, 1 (2007), 78–88. 2

**[AKKS99]** ANKERST M., KASTENMÜLLER G., KRIEGEL H.P., SEIDL T.: 3d shape histograms for similarity search and classiﬁcation in spatial databases. In SSD (1999), pp. 207–226. 1

**[CTSO03a]** CHEN D.-Y., TIAN X.-P.,SHEN Y.-T., OUHYOUNG M.: On visual similarity based 3d model retrieval. In Computer graphics forum (2003), vol. 22, Wiley Online Library, pp. 223– 232. 1

**[CTSO03b]** CHEN D.-Y., TIAN X.-P., SHEN Y.-T., OUHYOUNG M.: On visual similarity based 3d model retrieval. Comput. Graph. Forum 22, 3 (2003), 223–232. 2

**[DS14]** DONG J., SOATTO S.: Domain-size pooling in local descriptors: Dsp-sift. Eprint Arxiv (2014), 5097–5106. 6

**[EÖ98]** ETZION T., ÖSTERGÅRD P. R. J.: Greedy and heuristic algorithms for codes and colorings. IEEE Transactions on Information Theory 44, 1 (1998), 382–388. 5 [GJC∗14] GAO Y., JI R., CUI P., DAI Q., HUA G.: Hyperspectral image classiﬁcation through bilayer graph-based learning. TIP 23, 7 (2014), 2769–2778. 2 [GTH∗12] GAO Y., TANG J., HONG R., YAN S., DAI Q.: Camera constraint-free view-based 3-d object retrieval. TIP 21, 4 (2012). 2

**[JDS08]** JEGOU H., DOUZE M., SCHMID C.: Hamming embedding and weak geometric consistency for large scale image search. In Proceedings of the 10th European Conference on Computer Vision: Part I (2008), pp. 1.1–1.1. 6

**[LS88]** LAM L., SUEN C. Y.: Structural classiﬁcation and relaxation matching of totally unconstrained handwritten zip-code numbers. PR 21, 1 (1988), 19–31.

**[LZ13]** LIANG ZHENG SHENGJIN WANG Z. L. Q. T.: Lp-norm idf for large scale image search. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2013), pp. 1626–1633. 6

**[NLGS15]** NIE W.-Z., LIU A.-A., GAO Z., SU Y.-T.: Cliquegraph matching by preserving global & local structure. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), pp. 4503–4510. 3

**[SLW07]** SHIH J.-L., LEE C.-H., WANG J. T.: A new 3d model retrieval approach based on the elevation descriptor. PR 40, 1 (2007), 283–295. 2

**[TMN14]** TANIAI T., MATSUSHITA Y., NAEMURA T.: Graph cut based continuous stereo matching using locally shared labels. In CVPR (2014), pp. 1613–1620. 4

**[ZHS06]** ZHOU D., HUANG J., SCHÖLKOPF B.: Learning with hypergraphs: Clustering, classiﬁcation, and embedding. In Advances in neural information processing systems (2006), pp. 1601–1608. 6

**[Zis12]** ZISSERMAN A.: Three things everyone should know to improve object retrieval. In Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2012), pp. 2911–2918. 6

**[ZWLT14]** ZHENG L., WANG S., LIU Z., TIAN Q.: Packing and padding: Coupled multi-index for accurate image retrieval. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on (2014), pp. 1947–1954.