附录B

从虚拟三维草图到三维形状检索

**摘要**

越来越多的免费在线三维图形会进行二次或多次三维检索。然而，关于触发检索的最佳输入方式是什么，以及这种检索的最终使用场景，一直存在着激烈的争论。在本文中，我们提供了一个不同的角度来回答这些问题：我们研究了三维草图作为输入模态的使用，并提倡一个虚拟现实场景中进行检索。因此，最终的愿景是用户可以在虚拟现实环境中通过空中涂鸦自由检索3D模型。作为这个新的3D-VR草图对3D形状检索问题的第一个尝试，我们做了四个贡献。首先，我们编写了一个虚拟现实实用程序来收集三维虚拟现实草图并进行检索。其次，我们从ModelNet收集了第一组167个3D虚拟现实草图并把这些草图分为两类。第三，我们提出一种新的方法来产生一个不同抽象层次的人像三维草图的合成数据集来训练深层网络。最后，我们比较了常用的多视图和体绘制方法：我们发现，与三维形状到三维形状的检索相比，基于体绘制点的方法在三维草图到三维形状的检索中表现出更好的性能，这是由于三维VR草图的稀疏性和抽象性。我们相信，这些贡献将共同成为今后解决这一问题努力的助推器。

**1.介绍**

三维模型检索已经成为一个重要的话题，由于越来越多的免费在线三维存储库。它在CAD设计、3D打印、3D动画和电影制作中都有应用，通过从现有的3D形状集合中检索，可以大大减少3D形状建模所需的时间。已经尝试了各种输入模式——即使用3D扫描或现有的3D形状，但多数还是图像和粗略的2D草图，目前最新的的研究集中在3D到3D。

尽管取得了巨大的进步，但两个突出的问题仍然存在：什么是启动3D检索的最佳输入模式，以及在什么使用场景下可以最好地促进此类检索。对于前者，二维草图和图像都是二维的，因此无法达到所需的细节层次，二维-三维域的差距也可能是反直观的。另一方面，基于3D的范例要求现有的3D模型随时可用，这在某种程度上形成了一个“鸡和蛋”的问题（首先在哪里/如何获取输入模型）。至于后者，除了由用户自由定义的2D草图外，所有其他使用场景都不能提供所需输入方面的灵活性即不能轻易改变图像，更不用说3D模型或者扫描了。

在本文中，我们提供了一个新的视角三维形状检索——我们提倡使用三维草图作为一种新的输入模式。这种新的3D-VR从草图到3D形状的模式不仅能够根据常见的3D模型到3D模型的设置进行详细的检索，而且同时还提供了在基于2D草图的检索中其他地方发现的一定程度的灵活性。我们的最终愿景是：在考虑到一个特定的3D模型的情况下，一个人出现在一个虚拟现实环境中，使用手持控制器大致勾勒出模型，按住检索按钮，然后相关的模型将开始填充虚拟现实环境。

因此，我们的第一个贡献就是为上述目的编写一个虚拟现实环境。在这个虚拟现实环境中，我们收集了第一个人类虚拟现实草图数据集。10名没有艺术背景的用户被雇佣从ModelNet的椅子和浴缸两个类别中制作了167幅3D-VR草图。由于收集过程对时间和成本都很敏感，因此我们还向三维草图生成器提出了第一个三维模型，并构建了一个三维草图的合成数据集。我们的生成器的一个关键特性是，它可以生成不同抽象级别的三维草图，有效地模拟真实人体草图中的三维草图。通过使用真实数据集和合成数据集对一系列深三维虚拟现实草图进行三维检索模型的训练，我们得出了一些重要的启示：随着抽象程度的提高，检索性能下降；使用合成草图训练的模型在人体草图上进行测试时，已经可以达到相当好的性能水平。

最后，我们对不同的形状/草图表示进行了实验，并对三维模型检索中常用的最新损失进行了分析，但针对我们的问题进行了重新设计。由于草图的两个关键特性：稀疏性：完整的三维模型与稀疏的草图线，以及抽象性：三维模型在几何上是完美的，而草图是受变形影响的，因此我们重点研究了三维形状与三维草图之间的域差距。特别地，我们研究了基于多视角和基于点的三维表示，并表明后者对稀疏性和抽象程度的提高都更为稳健。我们进一步提出了一个具有重构路径的架构，以明确地允许抽象性的变化。

总之，我们的贡献包括：三维模型检索的新视角，首次使用三维虚拟现实草图进行检索；使用专门构建的虚拟现实环境收集的人类三维虚拟现实草图数据集；生成具有可变抽象度的合成三维草图的方法，使用最近的三维形状检索模型进行综合评价，这些模型被重新用于基于三维草图的检索任务，以获得见解，加上一种新的正则化轨道，以解决草图模型领域的差距。

**2.相关工作**

当处理单个图像或三维形状的三维形状检索时，现有的工作根据使用的形状表示分为两大组：基于视图的或体积的。体积表示可以进一步细分为点云[或基于体素的。在这项工作中，我们表明，在三维草图到三维模型检索的任务中，点云表示优于多视图方法，因为它能更好地处理图像三维草图的稀疏性。

基于二维图像和三维形状的检索。多类环境下的检索问题与形状分类问题密切相关，其中图像或三维形状的中间嵌入用于检索。多类形状分类的常规方法是使用软最大交叉熵损失。其他人的工作是专门处理检索，其中三重态丢失及其变体已成为一个共同的标准。其中，引用的文献14代表了3D模型检索的最新技术。它结合了三重态和中心损耗，通过使用类中心作为正样本来解决它们各自的缺点。在这项工作中，我们评估了基于三维草图的检索任务中的三重态和三重态中心损失，并将它们与一个额外的重建损失相结合，以适应基于三维草图的检索问题。

由于图像和三维形状之间存在域间隙，或者目标和查询三维模型来自不同的分布，暹罗或异构网络体系结构对于某个问题更为有利。在这项工作中，我们比较了这两种结构的多视图基线，并表明，由于三维草图和三维形状之间存在很大的域差距，异质结构远远优于暹罗结构。

基于三维草图的三维模型检索。到目前为止，基于三维草图的检索工作很少，尤其是近年来。他们中的大多数人使用的是使用微软Kinect收集的草图，这样不仅跟踪精度有限，而且收集界面也是反直观的，在可视化2D投影的同时进行3D草图绘制。因此，收集的草图大多保真度较低，显示的细节较少。我们的虚拟现实草图是完全不同的：可视化和素描都是在三维执行，和最新的虚拟现实技术的使用提供了高精度。它们共同确保了我们的虚拟现实草图具有高保真度和丰富的细节，因此更适合检索。一个值得注意的例外是Giunchi等人的工作，然而他们解决了一个不同的问题，即密集彩色VR草图（以及可选的基础3D形状）的3D模型检索，而我们的目标是更简单和抽象的形状表示——一组稀疏的单色线。需要注意的是，我们无法在开放访问中找到上述数据集，因此无法提供直接比较。然而，我们的NGVNN[13]基线已经优于和使用的最新技术。

非真实感渲染（NPR）。NPR是一个古老的图形和视觉问题，有关生成2D NPR渲染的现有方法的详细概述，请感兴趣的读者参阅最近的报告。Li等人提出了唯一一种尝试，即产生类似于三维草图的表示，自动形成三维形状，作为六个标准视点形状视图的串联。他们用这种表示法表明，他们的非学习方法在离线到三维模型检索任务中的性能明显优于三维草图到三维模型检索的性能。实验表明，这种简单的形状表示不足以作为基于三维草图的三维形状检索的训练数据。我们的实验证明：在抽象度较高的草图上训练的网络显著优于在干净草图上训练的网络。据我们所知，我们是第一个提出从三维模型生成抽象三维草图的方法。

**3.数据集合**

收集完整的三维人体草图数据集是一项劳动密集型任务。我们收集了一个人体草图的小数据集，用于指导合成数据集的生成。我们也使用它作为一个测试集来验证所提出的合成数据训练的网络能够很好地推广到人体草图。为了获得训练数据，我们采用了一种生成综合训练数据的通用策略。

**3.1 选择图形**

为了训练和测试基于草图的检索模型，我们使用ModelNet10中修复的干净流形网格1。我们使用ModelNet10中的所有10个类，总共有3958个形状。我们将数据集分为训练集、验证集和测试集，分别包含2847、317和792个形状，确保每个类的形状在三个集之间按比例分割。

**3.2 认为草图绘制**

**任务**。我们的目标是由新手创建的三维草图，可以将其视为来自QuickDraw数据集的快速二维草图的等价物。为了能够使用收集到的草图进行细粒度检索评估，我们构建了一个成对草图和三维模型的数据集。我们实验了一个场景，在这个场景中，一个人可以在虚拟现实中观察一个三维模型无限长的时间，然后被要求从记忆中绘制草图。我们观察到，在这种情况下，参与者有时会忽略对精确测试细粒度检索非常重要的特征，而是让参与者在参考三维模型上绘制草图。我们在补充中提供了一个定性评价的草图从记忆和检索性能等输入。为了模仿想象中的三维草图的细节层次，我们选择使用宽的带状线，但不对草图样式或细节层次构成任何限制。

**参与者。**我们从ModelNet10的测试集中选择了139把椅子和28个浴缸形状，并雇佣了10名没有艺术背景或3D素描经验的参与者。这些形状被随机分成10个亚组，每个亚组由13-14把椅子和2-3个浴缸组成。每个参与者被分配了其中一个子集。子集之间没有重复的形状。

**接口。**尽管有一些通用的虚拟现实绘画和设计软件可以让用户直接在3D中绘画（比如谷歌的Tilt brush2和Facebook的Quill3），但它们并不能完全满足我们的目的：我们希望记录基于笔划的详细信息；我们需要显示用于数据收集的参考模型的选项，以及（iii）我们希望为将来的任何附加功能（例如，基于草图的三维编辑）提供一个新的代码库。我们实现了基于Oculus Rift平台和Unity引擎的定制三维草图绘制环境。

**3.3 生成综合训练数据**

作为获得三维形状综合草图表示的第一步，我们使用Gori等人的方法提取了详细的曲线网络。这种方法的设计是为了产生一个曲线网络，它保留了良好的形状细节，这对于人类新手来说并不常见。我们观察到，新手不仅忽略了小细节，而且倾向于用单线表示薄的体积细节，而且往往忽略了特征线的子集。为了获得人体素描的外观，我们关注两个方面：细节层次和机械误差。我们首先执行细节过滤和行合并，然后执行行过滤。然后，我们将长曲线链分解为较短的笔划，并对其应用一组局部和全局变换。

**3.3.1 曲线网络提取。**

FlowRep方法要求输入是曲率对齐的四边形主导网格。为了提高处理效率，在使用Blender将三角网格转换为四元主导网格之前，我们先对三角网格进行抽取简化，但转换质量并不总是符合FlowRep的要求。因此，在这个阶段中，我们过滤了几乎20%的原始数据集，这些数据集没有被FlowRep处理。这可以在未来通过更先进的四边形网格算法得到缓解。

**3.3.2 过滤和合并。**

Gori等人的代码以输入网格的链边形式返回曲线网络。如果两条后继边之间的夹角小于135度，我们首先将这些链分成几个链。然后去除所有长度小于原始形状边界框最小高度和宽度10%的链。我们使用Ramer-Douglas-Peucker算法对所有链重新采样，精度参数设置为0.02dmin。最后，我们迭代地遍历所有链，并为每个链计算最近的、切向对齐的链。如果这两条链之间的距离小于这两条链最大长度的5%，则用它们的聚集曲线代替这两条链。这些步骤允许删除小细节并彼此靠近。然而，简化的曲线网络所包含的直线比新手在大多数三维草图中所能找到的要多得多。

**3.3.3 抽象化**

在本节中，我们将描述获得具有不同详细程度和机械不精确性的三维草图的方法，我们将其统称为草图抽象程度。我们将其限制在0和1之间。

细节层次为了减少线的数量，我们首先使用链之间的离散Frechet距离来计算相似矩阵。在计算Frechet距离时，我们首先通过平移其中一个笔划，将笔划在其端点处对齐。然后，我们使用凝聚层次聚类法根据相似矩阵进行分组，其中聚类数ncluster是其和草图中链数nchains的函数：ncluster=max（nchains（1− 0.8la）/2，10）。接下来，我们根据绝对位置为每个簇迭代地选择一对最远的两条线，并计算从簇中所有线的平均到这两条线的平均距离。我们删除所有的线在一个集群中的距离，其中任何两个选定的线小于平均。在得到链的约化集之后，我们把长链分解成几个短链，我们称之为笔划。我们在顶点分割每个链，其中曲率是原始链平均曲率的两倍。我们进一步过滤掉长度小于0.2smax的短笔划以避免微小细节，其中smax是输入网络的最大维数。

机械误差为了模拟人体草图的机械误差和透视误差，我们对每个草图应用一组全局和局部变形。我们首先根据给定的抽象级别应用全局平移、旋转和缩放。为了实现这一点，我们部署了一个辅助参数t，它是从[0,1.5]范围内随机采样的。然后从间隔中独立地随机取样每个轴的旋转角度[−10t，10t]度。每个轴的比例因子依赖于从[1]中随机取样− 0.1吨，1+0.1吨]。为了获得一个平移向量，我们从球面上随机采样，其半径值从[0，smax t]中随机采样，其中smax是输入网络的最大维数，如前所述。在全局笔划变形之后，我们对每个笔划顶点应用随机平移。从具有半径rvi的圆盘上随机取样平移向量，半径rvi位于与顶点vi笔划方向正交的平面上。半径rvi从范围[0，0.1lalstroke]中随机取样，其中lstroke∈ [0，1]是当前笔划的长度。我们将每个笔划的两端延伸p，p值从[0，0.1smax]中随机取样，以再现人类笔划的外观。

在全局和局部笔划变形后，采用三维样条插值，得到更平滑的笔划，与人体笔划的外观相匹配。

对于我们的实验，我们生成5个合成数据集，具有5个抽象级别，[0.0，0.25，0.0.75，1.0]，并通过混合所有抽象级别的草图或仅在中间具有3个抽象水平的草图来获得两个混合数据集。图3 c。显示了使用la参数的不同设置获得的示例草图。

**4.评价**

**4.1 采样和渲染策略**

我们对NGVNN使用了两种渲染样式，对PointNet++使用了两种采样策略。

多视图网络的三维草图绘制。我们产生12 224×224每个三维形状和三维草图的正交视图，每30度放置12个虚拟摄像机，如Lee等人提出的。摄像机被提升到离地平面30度的高度。对于三维形状和三维草图，我们使用两种类型的渲染样式进行实验：Phong着色和深度贴图。对于三维草图，我们将每条线表示为一个三维管。

点云采样。为了得到点云表示，我们首先从形状和三维草图中抽取10000个点。对于形状，我们使用蒙特卡罗采样4；对于草图，我们使用等距采样。然后从初始的10000个点开始采用两种采样方式，得到1024个点的稀疏集：随机采样或均匀采样。采用最远点采样法获得均匀采样。稀疏集是动态获得的，因此在不同的迭代中可能会有所不同。

**4.2 采样和渲染的效果**

我们比较了PointNet++的不同采样方法和NGVNN的渲染样式，当在其设置为0.5渲染的3D形状和3D草图对上使用分类损失和三元组损失进行训练时。PointNet++的统一采样和NGVNN的深度渲染性能最好，因此我们在其余的实验中使用这些设置。

当我们将暹罗体系结构用于NGVNN时，基于点的表示（虚线）远远优于多视图表示（实线）。因此，在剩下的实验中，我们使用了NGVNN的异构体系结构。

**5. 结论**

本文提出了将三维虚拟现实草图应用于三维模型检索的问题。我们首先介绍了一个专门构建的虚拟现实环境来收集虚拟现实草图并进行检索。然后，我们从ModelNet收集了两个形状类的子集的一组3D人体草图。我们进一步提出了一种生成具有不同抽象层次的合成三维草图的方法，并证明了在我们的合成数据上训练的方法可以很好地推广到人类草图。通过一系列的综合评价，我们发现与基于三维形状的检索相比，基于点的形状表示比基于多视图的检索更具优势。最后，我们提出了一个架构，该架构带有一个额外的草图正则化分支，可以在所有考虑的基线上获得更高的性能，说明了直接处理VR草图的抽象本质的好处。我们希望能对这个新问题提供一个有效的初步尝试。