

硕 士 研 究 生 读 书 报 告



题目对三维物体的六个自由度的两指手势操控

作者姓名 冯浩宇

作者学号 21551140

指导教师 李启雷

学科专业 移动互联网与游戏开发技术

所在学院 软件学院

提交日期 二○ 一五 年 十二 月

Two-Finger Gestures for 6DOF Manipulation of 3D Objects

Major Subject: Software Engineering

Advisor: Li Qilei

By

Feng Haoyu

Zhejiang University, P.R. China

2015/12

摘要

多点触控输入设备提供了对三维物体6DOF（六自由度）操纵的有效的解决方案。主要集中在大尺寸多点触摸屏幕，现有的解决方案通常需要至少三个手指和全6DOF操纵双手相互作用。然而单手两指的操纵方式是首选，特别是对多点触控的便携式设备（如流行的智能手机），这样可以使更少的手遮挡屏幕，减轻手的负担，另一只手可以用来完成必须的任务比如举着手机。对于两个手指控制6个自由度我们的关键想法是通过检查两个手指的移动特性来引入两个操作模式和两个对应的手势，而不是像以往的做法一样通过增加手指的数量或者判断各个手指间的关系，我们通过基于学习的方法解决了此想法中所产生的二元分类问题。我们的试点实验表明，我们的只有两个手指的通常只有单手互动接触的技术，可以跟现有的最先进的技术相媲美甚至比它们更好。

**关键字**：6个自由度 两指手势 三维物体的操纵

Abstract

*Multitouch input devices afford effective solutions for 6DOF (six Degrees of Freedom) manipulation of 3D objects. Mainly focusing on large-size multitouch screens, existing solutions typically require at least three fingers and bimanual interaction for full 6DOF manipulation. However, single-hand, two-finger operations are preferred especially for portable multitouch devices (e.g., popular smartphones) to cause less hand occlusion and relieve the other hand for necessary tasks like holding the devices. Our key idea for full 6DOF control using only two contact fingers is to introduce two manipulation modes and two corresponding gestures by examining the moving characteristics of the two fingers, instead of the number of fingers or the directness of individual fingers as done in previous works. We solve the resulting binary classification problem using a learning-based approach. Our pilot experiment shows that with only two contact fingers and typically unimanual interaction, our technique is comparable to or even better than the state-of-the-art techniques.*

***Key ：****6DOF、* Two-Finger Gestures、Manipulation of 3D Objects

1、引言

多点触控为显示屏提供了额外的输入宽带，从而为操纵三维物体提供了更有效的解决方法。这往往涉及到六个自由度即三个分别延x、y、z三个直角坐标轴的移动自由度和三个分别延x、y、z三个直角坐标轴的旋转自由度。要完全确定物体的位置，就必须清楚这六个自由度。然而由于触摸输入是由固有的二维并且每一个接触手指只能提供两个自由度，因此多点触控输入的组合的自由度和所需的三维操作任务的六个自由度之间没有直接的映射。着眼于大尺寸触摸板或者墙壁，所有现有的多点触摸交互技术，可以将全部的三维物体的六个自由度的映射于每个手指上，但是支持六个自由度通常要至少三个触摸手指。然而，三个或更多个指状物出现到小尺寸的触摸屏（例如手机）会太拥挤，可能会严重挡住视线。然而许多触摸传感器（例如，基于波引导红外线，表面声波等）只支持至多两个触摸点。触控与间接触控也是对屏幕上的物体大小是非常敏感的。本文介绍了一种完整的六个自由度操纵，它始终只需要两个手指操纵。 我们的技术的关键思想是要仔细对两个手指进行编码，区分出下面四种双指手势中的一种：平移，捏，旋转，按压平移（图1）。全部的6DOF操纵然后可通过下面的映射容易地实现：平移为XY轴的移动，捏用于z轴的移动，旋转用于物体绕z旋转，和按压平移为物体绕XY轴旋转。



通常二手指操作的平移，捏，旋转手势已经被广泛的用于2D操纵，这个种整体手势通常被成为旋转-缩放-平移（RST）手势。这种整体手势通常涉及两个移动手指，我们描述为2M模式，与此相反，按压移动涉及一个移动手指一个平移手指，对应于1m + 1 f模式。因此我们如何准确的区分出是2m模式还是1m+1f模式是主要的挑战所在。本文使用了监督学习方法解决这个二元分类问题，研究证明这个学习方法可以实现约95%的高识别率。我们进行了试点实验与现有的先进技术相对比来评估我们的技术，我们研究中发现我们用两个接触手指对手指的是否是直接接触没有要求，我们的技术在完成时间方面优于屏幕空间技术（Screen-Space）和DS3技术，跟至少要用三个手指的粘性工具技术可以相媲美。研究还表明，现有的所有技术都对两手的相互作用有强烈要求，而我们的技术只需要一只手就可以满足，这样更适用于各种不同的场景。

2、相关工作

传统的3D控制技术主要设计用于单点输入设备（如：pc中鼠标的输入），这种技术让用户可以在单次交互中可以分开的控制一个或两个自由度。这对约束到一个轴或者一个平面的操作是有益的，然而这种技术通常需要非常明确的模式转化，这种模式转换是对于复杂的物体转换操作来说是非常不方便的。

额外的交互宽带通过多点触控设备来实现，可以允许多个自由度并行的控制，因此提高的操作速度。然而如何优化六个自由度和多点触控提供的三维操作间的因为可分离输入结构的不匹配映射问题是主要的挑战。几个映射方法已经被提出来控制6DOF，总结如图2



这个被Reisman 等人提出的Screen-Space技术通过最小二乘优化的方式解决了平移和旋转的参数。优化是为了实现了直接操作能够让三维物体上的一个点始终再同一个指尖下，尽管这里有些自由度的操作（如xy轴的平移）可以仅仅由一个或则两个手指操作来实现，但是所有操作需要至少三个手指来实现。虽然直接操作确实自然直观，但是当在屏幕上的三维物体非常小非常薄或者是不规则的时候那直接触摸对象可能是非常棘手的。

接下来的由Hancock等人提出的Shallow-Depth技术，仅支持五个自由度的操作。这里就不展开介绍。

同样是Hancock等人提出的Sticky Tools技术实现了六个自由度的操作，x、y、z轴的平移和z轴的旋转用类似与我们常用的rst手势来实现，而第三个手指可以通过不直接接触物体的方式来实现剩下的两个自由度，但是是跟另外四个自由度的操作是分离的。我们通过这种分离的策略，并且这种分离策略是通过手指的操作是否有联系而不是手指个数来将自由度分离成两个子集。

Martinet等人提出的*DS3* (Depth-Separated Screen-Space) 技术将六个自由度分离，这个这个分离控制通过检查每个手指是否需要直接接触来实现。具体来说就是两个或者更多手指(Mode ≥ 2d)控制以一个整体的方式实现了三个自由度的旋转。完成了基于Screen-Space方法的一个变型。单个直接接触的手指控制xy轴的移动，第二个间接接触的手指控制z轴旋转（1d+1i模式）。试点实验表明，DS3比Screen-Space和Sticky Tools在特别的3D钉洞的任务更加有效，这个洞固定在一个是屏幕大小一半尺寸的长方体的中间。物体的大小会明显的影响到对物体上直接或者间接接触的性能（例如，当一个物体出现太小直接接触变得困难）。

3、设计原理

我们的设计原理受以下来驱动

**支持不同尺寸的多点触控屏幕；**理想的方法应该是能够适应不同尺寸的普通屏幕从几英寸如手机屏幕到几十英寸的如多点触控协作墙。

**两个手指操作**：原因是第一有些性价比较高的硬件只支持两点触控，第二更多手指会遮住更多屏幕第三，两个手指能够单手完成操作。

**支持单手操作：**在很多场景里优选单手操作

独立于手指的直接性：最好的技术应该能适应直接操作和间接操作，这样的技术在性能上可以很大程度上对屏幕尺寸不敏感。

**无缝衔接操作：**应该避免显示的模式切换，例如通过按钮切换应该避免，以保持相互作用的徒手性。

4、两指手势和操作

为了使我们的技术适用于大多数多点触摸设备，我们只假定每个手指与触摸屏接触使我们接触点产生唯一的2D位置信息。利用这样的信息，众所周知4DOF控制可以通过一个两指RST-风格手势来容易地实现（例如，参见在图1中Sticky Tools技术的2d模式）。我们的技术也采用了这种RST式的手势来控制4自由度。这里剩余的挑战是如何用相同的两个手指小心控制另外两个自由度，并且不取决于各个手指的接触的是否直接性，手指的联系性。

2m模式：这种模式涉及到两个移动的手指，利用RST式的手势控制四个自由度。

1m+1f模式：这种模式涉及到一个移动的手指和一个固定的手指，通过一个按压平移的手势来控制另外两个自由度。

下面我们简要地描述图一的具体实施，我们假设在触摸屏与xy轴平面平行，z轴平行于观察方向，并指向用户，我们还假设我们正在使用透视相机。向以前一样我们的技术集中于六个自由度和刚性变换。让(Tx, Ty, Tz, Rx, Ry, Rz)表示涉及的变换参数。其中Tx, Ty, Tz分别表示x、y、z轴的移动变换，Rx, Ry, Rz分别表示x、y、z轴的旋转变换。

在2m模式中，四个自由度通过RST手势对应于xy-轴的移动，z轴移动，z轴旋转，2m模式中两指必须是移动，但是的在以往的RST式手势中这不是必须的要求，由于这一要求，当持夹或者旋转物体时，通常手指会绕着两个接触点的中点旋转或者移动。在1m+1f模式中，一个手指被固定，一个手指移动，这种按压移动的的手势操作只有两个自由度的涉及，基本上靠控制移动的手指完成球型轨迹来实现。

平移、持捏、旋转的手势本质上是集成的，因此没有必要将xy轴的移动，z轴的平移和z轴的旋转这几个操作区分开依次操作就能完成复杂操作。对于另外xy轴的旋转这两个自由度，它们可以被分开控制，幸运的是，在视频显示，这两种模式之间的过渡是无缝的。甚至可以使用单个多触摸动作（例如单对的手指放下和抬起与触摸运动）来控制全部6个自由度。

我们的技术支持手动互动。由于拇指的对抗性性质，我们强烈建议用户使用拇指和其余四指从同一只手之一来执行两个手指操作（例如图一所示的拇指与食指）当进行按压移动时，可以使用无论是拇指或其他手指可以用来固定在某一位置。

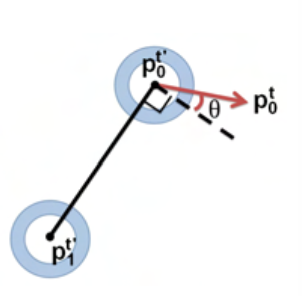
5手势识别

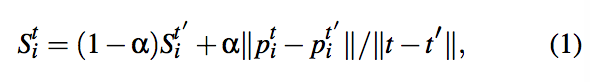
在本节中，我们提出了一个强大的算法，手势识别。由于平移、夹持、和旋转是集成，有必要单独识别这些手势，因此我们的关键问题是如果将RST式手势与按压旋转手势区分开，这实际上是一个二分分类问题。

简单的通过手指是否固定来分类并没有效果，原因有两个。第一，当一个手指在移动的时候，很难保证另一个是固定的。第二，即使是固定的手指在不精确的多点触控输入设备中也可能导致手指判定为移动的。

当用户与多触控屏幕交互时，设备会持续的产生触控事件流，我们的手势识别从手指与屏幕接触开始到手指从屏幕抬起结束。因此我们的探讨从有两个手指触摸开始，触摸点的顺序不重要，不失一般性，我们让*pti*表示接触手指的当前位置用t表示时间，用i表示接触点的编号。

5.1。线性分类

一个简单的分类解决方案是通过设置阈值，即显示的检查一个手指移动的速度是否高于某一个阈值。由于以每秒数百的速率产生触摸移动事件，仅仅在使用现在t和t’时间的先前位置和现在位置来计算手指的移动速度。因此我们用一个简单的指数加权移动平均值（EWMA）方法，该指数下降对速度的影响能使得近似值更接近真实速度，同时又能保持近似值的稳定，具体的讲，在时间t时触摸点的速度和触摸点的标号可以被估算为：



其中，α∈（0，1）是加权降低的程度（我们使用α=0.1）。

Let *Smin* = min(*S*0,*S*1) and *Smax* = max(*S*0,*S*1). Figure 3 (Left) plots pairs of min and max speed values in the *Smax*- *Smin* coordinate system, when several users were asked to perform specific gestures. As expected, the points for the RST-style gesture (especially for panning) are close to the line *Smax* = *Smin*, since the two fingers typically move at a similar speed.

让*Smin* = min(*S*0,*S*1)并且*Smax* = max(*S*0,*S*1)。图3（左）绘制了最大和最小速度在Smax- Smin坐标系中，当多个用户被要求执行某个特定的手势，正如所期待的，RST式手势的（尤其是对于平移手势）是接近SMAX= Smin的，因为两个手指通常会以相似的速度移动。对于将RST手势与按压移动手势的区分的一个解决方法是设置一个阈值，使得RST手势能被Stamina阈值捕获并且当Steaming Threshold & Stmax > Threshold按压移动手势就会被发现（例如一个手指固定一个手指移动）。当Stmax < Threshold任何手势都不能被识别出来，阈值的设定可以通过训练数据得知。然而就像是图3中左边所示，RST式手势与按压移动手势会高度重叠，造成在Smin-Smax空间中没有明显的分界线。经常被错误分类的手势，会产生糟糕的用户三维操控体验。

5.2。非线性分类

在上一小节讨论中我们对线性分离模式的初步试验，促使我们能够寻求更好的分离模式，以及更好的特征描述。

**特征描述。**虽然*Sti*提供了一个对分开的手指对移动的一个容易区别的特征，但是它未能捕捉到两个手指之间的运动的相关性。这样相关特性对手势识别是有用的，特别是当两个手指是来自同一只手，其中移动每个手指的肌肉块的动作是部分混合的，防止个别手指完全独立的运动。由于按压移动手势通常会有一个手指相对于另一个手指的旋转运动，通过实验我们发现，向心加速度的大小是另一个度量标准尤其是对于区分按压平移和旋转手势。具体地讲，我们当前的实现近似于心加速度的量值如下：



其中ωt1和ωt2是在时刻t的两个点的近似相对角速度，dt是两个接触点之间的距离。类似的计算如公式1的移动速度，相对角速度可以由EWMA的方法估计：



**分类。**通过检查我们的二维投影的三维特征，我们发现，虽然似乎有一个可行的二元分类的边界，但是我们不容易手动定义这样的边界，由于可能会有其他的分类方法可能会分类的更好，但是我们解决这样的问题通过采取基于支持向量机的基于学习的方法**。**由于其优良的经验性能SVM已广泛用于分类。通过使用非线性核函数，支持向量机的能够模拟非线性的依赖关系的特点是必要的，因为我们的三大特点就是非线性依赖。

通过训练SVM分类机，我们要求12个用户反复执行特定手势（即平移，捏，旋转，按压平移）因此得到了5000份标记训练样本。由于相比训练数据中实例的数量作为特征的数目比较小（只有三个），一个普遍方法是将数据映射到更高维空间。我们采用的是常用γ=0.5，ρ=-0.369的径向基核，其中非线性映射样本到高维空间中。我们用这种一般的模型达到了96.03％的很高的交叉验证准确度。图3绘出了分离边界的二维映射，这实质上是一个非线性超平面在三维空间中的二维投影。类似于线性分类模型，*Stmax* < *Sthreshold* 的点被标记为未分类。

验证。我们使用一个独立的测试数据集来验证我们的方法。测试数据集由数千个标记的特征矢量组成。这些数据集有些是从用户的训练数据集中收集的。我们SVM模型训练的分类精度因用户而变化，从79％至94％，平均为91.4％（SD=4％）。我们也尝试使用每个用户的用户量身定制的SVM模型的训练数据。这种个性化的SVM分类通常会有较高的识别率，达到平均为94.8％（SD=2.9％）。

相同的测试数据集还用于验证线性分离的模型（第5.1节），实现了平均89.7％（SD=4.1％）的分类率。需要注意的是由于触摸移动事件付出数百每秒的速率，单个触摸移动事件通常会导致小的转变。因此，偶尔会误判手势会引起轻微的差别。

虽然最近的作品上的多点触控交互都集中在直接操作3D对象，这是众所周知的直接控制和间接控制都有自己的长处和短处[MH08，SBG09。最近的一项研究Knodel和斧头[KH11]可见，间接互操作有利于效率和精度的RST对接任务。与DS3，我们的技术不依赖于手指的直接触发不同的操作模式。因此，我们的技术可应用于既直接操作和间接操纵。

7.初步研究

我们进行了试点实验，用我们训练的非线性分类器来评估我们的二指6自由度操作技术的有效性。

目标。在设计我们的实验中，当我们在目标如下：

要将我们的技术与显有的最先进的技术相比较，即Sticky Tools（ST），Screen-Space（SS）和DS3，特别是在完成时间方面的比较。

⋄为了评估对不同尺寸的屏幕上的屏幕和对象的不同技术的性能。

⋄要检查单手交互或双手交互的偏好。

参与者。十位大学生平均年龄23.6参加了实验。其中，8人为男性2人为女性。参与者有多点触控互动的经验，他们中间大部分人熟悉多点触控的应用，如移动设备上的图像和地图浏览工具。

设备。虽然我们有兴趣在不同尺寸的屏幕上实验，为了简单起见，我们通过使用一个单一的多点触摸的唯一设备改变不同的应用程序窗口大小来模拟，即一个12.1英寸惠普TouchSmart tx2笔记本模拟不同的屏幕尺寸。笔记本电脑被放置在桌面上（具有大约70厘米高度），与屏幕朝上。与会者坐在桌子前，使得两个手能轻易达到整个屏幕的正前方。与会者可以自由调整的笔记本电脑的位置和方向。

任务。参与者被要求尽可能快速准确地旋转和移动一个在虚拟3D世界中的对象目标的位置和方向。

如图4所示：



图4：左：简单的任务。右：复杂的任务。目标的位置和方向被示出在浅绿色。引入阴影效果是为了帮助确定的深度。

我们选择了一个CAD风格的物体。简单任务只需要围绕一个轴（x轴，y轴或z轴）旋转一个小于30度的旋转角并变换位置（例如，图4（左））复杂任务需要围绕多个轴旋转多次并移动位置（例如，图4（右））。物体尺寸大致分为三组：小尺寸（图4（左侧）），标准尺寸（图4（右）），和大的尺寸。

每个参与者都成功地完成了相同的一组试验中，在我们的情况下10次试验中，全部轮流采用这四种技术。由于智能手机和平板电脑的普及，我们关注以下两个窗口即屏幕尺寸，5英寸和11英寸。每次试验重复进行这两个尺寸。总之，我们的实验所涉及10×4×10×2= 800试验。

在开始实验一种新技术之前，学员们听取了技术和实践指导直到他们感觉很舒服（这段时间持续从几分钟到几分钟）。参与者被建议用单手的交互，除非他们觉得双手操纵灵活得多或高效。实验结束，从个别技术的可控性和可接受性的参加了质的反馈。在实验过程中，我们的系统记录下的信息进行定量分析：行驶距离单个试验的完成时间，操作的使用次数，接触时间，触摸按下事件数，和手指。

**跳过率。**跳跃率反映的易用性程度。我们发现，对于简单的实验（即涉及简单的旋转），我们的技术，ST和DS3没有参与者跳过，但SS的跳跃率为2％。对于复杂的任务，我们的技术，ST的SS和DS3的跳跃率分别为2％，0％，13％，和7.5％。 SS和DS3的较高跳跃率暗示这些两种技术都比较难以用于复杂转换。根据参与者的所报告，主要的原因是因为这两个SS和DS3需要直接操纵的对象，在小尺寸屏幕或小型对象精确的控制变得非常困难。

**技术。**重复测量的方差分析（ANOVA）发现，在完成一次四种技术之间的差异显著（F3,36=5.62，P <0.003）。成对比较表明，我们的技术（平均18.7s）为显著快于SS（平均38.0s; F1,18=9.21，P <= 0.01）和DS3（平均36.7s; F1,18=10.8，P <= 0.005） 。然而，有我们的技术和ST之间没有显著差别（平均21.2s; F1,18= 0.22，p值=0.65）。也就是说，我们的技术可媲美的ST，它需要至少三个接触指为种间动作。这里的结论是与为跳跃速率相一致。我们的发现与由马丁等人的不一致。 [MCG12]，其结论是DS3比ST快。这可能是因为DS3的敏感性筛选和对象大小不能暴露给他们的设置，中等大小的超大触摸屏上即固定目标。方差的重复测量分析(ANOVA)显示：在完成时间上四种技术差异显著（F3,36=5.62，P <0.003）。成对比较表明，我们的技术（平均18.7s）为显著快于SS（平均38.0s; F1,18=9.21，P <= 0.01）和DS3（平均36.7s; F1,18=10.8，P <= 0.005）。然而，有我们的技术和ST之间没有显著差别（平均21.2s; F1,18= 0.22，p值=0.65）。也就是说，我们的技术可媲美的ST，但是ST需要至少三个手指相互作用。这里的结论是与为跳跃速率相一致。

**复杂任务。**方差分析发现复杂的任务完成时间显著的效果（F1,78=28.5，P <9E - 07）。所有的技术都需要更多的时间才能完成复杂任务，如图5的（a）和（b）。特别是SS技术，这可能是因为该任务的感知结构和多点触控结构的不匹配造成的。



**屏幕尺寸。**正如预期的那样，方差分析显示在完成时间时发现屏幕尺寸有显著影响（F1,78=12.98，P <0.001）。5英寸的屏幕所有的方法都需要花更多的时间来完成试验（图5），但我们的方法提供了最好的性能。在其他方法中，DS3是最敏感的屏幕尺寸，可能是因为需要直接触摸的对象，很难在小屏幕上直接选中物体。

**对象的大小。**对于11英寸的屏幕，我们发现物体大小没有显著影响（图5（c））。这是可能是因为最小的对象在11英寸的屏幕仍然足够大。然而，对于SS和DS3技术在5英寸的屏幕上可以看出显著差异，（图5（D））（SS：F2,27=2.82，P =0.07; DS3：F2,27=2.85，P =0.08）。这是意料之中的，因为在5英寸的屏幕与尺寸更小的物体直接接触比较困难。

从与会者的反馈证实了我们的结论。当他们被要求投票最好的技术，他们六人来到了我们的技术，三为ST和一个用于SS。没有人投票给DS3，主要是因为它的小尺寸屏幕上的糟糕表现。相比ST特别是在5英寸的屏幕工作时当我们的技术是首选，因为用三根手指在小屏幕上进行交互会造成非常严重的手遮挡。

**8.结论和今后的工作**

这项工作首次引入了单手，两指多点触摸技术六自由度操纵3D对象。我们的主要贡献是引进了对于不同的操作模式手指不再需要相互联系，而不是向以往技术一样使用手指数或者每个的手指的是否为直接接触来区分。我们的技术用较少的手指和更少的手臂，我们的技术优于Screen-Space和DS3的方法和并足以媲美Sticky Tools的方法。在所有的技术，我们是对屏幕和物体的大小最不敏感的的方法。

如在一个小窗口我们的技术很难确定手指是处于固定或者在极为缓慢的移动，对于模型转换的微调控制我们的技术可能不是很适合。这个问题可以通过使用捕捉方法来减轻（参见[ATF12]和其中的参考文献）。我们目前的技术只在一个触摸设备上评估。我们感兴趣的是评估我们的技术在更广泛范围的大小屏幕的性能。我们也对如何在不同的响应率触摸传感器具上如何稳健地操作有兴趣。