

基于 Kinect 的民族弓弦乐器虚拟演奏系统*

张世程¹, 陈根方¹, 张明敏², 赵培涛², 潘志庚¹

¹(杭州师范大学 数字媒体与人机交互研究中心, 浙江 杭州 311121)

²(CAD&CG 国家重点实验室(浙江大学), 浙江 杭州 310058)

通讯作者: 张明敏, E-mail: zmm@cad.zju.edu.cn, http://www.zju.edu.cn

摘要: 利用 Kinect 相机结合增强现实技术和手势识别方法设计并实现了一个弓弦乐器虚拟演奏系统——以二胡为例, 将 Kinect 获取的现实场景和虚拟乐器融合在一起绘制成增强现实场景。通过 Kinect 得到的深度数据和贝叶斯肤色模型将用户的左手分割出来, 并再次绘制在增强图像上形成新的图像, 从而解决虚拟演奏场景中的虚实遮挡问题。利用基于反向动力学和马尔可夫模型的三维虚拟手势拟合方法, 对演奏过程中的左手手势进行识别, 并结合右手的运动状态完成乐器的虚拟演奏。

关键词: 虚拟演奏; Kinect; 增强现实; 手势识别

中文引用格式: 张世程, 陈根方, 张明敏, 赵培涛, 潘志庚. 基于 Kinect 的民族弓弦乐器虚拟演奏系统. 软件学报, 2015, 26(Suppl. (2)): 198-207. <http://www.jos.org.cn/1000-9825/15030.htm>

英文引用格式: Zhang SC, Chen GF, Zhang MM, Zhao PT, Pan ZG. Kinect-Based virtual performance system of national bowed string instruments. Ruan Jian Xue Bao/Journal of Software, 2015, 26(Suppl. (2)): 198-207 (in Chinese). <http://www.jos.org.cn/1000-9825/15030.htm>

Kinect-Based Virtual Performance System of National Bowed String Instruments

ZHANG Shi-Cheng¹, CHEN Gen-Fang¹, ZHANG Ming-Min², ZHAO Pei-Tao², PAN Zhi-Geng¹

¹(Digital Media & Interaction Research Center, Hangzhou Normal University, Hangzhou 311121, China)

²(State Key Laboratory of CAD&CG (Zhejiang University), Hangzhou 310058, China)

Abstract: Using Kinect camera combined with augmented reality and hand gesture recognition method, this paper designs and implements a virtual playing system for bowstring musical instruments (e.g. erhu). The fusion rendering of real scene captured by Kinect and virtual instrument forms the augmented reality scene. Through the depth data and a Bayesian skin model, the system segments the user's left hand region, then draws the hand image on the enhanced image, thus solving the problem of occlusion between virtual and real worlds in augmented reality system. Based on inverse kinematic methods and Markov model the presented design also provides a 3D virtual gesture fitting method to recognize left hand gestures, then completes the virtual instruments playing combined with right hand movement status.

Key words: virtual performance; Kinect; augmented reality; gesture recognition

在传统的音乐演奏中, 人们通过演奏各种真实的乐器发出美妙的音乐。然而, 由于购买困难、演奏技巧等原因导致很多人无法直接获得乐器去演奏。本文将增强现实技术和基于计算机视觉的手势识别技术与乐器演奏相结合, 用户无需手持真实的乐器, 仅需要步入场景之中做出乐器演奏的姿态动作, 就可以演奏出相应的音色。

增强现实技术不仅可以将虚拟物体和真实环境融为一体, 展现在人们的视野中, 还可以利用特定的设备使用户与虚拟物体进行互动。在虚拟乐器演奏中, 则需要用户能够和场景中的虚拟乐器进行互动, 使用户可以模拟演奏乐器。在这一过程中, 需要做的工作就是对演奏手势的识别。在以往的虚拟乐器演奏研究中, 采用的数据输

* 基金项目: 国家社会科学基金(12AZD120); 国家自然科学基金(61332017, 61173124); 浙江省自然科学基金(LY14F020037); 国家科技支撑计划(2015BAK04B00)

收稿时间: 2014-06-20; 定稿时间: 2014-08-20

入设备各有不同,所能达到的演奏效果也由于缺乏对手势识别的细节处理或者虚拟演奏场景的建模过于复杂或简单而各有差别.

本文讨论了虚拟乐器演奏场景的重建过程,解决在虚拟演奏场景中虚实遮挡的问题,使三维乐器模型融入到真实环境中之后具有真实感.对于乐器演奏使用基于 3D 模型的手势建模方法定义了 5 种手势,通过模型的参数,如关节角度和手掌位置等来推算手势,识别的精确度高、效果好,而且适合于解决乐器演奏手势识别过程中手指指尖丢失的情况.以往虚拟乐器演奏的研究大部分选择吉他等世界性的乐器作为虚拟演奏对象,而在国内关于民族传统乐器的虚拟演奏研究暂时还没有,因此我们选择了民族乐器中颇为广泛的弓弦乐器作为演奏对象,并以弓弦乐器中常见的二胡为例实现本系统.将本系统应用到民族传统乐器的虚拟演奏中,增加了乐器演奏的趣味性,不仅能够吸引更多的人来学习民族乐器的演奏,还可以为民族非物质文化遗产的保护和传承做出贡献.

1 相关研究

目前国内外在虚拟乐器演奏方面已有一些学术研究和商业应用,研究方法和技术各有不同.

文献[1]实现了一种以数据手套为交互工具的 PC 机虚拟现实系统,用来模拟对钢琴、小号、打击乐器等乐器的演奏.该文对虚拟乐器的演奏研究是一个创新,但却将重点放在了虚拟现实场景的构建上,对于手势的识别未进行详细的处理.文献[2]提出,将计算机视觉技术和数据手套相结合,实现虚拟场景里虚拟手对目标图像的动态跟踪,并根据数据手套的手势移动和指令动作实现对虚拟钢琴的演奏.以上基于数据手套的方法虽然具有识别率高、速度快、实时性强的优点,但是需要用户穿戴复杂的手套设备,不符合自然人机交互的要求,并且数据手套的价格昂贵,目前仅适用于实验室研究和一些特殊场合,不适合大量推广.文献[3]利用颜色手套和红外线照相机作为输入设备,通过确认手套之间的距离并综合弹奏者的手部动作形成一个音调,并以此制造了一具多功能虚拟吉他.文献[4]在之前虚拟空气吉他演奏工作的基础上,增加了 3 种手势的识别.文献[5]使用颜色手套和摄像头通过图像处理算法识别用户的动作意图,提出了一个虚拟吉他演奏的开源框架.使用颜色手套的方法,与数据手套相比设备价格便宜,但同样需要用户穿戴设备,使用起来不够方便.

文献[6]将基于计算机视觉的手势识别技术应用于虚拟场景的编钟演奏系统,使用单目摄像头获取操作者输入的手势信息,将肤色分割、手势模型特征提取的结果加以归类,以驱动模拟三维环境下的编钟模型动作,并发出悦耳的乐声.文献[7]使用单目摄像头获取操作者的手势信息,模拟三维环境下音乐演奏的动作,并发出基本的音阶音色.文献[8]基于 Kinect 传感器实现了 3 种乐器(鼓、吉他、蜘蛛王)的虚拟演奏.以上模拟虚拟乐器的演奏方式主要侧重在手势坐标的识别定位上,没有对手势手指等细节信息进行处理.文献[9]使用高清摄像头捕捉用户的手势模拟虚拟吉他演奏.但他们简化了吉他演奏的手势,将和弦手法映射为 12 种不同的手势,而且为了检测手势,操作者必须把左手放在一个特定的检测区域.虽然增加了对手势细节的处理,但是这样的演奏方式脱离了真实吉他的演奏方法,因而使得虚拟演奏操作显得不真实.

2008 年,日本任天堂公司正式发布了《Wii 音乐》模拟音乐演奏游戏,利用 Wii 遥控器与双截棍控制器,玩家可以《Wii 音乐》提供的超过 60 种乐器中选择所爱的,无论是打爵士鼓、吹萨克斯,还是弹吉他,都可以尽情发挥.这是虚拟乐器演奏在商业应用中的代表,它们的界面使用纯建模的方式,通过遥控器控制动画人物的运动.这种方式界面美观,但是建模工作量巨大,使用动画人物的方式来模拟乐器演奏者缺乏真实感.而与之相反,文献[7-9]中缺乏对乐器建模的部分,仅使用线条来模拟不同音阶或者琴弦的位置,同样失去了虚拟乐器演奏的真实效果.文献[10]使用 Kinect 来感应三维手势空间和自上而下的投影仪作为视觉反馈,配合 MIDI 钢琴键盘实现增强演奏.该系统是虚拟演奏与键盘硬件相结合的产物,手势控制用来增强演奏效果,而不是完全使用手势识别进行虚拟演奏,因此没有虚拟乐器的部分.

基于手势驱动的虚拟乐器演奏研究中,越来越侧重于使用计算机视觉技术进行手势识别的处理.早些年利用颜色手套的方式可以方便地检测出双手的位置,然后进行手势识别与跟踪.近几年,由于微软发布了新型的体感交互深度相机 Kinect,促进了虚拟乐器演奏研究的进一步发展.对于使用颜色手套还是 Kinect,文献[11]利用

一个基于用户手部位置跟踪的音乐游戏对两种设备进行了用户研究.虽然最后认为颜色手套具有更大的优势,但是该文中只应用到 Kinect 提供的骨骼数据,并没有考虑彩色图像和深度图像的数据处理;而且对于手势细节的识别并没有进行研究比较,具有一定的片面性.

本文采用深度相机 Kinect 作为输入设备,充分利用 Kinect 提供的彩色图像、深度图像和骨骼数据信息构建增强现实场景,模拟民族弓弦乐器的虚拟演奏——以二胡为例.本系统的研究内容主要分为两大部分.一是虚拟演奏场景的构建,利用 Kinect 相机捕获的彩色图像作为虚拟演奏的真实背景,在经过人体的演奏姿态识别之后,将乐器的三维模型通过 OpenGL 技术绘制在场景中人体的手部.在场景重建的过程中,需要对人体的手部和乐器模型的遮挡问题进行处理,利用深度信息和肤色模型对手部进行分割,然后将分割出来的手部图像绘制在正确的位置,同时对乐器模型绘制的坐标进行调整,使最终结果看起来是用户手持着乐器.二是虚拟演奏手势的识别,使用基于 3D 模型的手势建模方法定义了 5 种手势,利用基于反向动力学的手势拟合算法对二胡演奏中的 5 种基本手势进行识别.本系统采用了真实场景配合虚拟乐器部件的方法,建模工作量小且演奏效果更加真实.此外,对手势的手指细节识别进行了处理,能够实现对弓弦乐器演奏过程中基本音阶的模拟,表现力强.

2 虚拟演奏系统

2.1 概述

目前,国内外主要选择吉他和钢琴等世界性的乐器作为虚拟乐器演奏研究的对象.和他们不同,我们选择了中华民族传统的弓弦乐器.弓弦类乐器在中国大量存在,其中主要有汉族的二胡、京胡,蒙古的马头琴,新疆的萨它尔、艾捷克等等.本文以二胡为例构建虚拟乐器的演奏系统.在二胡的演奏过程中,需要左手握住琴杆,右手持弓,双手配合.左手不同的手指按下琴弦选择不同的音阶,右手选择内外弦拉弓配合演奏.演奏场景的重建则需要摄像机捕获的彩色图像叠加虚拟乐器模型来完成.系统的处理路线架构如图 1 所示.

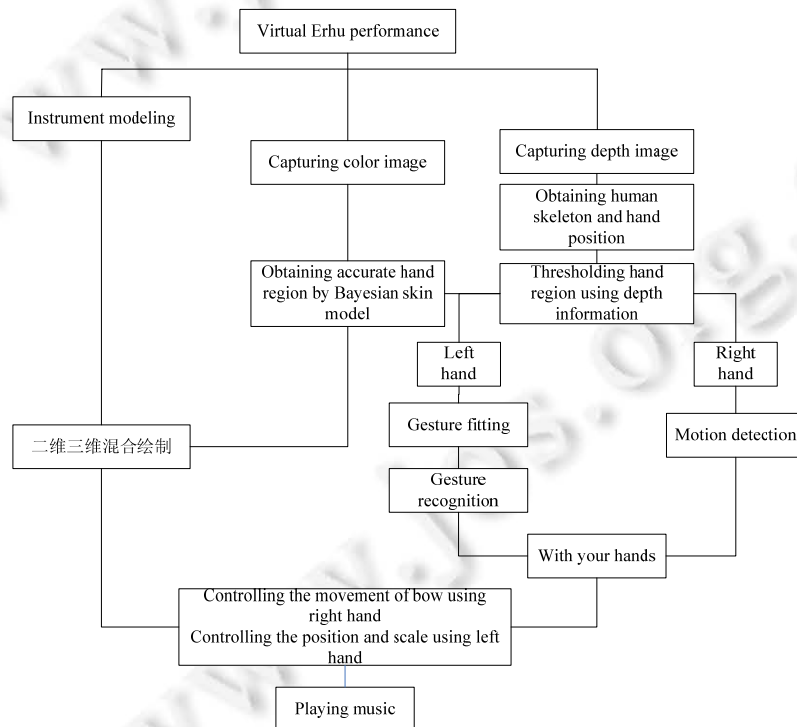


Fig.1 The system architecture diagram

图 1 系统路线架构图

2.2 虚实遮挡处理

在虚拟乐器演奏的研究中,部分研究者忽略了对虚拟乐器的建模,例如仅使用几条线段来模拟琴弦而没有琴体,或者仅有乐器的轮廓而缺少纹理,所以系统的运行效果并不真实,使用户不明确自己在演奏哪种乐器.我们吸取了前人的教训,本系统使用 3ds MAX 三维建模软件对二胡乐器进行建模,添加纹理和材质,导出标准 obj 三维模型格式的文本文件.然后在系统执行过程中读取模型文件,利用 OpenGL 技术重绘乐器模型,同时对模型的纹理和材质信息以及场景的灯光等条件进行设置,使乐器模型更加真实.

在增强现实应用中需要将计算机生成的虚拟物体、场景或系统提示信息叠加到真实场景中,从而实现对现实的增强.目前很多系统只是将虚拟物体简单地叠加在真实场景图像上,没有对虚实遮挡关系进行处理^[12].错误的遮挡关系容易导致观察者在感官方向上的迷失和空间位置上的错乱,长时间观察这种错误场景,有可能导致观察者的眼疲劳^[13].在弓弦乐器虚拟演奏中要求把虚拟的乐器绘制在摄像头捕获的真实影像中用户双手的位置,使看上去是用户真实地手持乐器演奏的效果.如果将虚拟乐器直接绘制在摄像机所拍摄的真实场景图像上,会导致乐器将用户的手遮挡住而缺乏用户手持乐器进行演奏的真实感.因此,为了得到逼真的虚实融合效果,必须解决场景中的遮挡问题.

虚实遮挡处理分为 3 个步骤:虚实遮挡关系检测、虚实遮挡边缘提取和虚实融合图像生成^[14].与其他 AR 系统相比,在二胡的演奏过程中,左手持琴杆,所以需要把乐器绘制在人的手部和身体之间,遮挡关系简单、明确;手部是唯一的前景物体,可以将手部图像从整幅场景图像中分割出来,应用图像平滑技术再绘制在增强的图像上,最终生成具有正确遮挡关系的虚实融合图像.该方法可以快速、鲁棒地解决虚实遮挡问题,效率比较高.下面将详细介绍本文的处理过程.

(1) 根据深度图像得到粗糙的手区域

通过 Kinect 可以获得 640×480 分辨率的深度图像,如图 2(a)所示,然后利用 NITE 接口函数跟踪左手的位置,即手的中心点,并映射到深度图中.在确定左手的位置以后,根据手的中心点截取感兴趣的区域 ROI,通过实验测试,我们定义 ROI 的分辨率为 140×140,如图 2(b)所示.然后,使用公式(1)根据左手中心点的 Z 坐标值对深度图进行分割得到手区域.

$$handmask = \begin{cases} 255, & z - \alpha < z < z + \alpha \\ 0, & \text{其他} \end{cases} \quad (1)$$

其中, z 表示深度图像中一点的深度值, α 为在图像分割时使用的深度阈值.深度阈值的设定,对手势识别的影响很大. α 是一个实验值,能够保证在 $z \pm \alpha$ 的范围内获取手部区域,目前在 Kinect 深度坐标系下根据实验将 α 设置为 5.

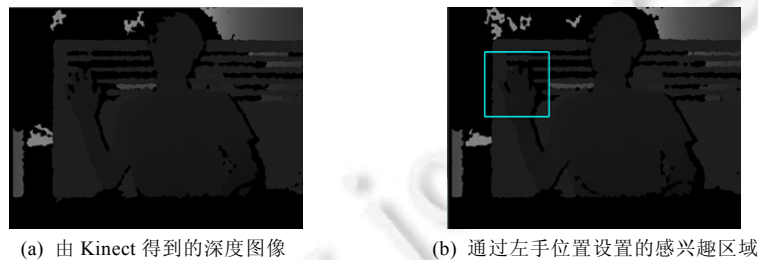


Fig.2 Depth image captured by Kinect and hand segmentation based on depth information

图 2 由 Kinect 得到的深度图像和基于深度的手区域分割

(2) 根据贝叶斯肤色模型获得优化的手区域

由于 Kinect 捕获的深度数据的特点,会导致得到的手区域边缘锯齿比较明显,如图 3(b)、图 3(c)所示,这样,在后面进行混合绘制时就会导致生成的场景不真实.因此在利用深度图像得到手区域之后,对它应用贝叶斯肤色模型,最后得到准确的手区域,如图 3(d)所示.

贝叶斯肤色模型是利用贝叶斯概率公式并结合直方图来计算给定的任意颜色 $color$ 属于肤色的概率

$P(\text{skin}|\text{color})$,如公式(2)所示.

$$P(\text{skin}|\text{color}) = \frac{P(\text{color}|\text{skin})P(\text{skin})}{P(\text{color}|\text{skin})P(\text{skin}) + P(\text{color}|\neg\text{skin})P(\neg\text{skin})} \quad (2)$$

首先需要建立肤色模型,采集大量包含人手的训练图像,并对每一幅图像制作一幅二进制掩码图像,其中肤色区域用 1 表示,非肤色区域用 0 表示.因为 HSV 颜色空间受光亮度的影响比其他颜色空间要小,所以我们采用 HSV 空间作为肤色分割的颜色空间.将训练图像从 RGB 颜色空间转换到 HSV 颜色空间,利用这些图像和掩码图像训练得到肤色直方图 H_{skin} 和非肤色直方图 $H_{\neg\text{skin}}$,从而得到公式(2)中各变量的值,如公式(3)、公式(4)所示.

$$P(\text{color}|\text{skin}) = \frac{V_{\text{skin}}}{T_{\text{skin}}}, P(\text{color}|\neg\text{skin}) = \frac{V_{\neg\text{skin}}}{T_{\neg\text{skin}}} \quad (3)$$

$$P(\text{skin}) = \frac{T_{\text{skin}}}{T_{\text{skin}} + T_{\neg\text{skin}}}, P(\neg\text{skin}) = \frac{T_{\neg\text{skin}}}{T_{\text{skin}} + T_{\neg\text{skin}}} \quad (4)$$

其中, V_{skin} 和 $V_{\neg\text{skin}}$ 分别表示 color 在肤色直方图和非肤色直方图中的高度, T_{skin} 和 $T_{\neg\text{skin}}$ 分别表示训练样本中颜色直方图和非颜色直方图的总高度.

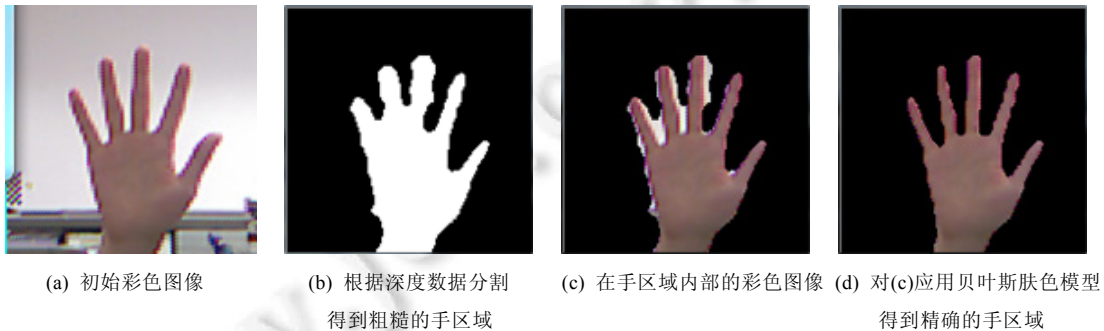


Fig.3 Hand segmentation based on depth and color data

图 3 基于深度和彩色数据的手区域分割

(3) 将得到的手区域绘制在增强的图像上

通过深度信息与贝叶斯肤色模型相结合的方式分割手区域,不仅减少了光照的影响,而且将贝叶斯模型应用在粗糙的手区域范围内,所需要操作的图像比 Kinect 直接获得的彩色图像小的多,所以处理速度也提高了很多.在得到精确手区域的彩色图像之后,利用保存的当前帧中手的坐标,将手区域内的所有像素重新绘制到合成图像上,获得虚实遮挡关系正确的新的合成图像.

虚实遮挡处理前后对比如图 4 所示.在图 4(a)中,虚拟物体二胡出现在真实的手之前,属于错误的虚实遮挡关系.经过以上方法处理之后,可获得正确的遮挡关系,如图 4(b)所示.

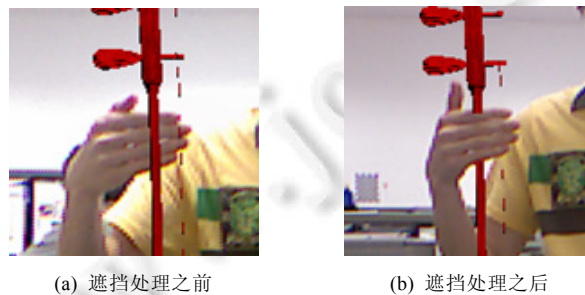


Fig.4 Comparison of before and after the occlusion handling

图 4 虚实遮挡处理前后对比

2.3 手势识别

2.3.1 手势定义

二胡的演奏技巧繁多且复杂,包括滑音、颤音等指法和颤弓、抛弓等弓法,这些高级的演奏方式很难用计算机来模拟,因此本系统只对二胡演奏中的基本方式进行模拟.二胡的传统把位共有 5 种,即上把、中把、下把、次下把和最下把^[15].左手的食指、中指、无名指和小指按弦控制音阶.例如 1=D(1-5 弦)调在上把位时的指法情况如图 5(a)所示.其中,“~”代表四指不按弦,“一”、“二”、“三”、“X”依次代表四指分别按弦的位置.基本手势如图 5(b)所示.

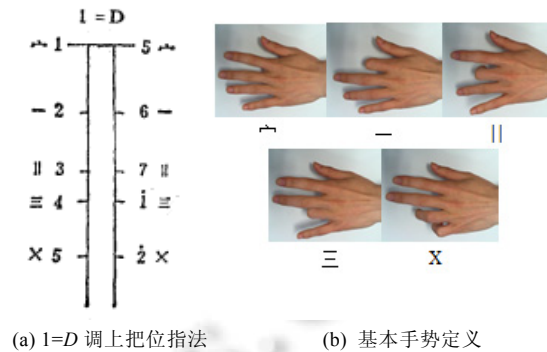


Fig.5 Fingering and basic gestures

图 5 上把位指法和基本手势图

2.3.2 手势识别

根据二胡演奏的手势特征决定,需要对左手的手指状态进行识别,因此需要对手指指尖进行检测.在演奏过程中,由于是用户的手背朝向 Kinect 摄像头,如果使用传统的方式进行指尖检测,则当某一手指按下琴弦的时候就会出现找不到指尖的情况,导致手势识别失败.所以我们采用拟合的方法解决这一问题,下面简单介绍处理过程.

(1) 利用长方体和胶囊体两种基本几何体来制作三维虚拟手模型,其中手掌用长方体来模拟,手指使用 3 根首尾相连的胶囊体组成;并为手指定义两种旋转操作来模拟手指的弯曲和运动.

(2) 根据能量最小思想,定义手指关节之间的弯曲约束条件,并使用基于约束的反向动力学算法实现虚拟手的手势还原.

(3) 定义手掌覆盖率和手指覆盖率,通过扫描线法计算手势图和虚拟手投影图之间的整体覆盖率来决定拟合程度.

(4) 在初始化状态下,手掌平铺背对着摄像机,取消手指的反向动力学解算,将虚拟手和分割出来的手势图进行基于面积的拟合,使拟合程度最高,同时获得虚拟手大小的比例系数.

(5) 随着手指的弯曲运动,手势图发生相应的变化,采用分层采样法,结合最速下降法和牛顿下山法的思想平移虚拟手,寻找最优拟合的掌心点位置.

(6) 根据反向动力学方法解算出手指各个关节的弯曲程度,以确定拟合程度最高的位置,从而将虚拟手还原为当前的估计手势形态.

(7) 将虚拟手投影到视平面上,根据和手势图的拟合程度不断调整掌心点和指尖点的位置,直至拟合程度达到最优为止.

通过以上拟合的思想去探索手势和指尖的位置,最终可以比较准确地将 3D 虚拟手还原为当前的手势.手势确定之后则通过虚拟手模型的参数(如关节角度)来推算手势.

从上一小节定义的 5 种二胡演奏的手势比较来看,手指在弯曲程度上具有显著的差别.在“~”状态下,食指、中指、无名指和小指的弯曲程度相似,而在“一”、“二”、“三”、“X”状态下,其中一指和另外三指相比具有最大

的弯曲度,而手指的弯曲程度可以通过虚拟手模型中手指各个关节的参数来获得.通过分析计算四指各自的弯曲程度并比较与其他手指的弯曲度差别,确定手势状态,从而达到手势识别的目的.

2.4 系统实现

在 Windows 7 操作系统下,使用 QT 作为图形界面开发框架,通过 OpenNI 2 和 NITE 2 开发接口获取深度相机 Kinect 采集的数据,利用 OpenGL 绘制三维模型,OpenCV 做图像处理,使用 Visual Studio 2010 开发工具和 C++ 编程语言实现了本系统.在 CPU 配置为 Core i5-3470QM,主频 3.2GHz,内存 4GB 的计算机上运行比较流畅.

在系统开始运行的初始化阶段需要做两项工作,一是上一节提到的手势拟合初始化,二是检测用户当前姿态是否为演奏二胡的姿态.演奏二胡时需要用户以坐姿手持乐器,通过 NITE 提供的 Kinect 编程接口可以获得人体的骨骼数据,包括 15 个关节点.坐姿识别可以根据人体各个关节点的相对位置、四肢躯干的角度计算,转化为几何三角的计算问题.选定人体的 3 个关节点:左肩、左臀部和左膝盖,分别定义为 S 、 H 、 K 这 3 点,同时可以获得它们的三维坐标值.通过三角几何知识可以计算出 SH 和 HK 两条边构成的夹角的角度值 α .如果 α 的值接近于 90° ,那么可以判定当前用户为坐姿.

系统初始化完成之后,虚拟乐器将会绘制在用户双手的位置,由于二胡演奏时从上到下分为 5 个把位,因此将初始化状态下双手的 Y 值坐标记下,之间的距离平分分为 5 段代表 5 个把位的位置.在某一固定把位状态下,对左手的手势进行识别,映射到正确的音阶,并配合右手的运动状态完成虚拟演奏.对于右手的运动状态判断,可以通过比较视频相邻两帧中右手在 X 轴的坐标差值来实现,定义一个阈值,当坐标差小于此阈值时判定为静止,大于此阈值时判定为运动.在专业二胡演奏人员演奏二胡时,使用录音设备录制各个把位及不同音阶状态下发出的声音,构成音色库,添加到系统资源中.双手的手势识别和运动状态在发生变化时需要立刻通知系统切换到相应的音色资源,这里我们借助了 QT 中信号和槽的通信机制.将左手的手势和位置以及右手的运动状态各自定义为一个信号,当状态发生变化时就触发此信号,与其相关联的槽函数将被立刻执行,完成音色资源索引调整和播放状态改变的操作.音色资源的播放使用了 QT 中的 phonon 多媒体框架.

系统运行初始化阶段如图 6(a)所示,初始化完成之后如图 6(b)所示,接下来就可以进行虚拟乐器的演奏,更多演奏状态如图 6(c)所示.

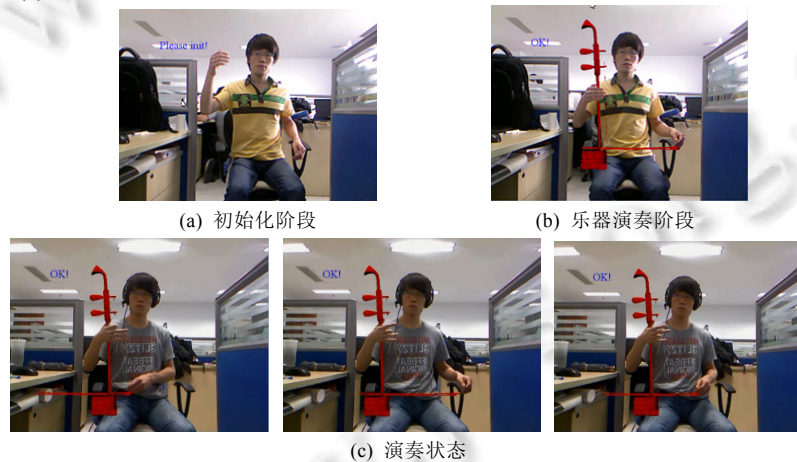


Fig.6 System operation interface

图 6 系统运行界面

2.5 实验分析

2.5.1 乐器绘制比例

在虚拟乐器演奏系统中,一个常见的问题是虚拟乐器模型大小的设定,即演奏者的双手和乐器的尺寸成一定比例,使其看起来比较自然.这需要根据演奏者的手的大小对乐器模型的绘制比例做出一定的调整.如果使用

普通摄像头,那么方法是将手分割出来,然后计算手区域的面积,再根据面积对乐器的绘制参数做出调整.使用 Kinect 设备,则可以根据手的深度距离来实现.

在 Kinect 坐标系中,感应器处于原点上,Z 坐标轴则与 Kinect 感应器的朝向一致.通过 NITE 获得用户的左手并记录当前左手的三维坐标中的 z 值,因此 z 值可以粗略地表示为左手距离 Kinect 的远近.根据 z 值调整乐器模型的绘制比例获得对应数据,见表 1.

Table 1 The relationship between left hand's z value and instrument's size

表 1 手的 z 值和乐器的比例关系

Z value of left hand	Instrument's size	Z value of left hand	Instrument's size	Z value of left hand	Instrument's size
700	0.5	950	0.42	1 200	0.3
750	0.5	1 000	0.4	1 250	0.27
800	0.5	1 050	0.38	1 300	0.25
850	0.47	1 100	0.35	1 350	0.25
900	0.45	1 150	0.33	1 400	0.25

通过分析数据,可以得到左手 z 值和乐器绘制比例之间的简单函数关系,如公式(5)所示.

$$scaleRate = \begin{cases} 0.5, & z \leq 800 \\ 0.4 - (z - 1000) \times 0.1 / 200, & 800 < z < 1300 \\ 0.25, & z \geq 1300 \end{cases} \quad (5)$$

由于绘制比例大于 0.5 时,乐器模型将超出窗口,同时手距离 Kinect 感应器过于接近,因此,当 z 小于 800 时将乐器绘制比例固定为 0.5;而当左手 z 值大于 1 300 时,手的大小和乐器模型之间产生较小的比例变化,而且此时 Kinect 捕获的手区域由于过小已经难以用来识别手势,因此,当 z 大于 1 300 时将乐器绘制比例固定为 0.25.

2.5.2 手势识别判断

本文通过拟合的方法进行手势识别,首先通过拟合将虚拟手还原为当前的手势,然后根据比较当前虚拟手指之间的弯曲度差别来判定为某一手势.

本文中三维虚拟手利用长方体和胶囊体两种基本几何体来制作.将手指分为基节骨、中节骨、末节骨 3 段,使用 3 根首尾相连的胶囊体来表示.每段手指关节都具有弯曲度的参数表示,因此可以获得每根手指的总弯曲度.在二胡演奏过程中,使用除大拇指之外的四指按弦来确定某一音调,因此,在手势判定部分只需要比较四指的弯曲度即可.通过不断地变换手势并记录各手指的弯曲度,计算得到平均弯曲度,见表 2.

Table 2 Average finger bending degree in each gesture state

表 2 各手势状态下手指平均弯曲度

Bending degree Fingers	Gestures				
	一	二	三	X	
Index finger	13	97	16	5	6
Middle finger	15	11	80	8	7
Ring finger	9	5	42	77	55
Little finger	19	2	3	10	101

通过分析比较各手势状态下的手指弯曲度,可以确定手势判断的分类参数.当在“一”、“二”、“三”、“X”手势状态,食指、中指、无名指和小指分别具有很高的弯曲程度,且数值都大于 70.在“一”手势状态,四指的平均弯曲程度并不是 0,这是因为,在手势拟合初始化阶段会根据当前的手势图面积确定固定的虚拟手大小,当在手势变化过程中,手和摄像头之间的距离可能会发生轻微的变化.当手距离摄像头相对之前略远时,即使在四指没有弯曲的状态下也会通过面积拟合的方式误判为手指发生了些许弯曲.在这种情况下,通过实验观察可以得到最大误差情况下四指的弯曲度都没有超过 50,因为在演奏过程中手和摄像头之间的距离变化不会太大.因此,可以设定一个阈值,当四指的弯曲度都低于该阈值时可判定为手指没有发生弯曲.

3 结 语

目前国内在虚拟乐器演奏方面的研究还较少,尤其是将虚拟演奏技术应用到民族传统乐器演奏方面.本文

设计并实现了一个手势驱动的二胡虚拟演奏系统.在该系统中使用深度相机 Kinect 来获取真实场景的彩色图像,结合乐器模型构建增强现实场景,并利用深度信息和贝叶斯肤色模型分割出用户的手部图像,解决了虚拟演奏场景中的虚实遮挡问题.通过手势拟合的方法进行二胡演奏过程中 5 种基本手势的识别,最后构成了虚拟演奏系统.这一系统将给音乐爱好者带来更多的便利和趣味性,经过一定的改进和完善之后,还可用于各类娱乐场所及展览馆等.另外,我们将把这一技术应用到新疆传统民族乐器萨塔尔的虚拟演奏中,丰富非物质文化遗产保护的思路,促进非物质文化遗产的保护和传承,具有重要的意义.

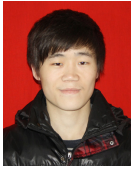
References:

- [1] Cheng XM, Hao Q, Wang YT, Yan DY. The virtual band on virtual reality technique. *Modern Scientific Instruments*, 2001,4: 38–40 (in Chinese with English abstract). [doi: 10.3969/j.jssn.1003-8892.2001.04.011]
- [2] Liang SL. A virtual performance based on computer vision gesture recognition [MS. Thesis]. Beijing: North China University of Technology, 2007 (in Chinese with English abstract).
- [3] Mäki-Patola T, Laitinen J, Kanerva A, Takala T. Experiments with virtual reality instruments. In: *Proc. of the 2005 Conf. on New Interfaces for Musical Expression*. Singapore: National University of Singapore, 2005. 11–16.
- [4] Karjalainen M, Mäki-Patola T, Kanerva A, Huovilainen A. Virtual air guitar. *Journal of the Audio Engineering Society*, 2006, 54(10):964–980.
- [5] Figueiredo LS, Teixeira JMXN, Cavalcanti AS, Teichrieb V, Kelner J. An open-source framework for air guitar games. In: *Proc. of the 2009 VIII Brazilian Symp. on Games and Digital Entertainment (SBGAMES)*. Rio de Janeiro: IEEE, 2009. 74–82. [doi: 10.1109/SBGAMES.2009.17]
- [6] Hu WJ. Research and realization of chime-bell playing system driven by gesture discrimination [MS. Thesis]. Hangzhou: Zhejiang University, 2007 (in Chinese with English abstract).
- [7] Wu Z. Research on gesture driving virtual performance system [MS. Thesis]. Kunming: Kunming University of Science and Technology, 2009 (in Chinese with English abstract).
- [8] Hsu MH, Kumara W, Shih TK, Cheng ZX. Spider King: Virtual musical instruments based on microsoft Kinect. In: *Proc. of the 2013 Int'l Joint Conf. on Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA)*. Tsuruga, 2013. 707–713. [doi: 10.1109/ICAwST.2013.6765529]
- [9] Yang G, Sung TP, Tsai WT. Air guitar on Altera DE2-70 FPGA architecture. In: *Proc. of the 1st Asia-Pacific Workshop on FPGA Applications*. Xiamen, 2012. 56–63.
- [10] Yang Q, Essl G. Augmented piano performance using a depth camera. *Ann. Arbor*, 2012. http://wwwweb.eecs.umich.edu/nime2012/Proceedings/papers/203_Final_Manuscript.pdf
- [11] Figueiredo LS, Teichrieb V. Kinect vs. Color gloves: A music game tool for hand tracking evaluation. In: *Proc. of the SBGames 2012*. Brasilia, 2012. 45–48.
- [12] Xu C, Li S, Wang JF, Peng T, Xie M. Occlusion handling in augmented reality system for human-assisted assembly task. *Intelligent Robotics and Applications*, 2008, 121–130. [doi: 10.1007/978-3-540-88518-4_14]
- [13] Zhu JJ, Pan ZG. Computer vision based occlusion handling algorithm for video based augmented reality. *Journal of Computers-Aided Design & Computer Graphics*, 2007,19(12):1624–1628 (in Chinese with English abstract).
- [14] Xu WP, Wang YT, Liu Y, Weng DD. Survey on occlusion handling in augmented reality. *Journal of Computers-Aided Design & Computer Graphics*, 2013,25(11):1635–1642 (in Chinese with English abstract).
- [15] Liu JX. *Erhu Playing Techniques and Exercises*. Ji'nan: Shandong Publishing House of Literature and Art, 1986 (in Chinese).

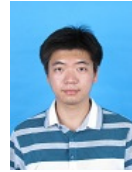
附中文参考文献:

- [1] 程雪岷,郝群,王涌天,闫达远.基于虚拟现实技术的虚拟乐队的实现. *现代科学仪器*,2001,4:38–40.
- [2] 梁善林.一种基于视觉手势识别的虚拟演奏[硕士学位论文].北京:北方工业大学,2007.
- [6] 胡文娟.手势驱动编钟演奏技术的研究与系统实现[硕士学位论文].杭州:浙江大学,2007.
- [7] 吴铮.手势驱动的二胡虚拟演奏系统的研究[硕士学位论文].昆明:昆明理工大学,2009.
- [13] 朱杰杰,潘志庚.用视觉计算实现视频增强现实遮挡处理. *计算机辅助设计与图形学学报*,2007,19(12):1624–1628.

- [14] 徐维鹏,王涌天,刘越,等.增强现实中的虚实遮挡处理综述.计算机辅助设计与图形学学报,2013,25(11):1635-1642.
[15] 刘建勋.二胡演奏技法与练习.济南:山东文艺出版社,1986.



张世程(1990-),男,山东临邑人,硕士生,主要研究领域为多媒体,虚拟现实,人机交互.



赵培涛(1987-),男,硕士,主要研究领域为图像处理,手势识别.



陈根方(1968-),男,博士,教授,CCF 会员,主要研究领域为图像处理,模式识别,数字音乐.



潘志庚(1965-),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为人机交互,多媒体技术,虚拟现实,数字娱乐.



张明敏(1968-),女,博士,副教授,CCF 会员,主要研究领域为虚拟现实,实时渲染,图像处理.

www.jos.org.cn

www.jos.org.cn