

标题: [像人脑一样思考 微软官博揭秘Kinect工作原理](#)

子标题: 人工智能快报

作者: mostai

日期: 01月12日

网址: <http://www.mostai.com/modules/article/view.article.php/c1/83>

关键词: kinect 微软 游戏

“你就是控制器。”(You are the controller。)如果你有在关注Kinect,相信已经听过这句给力的广告词了。从《Kinect Adventures!》中手脚并用堵漏水窟窿,到Zune播放界面中挥手换歌,Kinect开创了一种更加自然的娱乐交互方式。在这篇博客文章中,我将揭秘这款体感系统背后的秘密以及它如何让开发者创造Kinect体验。而Kinect团队的项目经理Arjun Dayal则将展示如何实现通过基于手势的方式来控制Xbox Dashboard和Kinect Hub。首先,让我们从指导Kinect研发的概念原理开始。

我们生活在一个模拟的世界

传统编程基于一系列的规则:原因和结果,非黑即白,非真即假。在为输入输出数目有限的简单系统建模时,这种方式工作得挺好。拿游戏《Halo》来说吧:按A键是让士官长跳,前拨左摇杆让他向前走,前拨右摇杆让他向上看。不是A,就是B。可惜的是,我们生活的真实世界并不是如此数字化,而是模拟的。

在模拟世界中,并不是只有简单的“是”和“否”,还有“也许是/否”;不仅有“对”和“错”,还有“对/错的可能性”。让我们想象一下挥手这一简单动作的所有可能性:身体运动的幅度,环境差异,衣服质地的不同,文化差异造成的动作差异等等。你可能需要研究10的23次方这么多的可能性,显然用传统编程方式来解决这类问题是不现实的。

我们从一开始就知道必须采用一种全新的,接近于人脑工作的方式来解决这一问题。当你遇到一个人的时候,你的大脑立即将注意力集中在他身上,并根据经验辨识出他的身份。这一过程并不是通过数百层的决策树来实现,人脑就是知道。婴儿很难区分出两个人的不同,但我们通过多年的学习和训练可以在几分之一秒内做到。事实上,你也许还能蛮准确地估摸出他们的年龄、性别、心情甚至个性。这也是让我们成就为人类的原因之一。

Kinect以类似的方法被创造出来。它观察身边的世界,它注意观察你的动作。即使Kinect从来没见过你挥过手,也能很快地从它学习过的TB级数据中猜测出你所做动作的含义。

Kinect传感器

Kinect骨架追踪处理流程的核心是一个无论周围环境的光照条件如何,都可以让Kinect感知世界的CMOS红外传感器。该传感器通过黑白光谱的方式来感知环境:纯黑代表无穷远,纯白代表无穷近。黑白间的灰色地带对应物体到传感器的物理距离。它收集视野范围内的每一点,并形成一幅代表周围环境的景深图像。传感器以每秒30帧的速度生成景深图像流,实时3D地再现周围环境。如果你玩过pin point impression 3D针模玩具可能更容易理解这一技术——将你的手(或者脸,如果你愿意的话)按压在这种玩具上,就可以产生你身体某一部位的简单3D模型。

像人脑一样思考 微软官博揭秘Kinect工作原理

寻找移动部位

Kinect需要做的下一件事是寻找图像中较可能是人体的移动物体，就像人眼下意识地聚焦在移动物体上那样。接下来，Kinect会对景深图像进行像素级评估，来辨别人体的不同部位。同时，这一过程必须以优化的预处理来缩短响应时间。

Kinect采用分割策略来将人体从背景环境中区分出来，即从噪音中提取出有用信号。Kinect可以主动追踪最多两个玩家的全身骨架，或者被动追踪最多四名玩家的形体和位置。在这一阶段，我们为每个被追踪的玩家在景深图像中创建了所谓的分割遮罩，这是一种将背景物体(比如椅子和宠物等)剔除后的景深图像。在后面的处理流程中仅仅传送分割遮罩的部分，以减轻体感计算量。

像人脑一样思考 微软官博揭秘Kinect工作原理

Kinect的大脑

真正的魔术在这里发生。分割化玩家图像的每一个像素都被传送进一个辨别人体部位的机器学习系统中。随后该系统将给出了某个特定像素属于哪个身体部位的可能性。比如，一个像素有80%的几率属于脚，60%的几率属于腿，40%的几率属于胸部。看起来这时候我们就可以把几率最大的可能性当作结果，但这么做未免太过武断了。我们的做法是将所有的这些可能性输入到接下来的处理流程中并且等到最后阶段再做判断。

看了上面的介绍，你也许要问我们如何教会Kinect辨识人体部位。开发这一人工智能(被称为Exemplar(模型)系统)可不是一件轻松的事情：数以TB计的数据被输入到集群系统中来教会Kinect以像素级技术来辨认手、脚以及它看到的其他身体部位。下图就是我们用来训练和测试 Exemplar的数据之一。

像人脑一样思考 微软官博揭秘Kinect工作原理

模型匹配：生成骨架系统

处理流程的最后一步是使用之前阶段输出的结果，根据追踪到的20个关节来生成一幅骨架系统。Kinect会评估Exemplar输出的每一个可能的像素来确定关节。通过这种方式Kinect能够基于充分的信息最准确地评估人体实际所处位置。另外我们在模型匹配阶段还做了一些附加输出滤镜来平滑输出以及处理闭塞关节等特殊事件。

骨架追踪系统的目标之一是为处理流程的各种输出提供一种菜单式的选择界面。游戏开发者可以选择任意的系统部件组合来开发各种游戏体验。比如，你可以仅仅使用分隔映射来制造一些惊人的华丽效果(《Your Shape: Fitness Evolved》是一个好例子)。

讲到这里，我们已经描绘出一个可用于控制游戏或娱乐的完全实时的体感系统。接下来，Arjun将介绍改进的Xbox Dashboard和Kinect Hub。他将向你展示这两个用户界面如何利用景深图像流和20关节骨架系统来创造一种基于自然手势的，访问游戏、电影、音乐和其他娱乐活动的全新方式。

像人脑一样思考 微软官博揭秘Kinect工作原理