·信号与信息处理·

基于3D体感技术的动态手势识别

涂 创

（北京航空航天大学,北京 100191）

摘 要:提出了一种基于3D体感机Kinect的图像处理手势识别算法,通过深度图像和骨骼图像的方法实现动态手势识别。首先在Kinect提供的骨骼图像中20个骨点中,选取2个离手部最近的骨骼点,通过追踪这两个骨骼点的位置来实现对手部的追踪,再通过判断手部的深度(即其相对于摄像头的距离)的变化来实现动态手势识别。

关键词:深度图像;骨骼图像;手部追踪;动态手势识别

中图分类号: TN94 **文献标识码**:A **文章编号**:1673-1255(2012)04-0055-04

Dynamic Gesture Recognition Based on 3D Kinect

GAN Chuang

(*Beijing University of aeronautics and astronautics, Beijing* 100191, *China*)

Abstract: A kind of gesture recognition algorithm of image processing based on 3D Kinect is proposed. The dynamic gesture recognition algorithm is performed by skeleton images and depth images. At first, two skeleton points which are nearest to hands are choosen from 20 skeleton points in a skeleton image. The process of tracking hands is performed by tracking the positions of the two skeleton points. Then the dynamic gesture recognition process is realized by the change of depths of hands (the distance between a hand and a camera).

Key words: depth image; skeleton image; hands tracking; dynamic gesture recognition

随着机器智能领域的迅猛发展,手作为人身体上最灵活的一个部位及人机交互的一个媒介,得到越来越多的应用。因此基于手势识别的各种应用也是层出不穷。手势是一种自然而直观的人际交流模式。手势识别也理所当然地成为了实现新一代人机交互不可缺少的一项关键技术。然而,由于手势本身具有的多样性(包括肤色、形态的差异性)、多义性(不同手势具有不同的意义)、以及时间和空间上的差异性(会受到光照等因素的影响)等特点,加之人手是复杂变形体及视觉本身的不适定性,因此基于视觉的手势识别是一个极富挑战性并具有很大应用空间的研究方向[1]。

1 手势识别技术的发展

手势识别分为两种,一种是静态的手势识别,即在摄像头下检测到某个手势时就给出命令。另一种是动态手势识别,即能够识别手做的一些动作。随着3D体感技术的出现,手势识别进入一个全新的领域。

1.1 静态手势识别

静态手势识别的常用方法主要有:基于模版匹配的,用边缘特征像素点作为识别特征,并利用Hausdorff 距离模板匹配完成静态手势识别[2];基于SVM 支持向量机,通过皮肤颜色模型进行手势分割,并用傅里叶描述子描述轮廓,采用针对小样本特别有效且范化误差有界的最小二乘支持向量机(LS-SVM)作为分类器进行手势识别[3]以及集合模版匹配和机器学习理论的手势识别方法[4]等。但由于静态手势识别技术应用的局限性较大,不够灵活,使用人数在减少。

收稿日期:2012-06-24

作者简介:涂创(1990-),男,辽宁锦州人,大学本科,研究方向为计算机图像处理.

1.2 动态手势识别技术

在静态技术基础上发展起来的是动态手势识别，即在视频流下能够对手部做出一些动作进行识别，这种识别的难度要比静态手势识别难度大很多，但却更具有实用性。动态手势识别的方法主要有：采用Camshift算法对手势进行分割，从而达到手势识别的功能[5]；通过双目视觉系统来建立数学模型，并结合图像分割技术进行手势判断[6]；基于机器学习进行手势识别，首先采用AdaBoost算法遍历图像，完成静态手势的识别工作，在动态手势的识别过程中，运用了光流法结合模板匹配的方法[7]等。

虽然手势识别方法取得了一些很好的效果，但这些现存方法都无法克服当光照条件变化较大或人体肤色差异性较大时会出现系统失灵的情况，这时往往需要重新调整各种参数来使得系统正常工作，从而大大降低了系统的稳定性。其不稳定原因主要在于根据人手的颜色进行图像分割的处理过程会受到光照、遮挡等各种因素的影响，进而对后续的手势识别产生干扰。因此提升手在摄像头下的识别精度成为了一个研究的重点。

1.3 基于Kinect体感技术的动态手势识别技术

Kinect是美国微软公司于2010年推出的XBOX360游戏机体感周边外设的正式名称，起初名为Natal，意味初生。它实际上是一种3D体感摄影机，利用即时动态捕捉、影像辨识、麦克风输入、语音辨识、社群互动等功能让玩家摆脱传统游戏手柄的束缚，通过自己的肢体控制游戏，从而实现与互联网玩家互动，分享图片和影音信息等交互功能[8]。

微软推出Kinect后，深度图像和骨骼图像技术使得手势识别进入一个全新的领域。由于Kinect在硬件上采用了CMOS红外感应设备，可以提供关于人的骨骼图和整个镜头下的深度图像，因此在对这两种类型的图像深入研究的基础上，提出了一种可以进行动态手势识别的方法，并在识别准确度上有了较大的改进。

1.3.1 深度图像的产生机理

Kinect采用了基于光编码(light coding)[9]理论的技术，可以直接获取物体与摄像头之间的距离。其基本思想是通过连续光(近红外线)对测量空间进行编码，再经过感应器得到编码的光线，在将数据传递给晶片进行运算解码后，产生一张具有深度的图像。其核心之一就是结构光技术，它与传统的技术有很大的差异性。它的光源打出去的并不是一幅周期性变化的二维的图像编码，而是一个具有三维纵深的“体编码”。这种光源叫做激光散斑(laser speckle)，是当激光照射到粗糙物体或穿透毛玻璃后形成的随机衍射斑点。这些散斑具有高度的随机性，而且会随着距离的不同变换图案，空间中任何两处的散斑都会是不同的图案，等于是将整个空间加上了标记，所以任何物体进入该空间以及移动时，都可确切记录物体的位置。

Kinect另一核心技术在于光源标定[10]，测量前对原始空间的散斑图案做记录，先做一次光源的标定，其采用的方法是每隔一段距离，取一个参考平面，然后把参考平面上的散斑图案记录下来；假设Kinect规定的用户活动范围是距离摄像头1~4 m，每隔10 cm取一个参考平面，标定后保存了30幅散斑图像；测量时拍摄一幅待测场景的散斑图案，将这幅图像和保存的30幅参考图像依次做互相关运算，得到30幅相关度图像；空间中有物体存在的位置，在相关度图像上就会显示出峰值。把这些峰值一层层叠在一起，经过插值运算，即可得到整个场景的三维形状[11]。

1.3.2 骨骼点追踪技术

Kinect骨架追踪处理流程的核心是一个不受周围环境的光照影响的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[12]。

骨骼点追踪采用了机器学习技术，通过建立了庞大的图像资料库，形成智慧辨识能力，尽可能理解使用者的肢体动作所代表的涵义。Kinect对深度图像进行像素级评估，来辨别人体的不同部位，其基本思想是先采用分割策略将人体从背景环境中区分出来，得到追踪对象背景物体剔除后的深度图，然后把深度图像传进一个可辨别人体部位的机器学习系统中，该系统将给出某个特定像素属于身体某个部位的可能，然后将这些数据输入到集群系统中，从而训练Kinect像素级辨认身体部位的能力。Kinect会评估Exemplar输出的每一个可能的像素来确定关节

点，然后根据追踪到的20个关节点来生成一幅人体骨骼图[13]。两幅经辨识的人体动态骨骼如图1所示。

2　基于深度图和骨骼图的动态手势识别技术

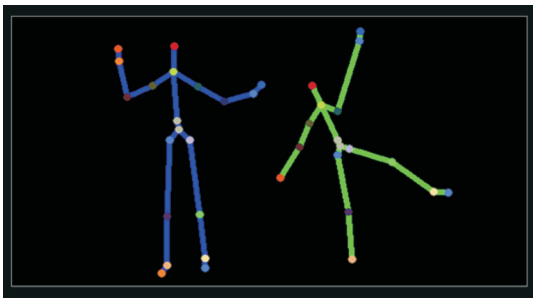

图1　人体动态骨骼图

动态手势识别技术主要分为两个步骤，第一个步骤是对手的部位进行追踪，即在视频流中每一帧中准确找到手的位置。第二个步骤是识别不同的手部动态动作。

2.1　基于Kinect骨点图手部追踪

要想进行手势识别，第一步要先在图像中找到手的位置，并在视频流中追踪手的位置。传统的手势识别方法大多数都是利用肤色分割并结合一些连通域的形状在图像中寻找手的位置，这种方法需要设定阈值。当光照变化很大或人的肤色差异性很大时都会出现问题，进而阻碍了手势识别技术的实际应用。而Kinect的出现解决了上述问题产生的识别干扰，通过Kinect的骨骼图像可知，手势追踪主要就是追踪Kinect人体控制点位置图2中的*A*点和*B*点。由于Kinect的平台本身可以提供骨骼点的地理坐标，因此可根据坐标来完成对手部的追踪。在应用方面，Kinect硬件可提供手在空间中的位置变化信息，可通过对该信息的比例变换，完成手势对目标物体的控制功能。

以鼠标控制为例，来验证此算法的实用性。可实现的功能有：(1)当镜头下的手上下左右移动时，鼠标也会跟随着手进行相应幅度的上下左右晃动，即完成手对鼠标的控制；(2)当手的前后变化距离达到一定程度时，可以完成对鼠标左键的按下与抬起的操作。

实现的方法是：Kinect本身具有可以提供骨点图的功能，通过控制(图2)所示*A*点在空间上下左右变化的值从而设定相应的鼠标上下左右变化的值，进而完成对鼠标的控制。

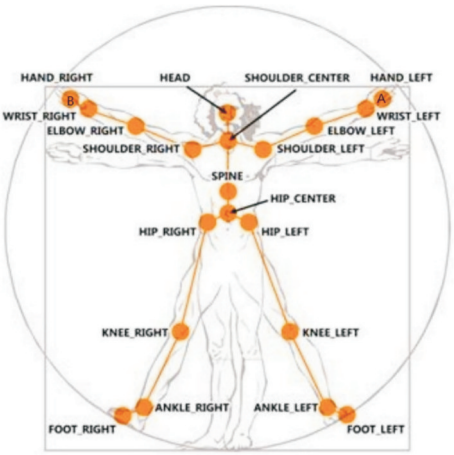

图2　Kinect人体控制点位置图

对20个身高、体重、年龄不同的人进行了10种比例的实验。手部与鼠标做上下移动的实验结果如表1所示，手部与鼠标做左右移动的实验结果如表2所示。从表1、表2可知，当人手掌的上下和左右移动距离与对应鼠标的移动距离之比分别为8:1和6:1时，体感交互满意度较高。

表1　手部与鼠标做上下移动的比例实验

上下移动比例	总人数	满意人数
5:1	20	9
6:1	20	14
7:1	20	17
8:1	20	20
9:1	20	20

表2　手部与鼠标做左右移动的比例实验

左右移动比例	总人数	满意人数
4:1	20	11
5:1	20	15
6:1	20	20
7:1	20	18
8:1	20	18

2.2　基于Kinect深度图像的动态手势识别算法

由Kinect的深度图像技术可知某点距离摄像头的距离，因此利用Kinect就可以完成对手势伸展的动作识别，并据此进行控制。

算法的核心思想是首先跟据应用的实际需要，设定一个骨点A或B的深度变化阈值H（单位：cm），通过判断骨点深度变化量的大小来进行手势识别。具体算法如下

$$Z=X-Y$$

式中，X代表实时测得手的深度值（单位：cm）；Y代表初始的测得的手的深度值（单位：cm）；Z代表两者的深度差值（单位：cm）。

（1）$Z>0$，且$Z>H$时，识别手的动作为前伸。

（2）$Z<0$，且$-Z>H$时，识别手的动作为后伸。

下面通过用手势进行对鼠标按键抬起和按下的控制为例，来验证算法的可行性。

Kinect提供了红外深度摄像头，可以测出物体的深度变化值。因此通过程序设置控制骨点A的深度变化阈值H就可以控制鼠标的抬起与落下。具体方法如下：

（1）$Z>0$，且$Z>H$时，识别为手的动作为前伸，可以设置鼠标左键按下。

（2）$Z<0$，且$-Z>H$，识别为手的动作为后伸，可以设置鼠标左键抬起。

在阈值的选取方面，由于深度摄像头的深度测量精度所限，在其稳定工作的状态，增大阈值可以使得对鼠标的控制成功率增大，但是如果选取较大阈值会使得用户的体验度大幅下降。

因此针对如何选取既可以提高鼠标控制成功率，并可以保证用户体验的问题，此实验针对每个阈值分别进行50次的独立实验。实验数据如表3所示，可以看到在阈值设为H=25 cm时，对鼠标左键按下的控制效果成功率较高。

表3 深度阈值实验数据

深度阈值/cm	试验次数	成功次数	成功率/%
5	50	15	30
10	50	24	48
15	50	36	72
20	50	46	92
25	50	49	98
30	50	49	98

2.3 算法优势

（1）摆脱了传统手势识别需要进行肤色分割（提取手的轮廓）的过程。因为这一过程会受到光照，人的肤色差异性大等各方面条件的限制，会严重的影响手势识别的稳定度，而骨点跟踪采用的是红外摄像头，受光照和颜色影响性不大，使整个系统的鲁棒性大幅度提升，也提高了手势识别的稳定性。

（2）通过人手深度的变化来完成动态识别手动作的前伸与后伸，具有很强的用户适应度，这个手势识别动作对于人来很容易操控，实用性更强。

（3）在动态手势识别的准确率上，此算法远远超过其他算法，高达99%的准确率使之具有很高的实际应用价值。

3 结束语

提出了一种基于深度图像和骨骼图像的手势识别算法，在手的追踪方面和动态手势识别方面的正确率和稳定度上超过了其他算法。

基于此算法的手势识别可以用于多种场合，例如：在远程操控中，通过手来控制远程汽车的前进与后退；在讲解PPT时，通过手来控制PPT翻页；在播放音乐时，用手的深度变化来控制音量的大小。在空中书写文字或符号时，通过手势的深度变化来区分抬笔和落笔等。

由于红外深度摄像头的精度所限，基于骨骼图像和深度图像的算法在深度变化不是很大时，还难以做到精准识别。但随着硬件功能的提升，动态手势识别会向更精确，更具有实用性的方向发展。

参考文献

[1] 任海兵，祝远新，徐光桔，等. 基于视觉手势识别的研究——综述[J]. 电子学报，2000，28(2)：11-12.

[2] 张良国，吴江琴，高文，等.基于Hausdorff距离的手势识别[J]. 中国图像图形学报，2012，7[A]:1-8.

[3] 刘江华，陈佳品.用于人机交互的静态手势识别系统[J]. 红外与激光工程，2002，6:499-503.

[4] 贾建军.基于视觉的手势识别技术的研究[D]. 哈尔滨：哈尔滨大学，2008.

[5] 唐文平，胡庆龙.基于多目标Camshift 手势识别[J]. 电子科技，2012，25 (2)：71-81.

[6] 谭同德，郭志敏.基于双目视觉的人手定位与手势识别系统研究[J]. 计算机工程与设计，2012，33(1)：259-264.

[7] 李文生，解梅，邓春健.基于机器视觉的动态多点手势识别方法[J]. 计算机工程设计，2012，5(8)：60-72.

[8] Microsoft Corp. Redmond WA. Kinect for Xbox 360[S].

（下转第63页）

5 结束语

将VIPA与衍射光栅结合实现二维成像,是光谱处理领域的一个重大进步。文中结合目前该技术的发展状况,详细介绍了其在光学滤波器、光谱处理、光学成像几个方面的具体应用,并对各项应用的未来发展做出展望。

参考文献

[1] Shirasaki M . Large angular dispersion by a virtually imaged phased array and its application to a wavelength demultiplexer[J]. Optics Letters ,1996 , 21(5) : 366-368.

[2] Shijun X, Andrew W.An Eight-Channel Hyperfine Wavelength Demultiplexer Using a Virtually Imaged Phased-Array(VIPA)[J]. Ieee Photonic Tech L,2005,17(2):372-374.

[3] Ghang-Ho L, Shijun X ,Andrew W. Optical Dispersion Compensator With >4000-ps/nm Tuning Range Using a Virtually Imaged Phased Array(VIPA) and Spatial Light Modulator (SLM)[J]. Ieee Photonic Tech L,2005,18(17):1819-1821.

[4] Shirasaki M. Filtering Characteristics of Virtually-Imaged Phased-Array[J]. Integrated Photonics Research (IPR), 1996,6 IMC3.

[5] Shijun X, Andrew W. 2-D wavelength demultiplexer with potential for $1000 channels in the c-band[J]. Opt. Express , 2004 , 12(13) : 2895-2902.

[6] Supradeepa V, Huang C, Leaird D, et al. Femtosecond pulse shaping in two dimensions:Towards higher complexity optical waveforms[J]. Opt. Express, 2008, 16(16) : 11878-11887.

[7] Shijun X, Andrew W.Optical Carrier-Suppressed Single Sideband (O-CS-SSB) Modulation Using a Hyperfine Blocking Filter Based on a Virtually Imaged Phased-Array (VIPA) [J]. Ieee Photonic Tech L, 2005,17 (7) : 1522-1524.

[8] Scarcelli G, Yun S.Multistage VIPA etalons for high-extinction parallel Brillouin spectroscopy[J]. Opt. Express, 2011, 19(11) :10913-10922.

[9] Cundiff S, Andrew W. Optical arbitrary waveform generation [J]. Nat. Photonics , 2010 , 4(11):760-766.

[10] Diddams S, Hollberg L, Mbele V. Molecular fingerprinting with the resolved modes of a femtosecond laser frequency comb[J]. Nature , 2007 , 445(7128) : 627-630.

[11] Keisuke G, Tsia K ,Bahram J.Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena[J]. Nature , 2009 , 458(7242) : 1145-1149.

[12] Shirasaki M.Compensation of chromatic dispersion and dispersion slope using a virtually imaged phased array[J]. TuS1OFC , 2001 (3):18-23.

(上接第50页)

[3] 闫丰,于子江,于晓,等.电晕探测紫外ICCD相机图像噪声分析与处理[J]. 光学精密工程,2006,14(4):0709-0713.

[4] 许强.军用紫外探测技术及应用[M]. 北京:北京航空航天大学出版社,2010.

[5] 张德峰.详解MATLAB数字图像处理[M]. 北京:电子工业出版社,2010.

[6] 赵玉环,闫丰,隋永新,等.紫外序列图像中目标的提取[J]. 光电工程,2007,34(11):0010-0013.

[7] 冯鹏,魏彪,米德伶,等.基于时域递归滤波的动态数字图像降噪[J]. 重庆大学学报(自然科学版),2005,28(2):0023-0025.

(上接第58页)

[9] J Salvi, J Pages, J Battle.Pattern codification strategies in structured light systems[J]. Pattern Recognition, 2004, 37 (4): 827-849.

[10] P Lavoie, D Ionescu, E Petriu. 3D reconstruction using an uncalibrated stereo pair of encoded images[C]// In Proceedings of the Int. Conf. on Image Processing, 1996.

[11] Chadi ALBITAR, Pierre GRAEBLING, Christophe DOIGNON. Robust Structured Light Coding for 3D Reconstruction[C]// In Proc. ICCV, 2007.

[12] P Lavoie, D Ionescu, E Petriu. 3D reconstruction using an uncalibrated stereo pair of encoded images[C]//In Proceedings of the Int. Conf. on Image Processing, 1996.

[13] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, et al. Real-Time Human Pose Recognition in Parts from Single Depth Images[C]//In Proc. CVPR, 2011.