**引言**

**—**基于深度摄像机的手势识别

手势识别是计算机科学和语言技术中的一个主题，目的是通过数学算法来识别人类手势，实现基于视觉的手势识别对构建自然方便的人机交互系统至关重要。目前摄像头提供的图像信息多为彩色图像，随着红外技术的进一步发展，使用新型传感器的深度摄像机进入市场，深度摄像机提供的深度图像也被用于行为理解研究。在研究深度摄像机的手势识别，参考了许多文献，对此有一些认识和看法，故编写了该篇文章，希望对这科研项目的研究能给自己带来一些新的认识与想法。本文主要内容如下：

1. 本文简单介绍了手势识别概念、其核心内容以及现阶段所发展的技术等，包括基于入侵式的手势识别，基于普通视觉的手势识别和基于深度摄像的手势识别技术。本文重点介绍了基于深度摄像的手势识别技术。
2. 对于人的行为理解算法，本文也稍微做了介绍，例如对于从数据流中获取行为的特征值方面，可以使用基于手部和脚步关节点速率的分割方法，可以有效减少计算量等。
3. 本文还介绍了现阶段深度摄像的市场所研发的几种摄像机，例如微软的Kinect，华硕的Xtion。以及技术研发的方向以及面临的问题，包括受限于高维度、快速手势运动，光影响等。

**正文**

**1.课题研究的背景和意义**

人机交互（Human—Computer Interaction，HCI）主要研究人与计算机之间的传递、交换信息的媒介和对话接口，是计算机系统的重要组成部分。随着计算机技术的快速发展，市场需求不断提高，人们不再只满足与交互的基本功能，开始从最起初的基本命令交互，发展到图像交互，人机交互的方式趋于多样化，逐渐开始追求更简约化的人机交互。随着计算机视觉、信号处理更加快速，符合人类自然的交互模式越来越广泛。

起初穿戴式传感器被利用于手势识别，PJuha等研究学者利用传感器以此来获取手势信息，再利用决策树和人工神经网络方法对数据分类，进而能够准确的测量手势运；MErmes等人在人体的手腕和臀部安装3D加速器采取人体运动信息并使用特殊的树结构方法对数据分类，还有很多研究例子。

但基于入侵式的手势识别的致命缺陷在于设备价格昂贵，不能普遍运用，以及不变携带等问题，让人们将焦点转向了视觉方面的手势识别—基于普通视觉的手势识别。当时很多高校和学者通过图像的方法研究人的行为，其中卡内基梅隆大学等十几所高校联合开展了VSAM项目，他们想通过摄像机技术、图像处理技术等实现对城市的自动监控；麻省理工Wren团队的Pfinder项目使用单一的摄像机和灰度传感器对复杂背景下的人体进行实时跟踪并理解人的行为。该技术明显比上一代的技术更符合人的自然表达，并且设备比较低廉，但同时带来了缺陷—不稳定性|。

对于如何在图像序列中准确地找到收的位置并且进行复杂的手势分割问题，人们研发了深度摄像头，这款摄像头与普通的摄像头的区别在于除了能够获取平面图像以外还可以获得拍摄对象的深度信息，也就是三位的位置和尺寸信息。这便意味着基于深度摄像可以减少目标遮挡，快速运动，光线变化等干扰，提高手势识别的准确性。上世纪八十年代就由蓝色巨人IBM提出相关概念；2005年创建于以色列的PrimeSense公司将这项技术民用化，当时深度摄像机仅适用于军事和工业领域，为机械臂、工业机器人等提供图形视觉服务；微软与PrimeSense合作开发了Kinect1成为深度摄像机在消费领域的开山之作；在2012年，微软又自行开发了Kinect2代......

1. **国内外研究现状**

基于入侵式的手势识别现在已经达到比较成熟阶段，准确性较高；而基于视觉的手势识别技术大部分还停留于从数字相机获取图像或视频；基于深度摄像的手势识别则处于起步阶段。现在深度摄像头主流有三个技术路线：单目结构光、TOF（Time of fligh飞行时间）和双目视觉。但是基于深度手势识别技术的研究尚处于起步阶段，手势识别仍然存在巨大的挑战，手势的高维度问题，手指间的自遮挡，快速的手势运动，以及手势识别处理速度，不可控制的环境等都影响手势识别的精度和效率。特别是大多研究还停留于静态手势识别的研究工作上，识别准确率还有待提高。总而言之，对于基于深度摄像的手势识别技术在未来各个领域中，具有非常大的研究意义和商业价值。

在我国，高校和科研单位也对该方向进行了一些研究，清华大学在2015年开展了“建设工程施工现场人机行为识别分析”项目，利用行为识别来保障施工现场的安全性；中科院自动化研究所生物识别与安全技术研究中心的“CBSR智能视频监控系统”重点项目针对目标异常行为的识别等课题做了深入的研究。

对于人的行为理解研究，人们通过深度摄像机获取具备深度信息的深度图，很多实验室或研究所还建立了人的行为的深度图像数据库，例如康奈尔大学计算科学中心的CAD-6、CAD-120数据库，微软前言技术研究所的MSR ACTion 3D Dataset、3D Online Action Dataset。而对于提取深度特征方面，JSung等研究对人体的深度图像使用了简单算法处理后直接稍作变换，可在彩色图像或灰度图像使用的HOG算法；L..Spinello等通过HOG算法启发，提出深度方向直方图算法提取特征，还有诸如此类的研究都广泛应用于手势识别中。

1. **基本内容**
2. 手势识别

手势识别可以来自人的身体各部位的运动，但一般是指脸部和手的运动。用户可以使用简单的手势来控制或与设备交互，让计算机理解人类的行为。其核心技术为手势分割、特征提取分析以及手势分析等。手势识别可以粗略地分为两个阶段：二维手势识别以及三维手势识别。二维手势识别仅限于识别几个静态的手势动作并且必须提前预设好，而三维手势识别增加y轴的信息，包含了一定的深度信息的手势识别 ，需要用传感器和光学摄像头来实现。

目前行业内主要采用的三种手势识别有：结构光、光飞时间以及多角成像。结构光(Structure Light)是通过激光的折射以及算法计算出物体的位置和深度信息，进而复原整个三维空间，其代表产品有微软的Kinect一代。光飞时间TOF（Time of Flight）加载一个发光原件，通过CMOS传感器来捕捉计算光子的飞行时间，根据光子飞行时间推算出光子飞行距离，便得到了物体的深度信息，其代表作品有Intel的三位摄像头。多角成像（Mulyi—camera）使用两个或两个以上的摄像头同时采集图像，通过比对这些不同摄像头在同一时刻获得的图像的差别，使用算法来计算深度信息，从而多角三维成像。

（2）手势分割

对于手势识别中的手势分割，可以利用简单的深度阀值将手分离出来，就是采用一个用户指定或者现场中最近的点作为手的质点，将质点附件深度值在阈值以内的像素认为手所包含的像素。但这种方法必须在特定假设下才能实现的。Cerlinca和Van den Bergh采用OpenCV中提供的鲁棒性的人脸检测方法，根据手势的深度接近任意的其他部位的假设，以及人脸的位置确定头部的位置，来估计其他可能的区域。Fujimura则利用Kahunen-Loeve分割方法来检测到人体区域，然后来估计收的位置。

手势分割可以看在个在Kinect传感器获得的RGB图像和深度图像中的手像素或非手像素标记问题。将RGB图像和对应的深度图像数据定义为： color为位置（i，j）对应的颜色信息，按照某种颜色空间定义，depth为该位置的深度信息， 进行手分割时，必须对Data加以处理，提炼特征。在文献中，把手分割看成一个深度聚类问题，像素分组于不同的深度水平。通过分析人姿势维度确定一个阈值，该深度水平对应的是手。文献利用k-means聚类算法和预定义阈值进行手检测，并对手型进行凸包分析定位手指。这两种方法都是假设手部距离传感器最近，此外算法的效果受到kinect深度数据的准确性影响非常大。文献，利用kinect SDK提供的关于手的位置信息，在深度图像的三个轴方向上设定阈值，小于该阈值的被看成是可能的手的区域。然后利用OPEN VC中的相关函数cvErode（腐蚀）、cvMorphologyEx、cvFindContours（找到边界）对手数据进行处理，计算边界和手中心。文献建立三个模型：深度模型（深度加上形态学），皮肤模型（肤色），背景模型（codebook算法），利用三层神经网络，输入是三个模型的两两重叠率，其反映了两个模型分割结果的一致性。文献直接在深度图中寻找管状或指间形状的物体，作为手掌和手指的候选对象，通过空间位置检测确认，因为手掌和手指是一起的。文献利用kinect产生的骨骼数据，确定手的位置，查找表存放了标准人类手在不同深度层的大小，给定手的3D位置，算法从查找表中找到对应手的大小，从而粗略定位手的区域。文献集成彩色信息进行手检测。将手检测问题看成是一个手像素或非手像素标记问题。RGB图像的皮肤检测算子和深度图像的聚类算子作为两个条件确认手像素。手区域是二者的交集。文献将不同的特征处理为不同算子，利用手部/非手部区域的相邻像素点的深度信息和颜色信息具有一定连续性的原理，从手心作为起始点进行搜索，保证了所有像素形成一个联通的有效区域。避免了传统简单基于深度算法中存在的手必须在最前方限制，或多物体匹配区分的问题，有效的处理了手部颜色和深度数据不均匀的问题。

（3）特征提取

特征提取包括对静态手势和动态手势识别。其对于基于视觉手势识别至关重要，所选特征对不同的手势是否具有有效的区分能力直接决定了手势识别的准确性。特征提取受限于目标尺寸较小、颜色均匀和缺乏明显纹理信息等因素。静态手势识别关注的是人手图像的静态信息（形状、结构等）。邓瑞等将深度图像投影到三维空间，则人手区域在该三维空间中形成点云，然后对人手区域的深度值进行直方图统计，由于这种特征不具有旋转不变性，在进行特征提取之前首先采用最小二乘法对于手势进行归一化处理。Liu则采用了基于Chamfer距离变换（Distance Transform，DT）图像与边缘图像做互相关提取图像的特征。Kurakin采用了分割后的深度图像进行网格划分，统计每个网格中确定为手势部分的面积占整个网格面积的比率作为特征，解手势的轮廓特征用于静态手势识别。张凯则采用了统计特征，轮廓特征，以及结构特征等多种特征描述手势。

而动态手势识别由包含手势的序列图像组成，需要提取由多帧图像中的手势共同提供的手势特征，现阶段对于动态手势识别技术还不完善，未来研究这项技术将会成为主流。目前的手势识别方法包括模版匹配法和基于状态转移的图模型方法等。通过手势样本为每种手势训练建立对应的模板。当识别一个新的手势时，首先计算其特征向量，然后与已知模板逐个进行匹配，相似度最高的模板对应的手势类型即为识别结果。一般使用3种函数来衡量输入手势和手势模板特征序列之间的相似度：平方差、相关系数和相关匹配。模板匹配方法建立和修改比较简单，但当动态手势变得非常复杂时，手势在时间和空间上的差异性，使得每种手势的模板阈值变大，即设定匹配范围较宽。当手势种类较多时，有可能出现一个模板同时与几种手势相匹配的情况，最终导致识别错误。模板匹配方法的识别速度随着手势种类增多而逐渐降低，因此，模板匹配方法不能解决手势的时空差异性问题，无法准确实现实时的多手势识别。

（4）手势分类

从深度图像中提取出了适合的特征或者将手分解为合适的部分，就可以通过机器学习的方法或者基于特征的特殊分离器对手势进行识别。其中包括隐马尔可夫模型（Hidden Markov，HMM）可对时间信息的数据进行分类，并且速度很快）、动态时间规整（Dynamic Time Warping，DTW）用于信号处理和模式识别，有效地衡量时间序列举例，多用于动态手势识别中和K近邻分类器常用于静态手势的分类，速度快且实现简单等分类方法。

（5）深度学习技术

　在手势识别基本流程中，特征对于手势建模起着非常关键的作用。良好的特征表达，对最终算法的准确性起着非常关键的作用。手工选取和设计特征非常费力，需要专业知识和大量的时间进行调节，成功极大地依赖于设计者的经验和运气。   
　　深度学习是机器学习研究中一个新的领域，其目标是建立、模拟人脑进行分析学习的神经网络，通过非监督学习，实现自动的学习特征。深度学习构建具有很多（5层、6层，甚至10多层）隐层的机器学习模型，通过海量数据训练，使得原样本空间逐层变换到一个新特征空间，以学习更有用的特征。这类似于人类从原始信号作低级抽象，逐渐向高级抽象迭代，最终的分类或预测在高级抽象层进行，从而提高准确性。与人工规则构造特征方法相比，利用大数据来学习特征，更能够刻画数据的内在信息。   
　　深度学习技术已经在图像处理领域取得了成功，许多学者尝试将深度学习引入到手势识别中。文献中]扩展了 CNN网络，自动学习时空特征，采用递归神经网络训练学习每个时间节点上的特征，在KTH数据集上取得了良好的效果。上述这些方法归属于基于时空的方法，主要缺陷是对复杂动作的表达能力较弱。

（6）目前主要的深度摄像头

Intel RealSense R200（参数规格：分辨率：1080p，深度有效距离：0.51-4，USB3.0）

乐视 LeTV Pro Xtion（参数规格：深度有效距离：0.8m 至 3.5m，USB2.0）

Orbbec Astr(参数规格：分辨率720p，深度有效距离：0.5 至 8m，USB2.0/USB3.0)

Intel RealSense F200（参数规格：分辨率：1080p，深度有效距离：0.2-1.2，USB3.0）这个主要是近距离的应用场景，比如在电脑笔记本前的应用。

1. **现阶段面临的问题**

手势识别是人机交互领域的重要内容。在彩色图像中，手势区域难以提取，这是制约于彩色视觉的手势识别的发展的一个重要方面。深度摄像为手势带来契机，使得手势检测更加容易，但还是会被背景噪音声影响，同时也仍然受限于手势区域面积小，颜色均匀，难以提取有效地纹理信息，手势的高维度问题，手指间的自遮挡，快速的手势运动，以及手势识别过程的处理速度，不可控制的环境等因素。因而这些问题促使更多人去加强研究这些方面。

**结论**

从各个文献来看，自然手势存在多样性、多义性、时空差异性等特点，使得手势识别成为一个非常具有挑战性的课题。传统的动态手势识别技术取得了一定的成功，但手工选取特征极大地限制了这些技术的发展。深度学习技术模拟人脑进行分析学习，可以从原始的数据空间层层进行特征抽象和变换，最终在高层特征空间进行，极大地提升了机器学习的准确性。学习和借鉴深度学习技术，无疑是动态手势技术发展的方向，动态手势的时空复杂性也将反过来进一步推进深度学习技术的进步和发展。手势区域面积小，颜色均匀，难以提取有效地纹理信息，手势的高维度问题，手指间的自遮挡，快速的手势运动，以及手势识别过程的处理速度，不可控制的环境因素待我们去解决。

**附录**

本文参考文献：《基于深度摄像的手势识别关键技术研究》—胡丽华

《基于Kinect的人的行为识别研究》—刘从文

《手势识别中手分割算法》—郭雷