**毕业设计(论文）开题报告**

|  |  |
| --- | --- |
| 论文题目 | 基于机器视觉的人体运动类型识别 |
| 1. 选题背景和意义：   现代社会快速的生活节奏和巨大的工作压力，严重影响着个人的身体健康。科学的运动可以提高身体素质增强运动能力，进而降低患病的风险(尤其是一些慢性疾病），例如糖尿病、血脂异常、高血压等。而进行科学运动的前提是实现人体运动类型的准确识别。  在信息安全领域，通过智能监控的方式利用计算机自动对视频中人体的运动类型进行识别从而为监控或者案件侦破提供依据也具有着重要的意义和广泛的用途。  在人机交互领域可以通过手势、身体姿态等信息对除了辅助交互输入设备以及自然语言分析进行补充，提高计算机和人进行交互的能力，使之更有趣。  在大型的图像数据库或者互联网上的信息中还可以利用人体运动类型的识别，对部分信息进行标注和理解，进而提供搜索和学习的能力。  本课题的任务是通过深度学习的相关理论，尝试用基于深度学习的机器视觉模型，应用于人体运动类型的识别，从而让模型对人体运动类型有更好的理解与识别，同时比较使用深度学习的人体运动类型识别算法和传统机器学习的人体运动类型算法的优劣。并通过毕业设计锻炼培养学生综合运用所学专业课、专业基础课知识的能力，了解深度学习的相关知识与应用，提高科研能力。 | |
| 二、课题关键问题及难点：   1. 课题关键问题：     由于一般的人体运动数据源都比较大，是一些色彩图像序列或者是深度图像序列。对于这些图像直接进行分类或者识别的话会造成输入到识别系统的数据的维度非常的高。如此高的数据维度意味着会消耗大量的计算资源或者难以收敛到最优解。为了解决输入数据的高维度问题，一般都需要对输入的数据进行特征的提取，也就是构造能够用来对数据进行某一个方面的识别或者分类的数据的一种表达方式。这样子就可以把数据的维度降低，从而对数据进行更有效的处理。   * 1. 先验的特征提取   其实就是人工地对图像进行预处理，去掉无关的信息，只保留对识别有用的信息。比如在人体运动的识别中，可以通过去除背景，去除人的其他特征(衣着、相貌），只保留人的骨架模型来减少数据的维度，降低数据识别的难度。   * 1. 后验(自动）的特征提取   其实先验的特征提取，就是把人的智能融入到识别的模型中去，这种过程很容易由于人工特征提取的不完整而造成数据无法被有效地表达，或者使得研究的模型与目标物体结合太过紧密而难以对其他的目标进行推广，比如专精于人体运动识别的模型可能难以应用于动物的运动识别上去。而后验的特征提取可以减少人工的干预，只由机器自己来进行特征的提取，这样子可以少人工的错误，也能将人的工作量转移到机器上去，让研究者能够把精力放在更加普适的理论的研究上去。   1. 课题难点：    1. 对数据进行有效的特征提取(人工和自动提取）。    2. 对数据特征进行识别分类。 | |

注：开题报告可单独装订，但在院(系）范围内，封面和装订格式必须统一。

|  |  |
| --- | --- |
| 1. 文献综述(或调研报告）： 2. 人体运动识别    1. 人体运动数据采集方式   首先，对于人体运动识别的数据的来源存在着几种不同的种类。例如由 RGB 摄像机、 距离传感器或其他遥感的方式。使用深度传感器来进行人体运动识别的发展始于 80 年代初。过去的研究主要集中在学习和认识到从视频序列(可见光相机）所采集的数据中。可见光视频的主要问题是从单目视频传感器捕捉得到的人体运动存在相当大的损失。由于视频天生的对的人体行为识别的限制，尽管已经有了过去几十年的努力，但通过视频来识别人体运动，仍然是非常具有挑战性的。  而得益于近期发布的成本低廉的深度传感器，我们看到了和 3D 数据相关的研究越来越多了。从过去的 20 年里，我们获得 3D 数据的方法，一共分为三类。一种方法是通过使用基于标记的动作捕捉系统，如 MoCap。 第二种方式是通过立体视觉： 从多个角度捕获 2D 图像序列，通过从多个视图来重建三维信息。第三种方式是使用距离传感器。深度相机在过去几年里取得迅速的发展。最近出现的深度照相机可以在相对低廉的成本和较小的尺寸里给我们提供较高的帧率和分辨率，这导致出现了许多新的研究中的动作识别都是采用的三维数据。   * 1. 人体动作识别的主要问题   在不同的环境中，人体运动可能有不同的形式，从简单的动作到复杂的活动。他们可以在概念上分为四个类别 [1]： 原子动作、 包含不同的动作序列的大动作、 包括人与对象之间的交互的相互作用，最后还有人与人之间的互动和小组活动。通过3D数据进行原子动作的研究已经进行了很多年，而复杂的活动和相互作用的研究由于最近 3D 数据已经变得非常容易获取，而变得开始多起来。至于使用 3D 数据的小组活动的研究非常有限，原因要么是由于数据难以获取，要么是由于传感器具有局限性。  基于视觉的人体动作识别里有四个主要的问题。第一个问题的挑战性比较小 [2] [3]：闭塞、 杂乱的背景、 阴影和不同光照条件会让运动难以分割或者被错误地识别。这是从 RGB 视频行为识别的一大难点。引入 3D 数据可以在很大程度上通过提供现场的结构信息，从而缓解这个问题。第二个是视角的变换[2] [4] [5] 和 [6]。相同的操作可以从不同的角度产生不同的"外观"。传统的 RGB 相机解决这一问题的方法主要是引入多个同步的摄像机，同时获得多个视角的图像，但对于某些应用程序，这不是件容易的事。不过对于三维运动捕捉系统，这不是一个严重的问题。而如果通过深度图像来进行识别的话，这个问题也会有部分被缓解，因为从轻微旋转的视角的外观可以推断深度的信息。这一点并不完全解决问题，因为摄像机始终还只是在对象的一侧上，这个距离图像只提供了部分的信息，还是没有人知道这个对象的另一面是什么样子的。如果可以运用单一深度相机来精确地推断出人的骨架模型，则可以通过骨架模型的信息来构造一种视图不变识别的算法。第三个问题是放缩上的差异，因为人离相机的距离的不同会影响主体的大小从而影响运动的识别。而在 RGB 视频中，这可以通过在多个尺度下的[7] 特征提取解决了。而在深度视频中，这可以很容易调整，因为真正的主体的 3D 尺寸直接是已知的。第四个问题是同一种类内的变异性和不同种类之间的相似性问题[8]。人可以通过不同的身体部位在不同的方向上做动作，但不同的方向和两个动作仅只有只由非常细微的细节来区分。而这个不管对于使用哪种数据的来源的算法来说都仍然是一个非常困难的问题。   1. 深度学习   模仿出人脑表征信息的高效和鲁棒性一直是近几十年来人工智能研究中的一个核心。而人类不仅每时每刻都暴露在由感官接收的无数的数据中，而且还能够通过某些方法捕捉到这些数据关键的部分来让自己能够以简单的方式在未来使用。早在五十年前，提出了动态规划理论以及开创了最优化控制领域的 Richard Bellman 就曾断言，数据的高维度是在人工智能科学和工程应用中的根本障碍。其中主要的困难，尤其是在模式分类的应用中，就是数据学习的难度会随着数据维度的线性增长，发生指数级的增长[9]。而克服这个“诅咒”的主流方法就是以一定的方式(比如分类引擎）对数据进行预处理来减少数据的维度，这样数据就可以被有效地处理。这种减少维度的做法一般被称为特征提取。因此，可以认为，在很多模式识别系统的背后的智能其实被转移到了人工的特征提取处理上去了，而这种人工的做法有时会很困难而且会高度依赖于具体的应用场景[10]。此外，如果不完整或者错误的特征被提取出来了，那么分类处理的性能就会从根本上受到限制。  [[1]](#endnote-1)最近，神经科学在哺乳类动物大脑上的新发现告诉我们，我们可以通过一个复杂的深层网络结构来对数据自动地进行预处理。这就是深度学习的基本来源。  而在深度学习中，卷积神经网络(CNNs）和深度信念网络(DBNs）(还有它们各自的变形）是最备受关注的。   * 1. 卷积神经网络(CNNs）   CNNs [11][12]是一个专门为二维数据(例如图片和视频）设计的多层神经网络的系列。 CNNs 是第一种真正意义上成功的深度学习方法，它能够成功地用一种鲁棒的方式训练出多层的神经网络。 单个 CNN 的基本构造是先通过多个参数可学习的卷积核进行卷积，得到了多个特征图之后，对它们进行重采样(池化）来降维，最后通过一个激活函数传播到后面。可以把多个 CNN 来得到不一样的特征的组合。数据经过 CNNs 后就相当于被提取了特征，之后再把它输入到经典的分类器比如多层神经网络里即可完成分类任务。   * 1. 深度信念神经网络(DBNs）   DBNs 由几层的限制玻耳兹曼机(RBM）组成。假设有一个二部图，每一层的节点之间没有链接，一层是可视层，即输入数据层(v)，一层是隐藏层(h)，如果假设所有的节点都是二值变量节点(只能取0或者1值），同时假设全概率分布 p(v, h) 满足Boltzmann 分布，我们称这个模型是Restrict Boltzmann Machine (RBM)。当输入v的时候，通过 p(h|v) 可以得到隐藏层h，而得到隐藏层h之后，通过p(v|h) 又能得到可视层，通过调整参数，我们就是要使得从隐藏层得到的可视层v1与原来的可视层v如果一样，那么得到的隐藏层就是可视层另外一种表达，因此隐藏层可以作为可视层输入数据的特征。由多个 RBM 组合起来就得到了DBNs。同样的，由 RBMs 描述出来的数据的特征可以在整个网络的最后一层里输入到结尾的一个分类器里完成分类任务。  **参考文献**  [1] V. Argyriou, M. Petrou, S. Barsky, Photometric stereo with an arbitrary number of illuminants, CVIU, 114 (2010), pp. 887–900  [2] D. Weinland, M. Özuysal, P. Fua, Making action recognition robust to occlusions and viewpoint changes, ECCV. Springer (2010), pp. 635–648  [3] L.C. Chen, J.W. Hsieh, C.H. Chuang, C.Y. Huang, D.Y. Chen, Occluded human action analysis using dynamic manifold model, ICPR, IEEE (2012), pp. 1245–1248  [4] M.B. Holte, T.B. Moeslund, N., Nikolaidis, I. Pitas, 3d human action recognition for multi-view camera systems, in: 3DIMPVT, 2011, pp. 342–349  [5] D. Weinland, E. Boyer, R. Ronfard, Action recognition from arbitrary views using 3d exemplars, ICCV, IEEE (2007), pp. 1–7  [6] M.C. Roh, H.K. Shin, S.W. Lee, View-independent human action recognition with volume motion template on single stereo camera, Pattern Recognit. Lett., 31 (2010), pp. 639–647  [7] M.Y. Chen, A. Hauptmann, Mosift: Recognizing human actions in surveillance videos, 2009.  [8] R. Poppe, A survey on vision-based human action recognition, Image Vision Comput., 28 (2010), pp. 976–990  [9] R. Bellman, Dynamic Programming. Princeton, NJ: Princeton Univ. Press, 1957.  [10] R. Duda, P. Hart, and D. Stork, Pattern Recognition, 2nd ed. New York: Wiley-Interscience, 2000.  [11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.  [12] F.-J. Huang and Y. LeCun, “Large-scale learning with SVM and convolutional nets for generic object categorization,” in Proc. Computer Vision and Pattern Recognition Conf. (CVPR’06), 2006. | |
| 四、方案(设计方案、或研究方案、研制方案）论证：   1. 人体运动的特征提取    1. 人工特征提取   从上的把握来说，人体运动中的类型很大一部分是由肢体的运动来完成的，因为人体的肢体的骨骼模型可以有效地对人体的运动进行简化、提纯，是一个行之有效的特征。而现在由于 3D 传感器的兴起，所以获取人体的深度信息，并且推算出人体的骨骼模型已经是一个非常成熟而且成功率极高的技术了，我们的人体运动的人工特征可以选择人体的骨骼模型。   * 1. 深度学习中的自动特征提取   我们可以利用深度学习中的众多技术来对人体运动的数据进行特征提取，从而得到一个优秀的识别模型。在这里面，我们可以通过多次的实验来使用 CNN 、 DBN 等多种网络来对已有的数据集进行学习，从中挑选一个性能最好的来满足我们对数据特征的提取的需求。   1. 特征分类   通过经典的分类器：MLP 或 SVM， 我们可以对上述方法提取出来的特征进行自学习的分类，从而得到我们最终想要的分类结果，也就是对人体运动类型的识别。 |
| 五、进度安排：   * 1. 2015年1月-2月 前期文献调研、理论学习与软件学习   阅读人体运动类型识别以及深度学习相关的文献，对相关的经典的课程进行学习，并动手实现相关的经典模型。   * 1. 2015年3月-4月 开展毕业设计的软件开发与仿真   对可能的多个模型进行仿真实验，对比结果，得到最优的方案并完善优化。   * 1. 2015年5月 撰写毕业论文   在老师的指导下对实验程序、数据进行总结整理并撰写论文。   * 1. 2015年6月 毕业论文答辩   进行毕业论文答辩。 |
| 六、指导教师意见：  签名： 年 月 日 |
| 七、开题审查小组意见：  签名： 年 月 日 |

1. [↑](#endnote-ref-1)