摘要

视觉人体跟踪是计算机视觉领域的热点研究内容。考虑到从图像特征空间到状态特征空间的映射具有多义性，本文基于混合专家模型(Bayesian Mixture of Experts, BME)，采用“分而治之”的思想，建立了从图像特征空间到三维参数空间的映射，从而实现三维人体跟踪。使用背景减法从图像序列中提取人体区域，采用PCA方法对人体区域特征进行降维，应用BME算法和k-means聚类算法建立图像特征空间和状态空间之间的对应关系，从图像特征中估计出人体模型的三维姿态。实验结果表明本论文方法能够准确地从单视频序列中重建三维人体运动。

关键词： 混合专家模型 三维人体跟踪 计算机视觉

ABSTRACT

3D human pose tracking plays an important role in vision-based human motion analysis. For the mapping from two-dimensional image observation to three-dimensional human configuration is multi-valued, Bayesian Mixture of Experts model (BME) with the idea of "Divide and Conquer" is introduced to build the one-to-many mapping function between image-feature space and 3D-parameter space. The foreground human region is extracted by background subtraction and then the silhouette is obtained by removing noise and filling in regions. Due to the high-dimensionality of silhouette descriptor, PCA (Principal Component Analysis) is used to reduce the dimension of image features. The mapping between image-feature space and 3D-parameter space is set up by using BME based on k-means cluster algorithm to infer 3D angles of human pose from image features. Experiment conducted on CMU Mocap database illustrates that the proposed method could track complex 3D motion of human in real monocular video sequences accurately and robustly.

**Keywords: Bayesian** **Mixture of Experts 3D human body tracking computer vision**

目录

[第一章 引言 1](#_Toc513984150)

[1.1课题背景 1](#_Toc513984151)

[1.2 典型应用 2](#_Toc513984152)

[1.3 研究现状 3](#_Toc513984153)

[1.4 主要跟踪方法概述 4](#_Toc513984154)

[1.5 本文研究内容 5](#_Toc513984155)

[第二章 混合专家模型 33](#_Toc513984156)

[2.1混合专家模型概述 33](#_Toc513984157)

[2.2 k-means聚类算法 34](#_Toc513984158)

[**2.2.1 k-means聚类算法工作原理** 34](#_Toc513984159)

[2.3 BME(Bayesian Mixture of Experts)算法 34](#_Toc513984160)

[**2.3.1 BME算法原理** 34](#_Toc513984161)

[**2.3.2 BME算法证明** 35](#_Toc513984162)

[第三章 特征提取 36](#_Toc513984163)

[3.1 背景减(background subtraction) 36](#_Toc513984164)

[**3.1.1 运动检测** 36](#_Toc513984165)

[**3.1.2 背景建模** 36](#_Toc513984166)

[3.2主成分分析 (Principal Component Analysis，PCA) 37](#_Toc513984167)

[**3.2.1 降维算法概述** 37](#_Toc513984168)

[**3.2.2 PCA算法原理** 37](#_Toc513984169)

[在高维数据分析中，维数间存在共线性，增加了分析的复杂性。但是若分别 38](#_Toc513984170)

[第四章 三维人体跟踪的实验 38](#_Toc513984171)

[4.1 实验概述 38](#_Toc513984172)

[**4.1.1 实验数据库** 38](#_Toc513984173)

[**4.1.2 图像特征概述** 38](#_Toc513984174)

[4.2 实验过程的实现 39](#_Toc513984175)

[**4.2.1 实验方法的选择** 39](#_Toc513984176)

[**4.2.2 实验流程** 39](#_Toc513984177)

[4.3 训练过程的实现 39](#_Toc513984178)

[**4.3.1 背景减** 40](#_Toc513984179)

[**4.3.2 PCA降维** 40](#_Toc513984180)

[**4.3.3 k-means聚类** 40](#_Toc513984181)

[4.4 跟踪过程的实现 41](#_Toc513984182)

[4.5　实验结果与分析 41](#_Toc513984183)

[第五章 结论与展望 43](#_Toc513984184)

[5.1 本文结论 43](#_Toc513984185)

[5.2 后续研究的展望 43](#_Toc513984186)

[**5.2.1 本文存在的不足** 43](#_Toc513984187)

[**5.2.2 后续研究工作** 43](#_Toc513984188)

[致谢 35](#_Toc513984189)

[参考文献 37](#_Toc513984190)

第一章 引言

**1.1课题背景**

人体运动的视觉分析是近年来计算机视觉领域中最活跃的研究主题之一，它对图像或视频进行分析，从而获得人体运动参数，并进一步进行姿态识别、语义分析及行为理解。通常涉及到运动检测、目标分类、人体跟踪及行为理解与描述几个过程[1]，见图1.1。人体运动分析的研究内容相当丰富，主要涉及到模式识别、图像处理、计算机视觉、人工智能等领域，在高级人机交互、安全监控、视频会议、医疗诊断及基于内容的图像存储与检索等方面具有广泛的应用前景和潜在的经济价值。同时，动态场景中运动的快速分割、人体的非刚性运动、人体自遮挡和目标之间互遮挡的处理等是今后研究的难点。







运动检测

目标分类

人体跟踪

行为理解和描述

图1.1 人体运动分析的一般处理框架

人体运动分析涉及的关键技术之一就是人体运动跟踪，是指采用立体视觉来

**1.2 典型应用**

三维人体跟踪具有广泛的应用前景，主要应用在智能监控(smart surveillance)、人机交互(HCI)、运动分析(motion analysis)、虚拟现实(virtual reality, VR)，基于模型的编码(modal-based coding)等领域。

（1）智能监控(smart surveillance)

智能监控系统的需求主要来自那些对安全要求敏感的场合，如银行、商店、停车场等。我们需要的监控系统应能够每天连续24小时的实时监视，并自动分析摄像机捕捉的图像数据，发现异常准确报告，从而避免犯罪发生，减少人力、物力等资源投入。在访问控制(access control)领域，也可根据三维人体运动的跟踪识别进行控制。

（2）人机交互(HCI)

传统方式下，计算机通过多种设定的程序与人交互，需要人去主动适应计算机。在新的交互模式中，计算机则能根据人的动作来主动适应人的要求。人们希望计算机能与人更加容易和便捷地交流，使人机交互自然和谐，这就要求未来的计算机必须具备感知外部环境的能力，进行人的识别和行为理解，结合面部表情、身体姿势和手势等的分析来与人进行相应的交流。

（3）运动分析(motion analysis)

分割图像中的人体部分并在图像序列中跟踪分析感兴趣的关节运动，对于建立人体的几何模型、解释人体的运动行为机制从而提高它的运动性能有着积极的推动作用，这可以应用于体育运动等训练中；医学步态分析可以提供人体正常步态建模的线索，分析病人的步态，从而做出有效的治疗；步态分析也可以作为独特的生物特征用于远距离的身份识别。

（4）虚拟现实(virtual reality，VR)

目前许多电脑游戏中逼真的人物形体，运动和行为得益于人体运动分析，包括人体跟踪和姿势恢复。为了在虚拟空间中产生生动逼真的人物动作，最好的办法就是先用视觉方法对物理世界中的人体运动进行分析，获取人体模型、人体姿态以及运动参数，再利用这些数据用图形学的方法生成动画。三维人体跟踪在电

* 1. **研究现状**

人体运动的视觉分析是计算机视觉中一个重要研究方向，也是一个很广泛的研究课题。它不仅在虚拟现实、智能监控、人机交互、运动分析、基于模型的编码等方面具有广泛的应用前景和潜在的经济价值，在计算机视觉理论研究的层次上也有很高的价值，而且对于人类认识自身有着重要的心理学和哲学上的意义。它涉及到计算机视觉中的很多基本问题，例如运动检测和分割、目标定位和识别、刚体和非刚体跟踪、时空推理、场景恢复与表示、行为分析和建模、语义理解、多摄像机数据融合，机器学习等。人体运动的视觉分析是具有挑战性的研究工作，它的很多研究成果都可以被其它领域借鉴，从而更广泛地对计算机视觉的整体研究工作产生推动作用。

因此，世界许多国家的政府、广大研究工作者以及商家对这个研究领域有着浓厚的兴趣，并争相展开大量的研究项目，美国、德国、日本的多所大学和科研机构长期获得各方面的资助，开展对人体运动分析的研究工作。在全世界范围内，尤其是美国和欧共体，已经开展了很多包含人体运动分析和动作识别关键技术的项目研究。1997 年，麻省理工大学的Wren等人实现了一个鲁棒的人体跟踪实时系统Pfinder；同年，美国国防高级研究项目署(Defense Advanced Research Projects Agency)设立了以卡内基梅隆大学(Carnegie Mellon University)为首、麻省理工学院(Massachusetts Institute of Technology)等高校参与的视觉监控重大项目VSAM(Visual Surveillance and Monitoring)[5]，主要研究用于战场及普通民用场景进行监控的自动视频理解技术；1998年起，Haritaoglu等人[6]研究开发了以”W4 :Who? When? Where? What?”为目的，能够对多人进行检测、跟踪和行为监控的实时系统W4；2002年3月到2005年3月，欧共体组织了一个大型的由近十个大学和研究机构参与的M4项目。另外，重要的学术会议如International Conference on Computer Vision (ICCV)、IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR)、European Conference on Computer Vision(ECCV)、IEEE International Workshop on Visual Surveillance (IWVS)等将人体运动分析研究作为主题内容之一，可见国际上对该项研究的高度重视。

**1.4 主要跟踪方法概述**

跟踪等价于在连续的图像帧间创建基于位置(position)、速度(velocity)、形状(shape)、纹理(texture)，色彩(color)等有关特征的对应匹配问题。按照在特征对应过程中是否采用反映先验知识的形状模型，可以把目前的人体运动的跟踪方法划分为两大类：基于模型的方法和基于学习的方法[7]。

基于模型的跟踪方法按照人体各部位的连接关系来建立人体模型，然后提取实际人体图像的特征，通过某种匹配算法，在模型与实际人体图像之间建立对应关系，最终确定人体的结构参数。这种方法需要预先对人体关节结构进行精确建模，并且需要较好的初始化。而且运动模型存在严重的准确性与适用范围之间的矛盾。在高维状态空间中搜索，从跟踪失败中自动恢复和计算复杂是模型跟踪方法面临的主要问题和挑战。基于学习的跟踪方法（也就是表象方法）是一类非模型匹配的姿态估计方法，通过对底层特征的跟踪、分析或推理获取人体姿态信息。它通过对各帧之间的位置、速度、色彩、纹理及经过各种变换后的特征参数预测和估计来建立帧之间的特征对应关系，通过建立图像特征空间和状态空间之间的对应关系，直接从可观察的图像特征推测出人体的状态参数。这种方法不依赖于先验知识，直接从图像序列中获得目标的运动信息并进行跟踪。其中以Marr David的视觉计算理论最具代表性。Marr认为视觉过程的主要任务是从2D图像中定量地恢复出图像所反映的场景中3D物体的形状和空间位置。基于学习的方法推理简单，能快速检测并跟踪运动目标，实现实时的跟踪，便于工程实现，目前已有大量实际系统利用该思路解决关节物体跟踪问题。同时此类方法既不需要对人体精确建模也不需要在图像上进行预先标注，因而在解决初始化和自动恢复问题上具有很大优势。所以，本文采用基于学习的跟踪框架，通过建立从图像特征空间到三维人体状态空间的映射关系来完成每帧图像的跟踪。此方法可以单独地看成一种跟踪方法，也可以用于解决基于模型跟踪方法中的自动初始化和跟踪失败恢复等问题。

**1.5 本文研究内容**

由于透视投影，从单目视频恢复三维人体参数存在着多义性问题，即从图像特征空间到状态特征空间的映射是“一对多”(one to many)，本文基于混合专家模型，采用“分而治之”的思想，把图像特征空间分为几个局部空间，每个空间对应一个专家，每个专家处理不同区域的局部线性映射。然后把多个专家的估计结果通过各自的权重组合起来，得到最终的三维人体姿态。这里要强调的是，每个专家的输出结果是三维状态向量的高斯分布，多个专家的混合结果是混合高斯分布，这更符合由多义性产生的多模态特性。本文采用的图像特征是人体剪影，使用背景减提取剪影图像，然后采用主成分分析(Principle Component Analysis, PCA)方法对剪影特征进行降维，避免了在高维状态空间的映射。应用混合专家模型算法和k-means聚类算法建立图像特征空间和状态空间之间的对应关系，从图像特征中推测出人体模型的三维角度，从而准确鲁棒地实现三维人体跟踪。

第一章讨论三维人体跟踪的背景、典型应用、研究现状和本文的研究内容。

第二章研究混合专家模型的原理及算法。

第三章研究基于背景减和主成分分析的特征提取方法。

第四章给出实验结果与分析。

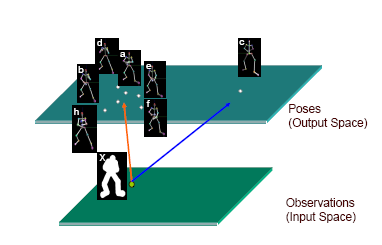
第五章总结全文的研究工作并对今后的研究工作进行展望。

第二章 混合专家模型

**2.1混合专家模型概述**

本文研究的内容是从单目视频中重建三维人体运动。由于人或物的遮挡，在现实世界中即使使用多摄像机也很难完整地观察到人体。而一个鲁棒的系统必须具有能够处理不完整的、模糊的、不确定的观察的能力。基于单目视频的3D运动跟踪对某些应用（如视觉监控及对已存档视频的运动分析）十分有用，而从单目视频恢复3D姿态，运动模型具有重要意义。因此，本文基于单目视频来进行研究。

从单目视频中重建三维人体运动要面对一个较为困难的问题[8]：从图像特征空间到状态特征空间(observation-to-state)的多义性问题。图2.1形象地解释了这个问题，可以看到，从图像特征空间中的一个剪影可以推测出若干个可能的人体运动。从图像特征空间到状态特征空间的映射是非线性的，多值的，这导致了从二维剪影重建三维人体运动是一对多(one to many)的关系，这使的我们无法准确地重建三维人体运动。人体模型是多元的、有着50个甚至更多人体关节角度的高维连续状态空间。这种临时的条件状态是多峰分布的，这导致了有效的卡尔曼滤波算法不再适用。混合的贝叶斯方法在这种情况下有着有效的作用。



(a)

**2.2 k-means聚类算法**

采用混合专家模型来完成对状态参数的估计，首先要将图像特征空间中的数据进行分类，即把他们分为若干个专家。为此，需要先选取若干个中心，然后将这些数据围绕中心进行聚类。

针对在不同领域解决不同的问题，人们已经提出很多聚类算法。常用的聚类算法包括k-means聚类、FCM聚类、密度函数聚类、分层聚类、计算机图论聚类、搜索算法聚类、神经网络聚类和核聚类等。

k-means聚类算法是聚类技术中非常著名的一个硬聚类算法，它算法简单，聚类速度快。与其它聚类算法一样，k-means聚类算法也是一个迭代寻优的过程。假设将n维样本集数据Z=(z1,z2,…zn)分成k类，首先选择一个k类的初始划分，计算这些类的均值向量，然后根据欧氏距离：

|  |  |
| --- | --- |
|  | 式(2-1) |

把剩余的样本分配到距离它最近的一个类别。

**2.2.1 k-means聚类算法工作原理**

k-means聚类算法接受聚类个数k（专家个数），然后将n个数据对象划分为k个聚类以便使得所获得的聚类满足：同一聚类中的对象相似度较高；而不同聚类中的对象相似度较小。

本文实验中采用三个专家对数据进行聚类，分别以圆形、星形、方形三种形状显示，见图2.3。可以看到k-means聚类算法准确地对数据分类，达到满意的效果。

**2.3 BME(Bayesian Mixture of Experts)算法**

**2.3.1 BME算法原理**

如图2.2所示，本文采用BME完成从提取特征后的图像特征数据z到状态特征数据y的映射，并从中推测出状态参数的估计。

BME算法的输入：，共n个训练数据；

BME算法的输出：由k个专家的输出结果组合而成的状态参数估计。

输出的状态参数用来重建三维人体运动，本文使用一个建立在图像特征数据z到状态特征数据y的映射上的映射函数，为了得到精确的状态参数估计，本文要求所有图像特征数据到相应状态特征数据的映射函数值最大，即最大。本文使用了正态分布和极大似然估计的方法以得到状态参数。

**2.3.2 BME算法证明**

本文采用k-means聚类算法将图像特征数据分成k类，即k个专家。分别用每个专家来建立从图像特征空间z到状态特征空间y的映射函数。前面提到，本文要求最大。反映到每个专家上，就要求每个专家上的图像特征数据映射到相应状态特征数据的映射函数值最大。即最大，其中j表示第j个专家。定义为：

|  |  |
| --- | --- |
|  | 式(2-2) |

其中表示第j个专家下由z推出y的概率，表示z中任何一个元素属于第j个专家的概率。

图像特征数据z和状态特征数据y满足正态分布：

|  |  |
| --- | --- |
|  | 式(2-3) |
|  | 式(2-4) |

在式(2-2)中，由k-means聚类算法直接得到，

|  |  |
| --- | --- |
|  | 式(2-5) |

那么所有图像特征数据到状态特征数据的映射函数可以表示为：

第三章 特征提取

**3.1 背景减(background subtraction)**

**3.1.1 运动检测**

运动检测是人体运动分析的第一步，是三维人体跟踪的基础。运动检测的目的是从序列图像中将变化区域从背景图像中提取出来。运动区域的有效分割对于目标分类、跟踪和行为理解等后期处理非常重要，因为以后的处理过程仅仅考虑图像中对应于运动区域的像素。目前，主要有三种常用的方法：背景减法、时间差分法和光流法。

（1）背景减法(background subtraction)是目前运动分割中最常用的一种方法，它是利用当前图像与背景图像的差分来检测运动区域的一种技术。它一般能够提供最完全的特征数据，而且速度快，能适合实时系统的要求，但对于动态场景的变化，如光照和外来无关事件的干扰等特别敏感。

（2）时间差分法(temporal difference)是在连续的图像序列中相邻帧间采用基于像素的时间差分并且阈值化来提取图像中的运动区域。时间差分方法对于动态环境具有较强的自适应性，但是一般不能完全提取出所有相关的特征像素点，在运动实体内部容易产生“空洞”现象。

（3）基于光流法的运动检测采用了运动目标随时间变化的光流场特性，提取和跟踪运动目标。该方法的优点是在摄像机运动存在的前提下也能检测出独立的运动目标。而缺点是计算方法相当复杂，抗噪性能差，如果没有特别的硬件装置不能被应用于全帧视频流的实时处理。

基于背景减法能够提供最完全的特征数据和计算快速性等优点，虽然它对环境的变化较敏感，但由于本文是对室内运动物体的跟踪，受光照等干扰较小，所以采用背景减法来进行跟踪。

**3.1.2 背景建模**

背景建模，即通过视频序列的帧间信息估计和恢复背景。对于背景的建模问题，常用的方法主要有：基于统计的背景模型、基于卡尔曼滤波的背景模型、基

**3.2主成分分析 (Principal Component Analysis，PCA)**

广义的特征提取包括特征生成、特征选择和特征变换（维数削减）[9]，在很多时候，特征提取就是指特征变换或维数削减。

在模式识别领域，对于一个对象x，人们经常使用一个向量来标志这个对象的各个属性，如，在实际使用时其维数往往很高。我们在做三维人体跟踪时，会使用视频序列中每一幅图像的所有像素点。每个人体图像被编码为一个人体图像中所有像素点连接而成的列向量，其维数等于图像所包含的像素数。一幅91×91大小的人体图像的维数即为8281，这是一个高维特征空间。而人体模型是有着50个甚至更多人体关节角度的高维连续状态空间。三维人体在图像特征空间和状态参数空间的特征表达都是高维的，使得直接在这两个空间上建立非线性的映射关系是非常复杂的，并且代价很高。为了克服这个问题，一个很自然的想法就是，将高维、非结构化的人体图像数据可靠地转化到低维空间中[10]，并在低维空间中进行特征选择。实际上这些高维数据之间存在着大量的冗余，如何消除数据间的冗余，寻找数据间的内在联系，将数据映射到一个低维空间中去，以利于后续处理，是数据降维算法所要解决的课题。

**3.2.1 降维算法概述**

降维算法是一个古老而又年轻的研究课题。经典的降维算法多数在寻找某个线性子空间，将高维数据投影到这个子空间中。这里，我们首先给出降维算法的数学描述：给定n个M维的数据点：，给出映射或者给出弱一些的结果。

**3.2.2 PCA算法原理**

早在1946年，Karhunen和Loeve就提出了K-L变换，指出了K-L变换是在线性条件下的最优变换。这就是PCA(Principle Component Analysis)方法的前身。PCA方法具有计算简单，去除噪声，彻底消除相关的优点，在各个相关领域中都有广泛而成功的应用，成为降维算法中应用最为广泛的算法。

在高维数据分析中，维数间存在共线性，增加了分析的复杂性。但是若分别

第四章 三维人体跟踪的实验

本文研究的是基于单视频的三维人体跟踪。分三步顺序完成：首先从图像帧中提取低层特征，并对之进行识别；然后是在各个帧的特征间建立对应关系；最后是从特征对应中恢复三维人体结构与运动信息。

**4.1 实验概述**

**4.1.1 实验数据库**

本文实验采用由美国卡内基梅隆大学提供的Mocap数据库，同时捕捉了人体运动的三维状态数据（关节角度向量）和与之对应的二维图像。本文选择了其中“Golf”运动数据中的“pick up”运动（见图4.1）共约270帧真彩色图像，其中80帧作为训练数据，190帧为测试数据。为降低计算时间，将图像大小调整为91×91。



图4.1 实验数据“pick up”运动

**4.1.2 图像特征概述**

特征提取是从图像中提取出边界、区域、轮廓，光流等底层特征，将前景与背景分离的过程。目前常用特征及提取方法[11]有：边缘。用边缘作匹配能得到较精确的位置。提取边缘不要求背景固定，可用简单边缘提取算子，但易受背景和人体本身纹理的干扰；剪影。即图像中的人体区域，用来与模型投影区域匹配。与边缘相比，剪影不易受噪声影响，但易丢失细节信息，一般可通过减背景、去除阴影、形态学滤波等操作得到；外轮廓。即剪影的边界，也是一种边缘特征。它可通过主动轮廓变形得到，在能量函数中可融入多种测度，因此比边缘方法鲁棒，但计算量大。

**4.2 实验过程的实现**

**4.2.1 实验方法的选择**

本文实验过程可以分为特征提取和混合专家模型两个部分，特征提取包括运动检测和维数削减。

基于背景减法能够提供最完全的特征数据和计算快速等优点，虽然它对环境的变化较敏感，但由于本文是对室内运动人体的跟踪，场景没有太大变化，因此采用背景减法提取图像剪影特征。

由于直接在高维图像特征空间和高维状态特征空间上建立非线性的映射关系是非常复杂的，因此，本文应用PCA方法将高维空间中的人体图像降维到一个低维子空间中，以利于后续处理。

图像特征空间到状态特征空间的映射存在着多义性问题，因此，本文基于混合专家模型，采用“分而治之”的思想，建立从图像特征空间到状态特征空间的映射，生成了状态参数非高斯分布的组合估计结果，准确地重建了三维人体运动。

**4.2.2 实验流程**

本文从单视频中读取图像序列，使用背景减提取特征，采用PCA方法对图像特征空间进行降维，应用BME算法建立图像特征空间和状态特征空间之间的对应关系，从图像特征中推测出人体的状态参数，并通过这些参数恢复人体三维角度，从而准确鲁棒地实现三维人体跟踪，如图4.4所示。

**4.3 训练过程的实现**

训练过程有三个模块：背景减，PCA降维，BME。背景减训练的是背景图像的像素高斯模型参数，PCA训练的是高维图像特征的降维矩阵，BME训练的是用于完成从图像特征空间到状态特征空间映射的模型参数。我们使用matlab2007b完成训练过程，具体实现步骤如下：

**4.3.1 背景减**

本文选取30幅背景图像、含有85幅图像的视频文件进行训练，背景有微小变化。

输入：包含运动目标的视频文件，以及相应的背景图像序列。

输出：背景图像的像素高斯模型参数。

步骤1：读取背景图像序列；

步骤2：获取背景图像序列的R,G,B空间三维颜色值；

步骤3：对R,G,B空间三维颜色值进行单高斯建模，生成高斯参数；

步骤4：读取包含运动目标的视频文件；

对于视频文件中的每一帧：

for i=1:nsmp

（1）获取每一帧的R,G,B空间三维颜色值；

（2） 基于概率统计分布将每一个像素的三维颜色值与高斯模型参数比较，减除背景。

**4.3.2 PCA降维**

实验中，图像是91×91的，那么反映到图像特征向量上就是8281维的，本文期望降维后的维数d=5。

输入：所有提取特征后的图像序列的数据，期望降维后的维数d。

输出：降维矩阵，低维图像特征向量。

步骤1：生成初始特征向量矩阵，特征值；

步骤2：对特征值降序排列，并取前d个最大特征值对应的特征向量；

步骤3：特征归一化，得到最终特征向量矩阵，也就是降维矩阵，使用和得到低维图像特征向量。

**4.3.3 k-means聚类**

输入：聚类个数k，以及包含n个数据对象的数据库。

**4.4 跟踪过程的实现**

我们使用VC++6.0，OpenCV开发平台实现跟踪过程。

输入：包含运动目标的视频文件、matlab中背景减过程训练的背景图像的像素高斯模型参数、PCA降维矩阵以及BME过程训练的模型参数。

输出：人体状态参数。

步骤1：读取包含运动目标的视频文件；

对于视频文件中的每一帧图像：

for i=1:nsmp

（1）获取图像中R,G,B空间三维颜色值；

（2）基于概率统计分布将每一个像素的三维颜色值与高斯模型参数比较，提取前景，生成图像特征向量；

（3）降维矩阵与图像特征向量相乘得到低维图像特征；

（4）依照式(4-1)，由低维图像特征向量和模型参数得到人体状态参数：其中，而求解见式(2-4)-式(2-8)。

|  |  |
| --- | --- |
|  | 式(4-1) |

**4.5　实验结果与分析**

本文采用RMS角度误差[12]来计算跟踪误差

|  |  |
| --- | --- |
|  | 式(4-2) |

其中，是估计的关节角度向量， 是真实角度向量。图4.5显示了在整个测试集上的跟踪结果。



图4.5 总体跟踪误差图

本文选取人体的整个运动中动作变化较大的关节进行局部跟踪误差测试，在实验中，由于本文选取的是人体的“pick up”运动，选择图4.2中的关节1：head来进行测试。本文显示关节绕x、y、z三方向旋转角度的比较结果，测试结果如图4.6所示。可以看出，本文准确鲁棒地完成了三维人体跟踪。



(a)

第五章 结论与展望

**5.1 本文结论**

从单视频中重建三维人体运动的主要难点在于从图像特征空间到状态特征空间映射的多义性。针对这个问题，本文基于混合专家模型建立图像特征空间和状态特征空间之间的对应关系，从单视频图像序列中跟踪三维人体，主要研究工作如下：

（1）采用单高斯建模的方法对图像序列进行特征提取，以便方便准确地进行后续处理；

（2）采用主成分分析法将图像数据的高维特征空间可靠地转化到低维空间；

（3）采用BME方法来完成从低维空间到高维状态空间的映射，训练出恢复人体三维角度所需的状态参数；

（4）对训练的状态参数测试，准确地重建三维人体运动。

**5.2 后续研究的展望**

**5.2.1 本文存在的不足**

由于时间有限，本文在三维人体跟踪的过程中还存在以下不足：

（1）虽然本文研究的是室内人体运动，背景没有太大变化，但仍有可能会有光照等影响，在背景减上可以采用自适应的方法来自动适应背景，更准确地提取图像特征；

（2）在单摄像机系统中，光照变化等因素将会影响背景减效果。由于仅用颜色信息进行背景减不稳定，因此我们可以在以后的工作中结合深度和颜色信息共同进行背景减，从而获得更可靠的人体区域。

**5.2.2 后续研究工作**

三维人体跟踪经过20多年的研究，已经取得了很大的进展，但还有很多难题有待解决。在后续研究中，除上述不足外，本文将在目前已有工作的基础上继续解决人体跟踪的这些问题：

（1）不加约束的人与背景的自动分离问题。在动态环境中，因受摄像机运动、光照变化、杂乱背景及人与背景颜色相近等影响，该问题一直未得到很好解决；

（2）遮挡问题。在遮挡情况下，特别是当上肢与躯干服饰的颜色纹理相近时，很难从图像中将它们区分开来。

致谢

参考文献

[1] L. Wang, W. Hu and T. Tan．Recent Developments of Human Motion Analysis．

Pattern Recognition．2003．page 2-4．

[2] Gavrila D,Davis L．3-D model based tracking of humans in action: A multiview approach．IEEE Proceedings of International Conference on Computer Vision and Pattern Recognition．1996．page 73-80．

[3] Wren C,Pentl and A．Dynaman: A recursive model of human motion．

Cambridge ,MA:MIT Media Lab．1998．page 1-2．

[4] Plaenkers R,Fua P．Articulated soft objects for video-based body modeling．IEEE Proceedings of International Conference on Computer Vision．2001．page 394-401．

[5] T.Kanade, R.T.Collins and A.J. Lipton．Cooperative Multi-Sensor Video Surveillance．DARPA Image Understanding Workshop．Nov,1998．page 3-24．

[6] Ismail Haritaoglu , David Harwood and Larry S.Davis．W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People．International Conference on Face and Gesture Recognition．April,1998．page 14-16．

[7] Xinxiao Wu, Wei Liang, Yunde Jia．Tracking Articulated Object by Learning Intrinsic Structure of Motion．2007．page 1-3．

[8] C. Sminchisescu, A. Kanaujia, Z. Li．Conditional models for human motion recognition．Technical Report CSRG-517．March,2005．page 2．

[9] 刘成林，谭铁牛．模式识别研究进展．模式识别学术会议．2007，第3页．

[10] 许明旺，施润身．维规约技术综述．计算机应用．2006年，10月，第26卷，第2页．

[11] 李豪杰，林守勋．基于视频的人体运动捕捉综述．计算机辅助设计与图形学学报．2006年，11月，第18卷，第5页．

[12] Z. Chen, H. J. Lee．Knowledge-guided visual perception of 3D human gait from a single image sequence．IEEE Trans. On Systems, Man, and Cybernetics．Feb 1992．page 336-342．