基于少量关键序列帧的人体姿态识别方法

蔡兴泉， 涂宇欣， 余雨婕， 高宇峰

(北方工业大学信息学院，北京 100144)

摘　　要：针对传统人体姿态识别数据采集易受环境干扰、难以解决人体运动姿态的相似性和人体运动执行者的特征差异性等问题，提出一种基于少量关键序列帧的人体姿态识别方法。首先对原有运动序列进行预选，通过运动轨迹取极值的方法构造初选关键帧序列，再利用帧消减算法获取最终关键帧序列；然后对不同人体姿态分别建立隐马尔科夫模型，利用 Baum-Welch 算法计算得到初始概率矩阵、混淆矩阵、状态转移矩阵，获得训练后模型；最后输入待测数据，应用前向算法，得到对于每个模型的概率，比较并选取最大概率对应的姿态作为识别结果。实验结果表明，该方法能够有效的选取原始运动序列的关键帧，提高人体姿态识别的准确性。

关 键 词：人体姿态识别；序列帧；帧消减；隐马尔科夫模型

中图分类号：TP 391　　**DOI：**10.11996/JG.j.2095-302X.2019030532

文献标识码：A　　**文 章 编 号：**2095-302X(2019)03-0532-07

Human Posture Recognition Method Based on Few Key Frames Sequence

CAI Xing-quan, TU Yu-xin, YU Yu-jie, GAO Yu-feng

(School of Information Science and Technology, North China University of Technology, Beijing 100144, China)

Abstract: This study focuses on the problems that the traditional human posture recognition data acquisition is easily disturbed by environment, and it's difficult to solve the similarity of human motion postures and the characteristics difference of the human motion executor. This paper proposes a human posture recognition method based on few key frames sequence. Firstly, the original motion sequence is pre-selected. The initial key frame sequence is constructed by taking the extremum of the motion trajectories, and the final key frames sequence is obtained by using frame subtraction algorithm. Then, we built the hidden Markov model for different human postures and trained the model. The Baum-Welch algorithm is used to calculate the initial probability matrix, the confusion matrix and the state transition matrix, and the post-training model is obtained. Finally, the probabilities for each model are achieved by inputting the measured data and applying forward algorithm, and the gestures corresponding to the maximum probability are compared and selected as what is identified. Experiment results show that our method can efficiently select key frames of the original motion sequence, and effectively improve the accuracy of human body gesture recognition.

Keywords: human posture recognition; frames sequence; frame subtraction; hidden Markov model

人体姿态识别是计算机视觉领域的重要研究难题，在智能监控、智能家居、人机交互、服务机器人等领域具有广阔的应用前景。随着社会的快速发展，智能服务机器人将会走进家庭，照顾人们的日常起居。面向智能服务机器人的人体姿态识别技术是当前急需解决的难题。

收稿日期：2018-11-12；**定稿日期：**2018-12-27

基金项目：国家自然科学基金项目(61503005)；北京市自然科学基金项目(4162022)；北方工业大学长城学者培养项目(NCUTCC08)

第一作者：蔡兴泉(1980–)，男，山东济南人，教授，博士，硕士生导师。主要研究方向为虚拟现实、人机互动。E-mail：xingquancai@126.com

当前，关于人体姿态识别的数据采集易受环境干扰，采集的数据掺杂噪声；人体运动姿态的相似性，比如打电话和吃东西、跑和跳等的过程相近；人体运动执行者的特征有差异，比如身体高矮胖瘦、动作快慢等。针对这些问题，本文主要研究一种基于少量关键序列帧的人体姿态识别方法，通过红外深度相机采集人体姿态序列帧数据，实现帧消减减少数据量，最终实现人体运动的各种姿态识别。

1　相关工作

对人体姿态识别技术运动序列进行特征提取及分类，按照不同的匹配方法可分为模板匹配法和概率网络法[1]。模板匹配法是通过将待测序列与已定义的动作模板进行匹配，完成人体姿态识别。此类方法普遍具有实现简单、运算复杂度低的特点，但存在对噪声和运动时间间隔变化敏感、抗干扰能力差、准确性不足等问题[2-4]。概率网络法是先定义人体的静态姿势，每个静态姿势为一个状态。运动序列就是在该类状态间进行一次遍历，相似动作的运动序列通常具有相似的转移概率[5]。进行人体运动识别时，通过计算各个遍历的联合概率，对相似的行为序列进行分类。该类方法通常计算复杂度较高，但准确性高、鲁棒性强，也使其成为目前研究的主流。

文献[6]提出了一种基于帧间距的关键帧提取的方法用于实现对运动序列关键帧的压缩。该方法通过提取代表运动数据内容的关键姿势作为关键帧，然后消除运动序列中差异较大的帧，最后采用四元数球面插值进行关键帧重构。该方法在提取人体边界姿势时准确率较低。文献[7]提出了一种基于部位检测的人体姿态识别，使用随机森林法将较小的人体部位合并，用均值偏移算法获取各部位关节点的位置。该方法可提高图像中的人体姿势的准确率，但无法进行实时的姿态识别。文献[8]提出了改进的基于支持向量机(support vector machine, SVM)的方法用于人体姿态识别。首先从运动序列中提取多种特征组成矩阵，之后利用奇异值分解等方法对运动采集数据进行特征降维，最后对数据降维后的子空间数据利用支持向量机进行特征分析，达到姿态识别的目的。该方法在数据规模较大时训练效果欠佳。文献[9]提出了基于三维人体关节特征和特定关节的隐马尔科夫模型的方法用于人体运动识别。首先通过对所有单个运动建立隐马尔科夫模型，然后通过对关节运动信息和时序判断运动过程中关节点的自由度信息，确定不同的运动特征。最后，针对不同活动中特定关节点的三维信息利用隐马尔科夫模型进行训练和匹配。该方法使用时需要较高的先验知识，存在需要预设隐状态数量以及维度灾难等问题。

随着深度学习技术的不断发展，基于深度学习的相关算法也广泛应用于人体姿态识别技术[10-12]。文献[11]提出了一种基于深度神经网络的人体动作识别方法。采用三维建模的方法建立人体模型，分层次地识别人体各部位的行为，从底层行为组合成为高层的动作和活动，应用隐马尔科夫模型完成人体行为的学习工作。该方法可以有效识别人体复杂姿态，但需要通过不断移动相机实现多角度的人体图像获取，易存在捕捉数据不够精准的问题。文献[12]提出了采用蒙版关节轨迹在深度视频序列的动作识别方法。本文也是受此启发，对视频序列进行处理，实现人体姿态识别。

为了更好地实现使用少量关键帧序列对人体姿态进行准确识别，本文提出了一种基于少量关键序列帧的人体姿态识别方法。

2　基于少量关键序列帧的人体姿态识别算法

为了实现人体姿态识别，本文首先定义人体运动姿态数据格式，然后基于深度相机获取人体姿态序列帧数据，接着基于预选策略实现帧消减，最后实现基于隐马尔科夫模型的人体姿态识别。

2.1　定义人体运动姿态数据格式

进行人体姿态识别前，需要保存序列帧的数据。针对不同数据源命名格式和内容格式的不同的问题，需要对序列帧的数据进行格式定义。数据集对 20 个骨骼点进行采集，分别对骨骼点数量、坐标方式和计量单位等重要参数进行定义，并构建相应数据库。在构建人体运动姿态数据库时，采用的数据存储格式为：

(1) 每 20 行数据存储为一帧数据，其中每一行代表一个骨骼点的数据；

(2) 每行包含 4 列数据，依次为空间坐标系下的 x 轴、y 轴、z 轴的坐标和置信度分数。其中置信度分数用来对数据中可能存在的遮挡等情况进行区分。数据集中已标注置信度分数的数据直接引用，未标注的部分采用手动标注和机器标注的方式共同进行。如此，定义人体运动姿态数据格式，构建相应的数据库，为后续训练和识别准备。

2.2 基于深度相机获取人体姿态序列帧数据

采用深度相机获取人体运动姿态序列帧数据，具体分以下 5 步完成：

步骤 1. 使用深度相机获取深度图像，由深度图像获取像素点的深度距离和用户索引值；

步骤 2. 根据深度相机有效距离设定深度阈值范围，保留深度值为 1.2~3.5 m 的点，舍去其余点；

步骤 3. 将深度图像进行二值化处理；

步骤 4. 采用人体目标识别方法，提取人体运动数据，进行人体像素点的判定；

步骤 5. 检测识别结果，并对识别结果进行膨胀腐蚀等预处理操作。

2.3 基于预选策略的帧消减

关键帧的选取是姿态识别中的关键步骤，主要对目标识别获取的实时数据进行处理，关键帧技术通过对原有动作序列的处理，可以有效减少姿态序列的数据量和后续计算，使用少数关键帧代表完整的原始动作序列，达到数据降维、减小数据量目的。针对关键帧选取时无法兼顾压缩比和重建误差、可移植性差等问题，本文提出了一种基于预选策略的帧消减算法。该方法通过对原始三维骨骼数据进行计算，分 2 步完成：

步骤 1. 对原有运动序列进行初选，通过运动轨迹取极值的方法构造初选关键帧序列。将人体各关节点的运动轨迹视作空间中的连续曲线，通过选取曲线上极值点的所在帧作为关键帧，构造预选关键帧序列。

步骤 2. 利用帧消减算法获取满足不同用户需求的最终关键帧序列。通过将关键帧选取过程中的 2 个优化目标分别设置为最大重建误差和关键帧数量，达到兼顾重建误差和压缩比的目的。利用帧削减算法逐帧删除重建误差小的预选关键帧，最终得到满足指定重建误差和压缩比要求的关键帧序列。

2.3.1 运动插值重建和重建误差

使用运动插值重建方法对关键帧进行插值以及重建原始运动序列。设 t_1 和 t_2 时刻关键帧序列中的对应关键帧分别为 $f(t_1)$和 $f(t_2)$，则 t 时刻的关键帧为

$$f(t)=\frac{t_2-t}{t_2-t_1}f(t_1)+\frac{t-t_1}{t_2-t_1}f(t_2) \tag{1}$$

本文选择使用各关节骨骼点位置信息来计算重建运动序列和原始运动序列之间存在的具体误差。设 p_i^f 为第 f 帧中骨骼点 i 在空间中的位置信息；f_1 和 f_2 分别为 2 个关键帧的运动数据。定义两帧之间所对应关节骨骼点间的最大距离为帧间距离 $d(f_1, f_2)$，即

$$d(f_1,f_2)=\max_{i\in skeleton}\left\{\left\|p_i^{f_1}-p_i^{f_2}\right\|^2\right\} \tag{2}$$

其中，$\|..\|^2$ 为对关节点坐标向量求取的欧拉距离。

对长度为 n 帧的运动序列 X，首先设置首帧和尾帧为默认关键帧。应用式(1)进行插值重建操作，并得到运动序列 X'。设 f_i 和 f'_i 分别为运动序列 X 和 X'中的第 i 帧。根据式(2)，运动序列 X 和 X'的位置误差为

$$e(X,X')=\max_{i=1,2,\cdots,n}\left\{d(f_i,f_i')\right\} \tag{3}$$

定义 $e(f_1, f_n)$为运动序列 X 的首尾帧重建运动的重建误差，即

$$e(f_1,f_n)=e(X,X')=\max_{i=1,2,\cdots,n}\left\{d(f_i,f_i')\right\} \tag{4}$$

设运动序列 $X=\{x_1,x_2,\cdots,x_n\}$的关键帧序列为

$$K(X)=\{f_1,\cdots,f_i,f_j,f_k,\cdots,f_n\} \tag{5}$$

其中，f_1 和 f_n 分别为运动序列 X 的首尾帧，f_i, f_j, f_k 为关键帧序列 $K(X)$中相邻 3 帧，$1\leqslant i<j<k\leqslant n$。

设运动序列 X'是对关键帧序列 $K(X)$的相邻关键帧按式(1)插值重建得到的运动序列，其中 $X'=\{x_1',x_2',\cdots,x_n'\}$。根据关键帧序列 $K(X)$重建原始动作序列的重建误差为

$$e\left(K(X)\right)=\max_{i=1,2,\cdots,n}\left\{d\left(x_i,x_i'\right)\right\} \tag{6}$$

$$e(f_j)=e(X(i:k),X'')=\max_{i=i,i+1,\cdots,k}\left\{d\left(f_n,f''_{n-i+1}\right)\right\} \tag{7}$$

对于运动序列的关键帧序列，假设将关键帧 f_j 删除，可以利用式(1)进行重建操作得到新的运动序列 X''，定义具体某一帧 f_j 的帧消减误差如式(7)所示，表示从当前的关键帧序列 $K(X)$中删除具体关键帧 f_j 后得到的重建运动序列与原始运动序列间的最大间距。由于消去关键帧 f_j，其前后相邻帧 f_i 和 f_k 的帧消减误差需要重新计算，导致相邻关键帧的重建误差随之改变。

2.3.2 基于帧消减算法的关键帧选取

在通过预选阶段获得预选关键帧序列后，对该关键帧序列应用帧消减算法以得到最终关键帧序列。帧消减算法通过每次对当前关键帧序列进行比较，删除最小消减误差的关键帧，直到满足预设的关键帧数量或最大重建误差时停止，输出符合需求的最终关键帧序列。具体步骤为：

步骤 1. 输入预选关键帧序列，以及关键帧数量或最大重建误差。

步骤 2. 对预选关键帧序列进行判定，若当前关键帧序列的关键帧数量小于等于预设或现有关键帧最大重建误差小于等于预设时，将当前关键帧序列作为最终关键帧序列输出。否则删除当前关键帧序列中最大重建误差的帧，并重新计算被删除关键帧前后关键帧的重建误差。其中预选关键帧序列首尾帧始终默认为关键帧。

2.4　基于隐马尔科夫模型的人体姿态识别

传统人体姿态识别一般采用手工标注，监督学习的方法，存在效率低、成本高等问题。目前，采用机器学习方法对人体姿态进行智能分析，此类方法效率高、成本低，但仍存在鲁棒性、准确性等方面需要提升的问题。针对目前现有人体姿态识别方法，本文提出一种针对人体运动特点的隐马尔科夫模型的方法进行人体姿态识别。

隐马尔科夫模型是动态模型，若在一个随机过程中某状态只与此状态之前的 n 个状态有关，而与其余所有状态无关，则称这个随机过程为 n 阶马尔科夫模型。一阶马尔科夫模型定义如下：当 n=1 时，对于 t+1 时刻，其状态仅与其前一个状态，即 t 时刻的状态有关。其数学表达式为

$$X(t+1)=f(X(t)) \tag{8}$$

$$\lambda=(\pi,M,N,A,B) \tag{9}$$

$$\pi_i=P(q_1=\theta_i)，\ 1\leqslant i\leqslant N \tag{10}$$

其中，X=(X_1,X_2,⋯,X_n)为随机序列；$f(X(t))$为 t 时刻的状态。通常来讲，隐马尔科夫模型可以具体数学表示为一个五元组。一个标准的隐马尔科夫模型数学表示如式(9)所示。

$\boldsymbol{\pi}$ 为初始概率矩阵 $\boldsymbol{\pi}$=$\{\pi_i\}$=(π_1,⋯,π_N)，表示在隐马尔科夫链开始时状态发生概率。其中 π_i 应满足公式(10)所示。π_i 为在一个随机过程中，首状态取状态 i 的概率。所有的 π_i 共同构成一个 1×N 的初始概率矩阵 $\boldsymbol{\pi}$。

N 为隐马尔科夫模型 λ 中隐状态的数量。记 N 个状态为 θ_1,⋯,θ_N，记 t 时刻马尔科夫链所处隐状态为 q_t。q_t 必定属于 N 个隐状态之一，即有 $q_t\in(\theta_1,\cdots,\theta_N)$。

M 为隐马尔科夫模型 λ 中可能的显状态的数量。M 个显状态分别记为 V_1,⋯,V_M，t 时刻可以观察到的显状态记为 O_t，其中 $O_t\in(V_1,\cdots,V_M)$。

$\boldsymbol{A}$ 为隐马尔科夫模型 λ 中各状态之间的转移概率矩阵，$\boldsymbol{A}$=(a_{ij})$_{N\times N}$。其中 a_{ij} 为从状态 i 转移到状态 j 的概率，即

$$a_{ij}=P(q_{t_{n+1}}=j)\big|q_{t_n}=i，\ 1\leqslant i,j\leqslant N,a_{ij}\geqslant 0 \tag{11}$$

$\boldsymbol{B}$ 为隐马尔科夫模型 λ 中的可观测概率输出矩阵，也就是显状态转换概率矩阵，$\boldsymbol{B}$=$\{b_j(k)\}$。其中 $b_j(k)$为在具体隐状态 j 下观测到相应显状态 k 的概率，即

$$b_j(k)=P[O_t=V_k\,|q_t=j]，\ 1\leqslant k\leqslant M \tag{12}$$

即不同状态观测到的不同事件能够取到的概率构成了的 N×M 的矩阵 $\boldsymbol{B}$。

因此，一个隐马尔科夫模型可以表示为或简写为 λ=(π,$\boldsymbol{A}$,$\boldsymbol{B}$)。

2.4.1　建立人体姿态的隐马尔科夫模型

隐马尔科夫模型通常用于解决解码、评估和学习问题。解码问题是已知显状态序列 O=O_1,O_2,⋯,O_T 和模型 λ=($\boldsymbol{\pi}$,$\boldsymbol{A}$,$\boldsymbol{B}$)，对特定显状态序列，求解对应最优隐状态序列的问题。即寻找能使 $P(O/)\lambda$ 达到最大值时，状态序列 O 对应的隐状态序列。目前，此类问题通常使用维特比(Viterbi)算法加以解决。

评估问题是针对一个待确定的模型与给定的观察序列，根据匹配程度选取最佳匹配的过程。已知观测序列 O=O_1,O_2,⋯,O_T 和模型 λ=($\boldsymbol{\pi}$,$\boldsymbol{A}$,$\boldsymbol{B}$)，对于某一具体未判定显状态序列，通过计算概率 $P(O/)\lambda$，对隐马尔科夫模型进行相应评估的过程。在人体姿态识别实际应用中由于针对不同行为序列分别建模，因此实际使用中通常将显状态序列 O=O_1,O_2,⋯,O_T 投入所有不同模型中，通过比较概率 $P(O/)\lambda$ 大小，找到最可能生成该显状态序列的模型，完成目标姿态识别。对于评估问题，通常采用前向后向方法加以解决。

学习问题主要存在对隐马尔科夫进行训练的阶段，已知显状态序列 O=O_1,O_2,⋯,O_T，模型参数 λ=($\boldsymbol{\pi}$,$\boldsymbol{A}$,$\boldsymbol{B}$)未知，通过调节模型参数希望使显状态序列概率 $P(O/)\lambda$ 尽可能大的过程。由于姿态识别数据量大、监督学习成本过高等问题，本文采用非监督性学习的 Baum-Welch 算法完成模型训练。人体姿态识别通过对每个动作或行为单独建模，运用输入关键帧序列计算每个模型符合概率的方法达成人体姿态识别的目的。

本文通过将人体姿态识别视作一种时变数据的分类。首先将所有人体的不同姿势定义为一种状态，然后用概率来表示各个状态之间的转化关系。可观测到的用户所做运动序列被视为显状态序列，每个动作又会组成运动序列。其中，对每一个运动序列的计算相当于遍历之前定义的各个状态。对这

个遍历的过程应用前向或后向算法计算对应的联合概率。最后把联合概率最大值对应的隐状态作为识别结果进行输出。具体流程如图 1 所示。

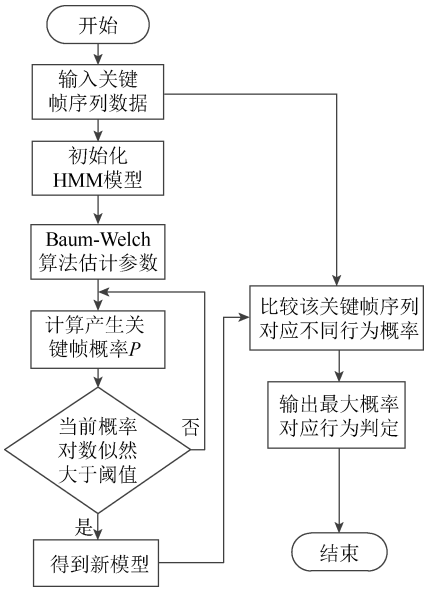

图 1 人体姿态识别的隐马尔科夫模型训练和识别流程

2.4.2 训练人体姿态的隐马尔科夫模型参数

本文算法进行人体姿态识别，使用 Baum-Welch 算法解决对隐马尔科夫模型的训练。Baum-Welch 算法是一种特殊的最大期望算法。其基本原理是通过找到一组能产生显状态序列 $O=O_1,O_2,\cdots,O_T$ 的模型参数，从而生成初始化模型 $\lambda_0=(\boldsymbol{\pi}_0,\boldsymbol{A}_0,\boldsymbol{B}_0)$，之后在此基础上不断寻找更好的模型参数。根据输入数据量的大小，经过不断的迭代最终可以得到最佳的模型参数。这种不断迭代寻找最大似然估计的方法又被称为最大期望算法(expectation-maximization algorithm, EM)。

Baum-Welch 算法可分为 E 步和 M 步。E 步为计算期望，即利用概率模型参数现有估计值计算隐状态期望。M 步为最大化期望，通过利用 E 步求得的期望对模型进行最大似然估计。E 步 M 步交替进行直到数据收敛或达到预设迭代次数结束。具体流程如下：

步骤 1. 输入显状态数据 $O=O_1,O_2,\cdots,O_T$，本文中为运动关键帧序列。

步骤 2. 对模型初始化，即当 n=0 时，选取 a_{ij}^0，$b_j(k)^0$，π_i^0，得到初始模型 $\lambda_0=(\boldsymbol{\pi}_0,\boldsymbol{A}_0,\boldsymbol{B}_0)$。

步骤 3. 迭代计算，对 n=1,2,$\cdots$ $a_{ij}^{(n+1)}$ 分别如式(13)所示，$b_j(k)^{(n+1)}$ 如式(14)所示，$\pi_i^{(n+1)}$ 如式(15)所示。

$$a_{ij}^{(n+1)}=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \tag{13}$$

$$b_j(k)^{(n+1)}=\frac{\sum_{t=1,ot=vk}^{T-1}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)} \tag{14}$$

$$\pi_i^{(n+1)}=\gamma_t(i) \tag{15}$$

实验结果表明，通常迭代 10 次以内数据已经收敛，因此设置迭代次数为 10。当迭代次数为 10 时，即 n=10 时停止，得到隐马尔科夫模型参数 $\lambda_i=(\boldsymbol{\pi}_i,\boldsymbol{A}_i,\boldsymbol{B}_i)$并输出。

3 实验结果与分析

本实验采用的计算机硬件环境为 Intel(R) Core(TM) i7-6700HQ CPU @2.60 GHz 2.59 GHz，16 GB 内存，NVIDIA GeForce GTX 980m 显卡，USB3.0 接口；软件环境为 Windows10 操作系统，Matlab，Visual Studio 2017，Untiy2017 等；所用主要开发语言为 C#。

3.1 基于预选策略的帧消减方法实验

为了验证本文提出的基于预选策略的帧消减的关键帧选取方法的可行性和有效性，设计本次实验。本实验所用的运动数据均来自开源的 HDM05 运动库和 MSR-Action3D 运动数据库。实验中针对 2 个数据库，分别选取了 high arm wave，draw x 等不同的运动类型进行测试。

图 2 是压缩比随最大重建误差变化数据图，各折线段分别代表了 high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle 等具体运动序列随最大重建误差设置而变化。

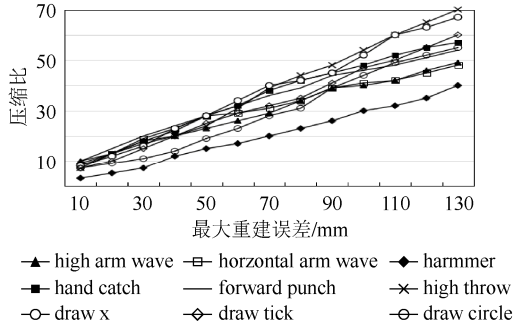

图 2 压缩比随最大重建误差变化数据图

由图 2 可以看出，随最大重建误差设置的增长，

动作压缩比普遍呈一种线性增长，与动作类型无关。当重建误差设置低于 10 mm 时，数据压缩率在 4~15 倍之间变化。这是因为对于不同的运动序列而言，在设置相同最大重建误差时，对于运动趋势缓慢的关键帧序列而言，压缩比较高，而对于 hammer 等剧烈的动作而言，压缩比相对较低。

对于处理后的关键帧序列而言，各关节点的平均重建误差往往远小于其最大重建误差。各关节点平均重建误差与最大重建误差基本呈线性变化，以 high arm wave 动作为例，如图 3 所示，其中横坐标为最大重建误差，纵坐标为各关节点的平均重建误差。与已有方法相比，基于预选策略的帧消减的关键帧选取方法可更好地控制误差，准确性更高。

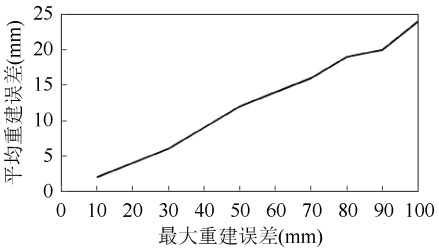

图 3　运动序列各关节点平均重建误差变化趋势

根据算法的运行效率，对已有数据库进行关键帧选取。图 4 表明算法处理数据速度随最大重建误差变化相关信息。

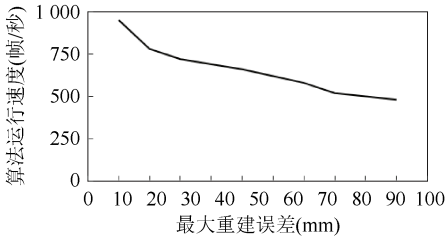

图 4　算法运行速率与重建误差变化示意图

从图 4 中可以看出，随用户输入的最大重建误差的增大，算法处理数据的速度随之降低。由于重建误差的变化需要删除更多原始序列中的关键帧，导致计算量增加，计算速度降低。随最大重建误差不断增大，系统处理速度逐渐降低但趋于平缓，并且处理速度远高于一般深度相机 60 帧/秒或 120+帧/秒的实时采样速度。本文提出的基于预选策略的关键帧选取方法可以用于人体实时数据的压缩和关键帧选取。其中，对于关键帧的预选部分时间运算效率高，通常占总时间的 5%以下。实验结果表明，本文所使用的基于预选策略的帧消减的关键帧选取方法是可行有效的。

3.2　基于隐马尔科夫模型的人体姿态识别实验

实验中建立关于人体姿态的隐马尔科夫模型，并用现有运动数据库的数据对选定动作进行训练学习以及测试。

对所有动作进行建模并训练后，可以得到如图 5 所示的模型信息。利用该模型输入人体运动数据可以对人体姿态进行识别和判定。

label	LL	prior	transmat	mu	sigma	mixmat
4	[-6.7687e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
5	[-5.6211e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
6	[-5.5880e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
7	[-7.3481e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
8	[-6.7368e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
9	[-7.9001e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
10	[-6.0941e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
11	[-8.0536e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
12	[-6.8648e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
13	[-8.3076e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
14	[-6.6081e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
15	[-3.6584e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
16	[-9.2257e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
17	[-6.8181e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
18	[-6.8630e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
19	[-7.7928e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double
20	[-8.5506e+...	[1;0;0;0;0;0...	9x9 double	60x9x2 double	4-D double	9x2 double

图 5　训练后的隐马尔科夫模型数据

在对人体姿态识别的过程中，从已有的运动数据库中，应用了 10 种姿势用于实验验证，包括 high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle, hand clap 等姿势。

通过应用数据集进行多次的训练和测试。最终实验结果表明，本文方法的平均识别率可达到 90%以上。对姿势的识别信息见表 1。

表 1　基于隐马尔科夫模型的人体姿态识别情况统计表

行为	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	37	0	0	1	0	0	1	0	1	0
A2	0	39	0	0	0	0	0	0	0	1
A3	0	0	35	0	1	2	2	0	0	0
A4	0	0	0	40	0	0	0	0	0	0
A5	0	0	0	0	40	0	0	0	0	0
A6	1	2	0	1	1	33	0	1	1	0
A7	0	1	0	0	2	0	35	0	1	1
A8	0	1	0	1	0	0	0	38	0	0
A9	0	1	0	0	2	0	0	0	36	1
A10	1	1	0	0	0	0	0	0	0	38

(1) 在表 1 中，y 轴表示每个具体动作的测试集，每个测试集含有 40 个测试数据。x 轴表示每一个对应姿态的隐马尔科夫模型。

(2) 表上的数据是对经过 10 次迭代获取的隐马尔科夫模型进行姿态识别的运行结果。A1~A10 分别代表了具体的行为，按照顺序依次为：high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle, hand clap。

(3) 对角线上的数据表示正确识别数量，数据越接近 40 代表识别效果越好，非对角线上的数字代表误识别的次数。

(4) 系统整体识别效果良好，但是对于 A3(hammer)，A6(high throw)以及 A7(draw x)动作识别效果相对较差，这是由于这几个动作运动过程相对剧烈，或与其他动作相对接近，因此容易产生误识别的问题。

3.3 与文献[12]对比实验

基于 MSR-Action3D 运动数据库，已经实现了本文的方法，同时也实现了文献[12]中的采用蒙版关节轨迹在深度视频序列的动作识别方法。本文方法与文献[11]方法进行准确率对比实验，选取 100 名成年男士、100 名成年女士、100 名儿童，对 10 组动作各进行 100 次识别实验，对识别准确率进行统计计算。绘制识别准确率折线图，如图 6 所示。

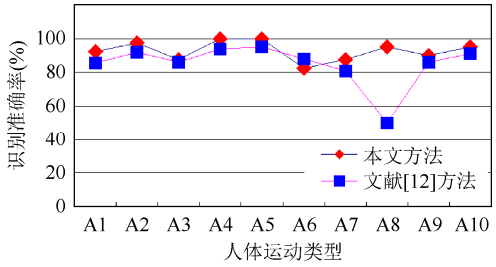

图 6 本文与文献[12]方法识别准确率对比

通过与文献[12]对比 A1~A10 动作的识别准确率，可以得出本文方法具有较高的识别准确率，具有较好的姿态识别效果。

4 结 束 语

针对传统人体姿态识别数据采集易受环境干扰、难以解决人体运动姿态的相似性和人体运动执行者的特征差异性等问题，本文提出了基于少量关键序列帧的人体姿态识别方法。遴选关键序列帧时，对原有运动序列进行预选，通过运动轨迹取极值的方法构造初选关键帧序列，再利用帧消减算法获取最终关键帧序列。该方法尽可能多的删除了冗余关键帧，尽可能多的保留了边界姿势帧，兼顾了重建误差和压缩比两个重要因素，节约了数据存储空间，又便于后续人体姿态识别计算。进行人体姿态识别时，首先对不同人体姿态分别建立隐马尔科夫模型；利用 Baum-Welch 算法计算得到初始概率矩阵、混淆矩阵、状态转移矩阵，获得训练后模型；输入待测数据，应用前向算法，得到对于每个模型的概率，比较并选取最大概率对应的姿态作为识别结果。实验结果表明，基于预选策略的帧消减算法可以有效对原始运动序列的关键帧进行选取；基于隐马尔科夫模型的人体姿态识别方法可以有效提高人体姿态识别的准确性，具有较好的识别效果。

下一阶段将研究特殊实验对象姿态识别和多目标人体姿态识别，以及拓展到智能服务机器人的应用中。

参 考 文 献

[1] BULBUL M F, JIANG Y S, MA J W. DMMs-based multiple features fusion for human action recognition [J]. International Journal of Multimedia Data Engineering and Management, 2015, 6(4): 23-39.

[2] 孟明, 杨方波, 佘青山, 等. 基于 Kinect 深度图像信息的人体运动检测[J]. 仪器仪表学报, 2015, 36(2): 386-393.

[3] RAMAN N, MAYBANK S J. Action classification using a discriminative multilevel HDP-HMM [J]. Neurocomputing, 2015, 154(C): 149-161.

[4] GAGLIO S, RE G L, MORANA M. Human activity recognition process using 3-D posture data [J]. IEEE Transactions on Human-Machine Systems, 2015, 45(5): 586-597.

[5] 景陈勇, 詹永照, 姜震. 基于混合式协同训练的人体动作识别算法研究[J]. 计算机科学, 2017, 44(7): 275-278.

[6] 李顺意, 侯进, 甘凌云. 基于帧间距的运动关键帧提取[J]. 计算机工程, 2015, 41(2): 242-247.

[7] 殷海艳, 刘波. 基于部位检测的人体姿态识别[J]. 计算机工程与设计, 2013, 34(10): 3540-3544.

[8] NASIRI J A, MOGHADAM CHARKARI N, MOZAFARI K. Energy-based model of least squares twin support vector machines for human action recognition [J]. Signal Processing, 2014, 104: 248-257.

[9] UDDIN M Z, KIM T S. 3-D body joint-specific HMM-based approach for human activity recognition from stereo posture image sequence [J]. Multimedia Tools and Applications, 2015, 74(24): 11207-11222.

[10] FEICHTENHOFER C, PINZ A, ZISSERMAN A. Convolutional two-stream network fusion for video action recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2016: 1933- 1941.

[11] 战青卓, 王大东. 基于深度神经网络的人体动作识别研究[J]. 智能计算机与应用, 2018, 8(2): 151-154.

[12] TEJERO-DE-PABLOS A, NAKASHIMA Y, YOKOYA N, et al. Flexible human action recognition in depth video sequences using masked joint trajectories [EBIOL]. [2018-09-17]. https://link.springer.com/article/10.1186%2Fs13640-016-0120-y.