ISSN 1000-1239/CN 11-1777/TP
50(2): 379-386, 2013

基于正弦级数拟合的行为识别方法

赵 绚[1,2] 彭启民[1]
[1](中国科学院软件研究所天基综合信息系统重点实验室 北京 100190)
[2](中国科学院大学 北京 100049)
(zhaoxuanhexue@163.com)

Action Recognition Based on Sine Series Fitting

Zhao Xuan[1,2] and Peng Qimin[1]
[1](*Science and Technology on Integrated Information System Laboratory*, *Institute of Software*, *Chinese Academy of Sciences*, *Beijing* 100190)
[2](*University of Chinese Academy of Sciences*, *Beijing* 100049)

Abstract This paper presents a sine series-based method for action recognition. The proposed method can be divided into three stages: feature extraction, series fitting and feature matching. In the stage of feature extraction, the method gets binary human silhouette through background difference and Gaussian background modeling at first, then uses the binary silhouette to represent the given image sequence and calculates the distance from the silhouette centroid to the boundary points according to a clockwise movement, thus changing the human silhouette into distance curve. In the stage of series fitting, the method fits the curve with the sine series and changes the distance curve into sine parameters with different amplitude, frequency and offset, which reduces the computational cost greatly and changes the process of action recognition into the matching of curve parameter features. Finally, in the stage of feature matching, the method classifies each frame in the given image sequence through calculating the minimum distance between it and the known action categories at first, and then gets the category of the given image sequence by voting. We adopt a leave-one-out scheme for experiment evaluation. The experiment results on a database of 90 short video sequences show that the promising performance is both effective and efficient.

Key words action recognition; silhouette distance curve; parameter fitting; sine series; feature matching

摘 要 提出了一种基于正弦级数拟合的行为识别方法.该方法利用二值轮廓序列来表示给定的运动图像序列,按照时针顺序计算从轮廓质心到轮廓边界点的距离,将人体轮廓转化为距离曲线,并将这一距离曲线利用正弦级数进行拟合,将距离曲线转化为正弦参数,从而极大地减小了计算量,将行为识别过程转化为曲线参数特征匹配的过程.在特征匹配过程中,通过计算待预测行为与已知类别行为的特征级数距离,对待预测行为中的每一个动作进行分类,最后通过投票决定该行为所属类别.在包含90个不同运动类别的视频数据库上进行留一交叉验证,实验结果表明,提出的方法能够有效地进行人体行为识别.

关键词 行为识别;轮廓距离曲线;参数拟合;正弦级数;特征匹配

中图法分类号 TP391.41

人体行为识别近年来受到了广泛的关注,成为计算机视觉和模式识别领域的研究热点,并且在人机交互、虚拟现实、智能监控、智能家居等方面得到了广泛的应用.

收稿日期:2011-01-19;**修回日期**:2011-11-16
通信作者:彭启民(pengqm@gmail.com)

尽管行为识别领域已经取得了一些进展,但是由于人体运动本身的复杂性(阴影、遮挡)和场景的多变性(光照、视角、分辨率),人体行为识别依然是一个具有挑战性的课题.

目前人体行为识别主要有 3 类:基于模板的识别方法、基于统计概率的识别方法和基于特征空间的方法[1].

基于模板的识别方法主要分为 2 类,一类是基于外形特征的模板,包括 2-D 模板[2-3],由于 2-D 模板要受到视角的限制,因此基于 3-D 模板[4-6]的方法逐渐受到关注.此类方法通过对人体姿态进行估计,从中抽取相关特征,将图像序列转化为静态形状模式,在识别过程中将人体姿态与相应的模板进行匹配.另一类是基于时空特征的模板,这类模板将人体行为视为整体进行估计和匹配. Bobick 和 Davis[7]将静态图像在时间轴上进行扩展,得到了运动历史图像(MHI)和运动能量图像(MEI). Gorelick 等人[8]利用泊松方程对人体轮廓特征进行分类和提取,得到了区分度较强的特征.通过模板对行为进行识别直观有效,缺点在于姿态估计的计算量大、运算时间较长.

基于统计概率的识别方法通过计算动作整体的概率和统计特征进行行为识别,以隐马尔可夫模型(HMM)[9-13]为代表.为了克服传统马尔可夫模型[9]存在的问题,近年来提出了很多改进的马尔可夫模型,如非参数化的马尔可夫模型[12],其消除了初始参数对模型的影响.该类模型结合了人体的运动特性,每一个动作姿态都被看作不可直接观测的内部状态,可以有效地用于动态时间系统,取得了比较好的识别效果,但是该类模型假定在时间序列上动作彼此独立,但事实上相似的动作经常发生在不同的时间点上且存在着长距离的相关性,因此会降低判别的可靠性.

基于特征空间的识别方法[14-17]将人体行为转化到特定的特征空间中,一般是对运动特征进行降维处理,降低行为识别需要的运算量,如线性判别分析法(LDA)[14]、局部保持映射法(LPP)[15]、主成分分析法(PCA)[16]、独立成分分析法(ICA)[17].通过将人体的运动特征转化到指定的特征空间中,运动特征被转化为特征空间中的点,通过计算这些点在特征空间的距离进行匹配.这一类方法极大地提高了行为识别的速度,但是识别的准确率有待提高.

本文提出了一种基于正弦级数的行为识别算法,将人体姿态特征转化为正弦参数,通过投票对人体行为进行识别,在降低识别运算量的同时提高了识别的准确率,具有很好的可扩展性和适应性.算法的流程如下:首先将运动序列转化为二值轮廓序列,再按顺序计算从轮廓质心到轮廓边界点的距离,将距离曲线利用正弦级数进行拟合,从而将人体行为识别过程转化为曲线参数特征匹配的过程,极大地减少了运算量.

1 基于正弦级数拟合的人体行为识别

1.1 处理流程

本文提出的行为识别方法主要包括 3 部分:特征提取、级数拟合和特征匹配,如图 1 所示:

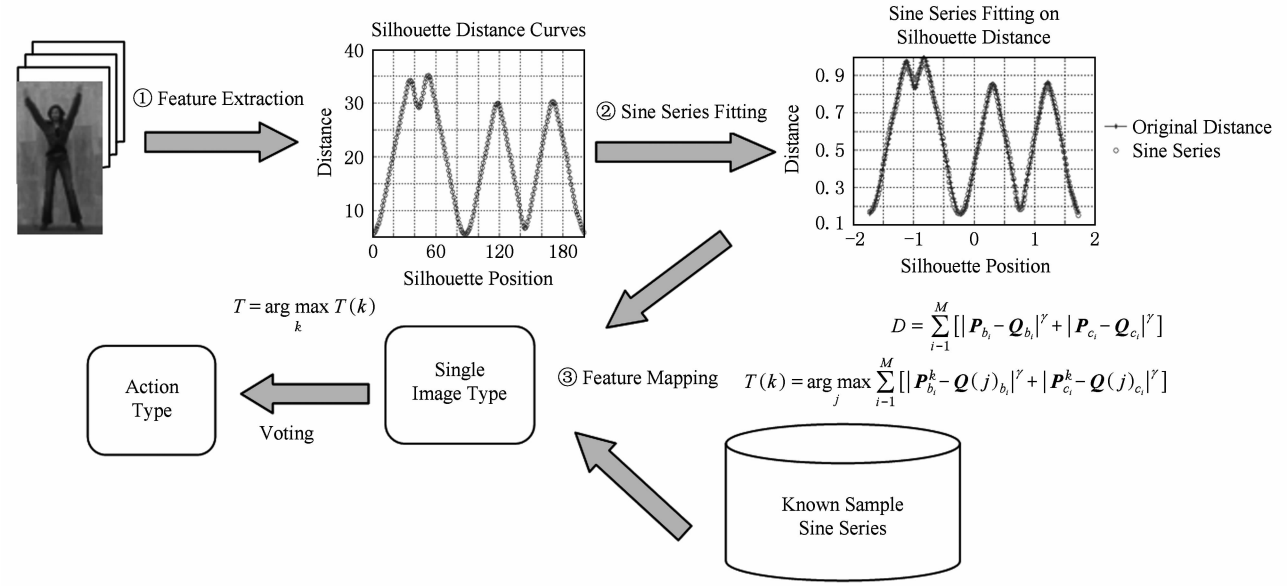


Fig. 1 Illustration of the system architecture.

图 1 处理流程

在特征提取阶段,对通过背景差分和高斯背景建模等方法分割得到的人体轮廓图像,按照指定的顺序计算轮廓质心到轮廓边缘点的距离,得到特征曲线.

级数拟合阶段对上一步得到的曲线利用多个正弦函数进行拟合,利用正弦函数的振幅、频率、偏移量等参数表示该曲线,从而将人体运动特征转化为正弦级数的参数.

特征匹配阶段包括特征匹配和投票 2 个环节,通过计算待预测行为与已知类别行为的特征级数距离,对待预测行为中的每一个动作进行分类,最后通过投票决定该行为属于哪一类动作,减少了个别动作异常引起的误差.

1.2　特征提取

特征提取的目的是得到对应特定动作的特征曲线,在本文是通过顺序求取运动轮廓图像的质心到轮廓边界每一点的距离实现的. 所谓顺序是指从给定的起点开始,按照顺时针或者逆时针方向沿着人体轮廓图像逐一遍历.

由于噪声的存在,输入的轮廓图像本身包含一些不准确的边界点,这些不准确的边界点会影响得到的距离曲线,因此需要对原始的距离曲线进行低通滤波,得到较为光滑的曲线,在去掉噪声点的同时保持了原始曲线的整体走向.

距离计算函数的具体流程如下:

输入:人体轮廓图像;

输出:距离曲线.

1) 计算轮廓图像的质心(x_o, y_o).

$$\begin{cases} x_o = \dfrac{1}{N}\sum_{t=1}^{N} x_t, \\ y_o = \dfrac{1}{N}\sum_{t=1}^{N} y_t, \end{cases}$$

其中 N 表示轮廓边界像素点数,(x_t, y_t)表示位于轮廓边界的像素点.

2) 指定开始遍历轮廓的起始点,默认为以质心为坐标原点的 0 度角点为起始点.

3) 按照逆时针方向从起始点开始遍历,对每一个边界点,寻找其 8 邻域中未被遍历的边界点作为下一个候选边界点.

4) 计算质心(x_o, y_o)到轮廓边界点(x_t, y_t)的欧氏距离 $d_t = \sqrt{(x_t - x_o)^2 + (y_t - y_o)^2}$,质心到整个人体轮廓的距离为 1 维距离向量 $\boldsymbol{d}(t) = d_t$.

5) 利用线性平滑滤波器(或者其他低通滤波器)对距离向量进行平滑处理,得到最终输出的距离曲线 $\bar{\boldsymbol{d}}(t)$.

本阶段得到的各类动作的部分轮廓距离曲线如图 2 所示,按照从左到右从上到下的顺序,动作类别依次为 jack,bend,jump,pjump,run,walk.

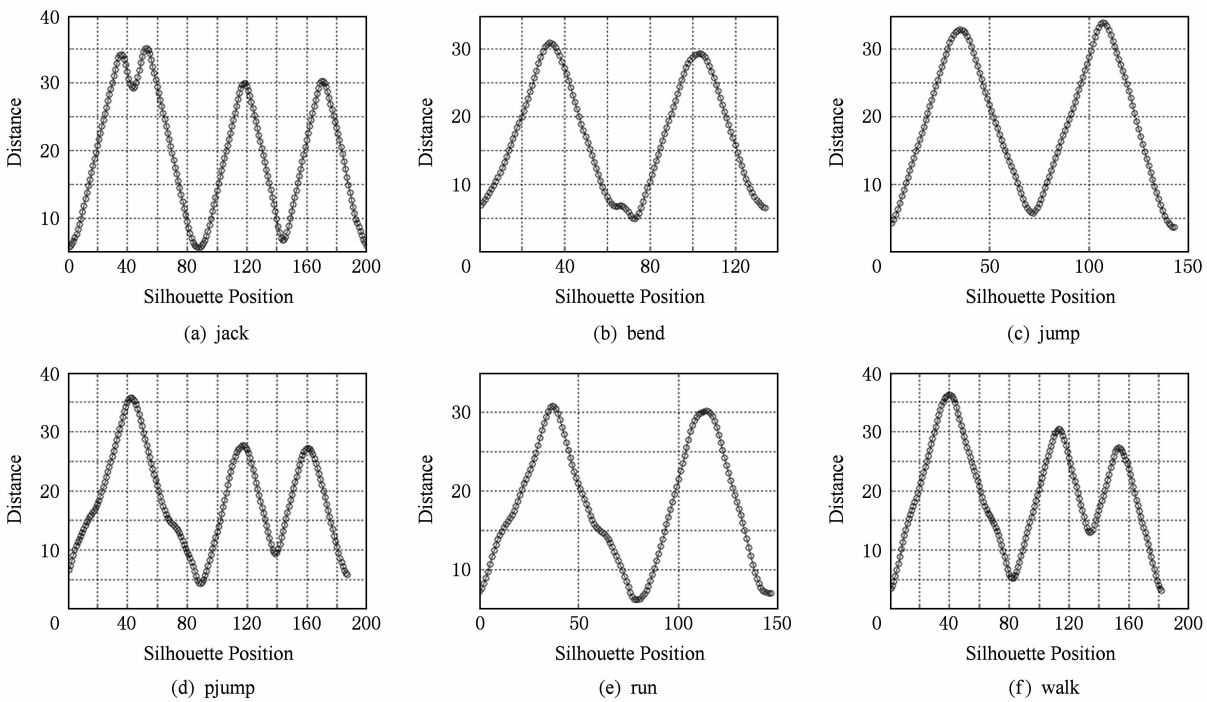


Fig. 2　Silhouette distance curves of different actions.

图 2　不同动作的轮廓距离曲线

1.3 正弦级数拟合

通过图 2 可以看出,不同轮廓距离曲线的形状呈现为不规则的波形,可以视为 1 维信号量.考虑利用若干个基函数叠加对上述信号进行拟合,如多项式、高斯函数、三角函数(正弦函数和余弦函数)、傅里叶级数等,从而将行为识别问题转化为特征空间上基底的距离计算问题.

为了确定拟合所用的基底类型及其有效性,从数据库中随机抽取 100 条轮廓距离曲线,分别利用 9 次多项式、高斯函数、正弦函数和傅里叶级数进行拟合,拟合得到的各项指标如图 3 所示.

图 3(a)(b)分别表示不同基底的残差平方和

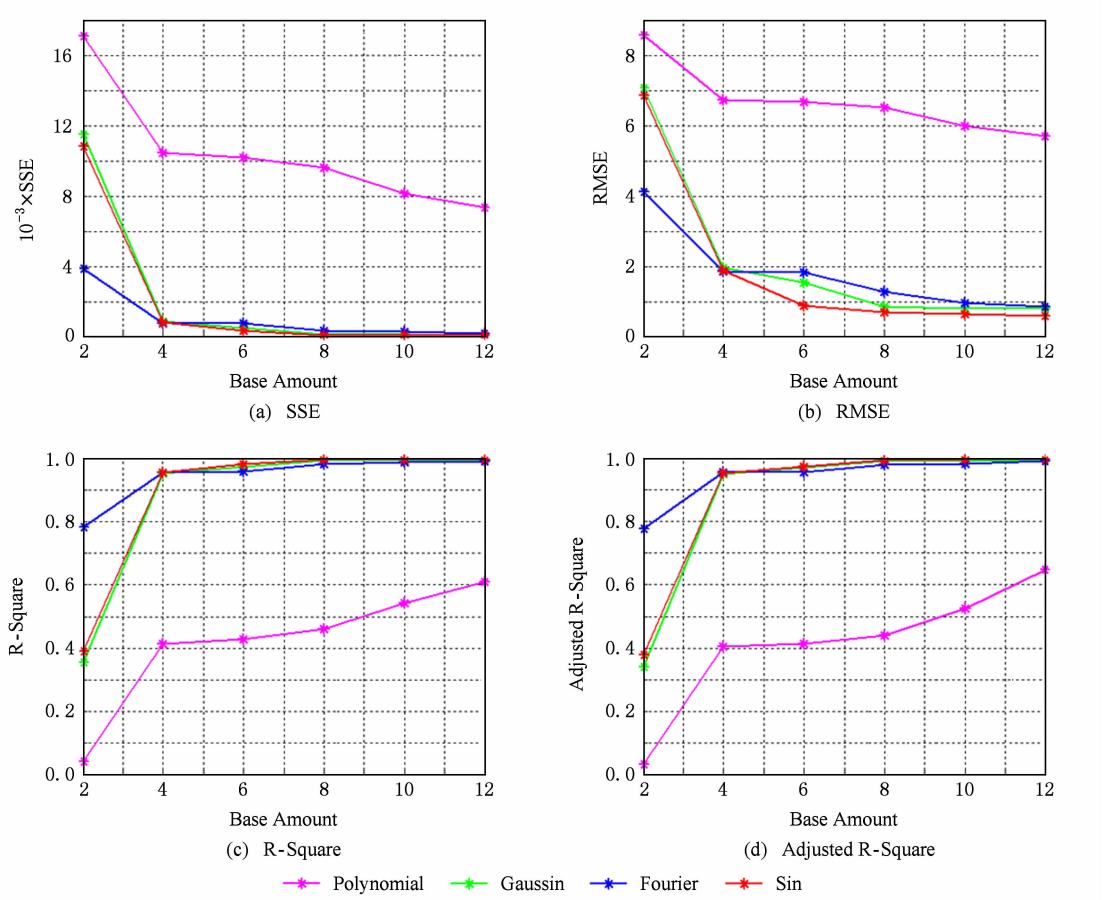


Fig. 3 Fitting performance on different bases.

图 3 不同基底的各项拟合指标

(SSE)与标准误差(RMSE),图 3(c)(d)分别表示拟合系数(R-Square)和调整后拟合系数(adjusted R-Square).图 3 中的多项式为 9 次多项式,这一次数的选取是综合考虑拟合复杂度和拟合误差得到的最优结果.由图 3 可以看出,多项式的拟合程度最差,各项指标明显低于其他 3 种拟合方式.傅里叶级数的拟合效率最高,在使用少数基底的情况下(如仅使用 2 个基底)拟合效果明显优于其他拟合方式,但进一步增加基底的个数对提高拟合精度的作用不明显.正弦函数与高斯函数都较好地拟合了实验数据.

多项式和傅里叶级数属于正交基函数,这类基函数一般揭示的是信号的整体性特征,而人体运动轮廓距离本身具有很多局部化特征,使用正交基对其进行拟合必然无法达到理想的效果,图 3 的实验结果证明了这一点.相比之下,正弦函数和高斯函数属于非正交基函数,与正交基函数相比,非正交基函数有一定的冗余性,但是它能够适用于局部性特征的精细刻画.相比之下,正弦函数在各项指标上表现更好,同时,经过观察发现,对一些双峰曲线进行拟合时高斯函数会出现过拟合情况,而正弦函数则比较稳定,不会出现上述情况.

综上所述,考虑采用正弦级数对距离曲线进行拟合,拟合公式如式(1)所示:

$$f(x)=\sum_{i=1}^{M}a_i\sin(b_ix+c_i), \tag{1}$$

其中,$f(x)$表示距离曲线函数,x 表示人体轮廓边界点的顺序值,a_i 表示第 i 个正弦函数的振幅,b_i 表示第 i 个正弦函数的频率,c_i 表示第 i 个正弦函数

的相位,M 代表共使用多少个正弦函数进行拟合.

拟合过程采用最小二乘法,同时考虑标准误差的相对变化量和拟合相似度.具体拟合停止条件如式(2)所示:

$$R_l > r,\ \frac{|\ S_l - S_{l-1}\ |}{S_{l-1}} < s, \tag{2}$$

其中 l 表示拟合次数,R_l 表示第 l 次拟合的相似度,S_l 表示第 l 次拟合的标准误差,r 表示符合条件的拟合相似度的最小值,s 表示符合条件的第 l 次与第 $l-1$ 次拟合的标准误差相对变化量的最大值.根据实验效果,本文取 $r=0.95$,$s=0.05$.

由图 3 可以看出,采用正弦级数进行拟合时各项指标随着基底的增加不断提高,根据式(2),在基底个数大于 8 时,进一步提高基底个数对指标的提高影响不明显,但是明显加大了拟合的复杂度,因此确定拟合所需的基底数 $M=8$,此时既保证了拟合系数对不同动作的分离度,又有效地控制了计算复杂度和所需的数据量.待检测序列中每一帧原始图像特征数量级为 10^4,经过正弦技术拟合后的特征数量为 24 个,极大地减少了运算量.

图 4 展示了对各类动作的轮廓距离曲线进行正弦级数拟合的情况,动作类别的排列与图 2 相同.由图 4 也可以看出,采用正弦级数对轮廓距离曲线进行拟合可以取得良好的拟合效果:

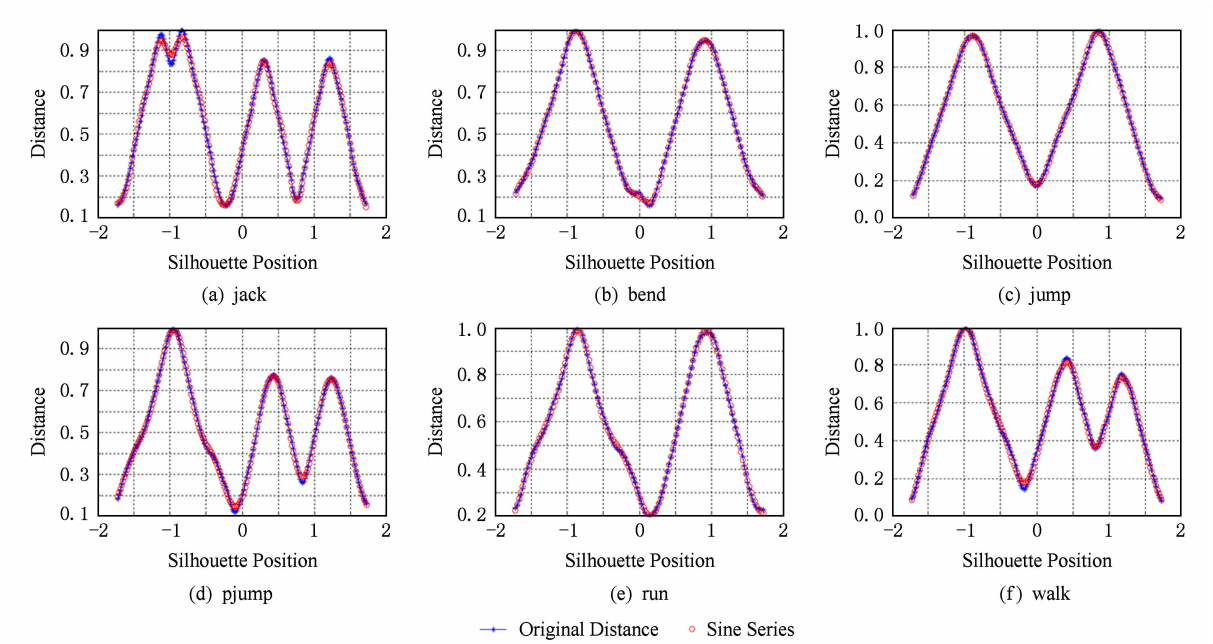


Fig. 4　Sine series fitting with different action types.

图 4　对各类动作的轮廓距离曲线进行正弦级数拟合

1.4　特征匹配

本文采用了一种新的计算行为相似性的距离公式,如式(3)所示:

$$D = \sum_{i=1}^{M} [\ |\ P_{b_i} - Q_{b_i}\ |^{\gamma} + |\ P_{c_i} - Q_{c_i}\ |^{\gamma}], \tag{3}$$

其中 D 表示 2 个正弦级数之间的相似性距离,$\boldsymbol{P}$,$\boldsymbol{Q}$ 表示 2 个不同的正弦级数,P_{b_i} 表示 $\boldsymbol{P}$ 所表示的正弦级数的第 i 个正弦波的 b 分量,同理,P_{c_i} 表示 $\boldsymbol{P}$ 所表示的正弦级数的第 i 个正弦波的 c 分量,Q_{b_i} 表示 $\boldsymbol{Q}$ 所表示的正弦级数的第 i 个正弦波的 b 分量,Q_{c_i} 表示 $\boldsymbol{Q}$ 所表示的正弦级数的第 i 个正弦波的 c 分量,γ 表示经验距离参数,由实验确定,本文中为 0.5,M 与式(1)中含义相同,表示距离曲线由 M 个正弦函数拟合而成.

在一个正弦函数的 3 个组成部分中,a 表示振幅,b 表示频率,c 表示相位,即偏移量,为了消除人体外形特征带来的影响,计算 2 个正弦级数相似性时只考虑频率和相位,而不考虑振幅可以提高匹配的准确率和精确性,同时,加入图像的宽高比率可以过滤掉不必要的误差,进一步提高准确率,实验结果证明了这一结论.

特征匹配的计算公式如下:

$$T(k) = \arg\min_{j} \sum_{i=1}^{M} [\ |\ P_{b_i}^{k} - Q(j)_{b_i}\ |^{\gamma} + |\ P_{c_i}^{k} - Q(j)_{c_i}\ |^{\gamma}], \tag{4}$$

其中 k 表示待预测行为序列中的第 k 帧图像,$T(k)$ 则表示其中的第 k 帧图像所属的类别,T 表示整个待预测行为序列的类别,为各个 $T(k)$ 的投票结果,

P^k 表示第 k 帧图像的正弦级数，$Q(j)$表示第 j 类动作的正弦级数，其他参数的含义同式(3).

2 实验与结果分析

本文的实验采用 Weizmann 数据库，该数据库包含 90 个固定背景的视频，大小为 180×144 像素，帧速率为 25 fps，属于低分辨率的视频图像. 其中包括 10 种动作类型，分别是 jack，bend，jump，pjump，run，walk，side，skip，wave1，wave2. 每种动作由 9 个人分别演示，这 9 个人具有不同的性别和外形特征，运动特征具有明显差异，同时动作有正向和逆向的演示，因此是目前比较权威的行为识别实验数据库. 数据库中的动作类别如图 5 所示，从上到下的演示动作分别为 jack，bend，jump，pjump，run，walk，side，skip，wave1，wave2.

首先采用留一交叉验证法对算法进行检验，对每一种行为类别，选取其中 8 个人的行为视频作为样本，另外 1 个人的行为视频作为测试，重复 10 次，得到的实验结果的混淆矩阵如表 1 所示.

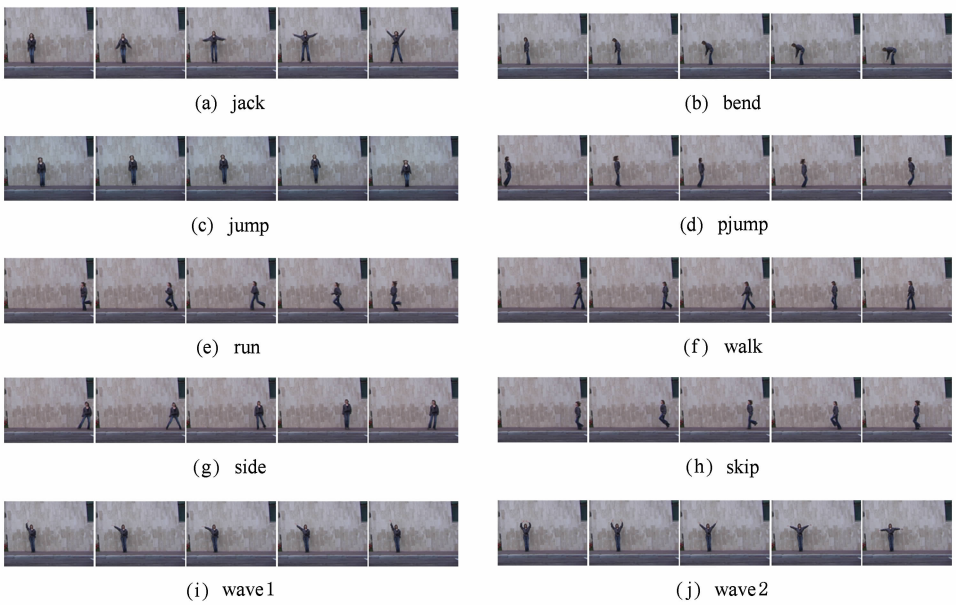


Fig. 5 Actions in Weizmann dataset.

图 5 Weizmann 数据库的动作

Table 1 Confusion Matrix for Recognition on the Testing Data (*CCR*=95.56%)

表 1 测试数据上的混淆矩阵(*CCR*=95.56%) %

Action Types	Cross Correction Rate (*CCR*) between Different Actions									
	jack	jump	pjump	run	side	skip	walk	wave1	wave2	bend
jack	100	0	0	0	0	0	0	0	0	0
jump	0	100	0	0	0	0	0	0	0	0
pjump	0	0	100	0	0	0	0	0	0	0
run	0	0	0	88.89	0	0	11.11	0	0	0
side	0	0	0	0	100	0	0	0	0	0
skip	0	0	0	11.11	0	88.89	0	0	0	0
walk	0	0	0	0	0	0	100	0	0	0
wave1	0	0	0	0	0	0	0	100	0	0
wave2	0	0	0	0	0	0	0	0	100	0
bend	11.11	11.11	0	0	0	0	0	0	0	77.78

为了进一步验证本算法的有效性，我们仅利用 1 个人的行为作为识别的样本，对其余 8 个人的行为分别进行实验，实验得到的混淆矩阵如表 2 所示. 由表 2 可以看出，在仅有少量数据作为识别样本的

情况下,本算法依然取得了良好的效果,平均识别率达到 77.78%,说明本算法能够在一定程度上消除体貌、颜色、动作频率等特征的影响,提取出了不同行为类别的主要特征,具有一定的鲁棒性.

Table 2　Confusion Matrix for Recognition on the Testing Data with Only One Person as Sample (*CCR*=77.78%)

表 2　仅用一人作样本的平均准确率(*CCR*=77.78%)　%

Action Types	Cross Correction Rate (*CCR*) between Different Actions									
	jack	jump	pjump	run	Side	skip	walk	wave1	wave2	bend
jack	100	0	0	0	0	0	0	0	0	0
jump	0	88.89	11.11	0	0	0	0	0	0	0
pjump	0	11.11	88.89	0	0	0	0	0	0	0
run	0	0	0	55.56	22.22	0	22.22	0	0	0
side	0	0	11.11	0	88.89	0	0	0	0	0
skip	0	0	0	11.11	11.11	77.78	0	0	0	0
walk	0	0	0	55.56	22.22	0	33.33	0	0	0
wave1	0	0	0	0	11.11	0	0	88.89	0	0
wave2	0	0	0	0	0	0	0	11.11	88.89	0
bend	0	33.33	0	0	0	0	0	0	0	66.67

表 3 展示了本文提出的算法与其他已经发表的降维算法在 Weizmann 数据库上的平均识别率的对比结果,通过对比可以说明本文提出的算法与同类算法相比具有一定优势.

Table 3　Average Class Accuracy on the Weizmann Dataset

表 3　不同方法在 Weizmann 数据库上的平均识别率

Recognition Method	Mean Accuracy/%
Klaser et. al.[18]	84.3
Chunhua Du et. al.[19]	94.97
Scovanner et al.[20]	82.6
Wei Huang et al.[21]	88.75
Our method	95.56

3　结论和下一步的工作

本文提出了一种基于正弦级数拟合的行为识别方法,利用正弦级数作为行为识别的特征,减少了行为识别所需的运算量,同时提高了单纯依靠特征空间进行识别的准确率. 在 Weizmann 数据库上进行的实验表明,本文的识别方法可以有效地对人体行为进行识别,具有良好的适应性和区分度以及空间尺度不变性. 下一步的工作将对其进行改进,使其可支持存在遮挡的情形.

参　考　文　献

[1] Ahad M D, Tan J K, Kim H S, et al. Human activity recognition: Various paradigms [C] //Proc of Int Conf on Control, Automation and Systems. Piscataway, NJ: IEEE, 2008: 1896-1901

[2] Polana R, Nelson R C. Detection and recognition of periodic, nonrigid motion [J]. International Journal of Computer Vision, 1997, 23(3): 261-282

[3] Dalal N, Triggs B, Schmid C. Human detection using oriented histograms of flow and appearance [C] //Proc of European Conf on Computer Vision. Piscataway, NJ: IEEE, 2006: 428-441

[4] Sigal L, Bhatia S, Roth S, et al. Tracking loose-limbed people [C] //Proc of IEEE Conf Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2004: 421-428

[5] Gu Junxia, Ding Xiaoqing, Wang Shengjin. Human 3D model-based 2D action recognition [J]. Acta Automatica Sinica, 2010, 36(1): 46-53 (in Chinese)
(谷军霞, 丁晓青, 王生进. 基于人体行为 3D 模型的 2D 行为识别[J]. 自动化学报, 2010, 36(1): 46-53)

[6] Fossati A, Dimitrijevic M, Lepetit V, et al. Bridging the gap between detection and tracking for 3D monocular video-based motion capture [C] //Proc of IEEE Conf Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2007: 1-8

[7] Bobick A F, Davis J W. The recognition of human movement using temporal templates [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(3): 257-267

[8] Gorelick L, Blank M, Shechtman E, et al. Actions as space-time shapes [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 12(29): 2247-2253

[9] Yamato J, Ohya J, Ishii K. Recognizing human action in time sequential images using hidden Markov model [C] //Proc of IEEE Conf Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 1992: 379-385

[10] Lv F, Navatia R. Recognition and segmentation of 3D human action using HMM and multi-class AdaBoost [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2006: 359-372

[11] Ren Haibing, Xu Guangyou. Human action recognition with primitive-based coupled-HMM [C] //Proc of Int Conf on Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2002: 11-15

[12] Jin N, Mokhtarian F. A non-parametric HMM learning method for shape dynamics with application to human motion recognition [C] //Proc of Int Conf on Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2006: 29-32

[13] Kulic D, Takano W, Nakamura Y. Representability of human motions by factorial hidden Markov models [C] //Proc of Int Conf on Intelligent Robots and Systems. Piscataway, NJ: IEEE/RSJ, 2007: 2388-2393

[14] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720

[15] He Xiaofei, Yan Shuicheng, Hu Yuxiao, et al. Face recognition using Laplacian faces [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340

[16] Turk M A, Pentland A P. Face recognition using eigenfaces [C] //Proc of IEEE Conf Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 1991: 586-591

[17] Bartlett M S, Movellan J R, Sejnowski T J. Face recognition by independent component analysis [J]. IEEE Trans on Neural Networks, 2002, 13(6): 1450-1464

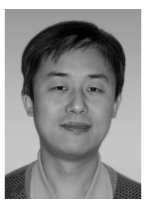
[18] Klaser A, Marszalek M, Schmid C. A spatio-temporal descriptor based on 3D-gradients [C] //Proc of British Machine Vision Conference. Leeds, UK: BMVC, 2008: 995-1004

[19] Du Chunhua, Wu Qiang, Yang Jie, et al. Subspace analysis methods plus motion history image for human action recognition [C] //Proc of Digital Image Computing: Techniques and Applications for IEEE. Piscataway, NJ: IEEE, 2008: 606-611

[20] Scovanner P, Ali S, Shah M. A 3-dimensional sift descriptor and its application to action recognition [C] //Proc of Int Conf on Multimedia. New York: ACM, 2007: 357-360

[21] Huang Wei, Wu Q M Jonathan. Human action recognition based on self organizing map [C] //Proc of IEEE Int Conf on Acoustics, Speech, and Signal Processing. Piscataway, NJ: IEEE Signal Proc Society, 2010: 2130-213

Zhao Xuan, born in 1987. Received her MSc degree from the Institute of Software, Chinese Academy of Sciences in 2012. Her current research interests include digital image processing and computer vision.



Peng Qimin, born in 1969. Received his PhD degree in computer science from Beijing Institute of Technology in 2005. Currently he is associate professor in the Institute of Software, Chinese Academy of Sciences. His current research interests include image understanding, pattern recognition, information fusion, etc.