

doi:10.16652/j.issn.1004-373x.2017.03.014

舞蹈视频图像中人体动作识别技术的研究

陈利峰

(郑州大学 体育学院, 河南 郑州 450044)

摘要: 主要研究采用图像预处理、3D Zernike矩、建立码书和支持向量机(SVM)等技术对动作视频中的人体动作进行识别。仿真试验结果表明,采用该方法可以有效地对KTH和A-go-go舞蹈视频数据库进行识别,且可以有效地识别舞蹈动作,进而对舞者进行动作纠正。

关键词: 舞蹈视频; 人体动作识别; SVM; 图像分割

中图分类号: TN911.73-34

文献标识码: A

文章编号: 1004-373X(2017)03-0051-03

Research on recognition technology for human-body action in dance video image

CHEN Lifeng

(College of Physical Education, Zhengzhou University, Zhengzhou 450044, China)

Abstract: The technologies of image preprocessing, 3D Zernike moment, code book establishment and support vector machine (SVM) used to recognize the human-body action in movement video are studied. The results of simulation experiment show that the method can identify the KTH and A-go-go dance video database effectively. The method can recognize the dance movement availablely, and correct the dancers' action.

Keywords: dance video; human-body action recognition; SVM; image segmentation

0 引言

视频中的人体动作识别是目前计算机研究中的一个热点,其目的是通过各种图像处理和识别分类技术对视频中的动作进行提取和分析,以判断视频中的人物所进行的动作,从而得出有用的信息,具有十分广泛的用途。视频人体动作识别的关键是对原视频图像进行合理的预处理,然后提取视频图像的特征并进行描述和分类^[1-2]。

视频动作识别技术与舞蹈艺术相结合的相关研究目前在国内刚起步。通过对舞蹈视频应用人体动作识别技术可以有效地识别舞蹈动作姿势。通过比对视频动作与标准动作,可以对舞者的舞蹈姿势进行评价,从而给出修改建议,是一种先进的辅助训练方式^[3]。

本文采用视频人体动作识别的相关技术,先以较为简单的KTH数据库训练出有效的SVM分类器,再利用更为复杂的A-go-go舞蹈视频数据库对SVM分类器进行强化训练,最后利用训练所得的成熟分类器对A-go-go舞蹈视频进行识别,并获得了较好的分类效果,本文的

研究思路如图1所示。

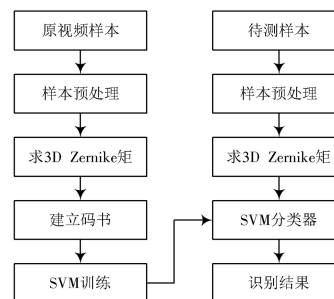


图1 本文研究思路

1 视频图像的数字化处理

1.1 灰度变换

对视频图像进行数字计算识别前,必须先进行灰度变换,图像阈值化和图像分割等预处理,以降低计算机的运算量,并提取出有效信息。

通常视频图像最初都是真彩图像,即RGB图像,它的每个像素都由R,G,B三原色构成。由于直接对真彩图像进行处理的运算量十分巨大,分析效率较低。因此,本文采用灰度变换的方法将视频图像处理为灰度图像,降低图像包含的色彩信息,灰度化后的图像如图2所示。

收稿日期:2016-06-20



图2 原图像和经过灰度变换后的图像

1.2 运动图像阈值化

图像阈值化的目的是为了获取运动图像的二值化图像,主要实现手段是选择合适的阈值,以阈值为区分标准,然后将像素灰度值分为两类,从而实现对运动图像的分割,阈值一般可写成如下形式:

$$T = T[x, y, f(x, y), p(x, y)] \quad (1)$$

式中: $f(x, y)$ 是在像素点 (x, y) 处的灰度值; $p(x, y)$ 是该点的灰度梯度函数。采用式(1)即可求出二值化后的图像。

1.3 运动图像分割

得到了动作视频当前时刻的二值图像后,需要将运动区域从场景中分离出来,这涉及到运动区域分割。本文采用 Matlab 软件自带的二值图像处理函数通过设置合适的阈值找到运动中的人体轮廓的边缘。

具体做法为:设当前时刻为 t ,检测视频在时刻 t 的帧为 $P(x, y, t)$,尺寸为 $M \times N$,得到二值图像为 $A(x, y, t)$ 。

对于 t 时刻,二值化图像 $A(x, y, t)$ 的前景灰度值为 255,背景灰度值为 0。设定沿着 A 的列向量进行扫描,统计每一列中前景像素的数目 $C(1 < C < n)$ 。从这 n 个频率数据中挑选出最大值 C_i ,该最大值对应的列序列号标定为 i 。如果 C_i 与帧之间的行数 m 之比大于 $1/6$,则判断当前帧包含运动人体的区域,如果 $1/16 < C_i < 1/6$,则判断当前帧包含部分运动人体;更新当前时刻值, $t = t + 1$,直到所有视频帧均被扫描为止^[4]。经过阈值化和图像分割后的图像见图3。

1.4 3D Zernike矩提取整体特征

获得运动视频图像的二值化图像后,本文采用 Zernike矩对运动图像的二值图像进行特征描述,从而实现对其分类识别。Zernike矩是一种对形状描述很有效的正交矩,已经在图像处理领域得到广泛应用,具有包含信息更加完整,冗余信息量少,对噪声不敏感等优点。对于一个图像序列,计算 Zernike矩的公式如下:

$$A_{nm\mu\gamma} = \frac{n+1}{\pi} \sum_{i=2}^{\text{images}} \sum_x \sum_y U(i, \mu, \gamma) [V_{nm}(r, \theta)] P_i(x, y) \quad (2)$$

式中: images 表示整个序列中图像的数目; $U(i, \mu, \gamma)$ 为引入的第三维度:

$$U(i, \mu, \gamma) = (\bar{x}_i - \bar{x}_{i-1})^\mu (\bar{y}_i - \bar{y}_{i-1})^\gamma \quad (3)$$

式中: \bar{x}_i 表示当前图像的重心; \bar{x}_{i-1} 表示前一幅图像的重心; \bar{y} 同理, μ 和 γ 是由用户自己定义的参数,根据不同序列中图像的数目可能不同。所以在计算出 3D Zernike后可以用下式对其进行归一化:

$$\bar{A}_{nm\mu\gamma} = A_{nm\mu\gamma} / (A * \text{images}) \quad (4)$$

式中: A 代表目标的平均面积(像素个数); images 与上文相同,表示序列中图像的数目。这样对于一个已经提取出人体剪影的图像序列,就可以使用式(4)得到其对应的 3D Zernike矩作为整体特征。



图3 经过阈值化和图像分割后的图像

1.5 建立码书

码书是由样本空间中最具有代表性的样本组合而成,这些样本能很好地区分本类别和其他类别。本文采用聚类分析的方法创建码书,首先定义各类别相似性测度,以此来描述类别之间的相似程度,本文采用基于 3D Zernike矩的相似测度建立码书。

相似测度是以两矢量的方向是否相近作为度量的基础,设已经求得两个图片的 3D Zernike矩的向量为 \mathbf{x}, \mathbf{y} , 他们的相似性系数可表示为:

$$\cos(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y} / [(\mathbf{x}^T \mathbf{x})(\mathbf{y}^T \mathbf{y})]^{1/2} \quad (5)$$

本文借鉴文献[5]构建码书的思想,将相似系数 $\cos(\mathbf{x}, \mathbf{y}) > 0.8$ 的一系列采样图片初步判断为一个动作,将 $\cos(\mathbf{x}, \mathbf{y}) > 0.9$ 所代表动作的图片提取出来作为一个标准描述符矩阵,这样数据库中包含几个不同动作就会提取几个描述符矩阵。

1.6 支持向量机分类

一旦码书创建成功,那么每个动作就可以被其中的一系列关键姿势来表示。每一帧图像都对应一个描述符,这些描述符可以被码书中与其距离最近的关键姿势表示,因此一个动作可以看成是由码书中一系列关键姿

势组合而成的,本文采用支持向量机实现。支持向量机(SVM)是一个基于类间距离最大化的判别算法,属于二分类模型,并能通过一对一法扩展成多类分类器。其具体表现形式如下:

(1) 设已知训练集:

$$T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l)\} \in (X \times Y)^l \quad (6)$$

式中: $\mathbf{x}_i \in X = R^n, y_i \in Y = \{1, -1\}, i = 1, 2, \dots, l, \mathbf{x}_i$ 为特征向量。

(2) 选取适当的核函数 $K(x, x')$ 和适当的参数 C , 构造并求解最优化问题。

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{j=1}^l \alpha_j \\ \text{s.t.} & \sum_{j=1}^l y_j \alpha_j = 0, 0 \leq \alpha_i \leq C \end{aligned} \quad (7)$$

得到最优解:

$$\alpha = (\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*)^T \quad (8)$$

(3) 选取 α 的一个正分量 $0 \leq \alpha_j^* \leq C$, 并计算此阈值:

$$\beta = y_j - \sum_{i=1}^l y_i \alpha_i^* K(\mathbf{x}_i - \mathbf{x}_j) \quad (9)$$

(4) 构造决策函数:

$$f(x) = \text{sgn} \left(\sum_{i=1}^l \alpha_i^* y_i K(x, \mathbf{x}_i) + b^* \right) \quad (10)$$

在利用支持向量机进行分类时,要先从分析对象样本集中抽取一部分样本集作为学习样本进行训练,获得合适精度的模型后,再对同一样本空间中的其他样本进行分类识别计算。

2 试验仿真

为了验证所提方法的有效性,本文采用KTH数据库和A-go-go舞蹈视频两类视频数据库进行验证^[6]。

KTH数据库共包含6类最常见的动作,如图4所示,分别是走路、慢跑、快跑、挥手、拍手和出拳,其包含的是一些较为基本和简单的动作。KTH中每种行为有100个视频,由25个人分别在4类不同的背景中完成。



图4 KTH数据库中包含的6类动作

采用第1节中的相关算法对这些视频逐帧进行预处理、特征提取和建立码书,利用获得的大量描述符矩阵作为训练样本对SVM进行训练,经过仿真比较,本文的决策函数直接调用Matlab软件自带的卡方核函数,采用SVM对这6种常见动作分类的结果如图5所示。

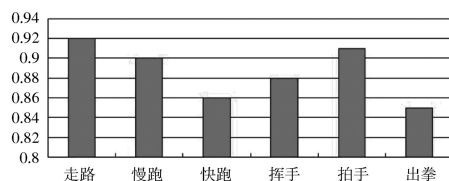


图5 KTH数据库动作的识别率

由图5可知,采用SVM对KTH动作视频数据库进行识别可以获得较高的识别率,其中对走路的识别率最高为92%,对出拳的识别率最低为85%,对整个KTH样本的平均识别率为88.7%。因此,该方法具有较高的识别正确率。

为了验证该方法在舞蹈视频中的识别效果,现利用该SVM分类器对A-go-go舞蹈视频数据库进行识别。A-go-go舞蹈一共包含19类基本的舞蹈动作,每类动作由5个具备不同舞蹈技能的被捕捉者采集,且每个被捕捉者对每类动作均执行3次,从而每类动作具有15个样本,总共捕捉285个动作模式样本。为了提高SVM的分类精度,先在285个动作样本中选出200组样本对之前获得的来自KTH的SVM分类器进行训练,再利用获得的SVM分类器对整个A-go-go舞蹈视频集进行分类。

采用SVM对A-go-go舞蹈的19个动作分类识别的结果如图6所示。

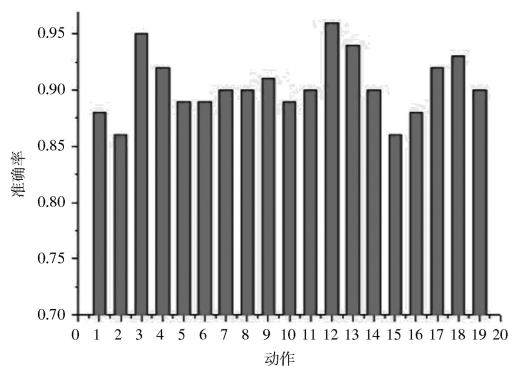


图6 对A-go-go舞蹈动作视频数据的识别率

由图6可知,采用加强训练后的SVM分类器对A-go-go舞蹈动作视频数据库进行识别可以获得较高的识别率,其中对动作3的识别率最高为95%,对动作2和动作15的识别率最低为86%,对整个样本的平均识别率为90.4%。因此,该方法对舞蹈动作视频有较高的识

(下转第57页)

间短,说明图像分类的实时性较好,结果验证了本文算法的优越性。

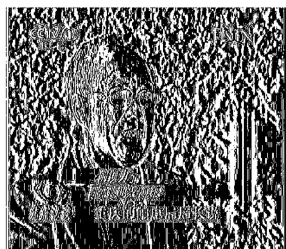


图4 Harris角点检测结果



图5 特征提取结果

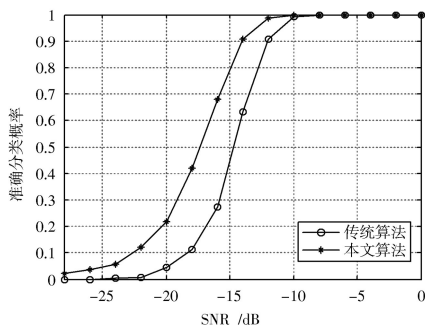


图6 分类精度对比分析

表1 图像分类算法性能对比结果

| 指标 | 本文算法 | 文献[4]算法 |
|-----------|------|---------|
| 峰值信噪比 /dB | 12.4 | 8.3 |
| 执行时间 /ms | 2.2 | 5.4 |

4 结 语

本文研究了多媒体信息数据库中的图像优化分类

作者简介:唐普霞(1982—),女,安徽太湖人,硕士,讲师。主要研究方向为计算机应用。
石丽怡(1982—),女,黎族,海南陵水人,讲师。主要研究方向为计算机应用。

(上接第53页)

别正确率。

3 结 论

本文主要对动作识别技术在视频图像,特别是舞蹈视频图像中的实现方法和效果进行研究。本文首先对视频动作图像进行灰度化、二值化和图像分割等预处理,接着计算了处理后的二值化图像的3D Zemike矩以作为动作分类的特征;然后采用基于3D Zemike矩相似性系数的码书作为分类标准模板;最后利用SVM支持向量机对KTH数据库和A-go-go舞蹈视频数据库进行分类识别,并获得了较好的识别准确率。

作者简介:陈利峰(1980—),男,河南郑州人,讲师。研究方向为体育舞蹈教学与训练研究。

问题,提出数据挖掘的图像分类算法,结果表明,采用该算法进行大型多媒体信息数据库的图像分类,对图像中特征点的检测准确度较高,输出峰值信噪比提升,提高了图像分类的准确性和鲁棒性。

参 考 文 献

[1] 陶淑苹,张续严,金光,等.基于方向WPS改进TDI CCD遥感图像清晰度评价函数[J].红外与激光工程,2013,42(8):2080-2084.

[2] 顾淑红,花均南,吕涛.激光主动探测成像中拼接算法设计与仿真研究[J].激光杂志,2016,37(2):30-33.

[3] 代具亭,汤心溢,王世勇,等.扫描型红外焦平面探测器图像实时传输系统[J].激光与红外,2016,46(4):476-480.

[4] 王小玉,张亚洲,陈德运.基于多块局部二值模式特征和人眼定位的人脸检测[J].仪器仪表学报,2014,35(12):2739-2745.

[5] 朱贺,李臣明,张丽丽,等.联合灰度阈值分割及轮廓形态识别的河道提取[J].电子测量与仪器学报,2014,28(11):1288-1296.

[6] 李武周,余锋,王冰,等.基于形态学滤波的红外图像背景补偿[J].红外技术,2016,38(4):333-336.

[7] 施晓东,刘格.一种光学遥感图像海陆分割方法[J].国外电子测量技术,2014,33(11):29-32.

[8] 罗泽峰,单广超.基于网络和虚拟多媒体技术的海战平台视景仿真实现[J].物联网技术,2015,5(3):91-92.

[9] 陆兴华,张晓军.人员图像跟踪过程中多人交叉区域防丢失方法[J].计算机仿真,2014,31(9):243-246.

[10] 张晓戎,王程成,胡光波,等.混沌差分优化数据聚类及在故障诊断中的应用[J].压缩机技术,2013(6):16-21.

参 考 文 献

[1] 傅莉,方帅,徐心和.基于计算机视觉的人体运动目标检测[J].兵工学报,2005(6):766-770.

[2] PERZANOWSKI D, SCHULTZ A C, ADAMS W, et al. Building a multimodal human-robot interface [J]. IEEE intelligent systems, 2001, 16(1): 16-21.

[3] 李昕迪.基于 Kinect 的人体姿势识别方法在舞蹈训练中的应用[D].南京:南京邮电大学,2015.

[4] 康钦谋.视频人体动作识别算法研究[D].成都:电子科技大学,2015.

[5] 刘刚.视频监控中的动作识别[D].南京:南京邮电大学,2013.

[6] 郝岩.视频数据中人体动作识别方法研究[D].南京:南京理工大学,2014.