

DOI:10.3969/j.issn.1673-4785.201305001
网络出版地址: <http://www.cnki.net/kcms/doi/10.3969/j.issn.1673-4785.201305001.html>

累积方向-数量级光流梯度直方图的人体动作识别

郭利¹, 曹江涛¹, 李平¹, 姬晓飞²
(1. 辽宁石油化工大学 信息与控制工程学院, 辽宁 抚顺 113001; 2. 沈阳航空航天大学 自动化学院, 辽宁 沈阳 110136)

摘要: 为了提高光流信息在人体动作识别系统中应用的效果和效率, 提出一种累积方向-数量级光流梯度直方图的人体动作特征表示方法。该方法首先利用 Horn-Schunck 光流算法计算图像光流, 然后将光流矢量按照不同的方向-数量级进行直方图统计, 得到单帧图像的方向-数量级的光流梯度直方图, 最后将单帧图像的直方图特征在时间维上进行累积来表示整个视频动作的特征。利用该特征在 KTH 动作视频库上进行动作识别测试, 4 个场景的混合测试得到了 87.5% 的平均正确识别率, 验证了算法的有效性。
关键词: 人体动作识别; Horn-Schunck 光流; 方向-数量级直方图; 梯度直方图
中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1673-4785(2014)01-0104-05

中文引用格式: 郭利, 曹江涛, 李平, 等. 累积方向-数量级光流梯度直方图的人体动作识别[J]. 智能系统学报, 2014, 9(1): 104-108.
英文引用格式: GUO Li, CAO Jiangtao, LI Ping, et al. Human action recognition based on accumulated orientation-magnitude histograms of optical flow[J]. CAAI Transactions on Intelligent Systems, 2014, 9(1): 104-108.

Human action recognition based on accumulated orientation-magnitude histograms of optical flow

GUO Li¹, CAO Jiangtao¹, LI Ping¹, JI Xiaofei²

(1. School of Information and Control Engineering, Liaoning Shihua University, Fushun 113001, China; 2. School of Automation, Shenyang Aerospace University, Shenyang 110136, China)

Abstract: In order to improve the recognition rate and efficiency of optical flow in the human action recognition system, a novel method for human action representation based on the accumulated orientation-magnitude gradient histograms of the optical flow is proposed in this paper. First the image optical flow is computed, and then every flow vector is counted according to the orientation-magnitude to obtain orientation-magnitude histograms of single frame image. Finally information of the video sequence can be represented by accumulating orientation-magnitude histograms in time dimension. The proposed feature is evaluated on a standard database of human actions: KTH. The experiment conducted on the four scenes demonstrates that this algorithm is effective and achieves a correct recognition rate of 87.5% with the KTH dataset.

Keywords: human action recognition; Horn-Schunck optical flow; orientation-magnitude histograms; gradient histograms

基于视频的人体动作识别是当今计算机视觉领域富有挑战性的热门研究方向。从视频序列中提取出相对有辨识力的动作特征是影响动作识别效果的重要研究内容。视频序列图像中可提取的动作特征有很多, 主要分为基于剪影的特征、基于兴趣点的特征、基于光流的特征等。
在动作识别中, 基于剪影的特征^[1-3]依赖于运

动人体的外观, 通常利用背景减除法获得精准人体剪影, 因此该类方法易受背景噪声的干扰, 对部分遮挡、视角的变化比较敏感。
兴趣点就是视频中动作突然发生变化的位置, 该位置包含丰富的运动信息。基于兴趣点的方法^[4-6]对视角、遮挡和噪声等干扰不敏感, 处理过程简单, 但此类方法的缺点是检测出来的稳定兴趣点数量太少, 且缺失全局信息。

基于光流的特征能够清晰地表示人体的动作, 受外观特征的干扰最小, 也不易受周围环境的影响, 有很好的鲁棒性, 因此该类方法较适用于实际环境

收稿日期: 2013-05-02. 网络出版日期: 2014-02-20.
基金项目: 国家自然科学基金青年基金资助项目(61103123).
通信作者: 曹江涛. E-mail: jiangtao.cao08@gmail.com.

下的应用^[7-8]。文献[7]提出了基于网格的光流表示方法,将兴趣区域按照人体比例大致划分为头颈、上半身、腿三部分,然后在每个部分利用径向直方图统计光流特征在横纵方向的光流幅度。此方法识别结果不够理想,用 Harris 角点检测运动人体,计算量大、计算速度慢。文献[8]将兴趣区域划分为均分的小区域,在每个子区域内仍然利用径向直方图统计横纵方向的光流幅度。该方法不需要对人体部分进行划分,识别结果较好,但是它在计算整帧图像的光流后再对提取的光流做处理,导致计算速度较慢。以上的光流特征表示方法均利用直方图统计小区域内光流在横纵方向的分量幅度作为特征,因此均忽略了光流的方向信息,而光流的方向信息对于区分相似动作是十分有益的。

基于以上的分析,本文提出了一种新的光流特征表示方法。该方法首先对单帧图像提取兴趣区域并进行网格化处理,然后对网格内光流矢量同时在方向和数量级上进行直方图统计,最后利用累积方向-数量级的光流梯度直方图表示整个动作视频的特征。该特征与其他方法最大的不同是光流矢量按照不同的方向和数量级进行直方图统计,而不是传统光流幅度的累加。

1 光流表示方法

通过光流计算所得到的帧间光流场,是从视频图像得到的动作原始特征。一个光流矢量可以用四维向量 (x,y,u,v) 表示,其中 (x,y) 代表了光流矢量在每帧图像中的位置信息, u,v 分别表示光流矢量的横向分量和纵向分量的大小。若一帧图像的分辨率为 $X \times Y$,则光流场中有 $X \times Y$ 个光流矢量。一个视频由多帧图像组成,提取的原始光流特征属于高维特征,计算量大,可用样本分布稀疏,不适用于动作分类识别。

本文所提出的累积方向-数量级的光流梯度直方图是对单帧光流矢量同时在方向和数量级上进行投票,然后在时间维上进行累积得到。其具体实现如图 1 所示,由以下 4 个步骤组成:1) 图像预处理得到兴趣区域;2) 计算运动估计(光流计算),平面坐标 (u,v) 转化成极坐标 (r,θ) , r,θ 分别为点 (x,y) 的光流大小和方向;3) 对所计算的光流数据进行处理得到单帧图像的方向-数量级的光流梯度直方图特征向量;4) 沿着时间轴进行特征累积,归一化处理得到累积方向-数量级的光流梯度直方图特征向量。

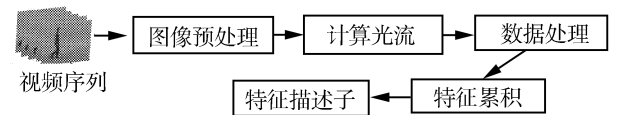


图 1 特征处理过程
Fig.1 Feature extraction process

2 累积方向-数量级光流直方图

2.1 图像预处理

提取图像中的兴趣区域是为了减少计算量。通常利用背景减除法确定出运动的大致区域,再利用中值滤波得到人体剪影,根据人体剪影信息确定兴趣区域,如图 2 中 4 张图像的虚线矩形框内所示。

根据背景减除法得到粗略的兴趣区域,有时不能完全包含运动人体,会出现兴趣区域内不包含头部、运动的手臂或腿部,如图 2(b)~(d)中虚线矩形框所示。为了避免这种情况,在确定兴趣区域时,将边界按照定宽高比例的方法进行扩展,使其包含完整的运动人体,如图 2 中实线矩形框内,本文方法只对兴趣区域内部进行处理。

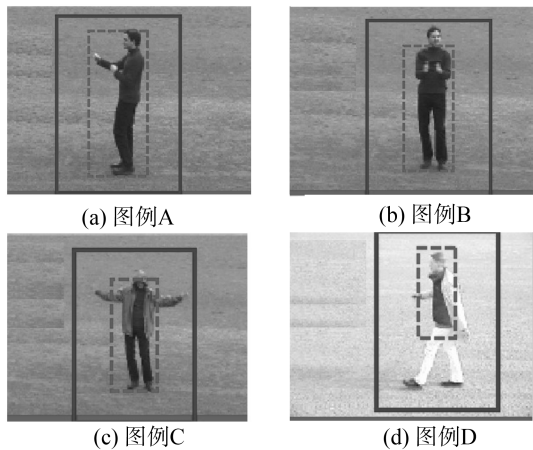


图 2 兴趣区域图例
Fig.2 The examples of region of interest

2.2 方向-数量级的光流梯度直方图

图像预处理后,将裁剪出的兴趣区域缩放到 40×40 像素大小,并转换为灰度图像。然后利用 Horn-Schunck 光流算法(HS 光流算法)计算运动光流,得出四维平面坐标的向量 (x,y,u,v) ,将其转换为极坐标 (x,y,r,θ) , r 为点 (x,y) 上光流的大小, θ 为该点的光流方向。经计算得出,光流方向的范围是 $[0^\circ, 360^\circ]$,单帧图像光流的大小范围基本在 $[0, 3]$ 之间。将光流方向分成 K 个方向柱,以 $360^\circ/K$ 为间隔,光流大小分成 M 个数量级。计算单帧图像的方向-数量级光流梯度直方图时,根据每个像素点上光流的方向和大小投票到相应的方向柱数量级上。

将光流方向分成 8 个方向柱,分别为 $[0^\circ, 45^\circ]$ 、 $[45^\circ, 90^\circ]$ 、 $[90^\circ, 135^\circ]$ 、 $[135^\circ, 180^\circ]$ 、 $[180^\circ, 225^\circ]$ 、 $[225^\circ, 270^\circ]$ 、 $[270^\circ, 315^\circ]$ 、 $[315^\circ, 360^\circ]$,光流大小分成 4 个数量级 $[0, 0.5]$ 、 $(0.5, 1.5]$ 、 $(1.5, 2]$ 、 $(2, +\infty)$ 。当极坐标为 $(0.3, 30)$,则投给第 1 个方向柱的第 1 个数量级一票。为了避免出现零的情况,将每个方向柱的数量级上的值初始化为 1。

为了使所提特征包含更多的细节,提高相似度较高的动作之间的识别精度,将每帧图像划分为 L 个无重叠的网格,在每个网格内计算方向-数量级的光流梯度直方图,最后把所有网格的直方图合并在一起就构成了整帧图像的方向-数量级光流梯度直方图。如图3所示,图3(a)为网格化的初始兴趣区域,图3(b)为子网格的光流,图3(c)为此网格的方向-数量级的光流梯度直方图,图3(d)为图3(a)对应的方向-数量级的光流梯度直方图。

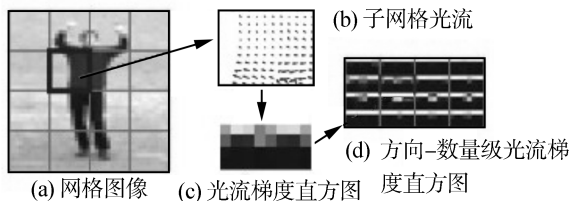


图3 方向-数量级的光流梯度直方图计算过程

Fig.3 The process of accumulating orientation-magnitude histograms of optical flow

2.3 光流梯度直方图特征提取

在获得每一帧图像方向-数量级的光流梯度直方图特征的基础上,计算累积方向-数量级的光流梯度直方图特征向量来表示 N 帧连续图像序列的运动特征。累积特征向量是第 t 帧图像的前 N 帧图像特征的累积^[9],具体计算过程如下。

一段人体动作视频 V 有 T 帧图像,即 $V = \{I_1, I_2, \dots, I_T\}$,第 t 帧图像的方向-数量级的光流梯度直方图用 $H(I_t)$ 表示。则从 I_1 到 I_N 连续 N 帧图像序列的累积方向-数量级的光流梯度直方图 H_N^a 为

$$H_N^a = \sum_{t=1}^N H(I_t) \quad (1)$$

每帧图像的方向-数量级的光流梯度直方图特征向量和累积方向-数量级的光流梯度直方图特征向量的维数相同。一个累积方向-数量级的光流梯度直方图的描述子由 L 个空间子网格、 K 个光流方向柱、 M 个光流大小数量级组成。这样就形成了 $L \times K \times M$ 维图像特征向量。

为避免不同动作或者不同人相同动作的特征之间的数量级差别较大,对所得到的上述特征进行如下的归一化处理。

$$H^a(l, k, m) \leftarrow H^a(l, k, m) / G_{lk} \quad (2)$$

$$G_{lk} = \sum_m H^a(l, k, m) \quad (3)$$

式中: $H^a(l, k, m)$ 表示第 l 个子网格、 k 个方向柱上、 m 个数量级。 G_{lk} 表示第 l 个子网格、 k 方向柱上所有数量级上值的和。

图4表示在 KTH 数据库中不同动作的累积方向-数量级的光流梯度直方图,第1行是当前帧图像,第2行是当前帧的前10帧图像进行计算得到的

累积方向-数量级的光流梯度直方图。

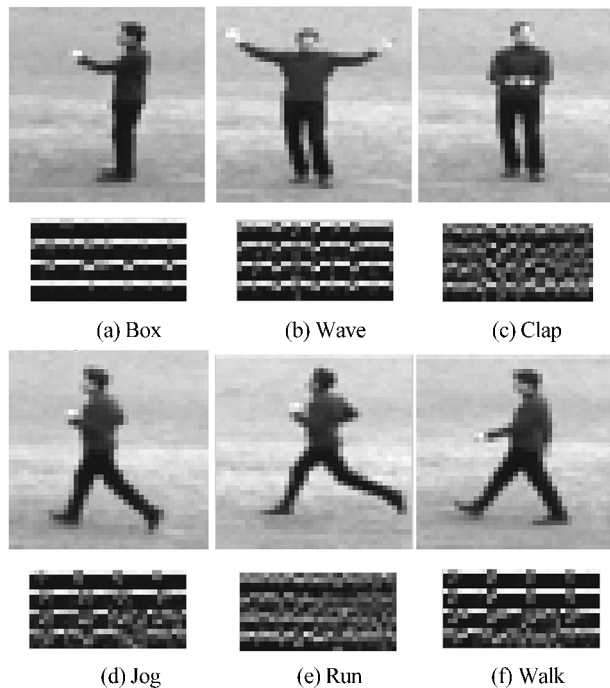


图4 累积方向-数量级的光流梯度直方图

Fig.4 The accumulated orientation-magnitude histograms of optical flow

3 实验结果与分析

目前有很多解决统计分类问题的方法,本文主要测试新特征的辨识能力,因此选用支持向量机 SVM 作为分类器,对视频中的每一帧进行类别标记,然后用投票表决的方法得到子序列的类别标号。实验利用台湾大学林智仁等^[10]开发设计的一个简单、易于使用和快速有效的 SVM 模式识别与回归的软件包 LIBSVM,在数据集上反复测试,其中的核函数采用径向基核函数。

采用公开的 KTH 动作识别数据库进行测试。KTH 数据库中有 6 种动作,分别为 Box、Handclap、Handwave、Jog、Run、Walk,每种动作由 25 个人在 4 个场景下完成,一共有 2 391 段视频序列,背景相对静止,除了镜头的拉近拉远,摄像机的运动相对轻微。在实验中,将 KTH 看作是 5 个视频库,每一个场景作为一个视频库,混合 4 个场景作为第 5 个视频库,以下所有的实验结果均是将其中 16 个人作为训练数据,余下的 9 个人为测试数据,循环 10 次得出的平均值。

3.1 参数设置

在不同的参数设置下进行测试实验。有 2 个主要的参数影响动作识别结果:网格的数目 L 和累积图像帧数 N 。

1) 网格数目 L 。预处理得到的兴趣区域被分割为 L 个网格,计算每个网格的方向-数量级光流直方图, L 的取值范围为 $[1, 64]$,取值决定精细程度及

向量的维数,在实验时累积帧数设定为 20,实验结果如图 5(a) 所示,其中横坐标的 1~9 分别对应网格数为 1×2、2×2、2×4、2×5、4×4、4×5、5×5、4×8、5×8。从图中可以看出,随着网格数的增加识别结果越来越好,但是到达某个峰值之后,随着网格数的增加识别结果反而变坏,实验在网格数为 4×4 时识别率达到最高。这表明网格数过少,得到的描述子所包含的信息量也相对少;网格数过多,每个网格对应的运动区域非常小,其中包含的运动信息量相应较少,提取的特征对识别的贡献不大。

2) 图像序列累积帧数 N 。视频序列的每一帧图像都包含动作的运动信息,但是多少帧图像能够完整地表达一个动作的属性,是一个非常重要的问题。由于动作数据库的各个视频的帧数不同,因此选取 5~35 帧作为累积帧数,分别进行动作识别的测试。如果某个视频图像序列少于累积帧数,则选取视频序列的所有帧进行累积。将网格数设定为 4×4,不同累积帧数对应的识别结果如图 5(b) 所示,可以看出识别率不是随着累积帧数的增加而增大,当累积帧数为 10 时识别率最高。这主要归因于人体动作执行过程是周期性的,累计帧数刚好对应一个人体运动周期时,得到的识别效果最好。

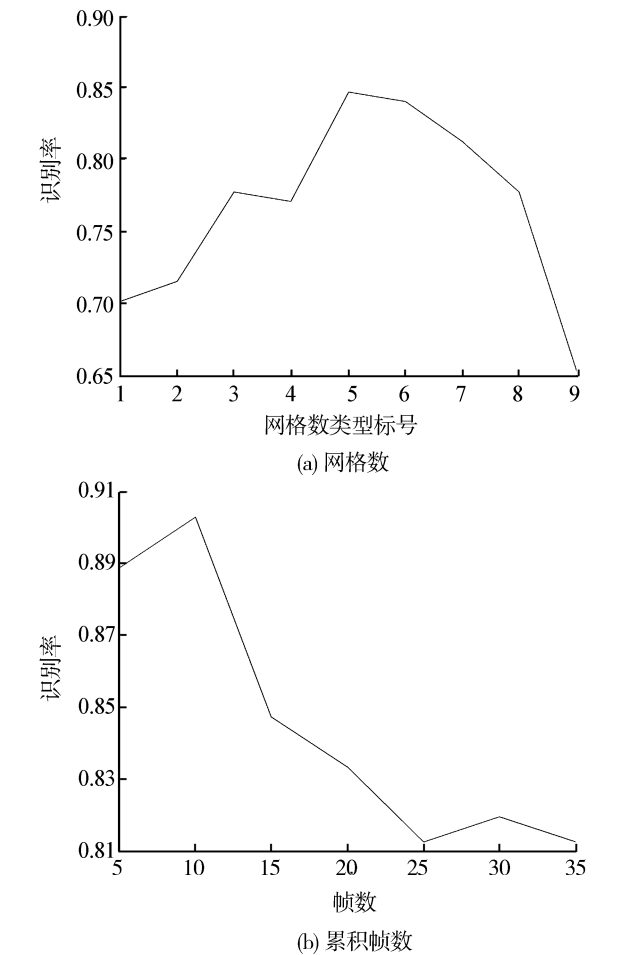


图 5 不同累积帧数及网格数的识别结果
Fig.5 The results of different frames and numbers of grids

3.2 最佳参数用于 KTH 数据库动作识别的结果

由 3.1 节得出最优参数,累积帧数为 10,网格数为 4×4。在最优参数设置下,进行不同场景的人体动作识别,其结果如表 1 所示。

表 1 不同的场景识别的结果

Table 1	Result of different Sceneries				%
场景	S_1	S_2	S_3	S_4	S_{1234}
识别结果	91.66	88.89	86.11	91.66	87.5

从表 1 可以看出,所提特征在 KTH 的数据库上得到了较理想的识别结果,场景 1 下得到了 91.66% 的正确识别率。在场景 2 和场景 3 识别结果相对低一些,主要是由于场景 2 的拍摄过程中摄像机有轻微运动或镜头拉远拉近,场景 3 下运动执行人的穿着有明显的变化,如穿上大衣、带上帽子或者背上书包等使人体外观有了较大的改变。在场景 2、3 下进行算法的测试具有一定的挑战性,本文方法分别得到了 88.89% 和 86.11% 的正确识别率,这样的识别结果是令人满意的。这也进一步证明了所提出的特征对于摄像机的轻微运动和人体外观的变化具有一定的鲁棒性。

动作识别的混淆矩阵如图 6 所示,可以看出累积方向-数量级的光流梯度直方图对挥手有较好的识别结果,正确识别率达 100%;对“跑”这一动作识别较差,多被识别为“慢跑”,因为“跑”和“慢跑”动作在运动过程中表现比较相近。

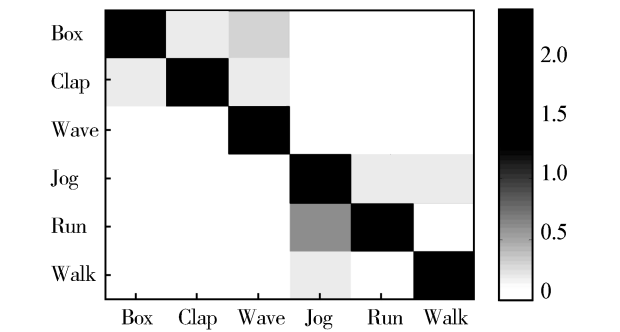


图 6 S_{1234} 不同动作识别的混淆矩阵
Fig.6 Confusion matrix on KTH: scenario S_{1234}

3.3 与其他方法比较

本文所提出方法与近期的相关方法在 KTH 数据库的识别性能比较见表 2。

从表 2 可以看出,新方法比文献[7,11]仅使用兴趣点特征和原始的分区域光流表示方法所得到的识别效果要好。与文献[12]得到的正确识别率相当,但文献[12]使用了 3 种混合特征用于动作识

别,其特征的复杂度远远高于本文方法。本文方法的特征易于提取和表示,且具有较高的可靠性。此外,利用新方法计算 100 帧图像特征仅需要 10.35 s,因此该方法可以应用于实际的系统中。

表 2 不同方法的识别结果

Table 2 Results of different methods		%
方 法	所用特征	识别结果
Liu 的方法 ^[11]	兴趣点	82.3
Danafar 的方法 ^[7]	分区域的光流直方图	85.0
Qin 的方法 ^[12]	PCRM+EOH+MOH 混合特征	87.3
本文方法	累积光流直方图	87.5

4 结束语

本文提出一种新的人体动作特征描述方法,即累积方向-数量级光流直方图。该描述方法与其他方法最大的不同是光流矢量按照不同的方向和数量级同时进行直方图统计,而不是传统光流幅度的累加。在 KTH 动作视频库上的大量测试实验证明,该特征描述方法提取速度快、识别率高,具有应用于实时系统的潜力。下一步的工作是研究如何将该人体动作特征描述方法应用在实际的智能监控系统中,进一步提高现有智能监控系统的性能。

参考文献:

[1] BLANK M, GORELICK L, SHECHTMAN E, et al. Actions as space-time shapes[C]//Proceedings of the International Conference on Computer Vision. Beijing, China, 2005: 1395-1402.

[2] GORELICK L, BLANK M, SHECHTMAN E, et al. Actions as space-time shapes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29 (12): 2247-2253.

[3] YILMAZ A, SHAH M. A differential geometric approach to representing the human actions[J]. Computer Vision and Image Understanding, 2008, 119(3): 335-351.

[4] LAPTEV I, LINDBERG T. Space-time interest points [C]//Proceedings of the International Conference on Computer Vision. Nice, France, 2003: 432-439.

[5] LAPTEV I, CAPUTO B, SCHÜLDT C, et al. Local velocity-adapted motion events for spatio-temporal recognition[J]. Computer Vision and Image Understanding, 2007, 108(3): 207-229.

[6] OIKONOMOPOULOS A, PATRAS I, PANTIC M. Spatio-temporal salient points for visual recognition of human actions[J]. IEEE Transactions on Systems Man and Cybernet-

ics—Part B: Cybernetics, 2006, 36(3): 710-719.

[7] DANAFAR S, GHEISSARI N. Action recognition for surveillance applications using optic flow and SVM[C]//Proceedings of the Asian Conference on Computer Vision. Tokyo, Japan, 2007: 457-466.

[8] TRAN D, SOROKIN A. Human activity recognition with metric learning[C]//Proceedings of the European Conference on Computer Vision. Marseille, France, 2008: 61-66.

[9] ZHANG Ziming, HU Yiqun, CHAN Syin, et al. Motion context: a new representation for human action recognition [C]//Proceedings of the European Conference on Computer Vision. Marseille, France, 2008: 817-829.

[10] FAN Rongen, CHEN Paihshuen, LIN Chihjen. Working set selection using second order information for training SVM [J]. Journal of Machine Learning Research, 2005, 6: 1889-1918.

[11] LIU Jingen, LUO Jiebo, SHAN M. Recognizing realistic actions from videos “in the wild”[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 1996-2003.

[12] QIN Yaohui, LI Hongliang, LIU Guanghui, et al. Human action recognition using PEM histogram[C]//IEEE International Conference on Computational Problem Solving. Singapore, 2010: 323-325.

作者简介:



郭利,女,1987 年生,硕士研究生,主要研究方向为模式识别、图像处理。



曹江涛,男,1978 年生,教授、博士,中国自动化学会机器人专业委员会委员和青工委委员,主要研究方向为智能方法及其在工业控制和视频信息处理上的应用。承担国家自然科学基金等项目多项,发表学术论文 40 余篇,其中被 SCI 检索 6 篇、EI 检索 22 篇。



李平,男,1964 年生,教授、博士生导师,IEEE 高级会员,中国自动化学会过程控制专业委员会委员,主要研究方向为工业过程的先进控制理论及其应用。承担国家“863”计划项目、国家自然科学基金等项目多项,发表学术论文 100 余篇,其中被SCI、EI 检索 50 余篇。