DOI:10.3969/j.issn.1673-4785.201410039
网络出版地址:http://www.cnki.net/kcms/doi/10.3969/j.issn.1673-4785.201410039.html

关键肢体角度直方图的行为识别

庄伟源[1,3],成运[2],林贤明[1,3],苏松志[1,3],曹冬林[1,3],李绍滋[1,3]
(1.厦门大学 信息科学与技术学院,福建 厦门 361005; 2. 湖南人文科技学院 通信与控制工程系,湖北 娄底 417000;
3. 福建省仿脑智能系统重点实验室,福建 厦门 361005)

摘 要:当前的姿态表示的行为识别方法通常对姿态的准确性做了很强的假设,而当姿态分析不精确时,这些现有方法的识别效果不佳。提出了一种低维的、鲁棒的基于关键肢体角度直方图的人体姿态特征描述子,用于将整个动作视频映射成一个特征向量。同时,还在特征向量中引入共生模型,用以表示肢体间的关联性。最后,设计了分层的 SVM 分类器,第 1 层主要用于选择高判别力的肢体作为关键肢体,第 2 层则利用关键肢体的角度直方图并作为特征向量,进行行为识别。实验结果表明,基于关键肢体角度直方图的动作特征具有较好的判别能力,能更好地区分相似动作,并最终取得了更好的识别效果。

关键词:角度特征;动作识别;关键肢体;角度直方图;姿态表示;行为分析;动作特征
中图分类号:TP391.4 **文献标志码**:A **文章编号**:1673-4785(2015)01-0020-07

中文引用格式:庄伟源,成运,林贤明,等. 关键肢体角度直方图的行为识别[J]. 智能系统学报, 2014, 10(1): 20-26.
英文引用格式:ZHUANG Weiyuan, CHENG Yun, LIN Xianming, et al. Action recognition based on the angle histogram of key parts[J]. CAAI Transactions on Intelligent Systems, 2014, 10(1): 20-26.

Action recognition based on the angle histogram of key parts

ZHUANG Weiyuan[1,3], CHENG Yun[2], LIN Xianming[1,3], SU Songzhi[1,3], CAO Donglin[1,3], LI Shaozi[1,3]
(1. School of Information Science and Technology, Xiamen University, Xiamen 361005, China; 2 Department of Communication and Control Engineering, Hunan University of Humanities, Science and Technology, Loudi 417000, China; 3. Fujian Key Laboratory of the Brain-Like Intelligent Systems, Xiamen 361005, China)

Abstract: The current pose-based methods usually make a strong assumption for the accuracy of pose, but when the pose analysis is not precise, these methods cannot achieve satisfying results of recognition. Therefore, this paper proposed a low-dimensional and robust descriptor on the gesture feature of the human body based on the angle histogram of key limbs, which is used to map the entire action video into an feature vector. A co-occurrence model is introduced into the feature vector for expressing the relationship among limbs. Finally, a two-layer support vector machine (SVM) classifier is designed. The first layer is used to select highly discriminative limbs as key limbs and the second layer takes angle histogram of key limbs as the feature vector for action recognition. Experiment results demonstrated that the action feature based on angle histogram of key limbs has excellent judgment ability, may properly distinguish similar actions and achieve better recognition effect.

Keywords: angle feature; action recognition; key parts; angle histogram; pose representation; action analyze; action feature

收稿日期:2014-10-24. **网络出版日期**:2015-01-13.
基金项目:国家自然科学基金资助项目(61202143);福建省自然科学基金资助项目(2013J05100,2010J01345,2011J01367);厦门市科技重点项目资助项目(3502Z20123017).
通信作者:林贤明. E-mail:linxm@xmu.edu.cn.

人体行为识别是计算机视觉领域的一个热门的研究课题,在智能视觉监控、视频检索、人机交互等领域有着广泛的应用前景,也受到了越来越多研究

学者的关注。在近 20 年的研究中,研究者们也提出了许多人体行为特征描述方法,如局部时空兴趣点[2]、密集点轨迹[3]、密集 3-D 梯度直方图[4]等,用于行为识别研究。虽然将这些方法用于行为识别研究也取得一定的成效,但是这些方法所采用的行为特征侧重于描述人体运动的底层或中层特征,缺乏语义性和直观性[5-14]。通过观察肢体在时间轴上的运动轨迹不难发现,现有这些方法对运动的描述与人类真实的运动是不相符合的。针对这些人体运动描述方法存在问题,研究者提出了基于姿态信息的方法。Sermetcan Baysal[6]提出的利用人体可见边缘信息,并转化为若干直线表示的直线姿态表示方法。L.Wang[7]提出了增强姿态估计进行动作识别。然而这 2 种方法存在部分局限性:Sermetcan 的方法中对于模糊边缘处理区分度欠缺,Li 的方法中对于近似动作如“慢跑”、“跑步”和“走路”判别性不强。

现有的基于姿态表示的行为识别方法通常是在对姿态正确分析的理想条件下进行的。而人体的姿态估计仍然是一个开放的研究问题,目前尚未得到很好地解决。而当姿态估计无法得到完整准确的结果时,目前现有的姿态估计方法也常常因此效果不佳[8]。当前姿态估计算法无法精确定位所有的身体部位时,如何利用提取到的正确的姿态信息来设计一个高判别力、有效的特征成为的本研究问题的核心。

通过对人体运动进行剖析可以发现:人体的行为动作可以分解为身体各个部位的运动,如:头部运动、手部运动、脚部运动等。但是,正如 W.Yang 在文献[1]所阐述的,各个身体部位在不同动作中所起的作用也是各不相同的。例如“拳击”动作是两只手在身体同一侧向前击出,而“挥手”动作是两只手在身体两侧左右挥动。除了这个区别外,其他身体部位的结构位置均是相似的。因而,要有效区分这 2 种动作,需要重点关注手部的运动信息。本文将这些具有高判别力的肢体称为关键肢体,并提出了一个基于关键肢体的鲁棒,有效的动作特征描述子,用于行为识别研究中。

姿态信息的动作识别方法,首先估计每一帧中人的姿态信息,然后将连续帧的姿态信息转化为沿着时间轴的姿态轨迹,再将姿态轨迹映射为动作特征,用于动作识别。随着当前姿态估计领域的发展,基于姿态的动作识别的准确率也在显而易见地提高。目前比较广泛使用的姿态估计方法包括 Poselet[9]、DPM[10]、Y.Yang[11-12]。Poselet 是一个基于实例的姿态估计方法,通过大量的模板匹配,在图像中找出与人体肢体部位姿态相一致的块。其中 Poselet 的模板数超过 1 000 个,计算复杂度远高于基于 DPM 和 Y.Yang 的算法。Y.Yang 在 DPM 和标准图案模型[15-17]的基础上,提出了一个通用的、灵活的混合模型来捕捉部位间的空间关系和共生关系,取得了很好的姿态估计效果;并且这个方法只用了 5 个模板,计算复杂度低、效率高,是当前姿态估计领域中的潮流方法。本文采用该算法来估计姿态信息。

由于当前姿态估计算法无法精确估计所有的身体部位,因此,合理设计的特征描述子可以更好地利用提取到的有效的姿态信息。以往的方法[7-13]利用部位位置信息表述姿态特征。实验证明,在对不同尺度下的动作视频做行为识别时,利用位置信息构建的姿态特征分类效果不佳,但是每个部位的角度信息具有尺度不变性。同时选用的姿态估计算法在部位间引入空间限制,这使得仅利用各部位角度信息表述姿态特征成为可能。因此,在设计动作特征时舍弃位置信息,仅提取角度信息。另外,在动作建模层面,文献[6-7]利用聚类算法在训练样本中生成一系列标准姿态,并在测试视频中每一帧的姿态信息中找出其最相近的标准姿态。这种方法在构建标准姿态时包含了所有部位信息,容易受到没有准确估计的部位信息的影响,不够鲁棒。考虑了另一种策略,即对每个部位单独构建特征,选取关键部位并级联组成动作特征向量。同时,受同一肢体的上部和下部(如大臂和小臂)的角度有相关性联系的启发,在设计特征时引入共生关系并称之为成对肢体特征。

1　关键肢体角度直方图的理论框架

图 1 显示了关键肢体角度直方图的基本流程。首先,采用 Y.Yang 提出的姿态估计算法对输入视频进行姿态估计,获取每一帧各个部位点的位置信息。然后,本文将具有生理关联性的部位点连接,并定义为肢体,利用部位点对的位置信息来计算肢体位置和角度信息。根据各个肢体对特定动作的判别力大小,选取判别力大的手臂肢体和腿部肢体共 8 个部位作为候选关键肢体。

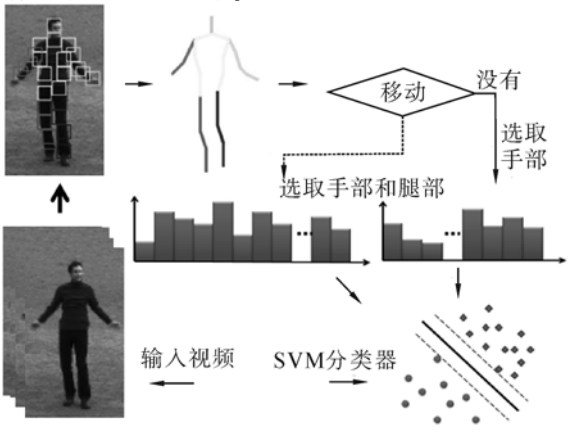


图 1　基于关键肢体角度直方图的动作识别算法基本流程

Fig.1　The basic flow of proposed method

设计一个2层的SVM分类器。第1层分类器用于从候选关键肢体当中选出关键肢体,将动作者躯干的水平位置信息分布直方图作为特征,根据这个特征将动作划分成两大类:非移动类和移动类。非移动类是指除了手部运动外,其他肢体运动较少的行为,只需选用手臂作为关键肢体;移动类则是腿部也有运动,因此需要将腿部也作为关键肢体。第2层SVM分类器中,为解决如图2显示的不同尺度下姿态表示问题,仅选用肢体的角度信息做姿态表示,并利用角度信息定义各个肢体在每一帧中的运动类型。然后,设计独立肢体特征和成对肢体特征2种运动类型直方图的统计策略,用以统计各个肢体在整个视频中的不同的运动类型的出现次数。最后,将级联的关键肢体的角度直方图作为动作特征,用于做动作识别。

图2 尺度变化实例

Fig.2 Example of scale variation

2 姿态特征与动作特征描述子

2.1 姿态信息估计

人体姿态通常呈现出高形变的特点,其类内表观差异性大,Y.Yang提出的姿态估计[11]具有表观变化一致性,允许姿态中人体部位发生轻微偏移,并可以利用少数的模板有效地估计姿态。该方法和图案结构模型一样,都运用了多成分的混合模型,其中每个成分表示训练数据集中某种姿态数据,并在此基础上引入共生模型来表示部位间的共生关系。该姿态估计模型包含3个模型:混合模型、成对弹簧模型和共生模型。混合模型是无方向图形结构的混合;成对弹簧模型是成对部位间的空间限制;共生模型是同一肢体上的部位在方向上的一致性限制。

姿态估计模型 输入一帧图像 l_{rh},输出所有部位的位置信息 L(设部位 i 的位置为 l_i)。其位置信息的计算公式为

$$S(I,L,M) = \sum_{i\in V} b_i^{t_i} + \sum_{i,j\in E} b_{i,j}^{t_i t_j} + \sum_{i\in V} \alpha_i^{m_i} \times \varphi(I,l_i) + \sum_{i,j\in E} \beta_i^{m_i m_j} \times \psi(l_i - l_j)$$

式中:$b_i^{t_i}$ 表示部位 i 的特定类型,$b_{i,j}^{t_i t_j}$ 表示部位类型的特定共生模型。$G = (V,E)$ 是一个相关联部位间设置了一致性关系的,K 个节点的关系图。$\varphi(I,l_i)$ 是在图片 I 中 l_i 像素位置提取的特征向量(如HOG特征[18-19])。$\alpha_i^{m_i}$ 是 m_i 混合的部位 i 的一元模板。$\psi(l_i - l_j)$ 是 l_i 和 l_j 的空间特征。$\beta_i^{m_i m_j}$ 是 m_i 混合的部位 i 和 m_j 混合的部位 j 之间的成对弹簧限定。

从图1左上图可见,在肢体间添加了中间部位点,以27个部位点位置信息取代通常的14个标准部位点(定位在身体14个关节点处,如肩、肘、手腕等)。其在肢体间添加了中间部位点(中上臂、中下臂等)。每一帧的部位点位置信息为 $l_v = (x_v, y_v)$,$v \in \{1,2,\cdots,27\}$。对于 N 帧的视频数据,可获得 $27 \times 2 \times N$ 的姿态信息矩阵。

2.2 姿态表示和候选关键肢体

获取到关节点位置信息矩阵之后,需要对其进行编码,映射为姿态特征,并从中选出关键肢体。应用姿态估计从不同动作获取的姿态信息如图3线段所示,相邻部位点间用线段连接后,在视觉上接近于骨架信息。将这些线段分别定义为小臂、大臂、躯干、小腿、大腿和头部等肢体,如用直线将右手的部位点 l_{rh} 和中下臂的部位点 l_{rla} 连接并定义为右小臂,设为 p_{rla}。计算其对应的线段中点位置 (x_{rla}, y_{rla}) 和相对于水平轴的角度 θ_{rla}。方法如下:

$$p_{rla} = (x_{rla}, y_{rla}, \theta_{rla}) = \left(\frac{x_{l_{rh}} + x_{l_{rla}}}{2}, \frac{y_{l_{rh}} + y_{l_{rla}}}{2}, \tan^{-1}\frac{y_{l_{rh}} - y_{l_{rla}}}{x_{l_{rh}} - x_{l_{rla}}}\right)$$

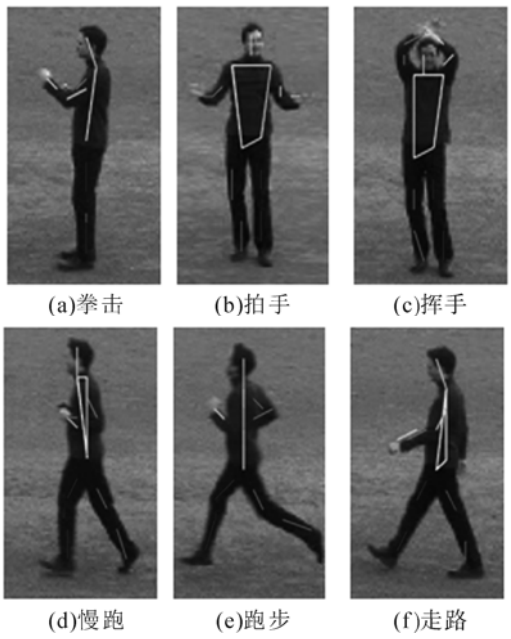


图3 应用姿态估计从不同动作获取的姿态信息

Fig.3 Using pose estimation to get different configuration for different human actions

实验分别对6种行为进行测试:拳击、拍手、挥

手、慢跑、跑步和走路,文中需要从中找出关键肢体。通过观察可以发现,拳击、拍手和挥手之间的区别主要集中在手部运动。而慢跑、跑步和走路的区别集中于手部运动和腿部运动中,剩下的肢体(头和躯干)区分度不高,提供很少的信息熵。因此,仅选取四肢的 8 个肢体作为候选关键肢体,设为 $p_i = (x_i, y_i, \theta_i)$, $i \in \{1,2,\cdots,8\}$ 。

2.3　肢体角度直方图

在第 1 层分类器中,主要任务是将动作分为非移动类和移动类两大类动作,本文提取了躯干的水平位置分布信息并用直方图特征表示,用来判断人是否发生移动。

在第 2 层分类器中,需要将每个部位的位置映射为特征向量。LI Wang[7] 和 WANG Jiang[13] 都采用了相对部位特征。将躯干部位 $p_{\text{torso}} = (x_{\text{torso}}, y_{\text{torso}}, \theta_{\text{torso}})$ 作为参照点,其他部位 p_i 映射为相对部位特征 $\triangle p_i = (x_i - x_{\text{torso}}, y_i - y_{\text{torso}}, \theta_i)$ 。这种特征在处理尺度不同的动作视频时(图 2 所示,与摄像头的距离不同导致的人在视频中尺度不同)分类效果不佳。同时由于不同动作中,人体姿态形变差异较大,目前没有有效的解决位置归一化的方法。注意到部位的角度具有尺度不变的特性,同时由于姿态估计模型中各个部位间存在空间限制关系,这也为仅用肢体角度信息表述姿态提供了基础,因此每个肢体表示为 $\bar{p}_i = (\theta_i)$ 。

一段十几秒的视频当中包含着三四百帧图像信息,如果将每一帧的关键肢体角度特征级联,所组成的特征向量维度很大,并且会降低姿态识别准确率。为了降低维度,实验当中分别对每个肢体定义 M 个运动类型 $t_{k,\bar{p}_i}$, $k \in \{1,2,\cdots,M\}$ 。其中具有相似角度的部位 $\bar{p}_i$ 被指定为同一运动类型。运动类型的判别方法是:首先,将部位的角度空间(手臂肢体的角度空间大小为 $0 \sim 4 \times 18 + 2 \times 9 \times 9$,腿部肢体的角度空间大小为 $\pi \sim 2\pi$)分为等长的 M 个区间;然后,假设第 f_n 帧中,部位 $\bar{p}_i$ 的角度值 $\theta_{\bar{p}_i}$ 落在角度空间的第 k 个区间内,那么它的运动类型就判定为 $t_{k,\bar{p}_i}$。具体公式如下:

$$t_{k,\bar{p}_i} = \text{floor}\left(\frac{\theta_{\bar{p}_i}}{2\pi/N}\right)$$

在一段视频当中,以帧为单位,判断每一帧上所有关键肢体的运动类型,然后通过对整个视频当中不同运动类型出现次数的统计,可以构建 2 种直方图特征:独立肢体特征和成对肢体特征。独立肢体特征也就是独立地统计每个部位的运动类型出现的次数。设部位 $\bar{p}_i$ 的独立肢体特征 $H^{\text{Ind}}_{\bar{p}_i}$ 的维度为 $M_{\bar{p}_i}$,视频中 $\bar{p}_i$ 的运动类型 $t_{k,\bar{p}_i}$ 出现的次数,统计在第 k 维。成对肢体特征是一种特征引入了共生关系的特征,通过统计成对肢体的运动类型来表示。所谓的共生关系是:属于同一个四肢的肢体 $\bar{p}_i$ 和肢体 $\bar{p}_j$ (如右大臂和右小臂都属于右臂),他们所对应的运动类型存在相关性。基于这种思想提出成对肢体特征,其具体步骤如下:设每个成对肢体特征 $H^{\text{pair}}_{\bar{p}_i,\bar{p}_j}$ 的维度为 $M_{\bar{n}} \times M_{\bar{n}}$,若第 f_n 中 $\bar{p}_i$ 的运动类型是 $t_{k,\bar{n}}$ 而 $\bar{p}_j$ 为 $t_{l,\bar{n}}$,则在 $((k-1) \times M_{\bar{n}} + l)$ 维统计。

对非移动类动作(包括拳击、拍手和挥手),使用每个部位设置 10 个运动类型,并使用成对肢体特征描述手臂部位。特征向量的维度为 $2 \times 10 \times 10$,为 200 维。对移动类动作(包括慢跑、跑步和走路),用独立肢体特征表示手臂部位其中每个部位包含 18 个运动类型,而腿部部位用成对肢体特征表示,其中每个部位包含 9 个运动类型。对手臂应用独立肢体特征而不是成对肢体特征的原因在于:经观察发现,移动类的动作中,脚部部位的姿态估计准确率更高,而由于手臂部位接近躯干,因此无法准确估计手臂的所有部位,在这种情况对成对肢体特征的干扰较大而独立肢体特征更具有鲁棒性。整个动作向量的维度是 $4 \times 18 + 2 \times 9 \times 9$,为 234 维。特征提取后同一进行归一化处理。

3　实验结果

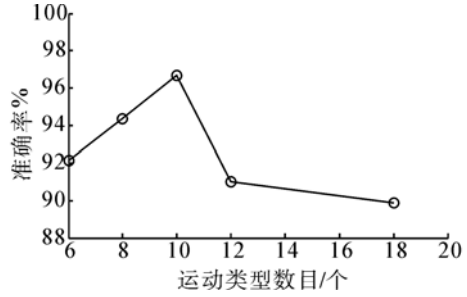
实验部分采用 KTH action dataset 数据集[20] 做测试。KTH 数据集包含了 600 个灰度视频,其中共 6 类动作:拳击、拍手、挥手、慢跑、跑步和走路。这些动作分别由 25 个参与者在 4 种不同的场景(户外、户外以及尺度变化、户外以及换其他服装和室内)完成。视频空间分辨率为 160 × 120。

选用 70%的视频作为训练集,并采用交叉验证的方法用对 SVM 模型参数进行优化。剩下 30%视频作为测试集,重复 4 次实验取平均实验结果。在姿态估计部分,人工对每个动作提取 15 帧图片,并标注部位点位置,用以训练姿态估计模型。在姿态估计中,尝试加入视频姿态估计[21] 的方法以引入时间限制模型。实验验证部位点在某一帧的定位情况会受到其他帧定位质量的影响,最终可能生成低判别力的特征。

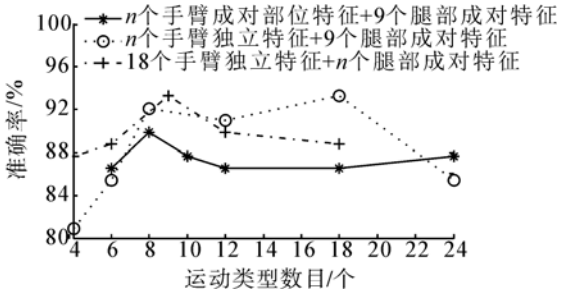
3.1　肢体运动类型的数目对比试验

为验证运动类型的数目对分类效果的影响,在第 2 层分类器中,分别改变非移动类(图 4(a))和移动类运动类型的数目。由于在移动类中选取手臂部位和腿部部位作为关键肢体,因此分别只改变腿部

运动类型数目(图4(b)点划线))或手臂运动类型数目(图4(b)虚线),保持另一关键肢体运动类型数目不变。



(a) 非移动类动作



(b) 移动类动作

图4 运动类型数目对实验准确率的影响

Fig.4 The affect of the number of movement types on our methods´s performance

图4中,如果类型数目低于某个值时,分类准确性随着类型数目的增加而提升;但如果类型数目超过某个值时,分类准确率就会下降。证明了更多的运动类型可以更好地表示动作,但当类型数目超过一定值时,鲁棒性会下降,姿态估计的质量对动作识别的影响更大。

3.2 移动类用成对肢体特征表示手臂

同时,为验证对手臂部位的定位误差对共生关系以及分类准确性的影响,在移动类中,用成对肢体特征表示手臂,替代原先的独立肢体特征。并且实验过程中,固定腿部肢体运动类型数目为9,改变手臂肢体运动类型数目。实验结果如图4(b)中实线所示,由于姿态估计算法尚不够获得完整准确的结果,成对肢体特征容易受到单个部位定位误差的影响,因此鲁棒性不如独立肢体特征。

3.3 用聚类算法替代角度直方图特征

除此之外,针对移动类动作,文中还对比了角度特征和相对位置特征2种姿态描述子的效果,由于相对位置特征不适合用本文的角度直方图特征,因此应用了Sermetcan[6]和LI Wang[7]的方法,利用K-Means聚类[22]对所有训练集的姿态描述子进行聚类,生成的聚类中心即为标准姿态。对于测试集,对每帧的姿态描述子利用KNN算法,寻找最相似的标准姿态,并用直方图统计各个标准姿态的出现次数。

表1 2种姿态描述结合K-Means聚类与K-NN在移动类的动作的准率

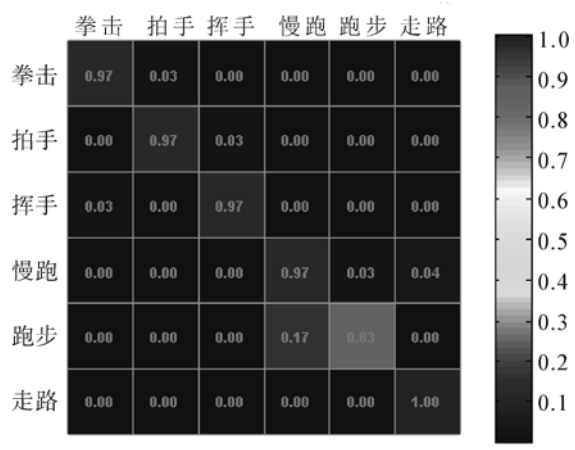
Table 1 Accuracy of pose feature together with K-Means and K_NN on Moving category: (a) angle feature; (b) relative position feature

姿态描述	慢跑	跑步	走路
相对位置特征	0.77	0.57	0.93
角度特征	0.80	0.69	0.90

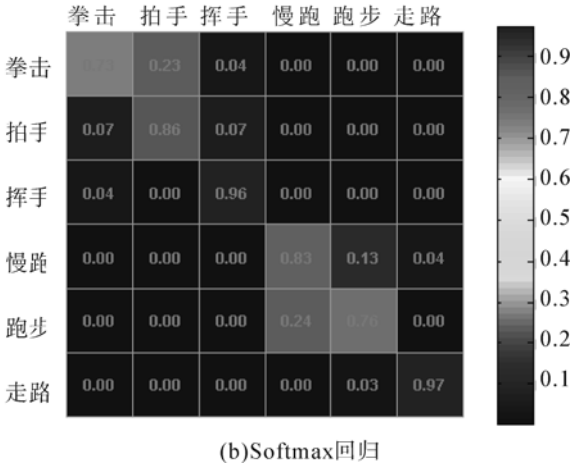
实验结果如表1所示,与相对位置特征相比,角度特征可以更好地区分不同尺度下的动作。但同时,这2种特征,与采用聚类算法生成标准姿态相比,本文方法中对每个部位独立构建特征并级联成行为特征的策略可以有效降低计算复杂度,且具有更高的判别力。

3.4 本文方法与当前行为识别算法对比

实验当中还对2种当前较为常用的分类器效果进行对比:SVM分类器和Softmax Regression分类器,实验结果如图5的混淆矩阵所示。其中,SVM分类器的动作识别的平均准确率达到94.9%,而Softmax Regression分类器的准确率为85.4%。



(a)支持向量机



(b)Softmax回归

图5 KTH数据集上识别效果的混淆矩阵

Fig.5 Confusion matrices generated by two classifiers

为了验证本文方法的准确率,分别与当前的主流算法进行对比。表 2 是本文方法和基于姿态的行为识别方法,在 KTH 动作数据集上具体动作的准确率,观察可得,本文方法在各个动作的识别中都有了较大的提升。表 3 是本文方法与当前经典的低维或者中维局部特征的动作识别方法在 KTH 数据集上的平均准确率实验结果对比。其中,在跑步动作中常无法准确识别,主要在于其骨架结构与慢跑近似,甚至肉眼也无法准确分辨。

表 2　基于姿态的动作识别算法在 KTH 动作数据集的准确率

Table 2　Recognition accuracy on KTH action dataset of pose-based method　/%

方法	拳击	拍手	挥手	慢跑	跑步	走路
Li Wang[7]	0.76	0.88	0.96	1.0	—	—
Sermetcan [6]	0.90	0.96	0.94	0.87	0.98	0.84
本文方法	0.97	0.97	0.97	0.97	0.83	1.0

表 3　动作识别算法在 KTH 数据集的平均准确率

Table 3　Recognition accuracy on KTH action dataset

方法	准确率 %
Laptev et al [23]	91.8
Bregonizo et al [24]	93.2
Liu and Shah [25]	94.3
Wu et al. [26]	94.5
Gilbert et al [27]	94.5
本文方法	94.9

4　结束语

由于姿态估计算法本身一直是一个复杂的研究问题,基于姿态的行为识别方法一直无法获得满意的效果。结合当前最优的姿态估计算法,我们设计了 2 层的分类器,第 1 层分类器用于选取关键肢体;在第 2 层分类器中,为解决不同尺度下的动作分类,仅用角度信息表示姿态,并提出了关键肢体角度直方图的动作特征,在姿态估计尚存在一定程度的估计误差时,依然能较为准确的识别动作。

当前对每帧独立地进行姿态识别,并且在构建动作特征时,仅用空间信息进行行为识别,已获得较精确的结果。如何在动作特征中引入前后时间关系,并保证特征的鲁棒性,使其可以应用于更为复杂的动作场景中,会是将来研究的重点方向。

参考文献:

[1] YANG Weilong, WANG Yang, MORI G. Recognizing human actions from still images with latent poses[C]//IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010: 2030-2037.

[2] LAPTEV I. On space-time interest points[J]. International Journal of Computer Vision, 2005, 64(2/3): 107-123.

[3] WANG H, KLASER A, SCHMID C, et al. Action recognition by dense trajectories[C]//IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 3169-3176.

[4] KLASER A, MARSZALEK M, SCHMID C. A spatio-temporal descriptor based on 3d-gradients[C]// British Machine Vision Conference. Leeds, UK, 2008: 275-285.

[5] SADANAND S, CORSO J. Action bank: a high-level representation of activity in video[C]//IEEE Conference on Computer Vision and Pattern Recognition. [s.l.], 2012: 1234-1241.

[6] BAYSAL S, DUYGULU P. A line based pose representation for human action recognition[J]. Signal Processing: Image Communication, 2013, 28(5): 458-471.

[7] LI Wang, LI Cheng. Human action recognition from boosted pose estimation[C]//International Conference on Digital Image Computing: Techniques and Applications. Sydney, AU, 2010: 308-313.

[8] 徐光祐,曹媛媛. 动作识别和行为理解综述[J]. 中国图像图形学报, 2009,14(2): 189-195.
XU Guangyou, CAO Yuanyuan. Action recognition and activity understanding: a review[J]. Journal of Image and Graphics, 2009, 14(2): 189-195.

[9] BOURDEV L, MALIK J. Poselets: body part detectors training using 3-D human pose annotations[C]// IEEE International Conference on Computer Vision. [s.l.], 2009: 1365-1372.

[10] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multi scale, deformable part model[C]//IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8.

[11] YANG Y, RAMANAN D. Articulated pose estimation with flexible mixtures-of-parts[C]//IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 1385-1392.

[12] YANG Y, RAMANAN D. Articulated human detection with flexible mixtures of parts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2878-2890.

[13] WANG Jiang, LIU Zicheng, WU Ying. Ming acionlet ensemble for action recognition with depth cameras[C]// IEEE Conference on Computer Vision and Pattern Recognition. [s.l.], USA, 2012: 1290-1297.

[14] 雷庆,李绍滋. 动作识别中局部时空特征的运动表示方法研究[J]. 计算机工程与应用, 2010, 46(34): 7-10.
LEI Qing, LI Shaozi. Research on local spatio-temporal features for action recognition[J]. Computer Engineering and Applications, 2010, 46(34): 7-10.

[15]EPSHTEIN B, ULLMAN S. Semantic hierarchies for recognizing objects and parts[C]//IEEE Conference on Computer Vision and Pattern Recognition.[S.l.], 2007: 1-8.

[16] FELZENSZWALB P, HUTTENLOCHER D. Pictorial structures for object recognition[J]. International Journal of Computer Vision, 2005, 61(1): 55-79.

[17]FELZENSZWALB P, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[18]DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE Conference on Computer Vision and Pattern Recognition.[S.l.], 2005: 886-893.

[19]曲永宇,刘清,郭建明. 基于 HOG 和颜色特征的行人检测[J]. 武汉理工大学学报, 2011, 33(4):134-141.
QU Yongyu, LIU Qing, GUO Jianming. HOG and color based pedestrian detection[J]. Journal of Wuhan University of Technology, 2011, 33(4): 134-141.

[20]LAPTEV I, CAPUTO B, SCHULDT Christian. Local velocity-adapted motion events for spatio-temporal recognition [J]. Computer Vision and Image Understanding, 2007, 108: 207-229.

[21]BURGOS-ARTIZZU X P, HALL D, PIETRO P, et al. Merging pose estimates across space and time[C]//British Machine Vision Conference. Bristol, UK, 2013: 58-69.

[22]王千,王成,冯振元. K-means 聚类算法研究综述[J]. 电子设计工程, 2012, 20(7): 21-24.
WANG Qian, WANG Cheng, FENG Zhenyuan. Review of K-means cluster algorithm[J]. Electronic Design Engineering, 2012, 20(7): 21-24.

[23]LAPTEV I, MARSZALEK M, SCHMID C, el al. Learning realistic human actions from movies[C]//IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8.

[24]BREGONZIO M, GONG S, XIANG T. Recognizing action as clouds of space-time interest points[C]//IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 1948-1955.

[25]LIU J, SHAH M. Learning human actions via information maximization[C]//IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8.

[26]WU X, XU D, DUAN L, et al. Action recognition using context and appearance distribution features[C]//IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 489-496.

[27]GILBERT A, ILLINGWORTH J, BOWDEN R. Fast realistic multi-action recognition using mined dense spatio-temporal features[C]//IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 925-931.

[28]凌志刚,赵春晖,梁彦. 基于视觉的人行为理解综述[J]. 计算机应用研究, 2008, 25(9): 2570-2578.
LING Zhigang, ZHAO Chunhui, LIANG Yan. Survey on vision-based human action understanding[J]. Application Research of Computers, 2008, 25(9): 2570-2578.

作者简介:

庄伟源,男,1990 年生,硕士研究生,主要研究方向为人体行为识别、计算机视觉、深度学习。

林贤明,男,1980 年生,助理教授,博士,主要研究方向为人体行为识别、移动视觉搜索、计算机视觉、模式识别。

李绍滋,男,1963 年生,教授,博士生导师,博士,福建省人工智能学会副理事长兼秘书长,主要研究方向为运动目标检测与识别、自然语言处理与多媒体信息检索等。发表学术论文 160 余篇,其中被 SCI 检索 16 篇、被 EI 检索 142 篇。