

DOI:10.11992/tis.201505006

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.tp.20151111.1700.010.html>

分层结构的双人交互行为识别方法

姬晓飞, 王昌汇, 王扬扬

(沈阳航空航天大学 自动化学院, 辽宁 沈阳 110136)

摘要:针对日常生活中双人交互行为因运动区域难以分割,造成无法准确识别的问题,提出了一种基于分层结构的双人交互行为识别方法。该方法首先按照交互行为双方身体是否接触作为分界点,将整个交互行为分为开始阶段、执行阶段和结束阶段。将开始阶段与结束阶段左右两侧人体所在矩形区域分别提取作为该兴趣区域,将执行阶段双人所在矩形区域整体提取作为感兴趣区域,分别提取 HOG 特征。使用 1NN 分类器获得每个阶段的每个对象的识别概率,最终通过加权融合各个阶段各个对象的识别概率实现对该交互行为的识别。利用 UT-interaction 数据库对该方法进行测试的实验结果表明,该方法实现简单,并具有良好的识别效果。

关键词:计算机视觉;交互动作;动作识别;方向梯度直方图;分层模型;最近邻分类器;UT-interaction 数据库;加权融合

中图分类号:TP391.4 **文献标志码:**A **文章编号:**1673-4785(2015)06-0893-08

中文引用格式:姬晓飞,王昌汇,王扬扬. 分层结构的双人交互行为识别方法[J]. 智能系统学报, 2015, 10(6): 893-900.

英文引用格式:JI Xiaofei, WANG Changhui, WANG Yangyang. Human interaction behavior-recognition method based on hierarchical structure[J]. CAAI Transactions on Intelligent Systems, 2015, 10(6): 893-900.

Human interaction behavior-recognition method based on hierarchical structure

JI Xiaofei, WANG Changhui, WANG Yangyang

(School of Automation, Shenyang Aerospace University, Shenyang 110136, China)

Abstract: To solve the problem of double interaction behavior recognition, in this paper we propose a novel interaction behavior-recognition method based on hierarchical structure. First, the interactive behavior of both areas of body contact is determined as the cut-off point. The interaction process is then divided into three stages—the start, execution, and end stages. We extract the left and right human body regions in the start and end stages, respectively. Both human body regions are extracted as a whole in the execute stage. Next, we utilize a histogram of oriented gradients (HOG) descriptor to describe information on the regions of interest of each stage. Thereafter, we use the nearest neighbor classifier to obtain the recognition probability of each object in each stage. Finally, we obtain the recognition result from the weighted fusion of this recognition probability. The experimental results, using the UT-interaction dataset, demonstrate that the proposed approach is easy to implement and has good recognition effect.

Keywords: computer vision; human interaction; action recognition; histogram of oriented gradient; layered model; nearest neighbor classifier; ut-interaction database; weighted fusion

基于视频的人体动作识别研究是计算机视觉领

域的重要研究方向,具有广阔的应用前景^[1-2]。根据行为的复杂程度,一般可以把人的行为分成 4 类:局部行为、单人行为、双人交互行为、人群行为^[3]。目前,在单人的行为识别领域已经有很多研究成果,而对于双人交互行为的研究还比较少。其主要原因

收稿日期:2015-05-06. 网络出版日期:2015-11-11.

基金项目:国家自然科学基金资助项目(61103123);辽宁省教育厅科技研究基金资助项目(L2014066).

通信作者:姬晓飞. E-mail:jixiaofei7804@126.com.

在于,双人交互行为的识别除了具有动作识别所面临的主要困难(图像采集设备的抖动、场景内光照强度的变化、非主要目标的遮挡等),其最主要的问题在于在双人交互行为中,如何描述动作执行双方的身体姿态以及交互行为的复杂的时空关系。目前,对于双人交互行为的识别方法,大致可分为以下 2 种,基于整体的交互动作识别和基于个体分割的交互动作识别。

基于整体的交互动作识别方法主要是将交互行为执行双方看做一个整体,通过描述该整体的时空关系来代表其所表示的双人交互行为的特征,通过将待测试样本与训练模板的特征进行匹配,来实现交互动作的识别和理解^[4]。此类方法多关注的是交互动作的时空特征表示及相应的时空匹配算法的研究。Yu 等^[5]采用金字塔时空关系匹配核对交互动作进行识别。Yuan 等^[6]提出构造时空上下文核函数进行交互视频的匹配识别。这些方法特征提取过程实现简单,但交互动作时空关系匹配的计算复杂度较高,且识别的准确度不高。近年来,一些研究者引入相对复杂的时空特征表示方法对交互行为整体进行描述,以期提高识别的准确度。Burghouts 等^[7]通过引入时空布局(spatio-temporal layout)描述来提高时空特征的类间区分能力,大大提高了交互行为识别的准确性;Li 等^[8]提出了基于 GA 训练的随机森林方法及有效的时空匹配方法实现交互行为的识别与理解。综上所述,基于整体的交互动作识别与理解方法将交互动作作为一个单人动作处理,无需对交互动作的特征进行动作个体的分割,处理思路简单,但是由于该类方法无法准确表示交互动作中交互的内在属性,因此其识别的准确性有限,往往需要十分复杂的特征表示及匹配方法来保证识别的准确性。

基于个体分割的交互行为识别与理解方法,就是将交互行为理解为单个人的子动作之间的时空组合,在识别的过程中先识别交互行为中单个个体动作的含义,再结合 2 个个体之间的时空关系,获得最终的识别结果。Kong 等^[9]提出训练基于 SVM 的识别模型对交互动作进行识别;Slimani^[10]等提出了一种基于共生视觉词典的方法。该类方法原理简单、容易实现,但识别准确性不高。总之,基于个体分割的交互动作识别与理解方法或者需要对人体的肢体部分进行跟踪检测,或者需要对原子动作进行识别,在复杂的交互行为场景下,由于具有遮挡等因素的影响,准确地得到人体部分区域并准确地识别原子动作是很难保证的。

基于以上分析,本文将能够完整表述全局信息

的基于整体的识别方法与能够准确描述交互双方行为的基于个体分割的识别方法相结合,提出了一种分层识别方法。该方法采用将整个交互行为过程分为 3 层,在交互行为的开始阶段与结束阶段分别识别交互行为双方的个体动作,再融合判断;在交互行为的执行阶段,将双人交互行为作为一个整体进行识别判断;最终加权融合 3 个阶段的分类概率。

1 分层处理

近年来,针对双人交互行为识别的特征提取,可以分成 2 种类型:1)将双人交互行为整体作为感兴趣区域进行处理^[4-8],这种方法很好地呈现了双人交互行为的全局特征,但是无法表现交互行为的内在属性,所以识别的准确性有限;2)将交互行为的双方分割开来,分别对其所进行的动作进行识别,然后再将识别结果融合判断^[9-10]。这种方法虽然很好地表达双人交互行为中人与人之间的运动关系,但是在复杂的交互场景下,由于交互在动作的执行过程中存在身体遮挡等因素,直接影响了运动目标区域获取的准确性,这对识别单人原子动作产生了较大影响。

为了充分表达双人交互行为中双方的运动关系,并且避免交互行为中的遮挡影响单人原子动作的识别,文中采用分层识别的方法对交互行为进行识别,其过程如图 1 所示。

首先,将交互行为分为 3 个阶段:交互行为开始阶段、交互行为执行阶段和交互行为结束阶段,对于各个阶段的感兴趣区域的分割提取,文中采用了不同的方法。

1)交互行为开始阶段:在交互行为开始阶段,交互行为双方的位置关系是由远及近的,在这一过程中,通过帧间差分的方式,可以获得交互行为双方的剪影信息,根据 2 个剪影的边界信息,可以分别获得以交互行为双方为主,冗余信息极少的感兴趣区域,其过程如图 2 所示。

2)交互行为执行阶段:在以交互行为双方为主的 2 个感兴趣区域间的距离为 0 时,交互行为双方身体出现接触,从这一时刻开始,意味着双人交互行为的准备阶段结束,双人交互行为进入执行阶段。在这一个阶段中,为了避免由于交互行为双方身体接触遮挡对于单人原子动作分割的影响,本文将双人交互行为整体所在区域作为感兴趣区域,进行分割提取操作,文中采用帧间差分的方法,基于剪影特征的边界信息分割提取双人交互行为的感兴趣区域。其过程如图 3 所示。

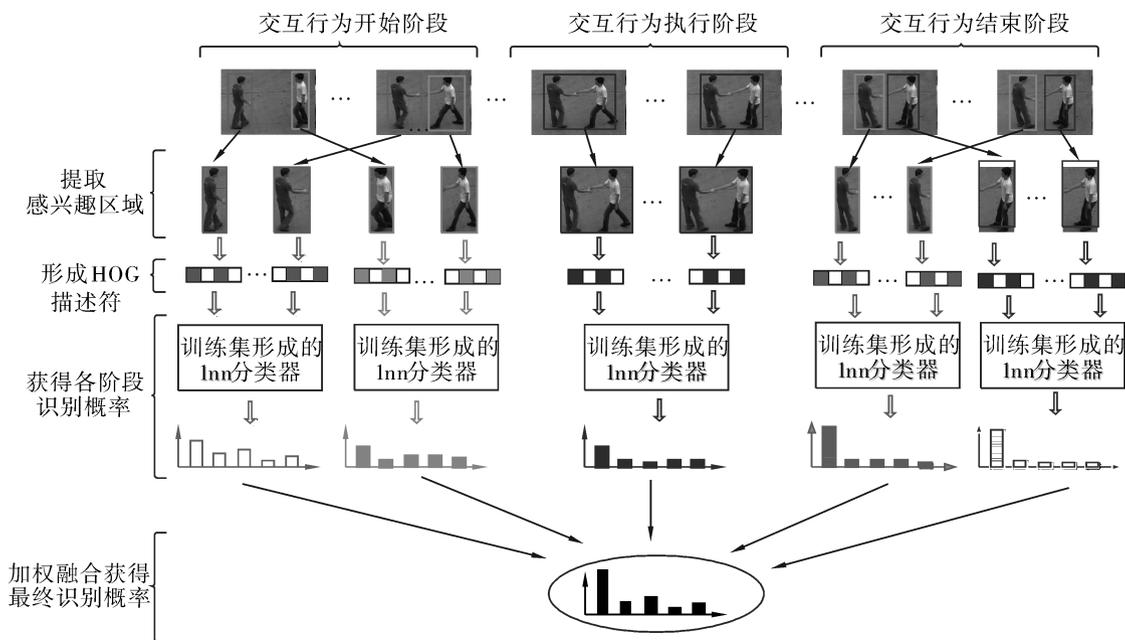


图 1 分层识别方法结构图

Fig.1 Structure diagram of hierarchical recognition method

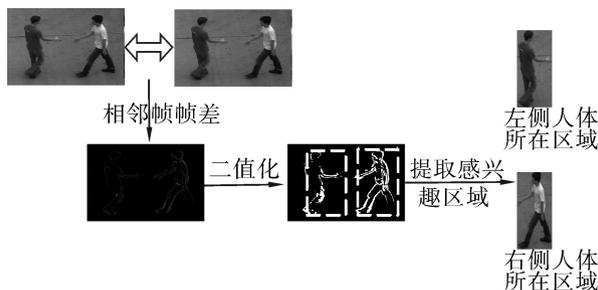


图 2 帧差法提取剪影所在感兴趣区域

Fig.2 Frame difference extract the ROI

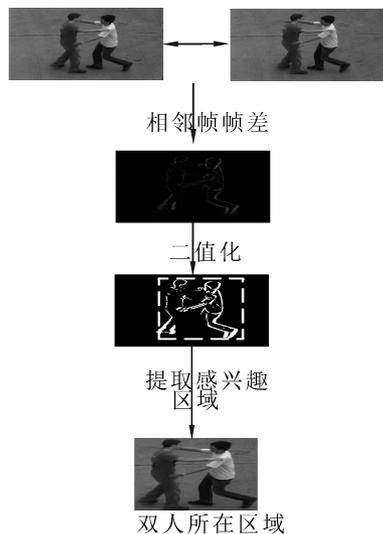


图 3 帧差法提取剪影所在感兴趣区域

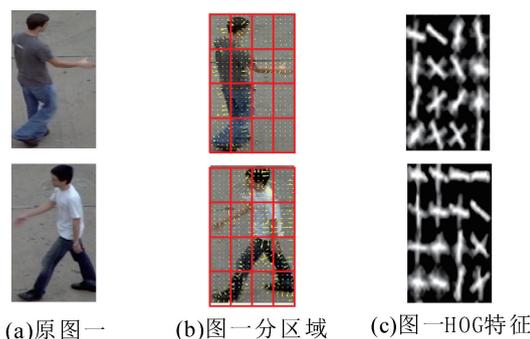
Fig.3 Frame difference extract the ROI

双方的距离逐渐拉开,但是双方的动作中仍然包含针对不同交互行为的单人原子动作的特性。在这一阶段,采用与开始阶段相同的感兴趣区域分割提取方法,分别切割提取,帧差后交互行为双方间躯体剪影的边界信息确定的矩形区域为感兴趣区域。

2 HOG 特征的提取

为了能够在有效地描述感兴趣区域内的全局信息的同时降低特征提取的复杂程度,文中采用 HOG 描述符对每帧图像的感兴趣区域进行表征。

HOG 特征^[11]是一种不需要在相邻帧间进行处理的简单全局特征表示法,其只需要在当前帧像素点间求取梯度的幅值和方向,并在不同方向区域上对像素点幅值大小进行直方图统计即可。在完成统计后,对当前帧各个方向区域的像素点幅值统计结果进行归一化处理,这样能够避免感兴趣区域变化带来的尺度干扰问题。本文在提取 HOG 特征时具体的流程如图 4 所示。



(a)原图一

(b)图一分区域

(c)图一HOG特征

3)交互行为结束阶段:当双人交互行为进行到一定程度时,交互行为双方结束身体接触,交互行为

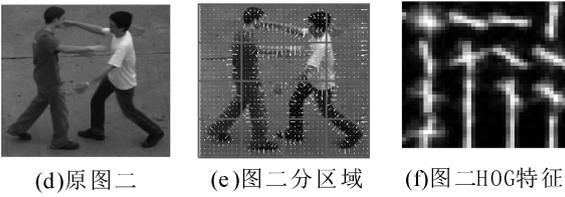


图4 不同阶段感兴趣区域 HOG 提取框图

Fig.4 HOG descriptor extraction of ROI in different stages

1) 计算图像的梯度

$$T(x_i, y_j) = \sqrt{P_x(x_i, y_j)^2 + P_y(x_i, y_j)^2} \quad (1)$$

$$\theta = \arctan \frac{P_x(x_i, y_j)}{P_y(x_i, y_j)} \quad (2)$$

式中: $T(x_i, y_j)$ 和 θ 分别表示像素点 (x_i, y_j) 梯度的幅值大小和方向, $P_x(x_i, y_j)$ 和 $P_y(x_i, y_j)$ 分别为图像上任意像素点 (x_i, y_j) 的水平 and 垂直方向梯度。如图 4 (b) 和 (e) 分别为图像中各个像素点上梯度的直观表示。

2) 梯度的直方图统计

将 1) 中求取的梯度图像均等地划分为 $P \times Q$ ($P = Q = 4$) 个不重叠的子区域, 并以每个子区域的中心点为原点的圆周内, 分割成大小相同的 K ($K = 12$) 个不重叠的扇形区间, 在每个扇形区域内统计该子区域的所有像素点梯度, 叠加所有子区域的直方图特征构成最后的 $P \times Q \times K$ 维特征向量。

3 动作识别

3.1 帧帧最近邻识别方法

帧帧最近邻识别方法是一种简单有效的识别方法, 识别单人动作速度很快^[12]。为了简化达到保证算法的实时性, 文中选取了帧帧最近邻的方法作为最终的识别方法。帧帧最近邻识别方法如式 (3) 所示:

$$g_i(\mathbf{x}) = \min \| \mathbf{x} - \mathbf{x}_i^k \|, k = 1, 2, \dots, c \quad (3)$$

式中: \mathbf{x} 为待测试的样本的 HOG 描述符, \mathbf{x}_i^k 为训练集中, 第 i 类动作的第 k 帧所形成的 HOG 描述符。通过式 (3) 可以求得与待测试样本距离最近的训练集样本帧所属动作类别, 就可以判断, 待测样本与距离最近的训练集样本帧属于同一动作类别。

统计待测试动作图像序列某一阶段的每一帧的识别结果, 能够得到该图像序列在这一阶段的分类票数结果, 将该分类票数结果进行归一化操作可得到该图像序列在这一阶段的识别概率:

$$V = v_1, v_2, \dots, v_c \quad (4)$$

$$p_i = \frac{v_i}{\sum_{i=1}^c v_i}, i = 1, 2, \dots, c \quad (5)$$

式中: v_i 代表分类票数, p_i 代表识别概率。

3.2 分层识别概率的加权融合

通过使用帧帧最近邻分类器可以分别获得交互行为开始阶段单人原子动作的识别概率、交互行为执行阶段整体的识别概率以及交互行为结束阶段单人原子行为的识别概率。

将 3 个阶段的识别概率进行融合的过程, 分为:

1) 交互行为开始阶段, 单人原子行为识别概率

的融合: 在交互行为开始阶段, 本文将交互行为为双方身体所在感兴趣区域进行分别提取和识别, 这样可以分别获得交互行为双方在开始阶段的识别概率, 通过将两者加权融合, 可以获得交互行为开始阶段的识别概率:

$$P_{\text{start}} = w_{sl} \times P_{sl} + w_{sr} \times P_{sr} \quad (6)$$

式中: P_{start} 为待测试动作图像序列开始阶段的最终的识别概率, P_{sl} 为交互行为开始阶段左侧执行人所做动作的识别概率, P_{sr} 为交互行为开始阶段右侧执行人所做动作的识别概率, w_{sl} 是融合过程中, 左侧执行人动作识别概率的权值, w_{sr} 是右侧执行人动作识别概率的权值。权值经由大量实验获得。

2) 交互行为结束阶段, 单人原子行为识别概率

的融合: 在交互行为结束阶段的单人原子行为识别概率融合的方法与在开始阶段融合的方法相同:

$$P_{\text{end}} = w_{el} \times P_{el} + w_{er} \times P_{er} \quad (7)$$

式中: P_{end} 为结束阶段的最终识别概率, 权值 w_{el} 与 w_{er} 经由大量的实验获得。

3) 三阶段识别概率融合: 分别获得 3 个阶段的

识别概率后, 通过加权融合的方式可以得到待测试动作图像序列的最终识别概率及结果:

$$P_{\text{Final}} = w_s \times P_{\text{start}} + w_m \times P_{\text{middle}} + w_e \times P_{\text{end}} \quad (8)$$

式中: P_{Final} 为待测试动作图像序列最终识别概率, P_{middle} 为交互行为执行阶段识别概率, 各个阶段的识别概率加权参数 w_s 、 w_m 以及 w_e 通过比较 3 个阶段的平均识别率获得, 分别为 25%、40% 及 35%。

4 实验与结果分析

4.1 数据库信息

本文所用实验数据均来自 UT-interaction 数据库^[13], 其中包含 2 个子数据库, 每个子数据库中包含握手、拥抱、踢、指、推、拳击 6 类动作, 每类动作下包含有 10 个动作视频, 除掉指动作, 整个数据库由 15 人在真实场景下两两完成。该数据库中的视频场景内大多含有杂乱的背景、相机抖动、变化的光照等挑战因素。该数据库被广泛用于双人交互行为识别研究中。本文的实验数据集由 UT-interaction 数

数据库中子数据库 1 中的握手、拥抱、踢、推、拳击 5 类动作的全部样本组成。在实验过程中, 采取留一交叉验证法对数据库进行了测试实验。

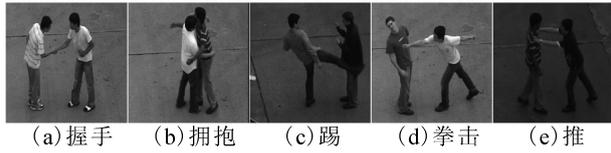


图 5 UT-interaction 数据库示例图

Fig.5 Exemplar frames from UT-interaction dataset

所有实验环境为主频 3.2 GHz, 内存 4 GB, 32 位 win7 操作系统下 MATLAB 2012 软件平台上完成。

4.2 不同阶段双人交互行为识别性能验证

在本次试验中, 利用 UT-interaction 数据库对分层识别的方法进行了测试。

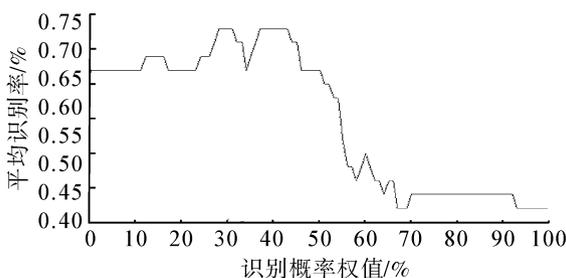
1) 在交互行为的开始阶段, 分别对左右两侧运动目标所在区域进行了分割提取, 使用 HOG 描述符分别对左右侧人体所在感兴趣进行描述, 通过帧帧最近邻分类器获得识别概率, 并最终进行加权融合。

表 1 交互行为开始阶段单人原子行为识别结果及加权融合识别结果

Table 1 The recognition result of single behavior and weight fusion in the start stages of interaction

区域	握手	拥抱	踢	推	拳击	识别率/%
left	60	50	20	50	30	42
right	70	80	50	60	50	62
final	80	100	60	60	40	68

由表 1 的实验结果可以看出, 在交互行为开始阶段, 动作的特性往往表现的并不明显, 有些交互行为的开始阶段动作相似度极高, 所以对这一阶段的单人原子动作的识别结果并不理想, 但是在对双人原子行为的识别概率进行加权融合后, 识别的效果有了明显的提升。其最优权值由大量的实验获得。如图 6(a) 所示, 在开始阶段的双人原子行为识别概率加权融合的最优权值分别为: 28% 和 72%。图 6(b)、(c)、(d) 中给出了交互行为开始阶段左右两侧人体单人原子行为以及识别概率加权融合后的混淆矩阵图及相应的整体识别率。



(a) 交互行为开始阶段最优权值实验结果

握手	0.60	0.20	0.00	0.20	0.00
拥抱	0.20	0.50	0.00	0.20	0.10
踢	0.00	0.00	0.20	0.50	0.30
拳击	0.00	0.10	0.20	0.50	0.20
推	0.00	0.00	0.20	0.50	0.30
	握手	拥抱	踢	拳击	推

(b) 交互行为开始阶段左侧人体单人原子行为识别混淆矩阵(识别率 42%)

握手	0.70	0.10	0.00	0.20	0.00
拥抱	0.00	0.80	0.10	0.00	0.10
踢	0.20	0.20	0.50	0.00	0.10
拳击	0.10	0.20	0.00	0.60	0.10
推	0.20	0.10	0.10	0.10	0.50
	握手	拥抱	踢	拳击	推

(c) 交互行为开始阶段右侧人体单人原子行为识别混淆矩阵(识别率 62%)

握手	0.80	0.10	0.00	0.10	0.00
拥抱	0.00	1.00	0.00	0.00	0.00
踢	0.20	0.10	0.60	0.10	0.00
拳击	0.10	0.20	0.00	0.60	0.10
推	0.20	0.20	0.10	0.10	0.40
	握手	拥抱	踢	拳击	推

(d) 交互行为开始阶段左右两侧人体单人原子行为加权融合识别混淆矩阵(识别率 68%)

图 6 交互行为开始阶段单人原子行为及加权融合后混淆矩阵以及最优权值测试图

Fig.6 The optimal weight and confusion matrixes for the interaction start stage

2) 在交互行为的执行阶段, 双人交互行为的识别是将交互行为双方人体所在区域作为一个整体进行分割提取, 使用 HOG 描述符对感兴趣进行描述, 通过帧帧最近邻分类器获得识别概率。交互行为执行阶段的识别混淆矩阵如图 7 所示。

3) 在交互行为的结束阶段, 分别对左右两侧运动目标所在区域进行了分割提取, 使用 HOG 描述符分别对左右侧人体所在感兴趣进行描述, 通过帧帧最近邻分类器获得识别概率, 并最终进行加权融合。由表 2 的实验结果可以看出, 在交互行为结束阶段, 动作的特性表现并不明显, 有些交互行为的可辨识

度较低,所以在这一阶段单人原子动作的识别结果并不理想,但是在对左右两侧人体单人原子行为识别概率进行加权融合后,识别的效果有了明显的提升。其最优权值由大量的实验获得。如图 8(a)所示,在开始阶段的单人原子行为识别概率加权融合的最优权值分别为:36%和 64%。图 8(b)、(c)、(d)中给出了交互行为结束阶段左右两侧人体单人原子行为以及识别概率加权融合后的混淆矩阵图及相应的整体识别率。

握手	1.00	0.00	0.00	0.00	0.00
拥抱	0.00	1.00	0.00	0.00	0.00
踢	0.10	0.00	0.70	0.20	0.00
拳击	0.00	0.00	0.10	0.50	0.40
推	0.00	0.30	0.00	0.10	0.60
	握手	拥抱	踢	拳击	推

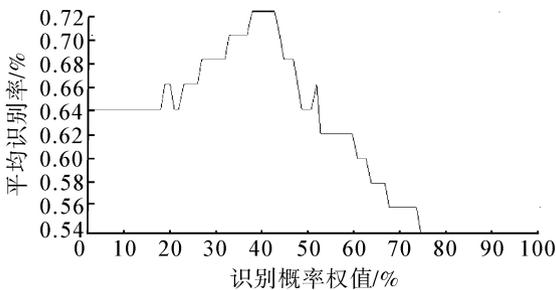
图 7 交互行为执行阶段双人交互行为识别混淆矩阵 (识别率 76%)

Fig.7 Confusion matrixes for execution of interaction

表 2 交互行为结束阶段单人原子行为识别结果及加权融合识别结果

Table 2 The recognition result of single behavior and weight fusion in the end stages of interaction

区域	握手	拥抱	踢	推	拳击	识别率/%
Left	90	60	20	30	70	54
Right	70	80	90	20	60	64
final	90	80	80	40	70	72



(a) 交互行为结束阶段最优权值实验结果

握手	0.90	0.10	0.00	0.00	0.00
拥抱	0.10	0.60	0.10	0.10	0.10
踢	0.30	0.10	0.20	0.10	0.30
拳击	0.20	0.00	0.20	0.30	0.30
推	0.00	0.10	0.10	0.10	0.70
	握手	拥抱	踢	拳击	推

(b) 交互行为结束阶段左侧人体单人原子行为识别混淆矩阵 (识别率 54%)

握手	0.70	0.20	0.00	0.10	0.00
拥抱	0.20	0.80	0.00	0.00	0.00
踢	0.10	0.00	0.90	0.00	0.00
拳击	0.00	0.50	0.10	0.20	0.20
推	0.10	0.00	0.10	0.20	0.60
	握手	拥抱	踢	拳击	推

(c) 交互行为结束阶段右侧人体单人原子行为识别混淆矩阵 (识别率 64%)

握手	0.90	0.10	0.00	0.00	0.00
拥抱	0.20	0.80	0.00	0.00	0.00
踢	0.10	0.00	0.80	0.10	0.00
拳击	0.00	0.20	0.10	0.40	0.30
推	0.00	0.00	0.10	0.20	0.70
	握手	拥抱	踢	拳击	推

(d) 交互行为结束阶段左右两侧人体单人原子行为加权融合识别混淆矩阵 (识别率 72%)

图 8 交互行为结束阶段单人原子行为及加权融合后混淆矩阵以及最优权值测试图

Fig.8 The optimal weight and Confusion matrixes for the interaction end stage

4) 交互行为三阶段识别概率融合:根据交互行为开始阶段以及交互行为结束阶段左右两侧单人原子行为识别概率加权融合的最优权值选择过程,可以发现,交互行为不同对象或不同阶段之间的识别概率加权融合的最优权值差值与它们之间的平均识别率差值正相关,如表 3 所示。

表 3 交互行为开始与结束阶段单人原子行为识别率与融合权值关系

Table 3 The relationship between recognition rate and weight of each stage

阶段	识别率		权值		识别率差	权值差/%
	left	right	left	right		
开始	42	62	28	72	20	44
结束	54	64	36	64	10	28

所以,在双人交互行为识别最终的三阶段识别概率融合过程,本文根据各个阶段的平均识别率设置的权值为:交互行为开始阶段识别概率加权参数 25% (该阶段平均识别率 68%);交互行为执行阶段识别概率加权参数 40% (该阶段平均识别率 76%);交互行为结束阶段识别概率加权参数 35% (该阶段平均识别率 72%)。实验结果对比如表 4 所示。通过将最终加权融合的识别结果与各个阶段识别结果

进行对比能够发现, 动作“握手”、“拥抱”以及“踢”这 3 种动作的识别率达到 100%, 动作“推”的识别率达到 70%, 优于各个阶段的识别效果, 动作“拳击”的识别率达到 80%, 优于各个阶段的识别效果。

表 4 各个阶段与最终融合识别效果

Table 4 Each stages and the final fusion recognition results

不同阶段	握手	拥抱	踢	推	拳击	识别率/%
开始阶段	80	100	60	60	40	68
执行阶段	100	100	70	50	60	76
结束阶段	90	80	80	40	70	72
最终融合	100	100	100	70	80	90

图 9 给出了交互行为三阶段识别概率加权融合后的最终识别混淆矩阵图。

握手	1.00	0.00	0.00	0.00	0.00
拥抱	0.00	1.00	0.00	0.00	0.00
踢	0.00	0.00	1.00	0.00	0.00
拳击	0.00	0.00	0.10	0.70	0.20
推	0.00	0.10	0.00	0.10	0.80
	握手	拥抱	踢	拳击	推

图 9 三阶段识别概率加权融合混淆矩阵

Fig.9 The three stage recognition probability weighted fusion confusion matrix

从混淆矩阵中可以看到, 本文方法对大部分双人交互动作均能够正确识别, 其识别率可达 90%, 尤其是对“握手”、“拥抱”及“踢”动作能够 100% 正确识别。从混淆矩阵的观察来看, 动作“推”和“拳击”2 组动作相互间存在较高的识别误差, 主要原因在于这 2 组动作的相似性较高造成。

4.3 不同方法识别效果比较

在表 5 中给出了近年来同样在 UT-interaction 数据库中进行双人交互行为的识别结果。将本文所提出的方法与其进行比较, 进而验证本方法对于双人交互行为优秀的识别性能。

从表 5 的实验结果对比可以看出, 本文所提出的方法在识别交互行为的准确率方面要明显优于其他的识别方法。在识别速度方面, 本文提出的方法平均速度达到 0.141 6 s/帧, 由于样本集帧数不同, 对数据库的平均识别速度达到 9.257 8 s/视频 (各个视频平均包含 65.38 帧)。相比于其他的识别方法, 本文提出的识别方法的优点在于并不需要建立复杂的概率模型, 同时实验表明识别准确率较高。

表 5 不同方法在 UT-interaction 数据库中的动作识别结果比较

Table 5 Comparison of different algorithms for interaction recognition in UT-interaction dataset

方法	识别率/%	用时/s
Three stage HOG+1nn	90	9.2578
global template +local 3D feature+ discriminative model ^[14]	85	—
3D XYT s-t volume + BoW + co-occurrence matrix ^[15]	41	—
KLT tracking detection+local descriptors+structured SVM ^[15]	84	—
Bipartite graph+key pose doublets ^[16]	79.17	—
2D+t tubes+spatio-temporal relationships graph model ^[17]	78.9	14.2

5 结束语

文中提出了一种将交互行为分阶段处理再融合的方法, 在交互行为的开始阶段和结束阶段, 对交互行为双方人体所在区域分别进行分割提取, 这样能够在最大程度的保留交互动作的动作特性的同时, 减少包含复杂环境等因素的冗余背景信息的影响; 在交互行为执行极端采用将接触的双方整体切割出来的方式, 即减少了冗余背景信息的影响, 也避免了由于人体遮挡导致误分割情况造成的错误识别。在 UT-interaction 数据库上的大量实验证明, 该方法实现简单, 对交互行为有较好的识别效果。本文所提出的方法对相似动作间的区分仍然存在一定误差, 下一步的研究重点将是进一步完善特征的代表和识别模型的构建, 进一步提高识别率。

参考文献:

- [1] WEINLAND D, RONFARD R, BOYER E. A survey of vision-based methods for action representation, segmentation and recognition [J]. Computer Vision and Image Understanding, 2011, 115(2): 224-241.
- [2] SEO H J, MILANFAR P. Action recognition from one example [J]. IEEE Transactions on pattern Analysis and Machine Intelligence, 2011, 33(5): 867-882.
- [3] 吴联世, 夏利民, 罗大庸. 人的交互行为识别与理解研究综述 [J]. 计算机应用与软件, 2011, 28 (11): 60-63. WU Lianshi, XIA Limin, LUO Dayong. Survey on human interactive behaviour recognition and comprehension [J]. Computer Applications and Software, 2011, 28 (11): 60-63.
- [4] YU Gang, YUAN Junsong, LIU Zicheng. Propagative hough

- voting for human activity recognition [C]//Proceedings of the 12th European Conference on Computer Vision, Florence, Italy. Berlin Heidelberg: Springer, 2012: 693-706.
- [5] YU T H, KIM T K, CIPOLLA R. Real-time action recognition by spatiotemporal semantic and structural forests [C]//Proceedings of the 21st British Machine Vision Conference. United Kingdom, 2010: 1-12.
- [6] YUAN Fei, SAHBI H, PRINET V. Spatio-temporal context kernel for activity recognition [C]//Proceedings of the 1st Asian Conference on Pattern Recognition. Beijing, China, 2011: 436-440.
- [7] BURGHOUTS G J, SCHUTTE K. Spatio-temporal layout of human actions for improved bag-of-words action detection [J]. Pattern Recognition Letters, 2013, 34(15): 1861-1869.
- [8] LI Nijun, CHENG Xu, GUO Haiyan, et al. A hybrid method for human interaction recognition using spatio-temporal interest points [C]//Proceedings of the 22nd International Conference on Pattern Recognition. Stockholm, Sweden, 2014: 2513-2518.
- [9] KONG Yu, JIA Yunde, FU Yun. Interactive phrases: semantic descriptions for human interaction recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(9): 1775-1788.
- [10] SLIMANI K, BENEZETH Y, SOUAMI F. Human interaction recognition based on the co-occurrence of visual words [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, Ohio, USA, 2014: 461-466.
- [11] JIANG Zhuolin, LIN Zhe, DAVIS L S. Recognizing human actions by learning and matching shape-motion prototype trees [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3): 533-547.
- [12] JI Xiaofei, ZHOU Lu, LI Yibo. Human action recognition based on adaboost algorithm for feature extraction [C]//Proceedings of 2014 IEEE International Conference on Computer and Information Technology. Xi'an, China, 2014: 801-805.
- [13] RYOO M S, AGGARWAL J K. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities [C]//Proceedings of the IEEE International Conference on Computer Vision. Kyoto, Japan, 2009: 1593-1600.
- [14] KONG Y, LIANG W, DONG Z, et al. Recognising human interaction from videos by a discriminative model [J]. Institution of Engineering and Technology Computer Vision, 2014, 8(4): 277-286.
- [15] PATRON-PEREZ A, MARSZALEK M, REID I, et al. Structured learning of human interactions in TV shows [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(12): 2441-2453.
- [16] MUKHERJEE S, BISWAS S K, MUKHERJEE D P. Recognizing interaction between human performers using "key pose doublet" [C]//Proceedings of the 19th ACM International Conference on Multimedia. Scottsdale, AZ, United states, 2011: 1329-1332.
- [17] BRENDEL W, TODOROVIC S. Learning spatiotemporal graphs of human activities [C]//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain, 2011: 778-785.

作者简介:



姬晓飞,女,1978年生,副教授,博士,主要研究方向为视频分析与处理、模式识别理论。承担国家自然科学基金、教育部留学回国启动基金等多项课题研究。发表学术论文40余篇,其中被SCI、EI检索20余篇。参与编著英文专著1部。



王昌汇,男,1991年出生,硕士研究生,主要研究方向为基于图模型的双人交互行为识别。



王扬扬,女,1979年生,工程师,博士,主要研究方向为视频分析与处理、模式识别理论。承担辽宁省教育厅科学研究一般项目等多项课题研究。发表学术论文20余篇,其中被SCI、EI检索8篇。