

# 基于3D骨架和MCRF模型的行为识别

刘皓<sup>1</sup>, 郭立<sup>2</sup>, 易波<sup>1</sup>, 王冠中<sup>1</sup>

(1. 中国科学技术大学物理系, 安徽合肥 230026; 2. 中国科学技术大学电子科学与技术系, 安徽合肥 230027)

**摘要:** 针对目前行为识别方法的不足, 提出一种基于人体3D骨架和多CRF模型(MCRF)的行为识别方法. 3D骨架数据量少且保留了行为关键信息的优点, 并具有融合多特征和上下文信息的优势. 为此, 首先基于3D骨架将人体动作划分为全局运动、手臂运动和腿部运动, 通过对动作序列进行多类特征提取, 形成多类特征集; 然后利用CRF模型对每一特征集建模, 再融合所有的CRF模型, 得到MCRF模型; 最后利用MCRF模型进行行为识别. 实验结果表明, 该方法具有较高检测率.

**关键词:** 行为识别; 3D骨架; MCRF; 特征提取

**中图分类号:** TP391.4      **文献标识码:** A      doi:10.3969/j.issn.0253-2778.2014.04.005

**引用格式:** Liu Hao, Guo Li, Yi bo, et al. Human activity recognition based on 3D skeletons and MCRF model [J]. Journal of University of Science and Technology of China, 2014, 44(4): 285-291.

刘皓, 郭立, 易波, 等. 基于3D骨架和MCRF模型的行为识别[J]. 中国科学技术大学学报, 2014, 44(4): 285-291.

## Human activity recognition based on 3D skeletons and MCRF model

LIU Hao<sup>1</sup>, GUO Li<sup>2</sup>, YI bo<sup>1</sup>, WANG Guanzhong<sup>1</sup>

(1. Department of Physics, University of Science and Technology of China, Hefei 230026, China;

2. Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230027, China)

**Abstract:** Considering the disadvantages of the traditional human activity recognition system, a human activity recognition system using an MCRF model and 3D skeletons was proposed. Its 3D skeleton data has less data and retains the key information, and the MCRF model has the advantage of being able to combine more features and utilizing adaptive contextual information. First, human activity was divided into global activity, arm activity, and leg activity. Several feature subsets were formed through more feature extraction. Then, CRF models were used on each feature subset to generate CRF units. Finally, all the CRF units were combined to produce the MCRF model which was utilized to recognize human activity. The experimental results indicate that the proposed method can improve detection accuracy.

**Key words:** human activity recognition; 3D skeleton; MCRF; feature extraction

收稿日期: 2013-09-11; 修回日期: 2014-01-01

基金项目: 国家自然科学基金(61071173)资助.

作者简介: 刘皓, 男, 1987年生, 博士生. 研究方向: 视频信息分析与处理. E-mail: lhnwos@mail.ustc.edu.cn

通讯作者: 郭立, 教授. E-mail: lguo@ustc.edu.cn.

## 0 引言

行为分析有广阔的应用前景,已成为多个领域的研究热点,如人机交互、公共安全、智能监控等.人的行为过程非常复杂,包含丰富的信息,这些信息多数难以提取,而且不易获得这些信息与行为之间的联系,所以行为识别是一个具有挑战性的课题.

行为识别包含以下两个基本问题:特征的提取和行为分类器的设计.

到目前为止,行为识别采用的特征多为二维特征<sup>[1-2]</sup>.基于二维特征的行为识别一般只能识别简单的行为.视频序列中人体朝向的不固定,相同动作从不同角度得到的观察结果差异度很大,这些问题给行为识别带来了很大的困难.基于高层人体结构的方法可以更精细地描述行为,清华大学谷军霞等<sup>[3-4]</sup>对行为识别进行了研究,但是仍然存在前景目标提取以及关节点获取困难的问题.

目前行为识别的算法可以概括为两类:一类是基于模板序列匹配的方法;另一类是基于状态空间的方法. Polana 等<sup>[5]</sup>提出一种基于模板的行为检测方法. Davis 等<sup>[6]</sup>提出了检测行为的两个模板:运动能量图(MEI)和运动历史图(MHI),这种行为描述方法用两幅模板图像表示了整个行为的过程.基于模板匹配的方法,计算量比较小,但是对时间间隔敏感,识别的准确性依赖于模板的选择. Yamato 等<sup>[7]</sup>将隐马尔科夫模型(hidden Markov model, HMM)引入行为检测中,展开了基于状态空间的行为检测研究. HMM 模型具有若干状态,并且按一定顺序和周期进行转移,比较符合人的行为过程,但是 HMM 的输出独立性假设,使得其不能联系上下文特征,限制了特征的选择.为了解决这一问题,Phillips 等<sup>[8]</sup>提出基于最大熵模型的行为检测方法,但是最大熵模型只能找到局部的最优值,同时也带来了标记偏见的问题.

近来,条件随机场(conditional random fields, CRF)<sup>[9]</sup>模型被用于行为检测,CRF 模型采用一种概率无向图的方式,直接对后验概率建模,不依赖于条件独立假设,有利用多特征和联系上下文的能力;然而随着特征数目增多,特征之间的干扰问题和过拟合问题会凸显出来,导致分类性能下降.

## 1 研究方法概述

针对上述问题,本文提出一种基于 3D 骨架特

征和多 CRF 模型(multiple CRF ensemble model, MCRF)<sup>[10]</sup>的行为识别方法,算法流程如图 1 所示.该方法针对人类行为动作的特点采用多种 3D 特征对行为进行描述,具有数据量少且保留了行为关键信息的优点,并在行为识别时引入 MCRF 模型,有效地利用了多特征与上下文信息,避免了特征之间的干扰和过拟合问题.

首先利用 Kinect<sup>[11]</sup>获取人体 3D 骨架行为序列,对行为序列进行多类特征提取,利用 CRF 模型对每一特征类型建模;然后融合所有的 CRF 模型得到 MCRF 模型,通过训练获取模型的参数,得到多特征融合分类器;最后对行为进行识别.

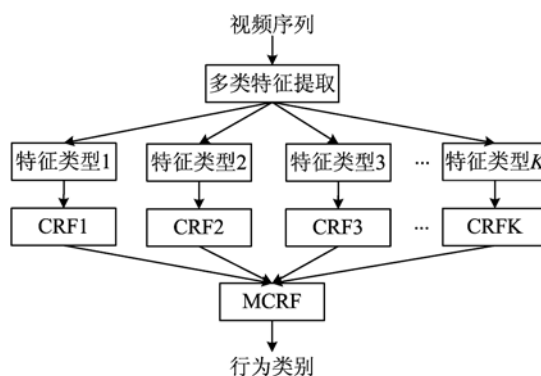


图 1 行为识别流程图

Fig. 1 The flow chart of our human activity recognition system

## 2 特征提取

我们使用 Kinect 获取行为序列中人体关节点在三维空间坐标系中的三维坐标序列,从 Kinect 骨架模型的 20 个骨骼点中选取 15 个关键点(头部,颈关节,臀部中心,两个腕关节,两个肘关节,两个肩关节,两个髋关节,两个膝关节,两个踝关节).定义一级关节点(两个肩关节,两个髋关节),二级关节点(两个肘关节,两个膝关节),三级关节点(两个腕关节,两个踝关节),并定义一级关节点与二级关节点中间骨骼段为一级骨骼段,二级关节点与三级关节点中间骨骼段为二级骨骼段.在此基础上,我们建立了人体骨架模型,以臀部中心为坐标系原点 O,臀部中心到肩膀中心的连线方向为 Z 轴, XOZ 平面平行于两个髋关节之间的连线,人体朝向为 Y 轴,建立动态人体坐标系(如图 2).在人体坐标系中,一级骨骼段绕一级关节点转动,二级骨骼段绕二级关节点转动.由人体结构连通性决定,一级关节点的位置信息包含于人体坐标系参数中;二级关节点的位置信

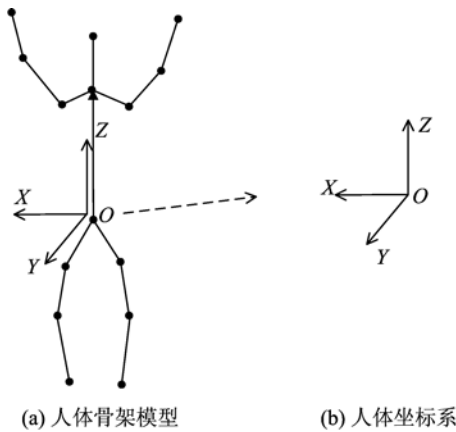


图 2 动态人体坐标系

Fig. 2 Dynamic body coordinate system

息包含于一级骨骼段绕一级关节的旋转参数和一级关节的位置信息中;三级关节的位置信息包含于二级骨骼段绕二级关节的旋转参数和二级关节的位置信息中;因此,用人体坐标系的参数和一级、二级关节的旋转参数即可完整地描述人体在三维空间中的姿态。

本文首先对行为序列按照身高比例进行归一化,然后在全局运动、手臂运动和腿部运动三种运动尺度上进行多类特征提取,共得到 11 种特征。

2.1 全局运动特征

全局运动包含人体身高的变化和人体坐标系在三维空间中的运动,表明人体在三维空间中位置与方向的变化.从中提取 5 个特征序列  $\{G_1, G_2, G_3, G_4, G_5\}$ ,分别为:身体高度特征序列  $G_1$ ,身体倾斜角特征序列  $G_2$ ,朝向角特征序列  $G_3$ ,速度特征序列  $G_4$ ,轨迹特征序列  $G_5$ .

身体高度特征序列为行为者在坐下、站起等动作中的身体高度.这里身体高度指三维空间中人体头部与脚部的距离.设  $G_1 = \{g_1(i), i=1, 2, \dots, N\}$  表示行为者高度序列,  $g_1(i)$  表示行为者第  $i$  帧的高度。

身体倾斜角特征序列表示行为序列中的身体倾斜度.设  $G_2 = \{g_2(i), i=1, 2, \dots, N\}$  表示行为者身体倾斜角序列,  $g_2(i)$  表示行为者第  $i$  帧中人体坐标系的 Z 轴与初始 Z 轴的夹角。

朝向角特征序列表示行为序列中行为者的朝向.设  $G_3 = \{g_3(i), i=1, 2, \dots, N\}$  表示行为者朝向角序列,  $g_3(i)$  表示行为者第  $i$  帧中人体坐标系的 Y 轴与初始 Y 轴的夹角。

速度特征序列描述了行为序列中行为者的速度.设  $G_4 = \{g_4(i), i=1, 2, \dots, N\}$  表示行为者速度

序列,  $g_4(i)$  表示第  $i$  帧中行为者人体坐标系原点的运动速度大小。

轨迹特征序列描述行为序列中行为者的三维空间位置,设  $G_5 = \{g_5(i), i=1, 2, \dots, N\}$  表示行为者的轨迹序列,  $g_5(i)$  定义为行为者第  $i$  帧的位置与初始位置的距离,行为者的位置用人体坐标系原点的位置表示。

2.2 手臂、腿部运动特征

本文在人体坐标系内描述手臂、腿部的运动过程,共提取 6 个特征序列  $\{A_1, A_2, A_3\}$  和  $\{L_1, L_2, L_3\}$ ,分别为:手臂旋转角特征序列  $A_1$ ,手臂关节点轨迹特征序列  $A_2$ ,手臂关节点速度特征序列  $A_3$ ,腿部旋转角特征序列  $L_1$ ,腿部关节点轨迹特征序列  $L_2$ ,腿部关节点速度特征序列  $L_3$ 。

2.2.1 手臂、腿部旋转角特征

旋转角特征序列描述了行为序列中从初始姿态到每一帧姿态的关节旋转角参数变化过程.设骨骼段绕人体坐标系 X 轴、Y 轴、Z 轴旋转矩阵分别为  $B_x, B_y, B_z, B_r$  如式所示,  $\phi_x$  为绕 X 轴旋转角参数,类似可以得到  $B_y, B_z$ 。

$$B_x(\phi_x) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \phi_x & \sin \phi_x & 0 \\ 0 & -\sin \phi_x & \cos \phi_x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1)$$

为适应各类行为中肢体的运动,共设置 16 个转动自由度<sup>[12]</sup>如图 3 所示(左肩关节  $\phi_1 \phi_2$ 、右肩关节  $\phi_3 \phi_4$ 、左肘关节  $\phi_5 \phi_6$ 、右肘关节  $\phi_7 \phi_8$ 、左髌关节  $\phi_9 \phi_{10}$ 、右髌关节  $\phi_{11} \phi_{12}$ 、左膝关节  $\phi_{13} \phi_{14}$ 、右膝关节  $\phi_{15} \phi_{16}$ ,每个关节 2 个旋转自由度).我们用  $\phi_j, j=1, 2, \dots, 16$  表示手臂、腿部的 16 个姿势参数.如初始

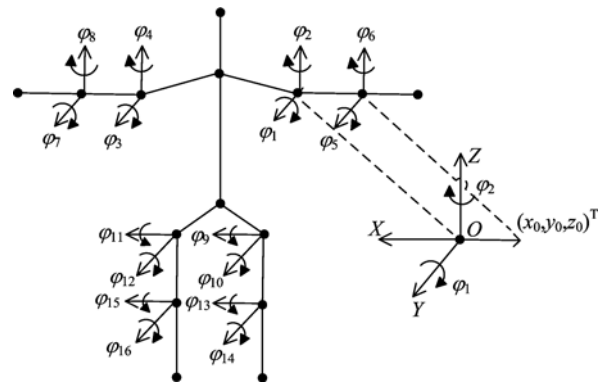


图 3 初始姿态及关节旋转自由度

Fig. 3 The initial gesture and joint-rotational degrees of freedom

姿态下左上臂骨骼段用向量表示为  $(x_0, y_0, z_0)^T$ , 在当前帧下左上臂姿态为  $(x, y, z)^T$ , 将两向量平移使两向量始点与人体坐标系原点重合, 则左上臂骨骼段绕左肩关节的旋转等效于骨骼段终点绕人体坐标系原点的旋转. 当前帧下左上臂姿态用旋转参数  $\varphi_1$  表示为

$$(x, y, z, 1)^T = \mathbf{B}_z(\varphi_2) \mathbf{B}_x(\varphi_1) (x_0, y_0, z_0, 1)^T \quad (2)$$

令  $A_1 = \{a_1(i), i=1, 2, \dots, N\}$  表示手臂旋转角序列,  $a_1(i) = \{\varphi_1, \varphi_2, \dots, \varphi_8\}_{t=i}$  表示行为者在第  $i$  帧下的手臂姿势参数,  $L_1 = \{l_1(i), i=1, 2, \dots, N\}$  为腿部旋转角序列,  $l_1(i) = \{\phi_9, \phi_{10}, \dots, \phi_{16}\}_{t=i}$  表示行为者在第  $i$  帧下的腿部姿势参数.

### 2.2.2 手臂、腿部轨迹特征

关节点轨迹序列表示行为序列中人体关节点在人体坐标系中的三维位置的变化过程. 设  $d_1, d_2, \dots, d_8$  分别表示左肘关节、右肘关节、左腕关节、右腕关节、左膝关节、右膝关节、左踝关节、右踝关节在人体坐标系内的三维空间坐标. 令  $A_2 = \{a_2(i), i=1, 2, \dots, N\}$  表示行为者手臂关节点轨迹序列,  $a_2(i) = \{x_j, y_j, z_j, j=1, 2, 3, 4\}_{t=i}$  表示第  $i$  帧手臂关节点在人体坐标系内的坐标,  $L_2 = \{l_2(i), i=1, 2, \dots, N\}$  表示腿部关节点轨迹序列, 其中  $l_2(i) = \{x_j, y_j, z_j, j=5, 6, 7, 8\}_{t=i}$ .

### 2.2.3 手臂、腿部速度特征

关节点速度序列描述了行为序列中人体关节点在人体坐标系中的运动速度的变化. 设  $u_1, u_2, \dots, u_8$  分别表示左肘关节、右肘关节、左腕关节、右腕关节、左膝关节、右膝关节、左踝关节、右踝关节的运动速度(前后相邻两帧对应关节点的三维空间欧式距离). 令  $A_3 = \{a_3(i), i=1, 2, \dots, N\}$  表示行为者手臂关节点速度序列,  $a_3(i) = \{u_1, u_2, u_3, u_4\}_{t=i}$ ,  $L_3 = \{l_3(i), i=1, 2, \dots, N\}$  表示行为者腿部关节点速度序列,  $l_3(i) = \{u_5, u_6, u_7, u_8\}_{t=i}$ .

通过上述步骤, 本文对行为序列提取了  $\{G_k\}_{k=1}^5, \{A_k\}_{k=1}^3$  及  $\{L_k\}_{k=1}^3$  共 11 种特征.

## 3 行为分类器

### 3.1 MCRF 模型

假设  $y = \{y_i, i=1, 2, \dots, N\}$  表示观测样本序列,  $y_i$  为第  $i$  帧的观测数据,  $x = \{x_i, i=1, 2, \dots, N\}$  为观测序列的标记,  $x_i \in X, X = \{1, 2, \dots, S\}$ . 其中,  $S$  表示标记行为类别的数目, 本文  $S=5$ . Lafferty 等<sup>[9]</sup> 给出了条件随机场模型的数学定义:

$$p(x | y, \theta) = \frac{1}{Z(y, \theta)} \exp\left\{ \sum_{c \in C} \phi_c(x_c, y_c, \theta) \right\}.$$

式中,  $Z(y, \theta) = \sum_x \exp\left\{ \sum_{c \in C} \phi_c(x_c, y_c, \theta) \right\}$  为归一化因子,  $\phi_c$  为势函数,  $\theta$  为势函数的参数.

对第  $i$  帧的观测样本  $y_i$ , 本文提取了全局特征集  $G_{1:5}^i(y) = \{G_1^i(y), G_2^i(y), G_3^i(y), G_4^i(y), G_5^i(y)\}$ 、手臂特征集  $A_{1:3}^i(y) = \{A_1^i(y), A_2^i(y), A_3^i(y)\}$  及腿部特征集  $L_{1:3}^i(y) = \{L_1^i(y), L_2^i(y), L_3^i(y)\}$ , 共 11 种特征.

可以推导出 MCRF 模型的表达式<sup>[13]</sup>:

$$p(x | G_{1:5}(y), A_{1:3}(y), L_{1:3}(y)) = \gamma \frac{1}{(p(x))^{10}} \cdot \prod_{k=1}^5 p(x | G_k(y)) \prod_{k=1}^3 p(x | A_k(y)) \prod_{k=1}^3 p(x | L_k(y)) \quad (3)$$

式中, MCRF 模型需要对  $p(x)$ ,  $p(x | G_k(y))$ ,  $p(x | A_k(y))$  及  $p(x | L_k(y))$  建模.  $p(x)$  包含标记序列的先验信息, 建模如下:

$$p(x) = \frac{1}{Z_0} \exp\left\{ \sum_{i=1}^N \sum_{j=1, j \neq i}^N \beta x_i x_j \right\} \quad (4)$$

式中,  $Z_0$  为归一化参数,  $\beta$  为模型参数.  $p(x | G_k(y))$ ,  $p(x | A_k(y))$ ,  $p(x | L_k(y))$  三者模型相同, 对于  $p(x | G_k(y))$  有:

$$p(x | G_k(y)) = \frac{1}{Z_g^k} \exp\left\{ \sum_{i=1}^N \phi_i(x_i, G_k(y), \lambda_g^k) + \sum_{i=1}^N \sum_{j=1, j \neq i}^N \phi_{ij}(x_i, x_j, G_k(y), \sigma_g^k) \right\} \quad (5)$$

式中,  $Z_g^k$  为全局特征第  $k$  个 CRF 模型的归一化函数,  $\theta_g^k = \{\lambda_g^k, \sigma_g^k\}$  为模型参数向量, 且

$$\begin{aligned} \phi_i(x_i, G_k(y), \lambda_g^k) &= \sum_{s=1}^S \delta(x_i = s) \log p(x_i = s | G_k(y), \lambda_g^k) \\ \phi_{ij}(x_i, x_j, G_k(y), \sigma_g^k) &= \sum_{s, h \in \{1, 2, \dots, S\}} (\sigma_{gsh}^k)^T \mathbf{g}_{ij}^k(y) \delta(x_i = s) \delta(x_j = h) \end{aligned} \quad (6)$$

$\lambda_g^k$  的维数等于第  $k$  类特征的维数乘以  $(S-1)$ ,  $\mathbf{g}_{ij}^k$  表示从整个观测序列数据提取的位置对  $(i, j)$  (共  $S^2$  对) 且维数为  $D_k$  的特征向量,  $\sigma_{gsh}^k$  为第  $k$  个 CRF 模型中双位置参数向量,  $\sigma_g^k$  表示将所有  $\sigma_{gsh}^k, s, h=1, \dots, S$  首尾链接得到的具有  $D_k \times S^2$  维的参数向量.

类似可以推导出  $p(x | A_k(y))$  和  $p(x | L_k(y))$  的建模表达式, 其参数分别为  $\{\lambda_a^k, \sigma_a^k\}_{k=1}^3$ ,  $\{\lambda_l^k, \sigma_l^k\}_{k=1}^3$ .

由式(3)~(7)可知, MCRF 模型的参数集合  $\theta = \{\beta, \{\lambda_g^k, \sigma_g^k\}_{k=1}^5, \{\lambda_a^k, \sigma_a^k\}_{k=1}^3, \{\lambda_l^k, \sigma_l^k\}_{k=1}^3\}$ .  $\theta$  需要通过训练获取, 极大似然(ML)估计方法是比较常用的参数估计法, 本文采用极大似然估计方法来估计参数. 给定  $M$  个训练样本, ML 估计选择使对数似然极大化的参数, 即

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \{ \log(p(x | y, \theta)) \} = \underset{\theta}{\operatorname{argmax}} \{ \log(\prod_{m=1}^M p(x^m | y^m, \theta)) \}.$$

文献[13]给出了详细的求解过程, 这里不再赘述.

### 3.2 多特征融合分类器

不同的特征对不同行为的区分度不同, 例如, 身高特征可以很好地区分 sitdown 和 walk, 却无法区分 walk 与 boxing, 权值可以很好地描述每一种特征对行为识别的贡献.

融合全局运动的身高特征、倾斜角特征、朝向角特征、速度特征、轨迹特征、手臂运动的旋转角特征、轨迹特征、速度特征和腿部运动的旋转角特征、轨迹特征、速度特征共  $K=11$  个特征的分类器如图 4 所示.

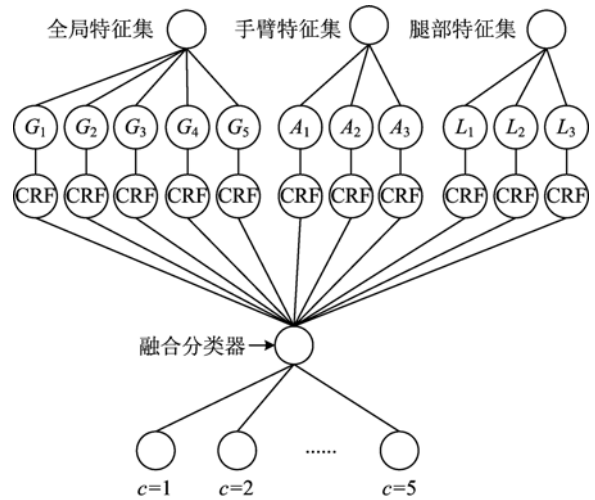


图 4 多特征融合分类器

Fig. 4 Multi feature fusion classifier

图 4 所示多特征融合分类器从全局运动、手臂运动及腿部运动中提取出的 11 种特征分别输入其对应的 CRF 模型中, 然后给每一种特征赋予权重. 通过联合 11 个 CRF 模型及其权重并利用式(3)得到多特征融合分类器模型, 利用该模型对行为识别得到行为类别  $c \in \{1, 2, \dots, S\}$ . 多特征融合分类器如下:

$$\begin{aligned} p(x | G_{1:5}(y), A_{1:3}(y), L_{1:3}(y)) = & \gamma \frac{1}{(p(x))^{10}} \prod_{k=1}^5 p^{g_k}(x | G_k(y)) \prod_{k=1}^3 p^{a_k}(x | A_k(y)) \prod_{k=1}^3 p^{l_k}(x | L_k(y)) = \\ & \exp\left\{ \sum_{k=1}^5 g_k \left( \sum_{i=1}^N \sum_{s=1}^S \delta(x_i = s) \log p(x_i = s | G_k(y), \lambda_g^k) + \right. \right. \\ & \left. \sum_{i=1}^5 \sum_{j=1, j \neq i}^N \sum_{s, h \in \{1, 2, \dots, S\}} (\sigma_{gsh}^k)^T g_{ij}^k(y) \delta(x_i = s) \delta(x_j = h) \right) + \\ & \sum_{k=1}^3 a_k \left( \sum_{i=1}^N \sum_{s=1}^S \delta(x_i = s) \log p(x_i = s | A_k(y), \lambda_a^k) + \right. \\ & \left. \sum_{i=1}^3 \sum_{j=1, j \neq i}^N \sum_{s, h \in \{1, 2, \dots, S\}} (\sigma_{ash}^k)^T a_{ij}^k(y) \delta(x_i = s) \delta(x_j = h) \right) + \\ & \sum_{k=1}^3 l_k \left( \sum_{i=1}^N \sum_{s=1}^S \delta(x_i = s) \log p(x_i = s | L_k(y), \lambda_l^k) + \right. \\ & \left. \sum_{i=1}^3 \sum_{j=1, j \neq i}^N \sum_{s, h \in \{1, 2, \dots, S\}} (\sigma_{lsh}^k)^T l_{ij}^k(y) \delta(x_i = s) \delta(x_j = h) \right) - \sum_{i=1}^N \sum_{j=1, j \neq i}^N \beta(1-K) x_i x_j - \log Z \} \quad (8) \end{aligned}$$

通过引入权重的方式, 既解决了推断过程中可能导致的过计数问题<sup>[14-15]</sup>, 又可以通过给不同的特征赋予权重达到更好的识别效果. 利用前述 MCRF

训练得到的参数  $\theta$ , 不断调整权重  $\{\{g_k\}_{k=1}^5, \{a_k\}_{k=1}^3, \{l_k\}_{k=1}^3\}$  对每一种行为在训练数据上进行验证, 进而得到最终的权值分布.

式(8)中所得分布形式的推断可以用 LBP<sup>[16]</sup>算法高效实现。

## 4 实验结果与分析

本文实验平台为 PC 机, CPU 为 Inter 双核 E8400, 主频 3.0 GHz, 内存 4.0 GB, 操作系统为 Windows 2008, 64 位。主要函数采用 C++ 编程实现, 并在 Matlab9.0 环境下对其调用。实验数据来自 Kinect 采集。数据库共包含 5 种行为(walk, jump, sitdown, pickup, boxing), 每种行为由 10 个行为者(每个行为者做 30 次)完成, 一共 1 500 个样本, 其中 1 200 个样本用于训练, 其余 300 个样本用于测试。

特征提取部分共提取了  $K=11$  种特征, 包括身高特征、倾斜角特征、朝向角特征、全局速度特征、全局轨迹特征、手臂旋转角特征、手臂轨迹特征、手臂速度特征、腿部旋转角特征、腿部轨迹特征以及腿部速度特征。

模型的最大迭代次数设为 300 次, 通过数据库中的 5 种行为, 1 200 个行为序列样本进行模型训练, 获取参数估计。将参数向量  $\{\lambda_g^k, k=1, 2, \dots, 5\}$ ,  $\{\lambda_a^k, k=1, 2, 3\}$  和  $\{\lambda_l^k, k=1, 2, 3\}$  初始化为 0, 向量  $\{\sigma_g^k, k=1, 2, \dots, 5\}$ ,  $\{\sigma_a^k, k=1, 2, 3\}$  和  $\{\sigma_l^k, k=1, 2, 3\}$  中的每一分量初始化为 0~0.5 之间的随机数, 算法经过平均 90 次迭代收敛。最终获得的 11 种特征对应 CRF 模型的加权参数, 如图 5 所示。

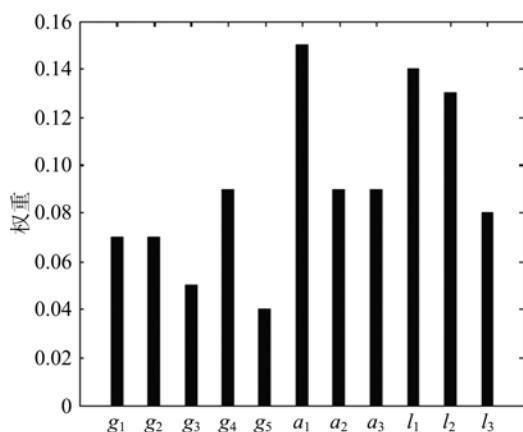


图 5 每个 CRF 模型的权重

Fig. 5 The parameters weighting the trained CRF compents

图 5 中, 横坐标对应的特征依次为: 身高特征  $G_1$ 、倾斜角特征  $G_2$ 、朝向角特征  $G_3$ 、全局速度特征  $G_4$ 、全局轨迹特征  $G_5$ 、手臂运动的旋转角特征  $A_1$ 、轨迹特征  $A_2$ 、速度特征  $A_3$  和腿部运动的旋转角特征  $L_1$ 、轨迹特征  $L_2$ 、速度特征  $L_3$ 。由图 5 可以看出,

在对包含 walk, sitdown, jump, pickup, boxing 这 5 种行为的数据库中的样本进行行为识别过程中, 手臂旋转角特征和腿部旋转角特征所占权重较大, 腿部轨迹特征所占权重次之。在 5 种行为中, walk, sitdown, jump 与腿部关系较大; pickup, boxing 与手臂关系较大。5 种行为中, 仅 walk 与全局轨迹特征相关, 因此图 5 中全局轨迹特征所占权重最低。整体上全局特征、手臂特征及腿部特征在行为识别中所占权重分别为 0.32, 0.33, 0.35。

本文针对人体行为运动提取的 11 种特征对数据库中的行为有较好的区分效果。例如, 身高特征与腿部特征易区分 sitdown 行为, 身体倾斜角特征易区分 pickup 行为, 手臂旋转角特征、腿部旋转角特征与全局速度特征区分 walk 时效果较好, 手臂速度特征在区分 boxing 时优势明显。本文识别算法检测率如表 1 所示, 在 walk, sitdown, jump, pickup, boxing 5 种行为中, sitdown 行为的识别准确率达到 100%, pickup 行为的识别准确率为 98.3%, walk 和 jump 的准确率是 96.7%, boxing 行为最复杂, 识别准确率为 95%。总体检测率达到 97.3%。

表 1 行为识别检测率

行为	检测率	总体检测率
walk	96.7%	97.3%
sitdown	100%	
jump	96.7%	
pickup	98.3%	
boxing	95.0%	

表 2 给出了本文方法和其他方法的检测率比较。其中, 线性分类 LDA<sup>[17]</sup>模型的识别准确率较低。文献[18]提出了基于关节角特征和 HMM 模型的行为识别方法, 然而 HMM 模型不能联系上下文信息、限制了特征的选择, 检测准确率为 92.5%。HCRF<sup>[2]</sup>模型的检测率高达 91.7%, 但其所用特征为 2D 特征, 只能描述简单的行为。与以上方法相比, 本文方法采用可以有效地描述人体行为的多种 3D 特征, 使用 MCRF 模型融合多特征和上下文信

表 2 本文方法与其他方法检测率比较

算法	检测率
HMM <sup>[18]</sup>	92.5%
HCRF <sup>[2]</sup>	91.7%
LDA <sup>[17]</sup>	81.5%
Our method	97.3%

息构建行为分类器,在特征选择和行为识别上均具有优势,检测率高于以上各种方法。

## 5 结论

本文提出一种基于3D骨架和MCRF模型的行为识别方法。首先基于Kinect提取人体行为序列,对行为序列提取多类3D特征,然后利用CRF模型对每一特征及其上下文信息进行建模,最后融合多个CRF模型得到MCRF模型,构建多特征融合分类器,对行为序列测试得到识别结果。实验结果表明,本文的方法获得了较高的识别率,比HMM、HCRF方法的检测准确率提升了5%左右。

本文方法只能识别单一动作,无法对边走路边招手这一类复合行为进行识别。本文实验中模型的训练过程和推断过程均是在线完成,无法完成实时识别的任务。如何解决这两个问题是我们下一步的研究方向。

### 参考文献(References)

- [1] Zhao L, Guo L, Xie J S, et al. Video abnormal target description based on CRF model[C]// International Conference on Audio, Language and Image Processing. Shanghai, China: IEEE Press, 2012: 519-524.
- [2] Wang Y, Mori G. Learning a discriminative hidden part model for human action recognition[J]. Advances in Neural Information Processing Systems, 2008, 21: 1 721-1 728.
- [3] Gu Junxia, Ding Xiaoqing, Wang Shengjin. A survey of activity analysis algorithms[J]. Journal of Image and Graphics, 2009, 14(3): 377-387.  
谷军霞,丁晓青,王生进.行为分析算法综述[J].中国图形图像学报,2009,14(3): 377-387.
- [4] Gu Junxia, Ding Xiaoqing, Wang Shengjin. Human 3D model-based 2D action recognition [J]. Acta Automatica Sinica, 2010, 36(1): 46-53.  
谷军霞,丁晓青,王生进.基于人体行为3D模型的行为识别[J].自动化学报,2010,36(1): 46-53.
- [5] Polana R, Nelson R. Low level recognition of human motion (or how to get your man without finding his body parts)[C]// Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects. Austin, USA: IEEE Press, 1994: 77-82.
- [6] Davis J W, Bobick A F. The representation and recognition of human movement using temporal templates[C]// IEEE Conference on Computer Vision and Pattern Recognition. San Juan, USA: IEEE Press, 1997: 928-934.
- [7] Yamato J, Ohya J, Ishii K. Recognizing human action in time-sequential images using hidden Markov model [C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Champaign, USA: IEEE Press, 1992: 379-385.
- [8] Phillips S J, Anderson R P, Schapire R E. Maximum entropy modeling of species geographic distributions [J]. Ecological modelling, 2006, 190(3-4): 231-259.
- [9] Lafferty J, McCallum A, Pereira F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]// Proceedings of the 18th International Conference on Machine Learning. Williamstown, USA: Morgan Kaufmann Publisher, 2001: 282-289.
- [10] Zhong P, Wang R S. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2007, 45 (12): 3 978-3 988.
- [11] Smisek J, Jancosek M, Pajdla T. 3D with Kinect[C]// Consumer Depth Cameras for Computer Vision Advances in Computer Vision and Pattern Recognition. Barcelona, Spain: IEEE Press, 2013: 3-25.
- [12] 谷军霞.行为表征与行为识别方法研究[D].清华大学,2010.
- [13] 钟平.面向图像标记的随机场模型研究[D].国防科学技术大学,2008.
- [14] Shotton J, Winn J, Rother C, et al. TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context [J]. International Journal of Computer Vision, 2009, 81(1): 2-23.
- [15] Shotton J, Winn J, Rother C, et al. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation [J]. Lecture Notes in Computer Science, 2006, 3951: 1-15.
- [16] Ramos F, Fox D, Durrant-Whyte, H F. CRF-Matching: conditional random fields for feature-based scan matching[A]// Robotics: Science and Systems, MIT Press, 2007.
- [17] Niebles J C, Wang H C, Li F F. Unsupervised learning of human action categories using spatial-temporal words[J]. International Journal of Computer Vision, 2008, 79(3): 299-318.
- [18] Uddin M Z, Thang N D, Kim J T, et al. Human activity recognition using body joint-angle features and hidden Markov model [J]. ETRI Journal, 2011, 33(4): 569-579.