

基于半监督聚类的3维肢体分割算法

谷军霞 丁晓青 王生进

(清华大学电子工程系智能技术与系统国家重点实验室, 北京 100084)

摘要 行为分析已经成为计算机视觉研究领域的热点,行为主体的肢体部件分割是行为分析中很重要的一部分同时也是一个难点问题,为了对3维肢体进行有效分割,提出了一种基于半监督聚类的肢体分割算法。该算法首先利用前一帧姿势估计反馈的时域信息来对3维主体进行初始的肢体部件分割;然后根据人体结构信息进一步确定行为主体上各个点与肢体部件之间的关系来得到半监督聚类的初始值;之后基于各个肢体部件的形状信息进行半监督聚类,迭代求解肢体部件分割的最优解;最后利用分割后的各个肢体部件进行行为主体的姿势参数估计。通过对IXMAS数据库中6种行为下的6个行为主体进行的肢体部件的分割实验结果证明,该新算法具有主体适应性、视角适应性以及行为适应性。

关键词 肢体分割 半监督聚类 姿势估计 行为分析

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2008)03-0558-08

A Semi-supervised Clustering-based Segmentation Algorithm of 3D Reconstructed Human Body Parts

GU Jun-xia, DING Xiao-qing, WANG Sheng-jin

(Department of Electronic Engineering, Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, Beijing 100084)

Abstract Human activity analysis is receiving increasing attention from computer vision researchers. One challenge is the segmentation of human body into meaningful body parts. A semi-supervised clustering-based body parts segmentation algorithm of 3D reconstructed human is presented in this paper. Firstly, we segment human body parts with the help of posture parameters of the previous frame. Then the structure information of human body is adopted to classify some points and initialize the centers of the semi-supervised clustering. Finally, based on the shape of body parts, semi-supervised clustering method is used to segment the body parts. In addition, body posture parameters are estimated with the segmentation result of body parts. The system is validated with IXMAS database, which includes 6 actors and 6 kinds of activity. The experimental results show that the presented algorithm can adapt the variety of views, actors and activities.

Keywords body parts segmentation, semi-supervised clustering, posture estimation, activity analysis

1 引言

目前行为分析已经成为计算机视觉领域的研究热点,它可广泛应用于智能视频监控、人机交互、3维游戏、虚拟现实等等。行为分析的难点之一就是视频序列中行为主体的肢体分割问题^[1,2],即如何

将前景目标分割为有意义的肢体部件,如头、躯干、胳膊、腿等等。

近几年,对于肢体分割问题,研究者已经提出了很多算法,如 Sangho Park 等人提出了一种交互行为中直立姿态下的肢体部件分割与跟踪算法^[1],该算法先利用 GMM(Gaussian Markov model)模型描述像素的颜色信息,之后用马尔可夫随机场分割肢体部件。

基金项目:国家高技术研究发展计划(863)项目(2006AA01Z115);国家自然科学基金项目(60472002)

收稿日期:2007-06-15;改回日期:2007-12-07

第一作者简介:谷军霞(1981~),女。2004年获西安电子科技大学学士学位,现为清华大学电子工程系硕博研究生。主要研究方向为模式识别、行为分析。E-mail: gujx04@mails.tsinghua.edu.cn

Ramanan 等人也提出了一种先利用颜色信息建立肢体部件模型,然后结合人体结构进行肢体分割的算法^[3]。之后又提出了基于三角剖分^[2]以及轮廓曲率^[4]的肢体分割算法。但由于上述这些算法都是针对2维图像的,因此一般只能处理限定视角下限定姿势的肢体分割问题,而3维肢体分割算法则可以适应不同的视角和不同的姿势。2000年,Cheung 等人用EM算法分割视频中3维肢体部件^[5],它首先根据前一帧的姿势估计结果对当前帧3维主体进行肢体部件的初步分割;然后用EM算法迭代求解肢体分割的最优解,该算法虽然利用了时域信息,但由于对初值比较敏感,因此可能造成误差累积。之后 Kehl 等人利用颜色信息分割3维肢体^[6],这种算法虽然速度较快,但是当各个肢体部件颜色一致时,该算法就会失效。

本文综合利用前一帧姿势估计反馈的时域信息、人体各个肢体部件之间存在的结构信息以及肢体形状信息提出了3维空间中基于半监督聚类的肢体分割算法(SSC-BPS: semi-supervised clustering-based body parts segmentation)。本文算法是首先根据前一帧姿势估计反馈的时域信息对当前帧的3维重构主体进行初步分割;然后利用先验的人体肢体部件之间存在的结构信息确定3维重构主体上的每个点与各个肢体部件的关系,这种关系分为以下3种情况:(1)该点属于某一肢体部件;(2)该点可能属于某一肢体部件;(3)该肢体点不可能属于某一肢体部件,由此得到的重构主体上的点与肢体部件的关系表就可以作为半监督聚类的初值;最后通过半监督聚类算法迭代求解,即可获得肢体部件的最终分割结果。实践表明,本文提出的SSC-BPS算法对视角、姿势和行为主体都有一定的适应性。

2 行为分析系统

肢体分割算法是行为分析系统的一部分,3维行为分析的系统框图如图1所示。整个系统的输入是由多个固定摄像头拍摄的多视角视频图像序列。该系统首先用背景减去法提取各个视角的前景目标,并由前景目标重构3维主体^[7,8];然后进行3维重构主体的肢体部件分割;之后再基于3维人体模型估计各个肢体部件的姿势参数,在获取了姿势参数序列之后就可以进行行为识别和分析了。

肢体分割是行为分析系统中非常重要的一部分,也是一个难点问题。如果不进行肢体部件分割,

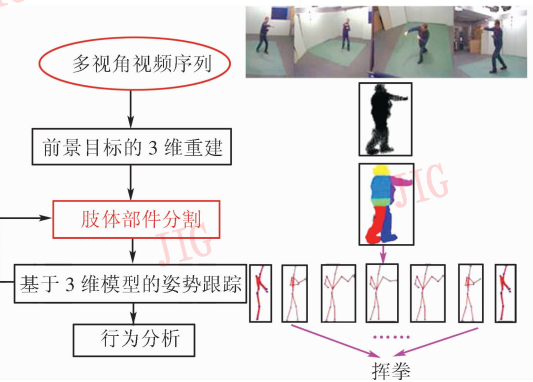


图1 行为分析系统框图

Fig. 1 Activity analysis system

而直接由3维重构主体的数据来估计各个肢体部件的姿势参数,则很容易陷入局部极小解,尤其当四肢和躯干距离比较近的时候,各肢体部件的数据往往难于区分,从而导致姿势估计失败^[6]。由此可见,肢体部件的精确分割是保证姿势参数估计正确的前提条件。另外,肢体部件分割也可以降低姿势估计的运算量。在姿势估计中,高维的自由度(30维左右)带来了极大的运算量^[9],搜索空间分解法^[10]是一种有效的降低自由度维数的方法,由于这个算法假设某些肢体部件的运动是独立的,而且由这些独立的肢体部件就可以分别估计姿势参数,从而可降低搜索空间的维数。由此可见,肢体部件分割是姿势估计中非常重要的一部分,它不仅可以降低姿势估计的运算量,也可以避免姿势估计算法陷入局部极小解。

3 肢体分割算法:SSC-BPS

根据人体关节的个数及结构,可将整个人体划分为11个肢体部件,它们依次为:head、torso、hip、left upper arm、left fore arm、right upper arm、right fore arm、left thigh、left shin、right thigh、right shin。

SSC-BPS算法框图如图2所示,图2的右侧是一个行为主体进行肢体分割过程的示意图。整个SSC-BPS算法由以下3个步骤组成:(1)根据前一帧姿势估计反馈的时域信息来对3维重构主体进行初步分割;(2)利用人体结构信息确定3维重构主体上的每个点和各个肢体部件的关系来获得半监督聚类的初值;(3)采用半监督聚类算法迭代求解肢体部件分割的最优结果。之后对每个肢体部件采用基于3维人体模型的匹配算法^[6]来估计姿势参数,进而得到根据姿势参数估计结果进行姿势变化后的3维人体模型,

该人体模型将会指导下一帧肢体部件的第 1 步分割。下面依次详细阐述肢体分割算法的 3 个步骤。

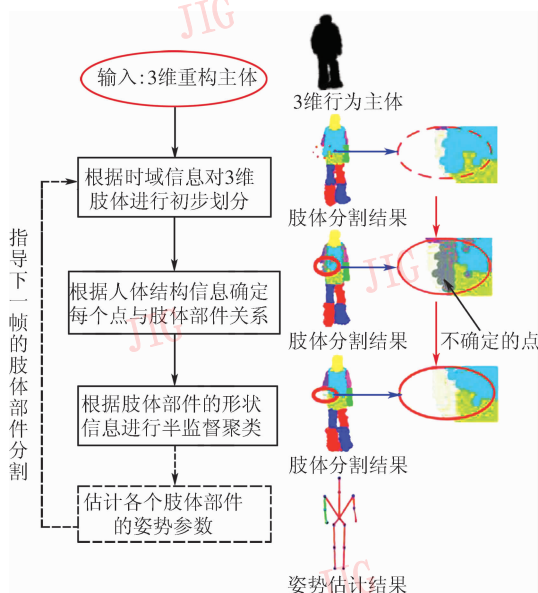


图 2 SSC-BPS 算法框图

Fig. 2 Chart of the SSC-BPS algorithm

3.1 3 维人体模型与 3 维重构主体

3 维人体模型是用 3 维点集表示的,它的初始姿势如图 3(a)所示。令 N_{body} 表示肢体部件的个数,本文将人体分割为 11 个肢体部件(如图 3(b)所示),即 $N_{\text{body}} = 11$ 。 $\mathbf{C} = \{c_i\}_{1 \leq i \leq N_{\text{body}}}$ 为各个肢体部件的类别标号集。

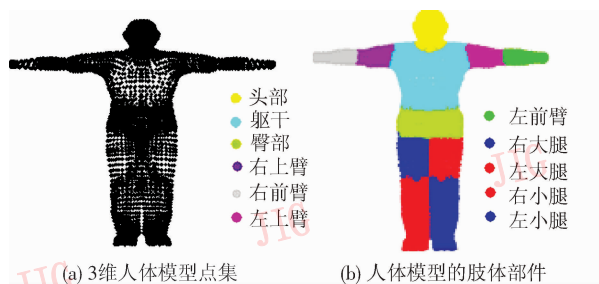


图 3 3 维人体模型及肢体部件

Fig. 3 3D human model and its body parts

3 维重构主体也是用 3 维点集表示的,令 N 表示重构主体上点的个数, $\mathbf{V} = \{v_n\}_{1 \leq n \leq N}$ 为相应的 3 维点坐标集。令 $\mathbf{Z} = \{z_n\}_{1 \leq n \leq N} (z_n \in \mathbf{C})$ 表示 3 维重构主体中每个点的肢体类别标记向量,则利用肢体分割算法分割肢体的目的就是获取肢体类别的标记向量 \mathbf{Z} 。

3.2 根据时域信息初步分割肢体部件

与文献[5]中的算法类似,本文首先利用前一

帧姿势变化后的 3 维人体模型来获取当前帧主体上的点的初始分割,即将主体上的点分割到人体模型上距离该点最近的肢体部件上。其中一例如图 4 所示。其中第 $k-1$ 帧时行为主体为直立姿势(如图 4(a)所示),第 k 帧时行为主体抬起了左前臂(如图 4(b)所示),根据第 $k-1$ 帧估计的姿势参数,人体模型发生相应的姿势变化(如图 4(c)所示)。本文将第 k 帧重构主体上的点分割到第 $k-1$ 帧人体模型上距离该点最近肢体部件上,其分割结果如图 4(d)所示,该圈上显示运动幅度较大的左上臂部分数据被错误地分割到了躯干上(红圈标示)。

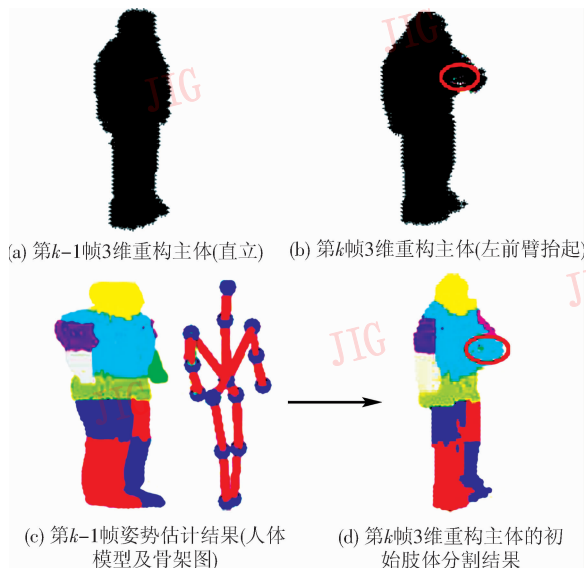


图 4 3 维重构主体肢体部件的初步分割

Fig. 4 The initial segmentation of the 3D actor

3.3 根据人体结构分割肢体部件

由图 4 可见,根据前一帧姿势估计反馈的时域信息来分割肢体部件时,姿势变化较快的部分可能会分割错误,但可以利用人体各个肢体部件之间存在的约束关系对肢体部件初始分割的结果进行修正。

令 $\mathbf{X} = \{x_{n,i}\}_{N \times N_{\text{body}}}$ 表示重构主体点集 \mathbf{V} 与肢体部件类别标号集 \mathbf{C} 之间的关系矩阵,其中,

$$x_{n,i} = \begin{cases} 1 & v_n \in c_i \\ 0 & v_n \notin c_i \\ \alpha & v_n \text{ 可能属于 } c_i \end{cases} \quad (1)$$

$0 < \alpha < 1$, 并且 $\sum_{i=1}^{N_{\text{body}}} x_{n,i} = 1$ 。对于点 v_n , 如果存在 $x_{n,i} = 1 (1 \leq i \leq N_{\text{body}})$, 则其肢体类别标号为 $z_n = c_i$ 。

大家知道,人体的各个肢体部件之间存在着多

种约束关系,如头部和躯干连接;上臂和前臂连接;小腿和大腿连接等等,这种约束关系可被用于确定关系矩阵 X 的某些元素。

下面以直立躯干为例来说明根据人体结构确定关系矩阵 X 的过程。由于在躯干上下左右分别连接着头、臀、左上臂、右上臂,同时在不同姿势下的躯干周围还可能存在着前臂、腿等等,所以初步分割结果中划分为躯干的点可能属于躯干周围其他肢体部件。本文用同中心的两个直立长方体 R_1, R_2 ($R_1 \subset R_2$) 来近似表示躯干数据所在的区域(如图 5 所示)。对于直立姿势下由初始分割划分为躯干的点,可进行如下的判断:①在 R_1 内的点一定属于躯干;②不在 R_1 内,而在 R_2 内的点可能属于躯干,也可能属于躯干周围的其他肢体部件,例如位于 R_1 下方的点 v_n 可能属于臀部;③在 R_2 之外的点一定不属于躯干。

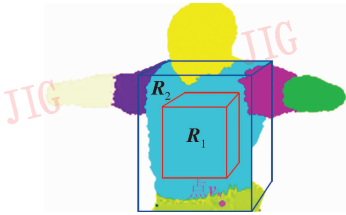


图 5 躯干结构
Fig. 5 Torso structure

根据这个规则,就可以确定关系矩阵 X 的部分元素,而根据人体肢体部件结构确定的关系矩阵 X 就将作为半监督聚类的初值,记为 $X^{(0)}$,并将由此确定的肢体类别标记向量记为 $Z^{(0)}$ 。设躯干的类别标号为 c_{torso} ,则对于初始分割中划分为躯干的点 v_n^{torso} 可以进行如下推断:

$$x_{n,c_{torso}}^{(0)} = \begin{cases} 1 & v_n^{torso} \in R_1 \\ 0 & v_n^{torso} \notin R_2 \\ \alpha & v_n^{torso} \notin R_1 \cap v_n^{torso} \in R_2 \end{cases} \quad (2)$$

如果 $x_{n,c_{torso}}^{(0)} = \alpha$,令 $\hat{C} = \{\hat{c}_i\}_{1 \leq i \leq \hat{N}}$ 表示可能存在于 v_n^{torso} 周围的其他肢体部件,则 $x_{n,\hat{c}_i}^{(0)} = \alpha$, ($1 \leq i \leq \hat{N}$)。

由于在不同行为姿势下的躯干不一定是直立的(如图 6(a)所示),所以首先需要把躯干数据纠正为直立姿势。而根据前一帧躯干的姿势参数对当前帧的躯干数据进行逆变换,就可以得到近似直立的躯干(如图 6(b)所示)。

对于点 v_n ,如果存在 $x_{n,i}^{(0)} = 1$ ($1 \leq i \leq N_{body}$),则称该点为确定点,并且可以确定其肢体

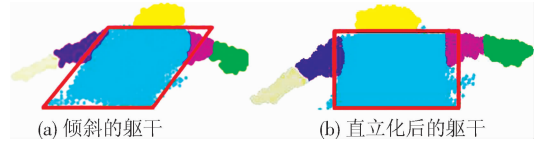


图 6 躯干直立化
Fig. 6 Torso modification

类别标号为 $z_n^{(0)} = c_i$;否则称该点为不确定点。图 7(b)为根据人体结构进行分割的结果,其中深灰色部分的点是不确定点。由此图可见,不同肢体部件的相邻边缘的点都变成了不确定体点。这些不确定点将会通过下一步半监督聚类算法划分到各个肢体部件中。

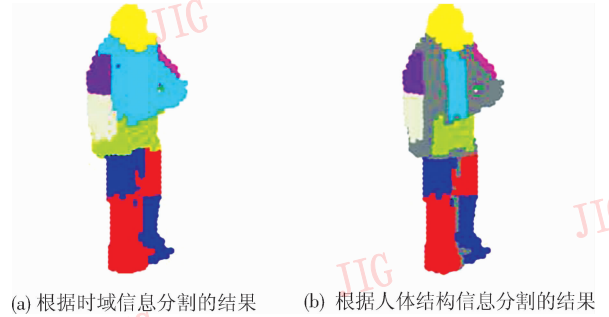


图 7 根据人体结构分割肢体的结果
Fig. 7 Body parts segmentation result with the body structure

3.4 基于半监督聚类的肢体部件分割

根据人体结构进行肢体分割使得每一个肢体部件由以下两部分组成:①确定属于这个肢体部件的点;②可能属于这个肢体部件的点。于是将这些不确定的点分类到各个肢体部件的问题就变成一个典型的半监督聚类^[11]问题。

如果用球体近似表示人体的头部,而用圆柱体近似表示其他肢体部件,则肢体部件分割问题就成为一个球体数据集和若干个圆柱体数据集的聚类问题。

令 $U = \{u_1, \dots, u_{N_{body}}\}$ 表示各个肢体类别的中心。对于用球体表示的头部,其类别的中心就是该类数据的欧氏中心;对于用圆柱体表示的其他肢体部件,其类别中心则是圆柱体的中心轴线段,本文用该类数据 KL 变换得到的主轴来近似表示该类中心轴,记为 $u_i = \{x_1, x_2\}$,其中 x_1, x_2 分别为聚类中心轴的两个端点坐标。令 $p_{n,i} = p(x_{n,i} = 1 | V)$ 表示重构点 v_n 属于第 i 个肢体部件的条件概率。

迭代法完成肢体部件分割的步骤如下:

(1) 设定初值

设迭代次数的初值为 $k = 0$, 关系矩阵 X 的初值为 $X^{(0)}$, 肢体类别标记向量的初值为 $Z^{(0)}$, 聚类中心的初值为 $U^{(0)}$, 其中 $X^{(0)}$ 和 $Z^{(0)}$ 已由上一步的肢体分割方法获得, 而聚类中心的初值 $U^{(0)}$, 则可以根据肢体类别标记向量的初值 $Z^{(0)}$ 来获取。令 $D_i^{(0)} = \{v_n \mid z_n^{(0)} = c_i, 1 \leq n \leq N\}$ 表示初始时属于第 i 类的点集。头部类别的中心 $u_{\text{head}}^{(0)}$ 为

$$u_{\text{head}}^{(0)} = \frac{\sum_{n=1, v_n \in D_{\text{head}}^{(0)}}^N v_n}{|D_{\text{head}}^{(0)}|} \quad (3)$$

其他肢体部件的聚类中心是一条线段, 令 $u_i^{(0)} = \text{KL}(D_i^{(0)})$ 表示点集 $D_i^{(0)}$ 由 KL 变换获得的主轴线段, 其中线段的两个端点由 $D_i^{(0)}$ 在中心轴投影的端点确定 (如图 8 所示)。由于有时在点集 $D_i^{(0)}$ 中存在噪声, 所以在选择端点时要滤除噪声点。

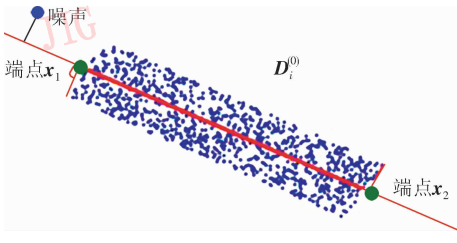


图 8 聚类中心计算示意图

Fig. 8 Sketch map of the clustering centers

(2) 更新条件概率 $p_{n,i}^{(k)}$

令 $d(v_n, u_i^{(k)})$ 表示点 v_n 到第 i 类中心 $u_i^{(k)}$ 的距离。若 $u_i^{(k)}$ 为 3 维点坐标 (头部的中心), 则 $d(v_n, u_i^{(k)})$ 为 v_n 到 $u_i^{(k)}$ 的欧氏距离; 若 $u_i^{(k)}$ 为线段的两个端点坐标, 则 $d(v_n, u_i^{(k)})$ 定义为 v_n 到线段 $\overrightarrow{x_1 x_2}$ 的距离 (如图 9 所示), 当 v_n 处于区域 I_0 时, $d(v_n, u_i^{(k)})$ 为点 v_n 到线段 $\overrightarrow{x_1 x_2}$ 的最短距离; 当 v_n 处于区域 I_1 时, $d(v_n, u_i^{(k)})$ 为点 v_n 到点 x_1 的距离, 当 v_n 处于区域 I_2 时, $d(v_n, u_i^{(k)})$ 为点 v_n 到点 x_2 的距离。

由于不同肢体部件的半径并不相同, 可以理解为不同类别的方差不同, 所以要将 $d(v_n, u_i^{(k)})$ 归一化。令 $\hat{d}(v_n, u_i^{(k)})$ 表示点 v_n 到第 i 类中心 $u_i^{(k)}$ 的归一化距离, 则

$$\hat{d}(v_n, u_i^{(k)}) = \frac{d(v_n, u_i^{(k)})}{r_i} \quad (4)$$

其中, r_i 表示第 i 个肢体部件的半径。

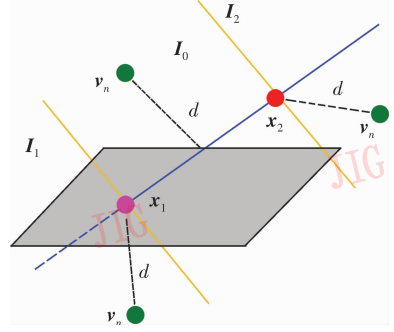


图 9 点到线段距离示意图

Fig. 9 Sketch map of the distance between segment and point

由点 v_n 到聚类中心 $u_i^{(k)}$ 的归一化距离 $\hat{d}(v_n, u_i^{(k)})$ 就可以近似计算以下条件概率:

$$p_{n,i}^{(k)} = \begin{cases} \frac{1}{\sum_{j=1, 0 < x_{n,j}^{(k)} < 1}^{N_{\text{body}}} \frac{\hat{d}(v_n, u_i^{(k)})}{\hat{d}(v_n, u_j^{(k)})}} & 0 < x_{n,i}^{(k)} < 1 \\ 0 & x_{n,i}^{(k)} = 0 \\ 1 & x_{n,i}^{(k)} = 1 \end{cases} \quad (5)$$

(3) 更新关系矩阵 $X^{(k+1)}$ 及肢体类别标记向量 $Z^{(k+1)}$

当点 v_n 属于第 i 类的条件概率 $p_{n,i}^{(k)}$ 大于某一阈值 T_1 ($0 < T_1 < 1$) 时, 则将 v_n 划分为第 i 类。即 $p_{n,\hat{i}}^{(k)} = \max_{1 \leq i \leq N_{\text{body}}} (p_{n,i}^{(k)})$; 若 $p_{n,\hat{i}}^{(k)} \geq T_1$, 则 $v_n \in c_{\hat{i}}$, $x_{n,\hat{i}}^{(k+1)} = 1$, $x_{n,i}^{(k+1)} = 0$ ($i \neq \hat{i}$), $z_n^{(k+1)} = c_{\hat{i}}$ 。

当点 v_n 属于某一类的条件概率 $p_{n,i}^{(k)}$ 小于某一阈值 T_2 ($0 < T_2 < T_1$) 时, 则认为 v_n 肯定不属于第 i 类, 即若 $p_{n,i}^{(k)} \leq T_2$, 则 $v_n \notin c_i$, $x_{n,i}^{(k+1)} = 0$ 。

实验中的两个阈值根据经验选择为 $T_1 = 0.8$, $T_2 = 0.1$ 。

(4) 更新类别中心 $U^{(k+1)}$

头部类别的中心 u_{head}^{k+1} 为

$$u_{\text{head}}^{(k+1)} = \frac{\sum_{n=1}^N p_{n,c_{\text{head}}}^{(k)} v_n}{\sum_{n=1}^N p_{n,c_{\text{head}}}^{(k)}} \quad (6)$$

其他肢体部件的类别中心为

$$D_i^{(k+1)} = \{v_n \mid z_n^{(k+1)} = c_i, 1 \leq n \leq N\} \quad (7)$$

$$u_i^{(k+1)} = \text{KL}(D_i^{(k+1)}) \quad (8)$$

(5) 迭代终止条件的判断

当相邻两次迭代的聚类中心 $U^{(k)}$, $U^{(k+1)}$ 的距离小于某一阈值 T_3 或者达到最大迭代次数时,则迭代终止,否则 $k = k + 1$, 转向步骤(2)。当相邻两次迭代的聚类中心是两条线段时,则用以下两个距离来表示聚类中心之间的距离:第1个距离是这两条线段的空间夹角;第2个距离是这两条线段端点之间的距离,相应的就有两个截止阈值 T_3^{ang} , T_3^{dis} 。 T_3^{ang} 由姿势空间的量化间隔决定,实验中取 $T_3^{\text{ang}} = 3^\circ$; T_3^{dis} 由3维空间的量化间隔决定,3维空间的量化间隔为 3.125cm,本文取 $T_3^{\text{dis}} = 3.5\text{cm}$ 。

经过半监督聚类迭代求解后,肢体分割的结果如图 10(c)所示,由此可见,以前划分错误的左上臂基本分割正确了。



(a) 时域信息分割结果 (b) 人体结构分割结果 (c) 半监督聚类分割结果

图 10 肢体分割结果图

Fig. 10 Segmentation result of the body parts

4 实验结果与分析

4.1 实验数据

为验证本文算法的分割效果,利用法国国家信息与自动化研究所(INRIA)的 IXMAS 数据库进行了分割实验^[12],该数据库已被用于基于模板的行为识别之中^[13, 14]。实验就是对这个数据库中6个行为主体进行肢体分割。

数据库中的行为主体由3维点集表示,整个3维空间的大小为 $2\text{m} \times 2\text{m} \times 2\text{m}$,划分为 $64 \times 64 \times 64$ 个体素,其中每个行为主体都进行了以下6种行为:①在 stand 序列中行为主体站在原地不动;②turn 序列中行为主体在原地转圈;③check watch 序列中行为主体抬起前臂看手表;④wave 序列中行为主体挥手;⑤kick 序列是指踢腿的序列;⑥point 序列是抬胳膊向前指的序列,各个行为主体自由选择视角。每一行为序列的平均长度为30帧,每一个行为序列都是以站立姿势开始,并以站立姿势结束的。

4.2 初始化

初始化包括人体模型的形状初始化和姿势初始

化。由于不同行为主体的身高、胖瘦各不相同,所以需要将人体模型的形状进行初始化。同时由于 SSC-BPS 算法是要根据前一帧姿势估计的结果来获得当前帧肢体的初步分割的,所以也需要对人体模型进行姿势的初始化。

本文假设行为序列第1帧中主体是直立姿势的,则由此就可以获取行为主体的身高、胖瘦、初始姿势等参数,并据此对人体模型进行初始化。

4.3 实验结果与分析

在进行肢体分割时,可将人体划分为11个肢体部件,但是同一侧的前臂和上臂,以及同一侧的大腿和小腿之间往往比较难于区分,所以应先将左前臂和左上臂、右前臂和右上臂、左大腿和左小腿、右大腿和右小腿的数据分别合并,然后同时优化合并后的肢体部件姿势参数。

图 11 为行为主体 daniel 在6种行为序列中肢体分割及姿势估计结果,其中人体骨架中的绿色线段表示的是右臂。由此图可见,本文提出的 SSC-BPS 算法可以适应不同行为序列下的姿势估计,并且可以分割距离很近或者粘连在一起的肢体部件。

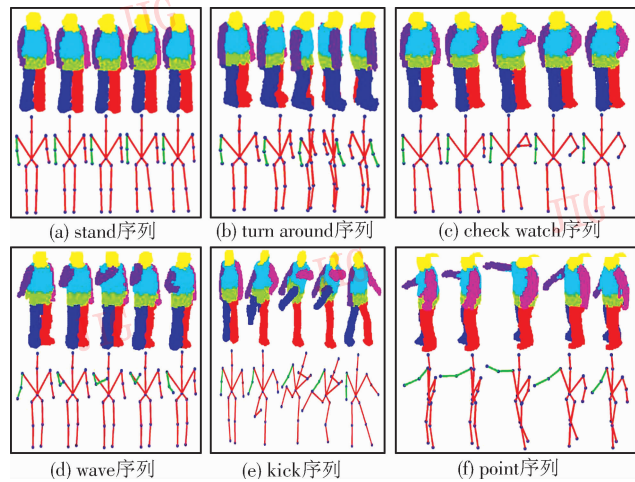


图 11 不同行为下行行为主体 daniel 的肢体分割结果

Fig. 11 Segmentation result of different activities

为了与只利用姿势估计反馈的时域信息算法进行比较,图 12 中列出了在 check watch 行为序列中相邻3帧只利用时域信息进行肢体部件分割的结果。由此图可见,在后两帧中主体的前臂已经放下了,但是由于时域的延迟效应,使得前臂部分分割错误。

图 13 为6个行为主体在6种行为序列中的某一帧的肢体分割及姿势估计结果。由此图可见,本文的肢体分割算法可以适应不同的行为主体。

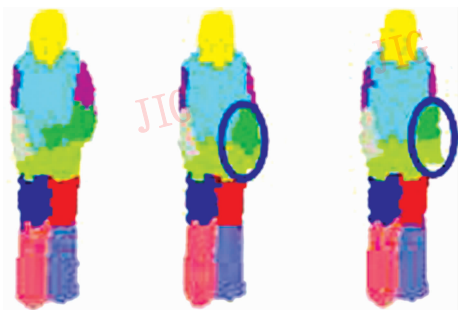


图 12 累积误差

Fig. 12 Cumulative error

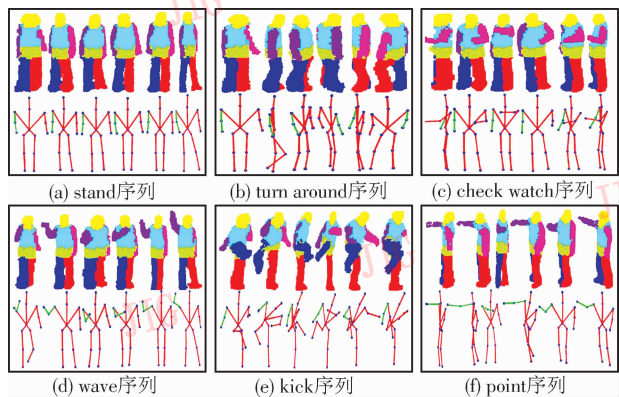


图 13 不同行为主体的肢体分割结果

Fig. 13 Segmentation result of different actors

由于原始实验数据集中没有提供各个肢体部件分割的真实结果,且 3 维数据分割的手工标定是非常困难的,而进行肢体分割的目的则是要进行姿势估计和行为分析,所以本文用姿势估计的精度来间接分析肢体分割算法的性能。姿势估计有以下两个量化评测准则:①关节的位置误差;②关节的角度误差。文献[15]在 2 维图像上估计姿势,walk 序列中关节角度的平均误差为 5.5°。文献[16]也是在 2 维图像上估计姿势,在 CMU 的 MoBo 数据库上关节位置的平均误差为 19.0 个 pixels。由于 3 维真值标定比较困难,因此提供 3 维量化性能评价的算法并不多,例如文献[5]、[6]只提供了一些样本示例。文献[17]先将 3 维空间中关节角度估计值的 Savitzky-Golay 光滑值作为真值,然后计算光滑值与估计值之间的误差,在 walk 序列中,关节角度的平均误差为 3°。

本文以手工标定的 3 维重构主体上的关节位置和姿势估计后人体模型上关节的位置误差来量化分析算法性能。其中手工标定 3 维关节点的算法如下:首先将 3 维重构主体投影到多个视角(本文以

30°间隔为一个视角);然后在各个视角的 2 维图像上标定关节的位置;最后用最小二乘法求得各个关节的 3 维坐标(如图 14 所示)。

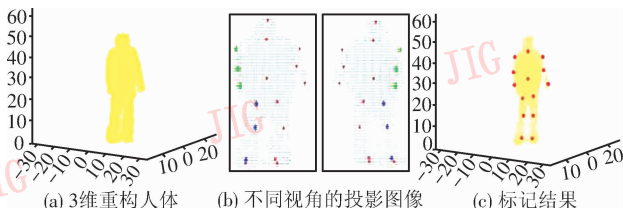


图 14 手工标定 3 维关节点

Fig. 14 Label joints

图 15 为行为主体 daniel 在 6 种行为序列中的关节位置误差曲线图。关节的位置误差主要由以下 5 个因素引起:(1)体素的空间量化误差,空间体素的量化间隔为 3.125cm;(2)姿势空间的量化误差,姿势空间的量化间隔为 3°;(3)人体模型的各个肢体部件与重构的 3 维主体各个肢体部件之间的形状误差;(4)手工标定 3 维关节点不精确造成的误差;(5)肢体分割错误引起的误差。当仅考虑肢体分割错误造成的影响时,则可以认为由前 4 个因素造成的误差为噪声。由于在行为主体直立姿势下,肢体分割错误率是很低的,所以可以由 stand 序列行为下的关节误差来近似推断出由前 4 个因素带来的噪声值。

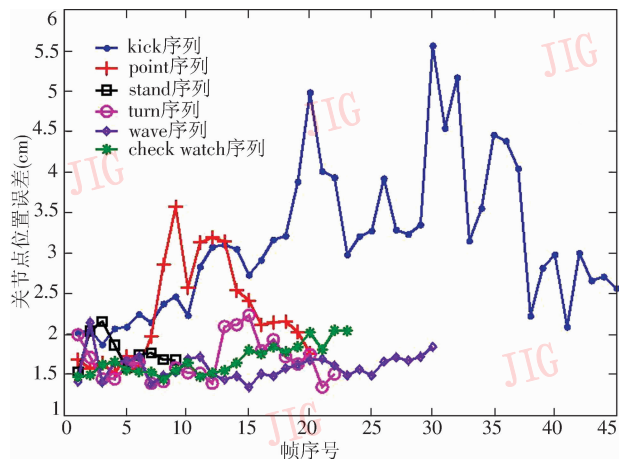


图 15 关节点位置误差曲线图

Fig. 15 Curve of the joint position error

由图 15 可见,对于关节位置变化较小的序列,如 turn、wave、check watch 等序列,其关节的位置误差与 stand 序列的关节位置误差相差不大;而对于关节位置变化较大的 point 序列和 kick 序列,其行为序列中间的若干帧的误差要大于 stand

序列的误差。kick序列中关节位置的最大误差为5.56cm,大约相当于 5.5° ;kick序列的平均关节位置误差为3.14cm,约相当于 4° 。而行为分析是可以容忍一定的姿势估计误差的,例如本文以kick序列的行为作为异常行为,同时所有6个主体的6种行为中检测该异常行为,其检测正确率为100%。另外,由图15也可以看到,关节估计的误差并不会随着时间的推移而逐渐增大,而只是随着行为复杂度的增加而增大。

本文实验平台为Pentium 4, 3.0GHz CPU, 1.0G内存,肢体分割算法的运算速度可以达到5fps,目前还达不到实时处理的水平,这也是以后需要改进的方向。

5 结论

本文提出了一种适用于行为分析的基于半监督聚类的3维肢体分割算法,该算法综合利用了前一帧姿势估计反馈的时域信息、人体肢体部件之间的结构信息以及人体肢体部件的形状信息,取得了较好的分割效果。实验证明,该算法不仅对不同视角、不同主体、不同行为都具有良好的适应性,并且能够分割距离很近或者粘连在一起的肢体部件。

在本文算法中,人体模型的点数为4964个,3维重构行为主体的平均点数约为2500个,并且要将人体肢体分割为成11个部件,这样详细的部件分割给本文算法带来了很大的计算量,目前还达不到实时处理的水平。因此如何降低复杂度,提高速度是本文算法需要改进的一个方向。

参考文献 (References)

- 1 Park Sangho, Aggarwal J K. Segmentation and tracking of interacting human body parts under occlusion and shadowing [A]. In: Proceedings of the IEEE Workshop on Motion and Video Computing [C], Orlando, Florida, USA, 2002;105 ~ 111.
- 2 Chen Chih-chiang, Hsieh Jun-wei, Hsu Yung-tai, *et al.* Segmentation of human body parts using deformable triangulation [A]. In: Proceedings of the 2006 International Conference of Pattern Recognition [C], Hong Kong, China, 2006;355 ~ 358.
- 3 Ramanan Deva, Forsyth D A, Zisserman Andrew. Strike a pose: tracking people by finding stylized poses [A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C], San Diego, CA, USA, 2005: 271 ~ 278.
- 4 Cao Dan-hua, Tang Jin-kuan, Wu Yu-bin. Algorithm for fast extracting human limb contours using searching NCM points [J]. Journal of Huazhong University of Science and Technology (Nature Science

- Edition), 2007, 35(5): 16 ~ 18. [曹丹华, 汤金宽, 吴裕斌. 基于最小负曲率点搜索的快速人体肢体提取算法. 华中科技大学学报(自然科学版), 2007, 35(5): 16 ~ 18.]
- 5 Cheung K, Kanade T, Bouguet J Y, *et al.* A real time system for robust 3D voxel reconstruction of human motions [A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C], Hilton Head Island, South Carolina, USA, 2000: 714 ~ 720.
- 6 Kehl R, Bray M, VanGool L. Full body tracking from multiple views using stochastic sampling [A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C], San Diego, CA, USA, 2005: 20 ~ 25.
- 7 Slabaugh Gregory G, Culbertson W Bruce, Malzbender Thomas, *et al.* A survey of methods for volumetric scene reconstruction from photographs [A]. In: Proceedings of the International Workshop on Volume Graphics [C], Stony Brook, New York, USA, 2001: 21 ~ 22.
- 8 Franco Jean-Sebastien. EPVH [CP/OL]. <http://perception.inrialpes.fr/~Franco/EPVH/>, 2007-12-01.
- 9 Deutscher J, Blake A, Reid I. Articulated body motion capture by annealed particle filtering [A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C], Hilton Head Island, South Carolina, USA, 2000:126 ~ 133.
- 10 Gavrilu D, Davis L. 3D model-based tracking of humans in action: a multi-view approach [A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C], San Francisco, CA, USA, 1996: 73 ~ 80.
- 11 Dong Anlei, Bhanu Bir. A New semi-supervised EM algorithm for image retrieval [A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C], Madison, Wisconsin, USA, 2003: 662 ~ 667.
- 12 Weinland Daniel. The Multiple-video Data Used Here are From INRIA Rhone-Alpes' Multiple-camera Platform Grimace and PERCEPTION Research Group [DB/OL]. <http://charibdis.inrialpes.fr>, 2007-12-01.
- 13 Weinland Daniel, Ronfard Remi, Boyer Edmond. Free viewpoint action recognition using motion history volumes [J]. Computer Vision and Image Understanding, 2006, 104(2 ~ 3): 249 ~ 257.
- 14 Weinland Daniel, Boyer Edmond, Ronfard Remi. Action recognition from arbitrary views using 3D exemplars [EB/OL]. <http://perception.inrialpes.fr/Publications/2007/WBR07/WeinlandBoyerRonfard-ICCV07.pdf>, 2007-12-01.
- 15 Patrick Peursum, Svetha Venkatesh, Geoff West. Tracking-as-recognition for articulated full-body human motion analysis [A]. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition [C], Minneapolis, Minnesota, USA, 2007: 1 ~ 8.
- 16 Alireza Fathi, Greg Mori. Human pose estimation using motion exemplars [EB/OL]. http://www.cs.sfu.ca/~mori/research/papers/fathi_motion_exemplars_iccv07.pdf, 2007-12-01.
- 17 Mikić I. Human body model acquisition and tracking using multi-camera voxel data [D]. Thesis of PhD. University of California, San Diego, CA, USA, 2002.