

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2016)02-0225-10

论文引用格式: Luo P, Liu X P. Robust foot plant detection for Kinect captured motion data[J]. Journal of Image and Graphics, 2016, 21(2): 0225-0234. [罗飘, 刘晓平. 面向 Kinect 运动数据的鲁棒足迹检测[J]. 中国图象图形学报, 2016, 21(2): 0225-0234.] [DOI: 10.11834/jig.20160212]

面向 Kinect 运动数据的鲁棒足迹检测

罗飘, 刘晓平

合肥工业大学计算机与信息学院, 合肥 230009

摘要: 目的 Kinect 可实时获取运动数据且较传统的运动捕捉设备采集成本低廉, 在运动数据捕捉方面得到了广泛应用。但 Kinect 获取的运动数据精度较低, 现有运动数据处理算法难以适用。方法 针对运动数据处理的关键步骤足迹检测问题, 提出面向 Kinect 运动数据的鲁棒足迹检测算法。首先使用自适应的双边滤波算法降低 Kinect 运动数据中的噪声; 其次定义多种脚部运动特征并用于分类, 优化分类效果; 最后使用支持向量机(SVM) 算法训练决策函数并用于足迹检测。结果 本文算法应用于多种类型运动数据后, 可以有效地减少 Kinect 运动数据中的噪声, 足迹检测的时间性能以及准确性良好, 其中足迹检测的准确率比经典的基准线方法提高了 10% 左右, 比 K 近邻方法提高了 8% 左右, 检测一帧运动足迹的速度为 K 近邻方法的 7 倍左右。结论 对实验结果的分析证明算法具有良好的鲁棒性、时间性能以及准确率, 可广泛应用于运动数据的处理之中。

关键词: 计算机动画; Kinect; 运动编辑; 足迹检测; 支持向量机

Robust foot plant detection for Kinect-captured motion data

Luo Piao, Liu Xiaoping

School of Computer and Information, Hefei University of Technology, Hefei 230009, China

Abstract: **Objective** Kinect can be utilized to capture motion data in real time. Given that its cost is lower than that of traditional motion-capture devices, Kinect is widely used to capture motion data. However, the noise in Kinect-captured motion data makes the quality of motion data relatively unsatisfactory. Thus, previous data-processing methods failed to handle such data well. **Method** Foot plant detection is a key procedure in motion editing; it detects whether the character's foot is on the ground. A robust foot plant detection algorithm for Kinect-captured motion data is proposed in this study. First, an adaptively bilateral filtering method is proposed to reduce the noise in Kinect-captured motion data. Second, multiple features of the motion data are defined and utilized to optimize the effect of foot plant detection. Finally, a decision function is trained with the support vector machine algorithm and applied to foot plant detection. **Result** After being applied to a dataset that consists of various types of motion, the noise in the Kinect-captured motion data was reduced effectively. The accuracy of foot plant detection increased by 6% after applying the proposed adaptively bilateral filtering method. Good time performance and high accuracy of foot plant detection were acquired as well. The foot plant detection accuracy of the proposed detection algorithm increased by 11% and 8% compared with that of the baseline method and the K nearest-neighbor

收稿日期: 2015-07-09; 修回日期: 2015-08-20

基金项目: 国家自然科学基金项目(61370167, 61305093); 国家科技支撑计划项目(2012BAJ08B01)

第一作者简介: 罗飘(1991—), 男, 现为合肥工业大学计算机与信息学院硕士研究生, 主要研究方向为计算机动画、机器学习等。E-mail: ropeal@163.com

通信作者: 刘晓平, 教授, 博士生导师, E-mail: lisjhfut@hfut.edu.cn

Supported by: National Natural Science Foundation of China(61370167, 61305093); National Key Technology Research and Development Program of the Ministry of Science and Technology of China (2012BAJ08B01)

method, respectively. The time consumed in the detection of the motion data of one frame is a seventh of that of the K nearest-neighbor method. **Conclusion** Experimental results proved the effectiveness and robustness of the proposed foot plant detection algorithm. Thus, this algorithm can be widely utilized in motion data processing.

Key words: computer animation; Kinect; motion edit; foot plant detection; support vector machine

0 引言

使用运动捕捉仪获取的运动数据精度较好但是成本高昂,通常仅能在大型影视或游戏公司得到应用,普通用户的需求受到了限制。微软在2010年发布的Kinect可以实时获取到人运动时主要关节的空间坐标信息,而且该设备价格低廉,使得普通用户获取实时运动数据来操控游戏中的动画角色成为可能。但是Kinect采集到的运动数据存在较多的噪声,图1示意了足部运动时常见的悬空和穿地噪声(骨骼整体为红色,为区分骨骼正反面将左右脚分别着色为红色和蓝色,下同),传统的针对运动捕捉仪的运动数据处理算法难以适用。

用运动捕捉设备获得的运动数据只能在特定场景下应用,当场景或用户需求改变时就需要对原始运动数据进行编辑以后才能使用。运动编辑算法的核心是在保持运动数据中的时空约束的同时对新运动进行求解。足迹是生物运动的一种典型时空约束,表现为脚部接触地面并保持固定的一段时间^[1]。

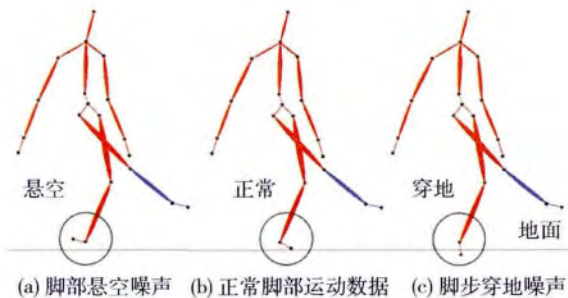


图1 Kinect运动数据中的噪声

Fig. 1 Noise in Kinect captured motion data((a) dangling noise of foot; (b) the correct foot; (c) the across ground noise)

几乎所有的运动编辑算法都需要首先检测并记录原始运动的足迹,以便于编辑后的足迹修正。足迹检测的目标正是从运动数据中检测出属于足迹的数据帧。现有的足迹检测方法主要包括人工检测法、基准线法以及分类器法,并且已经在运动重定

向^[2]、运动合成^[3]以及运动压缩^[4]等问题中得到了应用。

人工检测方法的正确率很高,但其需要大量的人工工作,难以处理大批量的运动数据。基准线方法一般通过人工指定角色脚部关节与地面的距离以及脚部运动速度(不考虑方向)的阈值进行足迹检测。Kovar等人^[1]通过人工指定脚部关节与地面距离的阈值来确定足迹;Ménardais等人^[5]规定距离和速度同时小于各自指定的全局阈值的运动时段为足迹,然而这类方法对足迹的边缘判断不够准确,过大的噪声还可能使得足迹发生断裂。Sung等人^[6]将指定全局阈值的方法改为结合运动模式的自适应阈值方法,但该类方法依赖于预设的运动模式,不能有效处理种类繁多的运动类型。分类器方法主要以Ikemoto等人^[7]提出的基于K近邻(KNN)的足迹检测方法为代表,但由于K近邻算法本身为一种懒惰的学习算法^[8],对新数据进行分类时每次都需要扫描运动数据库中的所有运动数据再进行判断,这使得该方法的时间性能受到了影响。所以,使用上述3种方法处理运动捕捉仪获取的数据时在鲁棒性和时间性能上均存在问题,而由于Kinect采集到的运动数据精度较低,使用上述3种方法处理时出现的问题更为明显。

对Kinect运动数据进行足迹检测难点在于数据噪声较大,所以必须要对采集到的数据进行有效降噪,在检测时还需要考虑采用多特征融合的方法以提高检测精度。因此,提出一种面向Kinect运动数据的鲁棒足迹检测算法来解决因Kinect数据精度低而凸显的一些问题。如图2所示,首先使用自适应的双边滤波算法降低Kinect运动数据中的抖动;其次,对滤波后的数据进行类别标记,并选择合适分类特征;最后使用支持向量机(SVM)算法训练决策函数并用于足迹检测。在多种运动类型实验数据上的实验结果均表明,本文算法能够有效去除Kinect运动数据中的噪声,鲁棒性较以往算法更好,并且在识别率和识别速度上均优于现有足迹检测算法。

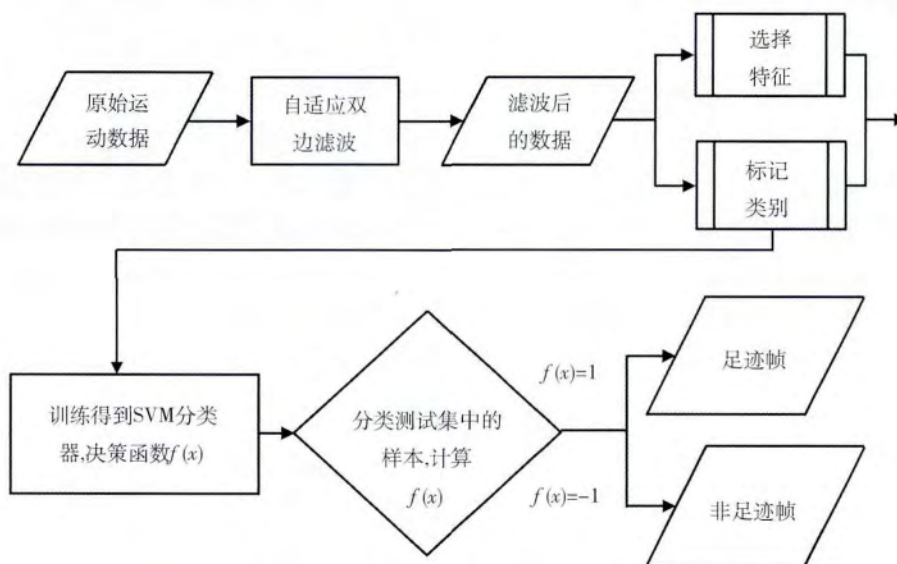


图2 系统流程示意图

Fig. 2 System overview(First , the proposed adaptive bilateral filter is applied to reduce the noise in the Kinect captured motion data; secondly , select features and label the denoised data; thirdly , train a classifier with SVM algorithm; finally , detect the footprint of the test sample with the trained classifier)

1 Kinect 运动数据相关记号

在 Kinect 官方开发包中, 标准人的骨架是由 20 个关节组成的, 其骨架结构如图 3 所示。设第 j 个关节用 o_j 表示, 每个关节的具体含义如表 1 所示。为描述 Kinect 运动数据的运动特征, 定义如下记号:

- 1) 骨骼 称连接两个关节的线段为骨骼, 将由关节 o_j 和 o_k 构成的骨骼记为 $b_{j,k}$;
- 2) 骨骼平面 称 3 个关节构成的平面为骨骼平面, 将由关节 o_j, o_k, o_l 构成的骨骼平面记为 $s_{j,k,l}$;
- 3) 关节坐标 第 i 帧第 j 个关节的空间坐标记为 $p_{i,j}$;
- 4) 姿态 称第 i 帧时骨架的各个关节的坐标集合为一个姿态, 记为 $M_i = \{p_{i,j} | j = 0, 1, \dots, 19\}$;
- 5) 运动 称由 n 个姿态组成的运动序列为一段长度为 n 的运动, 记为 $M = \{M_i | i = 1, 2, \dots, n\}$ 。

2 Kinect 运动数据去噪

Kinect 利用红外线摄像机来测量空间物体的深度信息, 由于环境光的影响和物体本身材料的影响,

在采集深度数据时会出现噪声, 表现为数据的抖动^[9]。这些抖动不仅给计算用于分类的特征带来了麻烦, 甚至也使得人在标定足迹时产生混乱, 所以必须首先对得到的数据进行有效的去噪。

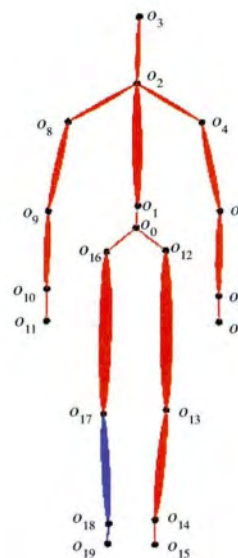


图3 Kinect 骨架(红色为左脚, 蓝色为右脚)

Fig. 3 Skeleton in Kinect(red denotes left , blue denotes right)

在处理 Kinect 运动数据时, 发现其结构与图像数据的结构有相似之处。数字图像由像素组成, 每个像素中都保存着该像素的颜色信息(如 RGB), 类

似地, Kinect 骨架是由人体各个关节组成, 每个关节的数据包含了其空间 3 维坐标(X、Y、Z) 信息。将运动的帧数看成图像的宽度, 将角色的关节数看成图像的高度, 每一个关节的 3 维坐标 X、Y、Z 可以映射到图像中的 RGB 空间。

在 Kinect 运动数据中, 一个关节某一帧的运动数据与其时间邻域和空间邻域内数据有直接联系, 在对运动数据去噪的同时, 需要保留运动的细节信息, 在图像中这种细节信息就表现为图像的边缘

特征。在图像处理中, 双边滤波算法^[10] 结合像素间的空间邻域信息和灰度相似度信息对图像进行滤波, 同时还能保留图像的边缘特征。为了在保留运动数据边缘特征的同时达到良好的滤波效果, 借鉴图像中的双边滤波算法对运动数据进行处理。由于传统的双边滤波算法的参数是由经验给定的, 对于种类繁多、数据量庞大的运动捕捉数据, 这种参数选择方法带有很大的盲目性, 因此提出了一种用于处理运动数据的自适应双边滤波算法(ABF)。

表 1 Kinect 骨架关节点具体含义

Table 1 Description of joints in Kinect skeleton

符号	语义	符号	语义	符号	语义	符号	语义
o_0	根关节 HIP-CENTER	o_5	左肘 ELBOW-LEFT	o_{10}	右手腕 WRIST-RIGHT	o_{15}	左脚 FOOT-LEFT
o_1	脊柱 SPINE	o_6	左手腕 WRIST-LEFT	o_{11}	右手 HAND-RIGHT	o_{16}	右臀 HIP-RIGHT
o_2	肩膀中心 SHOULDER-CENTER	o_7	左手 HAND-LEFT	o_{12}	左臀 HIP-LEFT	o_{17}	右膝盖 KNEE-RIGHT
o_3	头 HEAD	o_8	右肩膀 SHOULDER-RIGHT	o_{13}	左膝盖 KNEE-LEFT	o_{18}	右脚踝 ANKLE-RIGHT
o_4	左肩膀 SHOULDER-LEFT	o_9	右肘 ELBOW-RIGHT	o_{14}	左脚踝 ANKLE-LEFT	o_{19}	右脚 FOOT-RIGHT

2.1 双边滤波

对于第 i 帧运动中第 j 个关节的空间坐标 $p_{i,j}$, 定义其 R 邻域为

$$NP_{i,j}(R) = \{p_{i,j} | |i - t| \leq R, i, t \in \mathbf{N}^+\} \quad (1)$$

对于第 i 帧运动中的第 j 个关节, 设其滤波后的空间坐标为 $bfp_{i,j}$, 则有

$$bfp_{i,j} = \frac{1}{W_p} \times \sum_{p_{t,j} \in NP_{i,j}(R)} (G_{\sigma_t}(\|t - i\|) \times \quad (2)$$

$$G_{\sigma_d}(\|p_{t,j} - p_{i,j}\|) p_{t,j})$$

式中, 归一化因子 W_p 为

$$W_p = \sum_{p_{t,j} \in NP_{i,j}(R)} (G_{\sigma_t}(\|t - i\|) \times \quad (3)$$

$$G_{\sigma_d}(\|p_{t,j} - p_{i,j}\|))$$

时间临近度因子和坐标临近度因子 G_{σ_t} 、 G_{σ_d} 是由形如

$$G_{\sigma}(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (4)$$

的标准正态分布所确定, 决定了第 i 帧运动中第 j 个关节 R 邻域 $NP_{i,j}(R)$ 内的关节空间坐标对该关

节点空间坐标的影响权重, 高斯函数的参数 σ_t 、 σ_d 则决定了双边滤波器的性能, 在实验中发现 $R = 3$ 时算法的时间性能和准确度能得到较好的平衡。

2.2 自适应参数选择

高斯滤波中 σ 的取值决定了滤波窗口中高斯曲线的扩散程度, σ 取值越大, 高斯曲线下降越慢, 平滑效果越明显, 但太大的 σ 取值可能导致过平滑现象。传统的双边滤波的参数是由经验给定的, 对于种类繁多、数据量庞大的运动捕捉数据, 这种参数选择方法带有很大的盲目性, 为此提出了一种用于运动数据的参数自适应双边滤波算法。

图像边缘和细节的清晰度与 σ_d 的取值相关, 而运动中数据变化大的帧的看成是运动中的边缘和细节, 为了在降噪的同时保护好这些边缘和细节信息, 需要选取合适的 σ_d 。 σ_d 的选取和数据中本身噪声的水平存在直接的联系, 设数据中本身噪声的标准差为 σ_n , 通过实验得出两者的比值 $\sigma_d : \sigma_n$ 应该在 $[2, 3]$ (本文取 $\sigma_d = 2\sigma_n$)^[11], 因

此要获得合适的 σ_d , 首先需要估计数据中噪声本身的标准差 σ_n 。

噪声估计有基于数据块的方法和基于滤波器的方法^[12], 采用的是基于数据块的方法。将运动数据 M 根据关节的编号 j 分为 20 块, 即

$$Q_j = \{p_{i,j} \mid i = 1, 2, \dots, n\} \\ j = 0, 1, \dots, 19 \quad (5)$$

对每一块求其标准差, 标准差最小的块中的坐标基本一致, 可近似为噪声标准差 σ_n 。

一个窗口中数据方差的大小反映了窗口中数据的平滑程度, 数据点与其邻域边界存在一个相似度^[13-14], 基于相似度提出了一种自适应 σ_i 的方法。为了方便叙述, 定义第 i 帧时第 j 个关节空间坐标 R 邻域的边界为

$$B_{i,j}(R) = \{p_{i,j} \mid p_{i,j} \in (NP_{i,j}(R) - NP_{i,j}(R-1))\} \quad (6)$$

则关节 j 在第 i 帧时与其 R 邻域的边界的相似度为

$$U_{i,j}(R) = \frac{\sum_{p_{i,j} \in B_{i,j}(R)} \exp\left[-\frac{(p_{i,j} - p_{i,j})^2}{2\sigma_u^2}\right]}{|B_{i,j}(R)|} \quad (7)$$

式中, $|B_{i,j}(R)|$ 表示 $B_{i,j}(R)$ 中元素的个数, σ_u 是一个与坐标差异 $\|p_{i,j} - p_{i,j}\|$ 有关的局部统计量, 其定义和计算方法在^[14]中有详细介绍, 由此可以计算出在半径为 3 的滤波窗口中, 关节 j 在第 i 帧时与其 1、2、3 邻域边界的相似度为 $U_{i,j}(r)$, ($r = 1, 2, 3$)。

显然相似度越高, 说明其与关节 j 的坐标越相似, 相似度的变化程度也反映了在这个窗口中关节 j 的坐标的平滑程度, 方差可以用来度量数据的变化程度, 因此可以得到第 i 帧的数据在滤波窗口中的平滑程度为

$$\sigma_s^2 = \text{var}\{U_{i,j}(r) \mid j \in \{0, 1, \dots, 19\}, \\ r \in \{1, 2, 3\}\} \quad (8)$$

式中, var 表示求方差。 σ_s 与 σ_i 密切相关, 类似于前文 σ_d 的选择, 本文取 $\sigma_i = 2\sigma_s$ 。

3 Kinect 运动数据特征定义

在分类中, 特征的好坏直接影响到分类的效果, 一组好的特征要能够反映不同类型数据的本质属性。在处理种类繁多、数据量庞大的运动数

据时, 特征的选取不可避免会成为其中的一个焦点^[15]。要对足迹进行检测, 必需选择一些有效的特征。在基准线方法中, 使用了脚与地面的距离以及脚的运动速度两个特征^[5]; 在 K 近邻方法中, 直接使用脚的空间 3 维坐标作为特征^[7]; 在使用 Kinect 捕捉的运动数据时, 噪声的存在会导致关节点坐标数据发生抖动, 在人脚落地之后, 脚的速度和高度也不能保持为零, 因此仅仅使用脚的运动速度和脚与地面距离进行足迹检测是不够的。在使用上述脚与地面距离、脚的运动速度以及脚的空间位置坐标作为足迹分类特征的同时, 将运动时骨骼之间的角度变化加入考虑, 提出了大腿与小腿夹角、大腿与盆骨平面夹角, 以及小腿与脚掌之间夹角等一些新的特征作为足迹分类特征, 以此降低运动数据噪声给足迹分类带来的困难。参照陈松乐等人^[16]的特征分类方法, 将上述特征分为空间特征和时间特征两类。

3.1 空间特征

人体运动是由一系列姿态组成的, 姿态可由人体关节点形成的点、线、面进行描述。几何元素点、线、面之间的关系主要表现为相互之间的距离和角度, 通过分析这些几何元素的关系, 可以提取出姿态的空间特征。由于人在运动时上半身动作有很强的随意性, 所以在对足迹进行检测时主要考虑人体下半身的运动特征。不失一般性, 仅对右脚的的特征进行描述, 左脚的特征可以用相同的方法进行计算。

1) 右脚尖 o_{19} 与地面的距离 F_1 。设地面的方程为 $Ax + By + Cz + D = 0$, 第 i 帧时右脚尖的空间坐标为 $p_{i,19} = (x_{i,19}, y_{i,19}, z_{i,19})$, 则有

$$F_1 = \frac{|Ax_{i,19} + By_{i,19} + Cz_{i,19} + D|}{\sqrt{A^2 + B^2 + C^2}} \quad (9)$$

脚尖与地面的距离是足迹帧最明显的特征, 也是大多数文献中用于检测足迹的特征。

2) 右脚踝 o_{18} 与地面的距离 F_2 。第 i 帧时右脚踝的空间坐标为 $p_{i,18} = (x_{i,18}, y_{i,18}, z_{i,18})$, 则有

$$F_2 = \frac{|Ax_{i,18} + By_{i,18} + Cz_{i,18} + D|}{\sqrt{A^2 + B^2 + C^2}} \quad (10)$$

在同一帧中, 脚踝与脚尖的空间坐标具有很强的相关性。Kinect 数据的抖动较大, 使用 F_2 与 F_1 联合识别足迹, 可以有效降低噪声的干扰。

3) 右大腿 $b_{16,17}$ 与右小腿 $b_{17,18}$ 的夹角 F_3 。构成右大腿的两个关节为 o_{16} 和 o_{17} , 构成右小腿的两个关节为 o_{17} 和 o_{18} , 定义右大腿与右小腿的夹角为

$$F_3 = \arccos\left(\frac{b_{16,17} \cdot b_{17,18}}{\|b_{16,17}\| \times \|b_{17,18}\|}\right) \quad (11)$$

一般来说,人在运动过程中,若脚离开地面,大腿和小腿的夹角会变小,而脚与地面接触时大腿和小腿的夹角会变大,所以大腿和小腿之间的夹角可以用于区分当前帧是否为足迹帧。

4) 右大腿 $b_{16,17}$ 与盆骨平面 $s_{0,12,16}$ 的夹角 F_4 。定义关节点 o_0, o_{12} 和 o_{16} 构成的平面为盆骨平面, 记为 $s_{0,12,16}$ 。根据向量的叉乘可计算出 $s_{0,12,16}$ 的法向量 $n = b_{0,12} \times b_{0,16}$, 再计算右大腿 $b_{16,17}$ 与 $s_{0,12,16}$ 的法向量 n 之间的夹角

$$F_4 = \arccos\left(\frac{b_{16,17} \cdot n}{\|b_{16,17}\| \times \|n\|}\right) \quad (12)$$

人在运动的时候盆骨平面通常运动幅度比较小,而运动过程中大腿和盆骨平面之间的夹角在不断的变化,所以大腿和盆骨平面之间的夹角可以用于区分当前帧是否为足迹帧。

5) 右小腿 $b_{17,18}$ 与右脚掌 $b_{18,19}$ 的夹角 F_5 。构成右脚掌的关节点为 o_{18} 和 o_{19} , 定义右小腿与右脚掌的夹角为

$$F_5 = \arccos\left(\frac{b_{17,18} \cdot b_{18,19}}{\|b_{17,18}\| \times \|b_{18,19}\|}\right) \quad (13)$$

小腿与脚掌的夹角也会随着脚的运动发生变化,而足迹最明显的反映就是脚掌的运动状态。

6) 右脚尖空间坐标 F_6, F_7, F_8 。在 K 近邻方法中,直接使用了脚尖空间坐标的 x, y, z 分量作为分类特征进行足迹检测,本文也将脚尖的空间坐标作为分类特征。

3.2 时间特征

空间特征反映了运动姿态的静态特点,而时间特征则能反映姿态与姿态之间的联系。运动速度的大小通常可以用于检测人的运动状态。

1) 脚尖的的运动速度 F_9 。脚尖的速度可以通过计算前后帧运动脚尖的空间坐标差与时间的比值得到,根据前面的定义,右脚尖关节在第 i 帧的速度可以表示为

$$F_9 = \|p_{i,19} - p_{i-1,19}\| \times f_{ps} \quad (14)$$

式中 f_{ps} 是指 Kinect 采样的频率。

2) 脚踝的运动速度 F_{10} 。与脚尖速度类似,脚踝的速度通过计算脚踝前后帧的空间坐标差与时间的比值而得

$$F_{10} = \|p_{i,18} - p_{i-1,18}\| \times f_{ps} \quad (15)$$

当人的脚在地上时,通常脚尖和脚踝的速度接近于零,故可以将这两个速度特征选作分类特征。

同样地可以定义左脚的特征。图 4 对上述的 10 个特征进行了描述。

4 Kinect 运动数据足迹检测

支持向量机(SVM)是在统计学理论上发展起来的一种通用的机器学习方法,该方法由 Vapnik 等人^[17]在 1995 年提出。SVM 方法较好地解决了小样本、非线性、高维数据和局部极小等实际问题,且具有结构简单、全局最优、泛化能力好、计算复杂度低等优点^[18],SVM 算法本身作为一种饥渴学习算法(eager learner)^[18]时间性能上较 K 近邻方法有较大的优势。

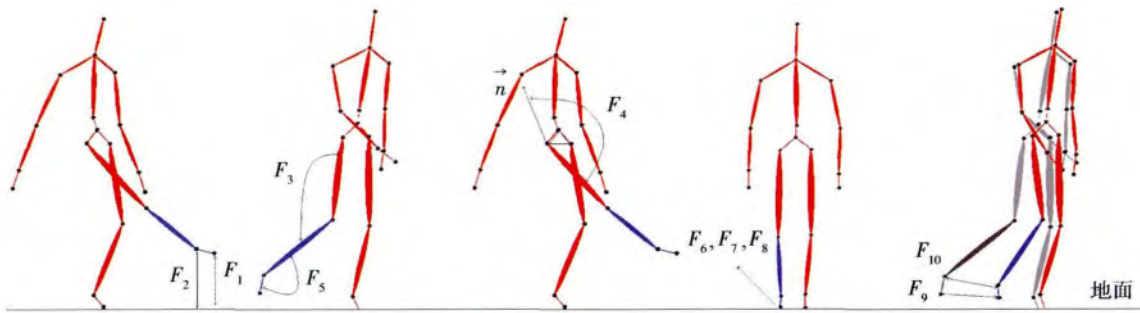


图 4 脚部运动特征 (F_1 至 F_8 为空间特征, F_9, F_{10} 为时间特征)

Fig. 4 Motion features of foot (F_1 to F_8 denotes spatial features F_9, F_{10} denotes temporal features)

足迹检测的目标是将脚分为着地和离地两种状态, 这可以看成是一个二分类问题。二分类的 SVM 分类器的目的是寻找一个分类超平面, 使得训练样本中的两类样本点能够被区分开, 并且使得这个超平面距离两类样本点尽可能地远。对于线性不可分的情形, 通过核函数将低维样本数据映射到高维线性可分的空间中, 在高维空间中对原问题进行求解。本文选择径向基函数序列最小优化 (RBF) 作为核函数

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2}{2\sigma^2}\right) \quad (16)$$

式中, σ 是一个自由参数, $\mathbf{x}_1, \mathbf{x}_2$ 为特征向量。在进行足迹检测时, 实验样本集为 T , 即

$$T = [X, Y] \quad (17)$$

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} F_1^1 & \cdots & F_j^1 & \cdots & F_{10}^1 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ F_1^i & \cdots & F_j^i & \cdots & F_{10}^i \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ F_1^n & \cdots & F_j^n & \cdots & F_{10}^n \end{bmatrix} \quad (18)$$

$$Y = [Y_1 \cdots Y_i \cdots Y_n]^T, \quad (19)$$

$$Y_i \in \{-1, 1\}$$

式中, F_j^i 表示第 i 帧时第 j 个特征, X_i 表示第 i 帧时各个特征的集合, Y_i 表示第 i 帧时人工标记的足迹标签, $Y_i = -1$ 为非足迹帧, $Y_i = 1$ 为足迹帧。SVM 方法就是寻找这两类运动帧之间的一个最优分类超平面 $\mathbf{w}^T \Phi(X) + b = 0$, \mathbf{w} 是一个垂直于超平面的向量, Φ 将样本集映射到一个线性可分的高维空间之中。SVM 方法中的参数 C 通常很难选择^[19], 使用的 ν -SVM 方法是 Schölkopf 等人^[20]提出的一种新的 SVM 方法, 该方法引入了一个新的更好确定的参数 ν , 将寻找最优分类平面的问题等价于求解一个凸二次优化问题

$$\begin{aligned} \min & \frac{1}{2} \mathbf{w}^T \mathbf{w} - \nu \rho + \frac{1}{l} \sum_{i=1}^l \xi_i \\ \text{s. t.} & y_i (\mathbf{w}^T \Phi(X_i) + b) \geq \rho - \xi_i \\ & \xi_i \geq 0 \quad i = 1, \dots, l \quad \rho \geq 0 \end{aligned} \quad (20)$$

式中, $\nu \in [0, 1]$, ξ_i 是松弛变量, 使用拉格朗日乘子和对偶原理可以将原问题化成对偶问题

$$\begin{aligned} \min & \frac{1}{2} \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} \\ \text{s. t.} & Y^T \boldsymbol{\alpha} = 0 \quad \boldsymbol{\rho}^T \boldsymbol{\alpha} \geq \nu \\ & 0 \leq \alpha_i \leq 1/l, \quad i = 1, \dots, l \end{aligned} \quad (21)$$

再使用序列最小优化 (SMO) 算法对式 (21) 进行求解, 可以得到最优分类平面

$$h(\mathbf{x}) = \sum_{i=1}^l \alpha_i Y_i K(\mathbf{x}_i, \mathbf{x}_2) + b = 0 \quad (22)$$

由此可以构造足迹检测的决策函数

$$f(\mathbf{x}) = \begin{cases} 1 & h(\mathbf{x}) > 0 \\ -1 & h(\mathbf{x}) < 0 \end{cases} \quad (23)$$

即当 $f(\mathbf{x}) = 1$ 时认为是足迹帧, 当 $f(\mathbf{x}) = -1$ 认为是非足迹帧。

5 实验

实验使用的计算机硬件环境为 Intel(R) Xeon (R) CPU W3520@2.67 GHz、4 GB 内存, 软件环境为 Visual Studio 2012, 实验中用到了 dlib^[21] 机器学习库。

实验数据是由 Kinect 采集得到的一系列运动数据。在采集数据的时候, 考虑到 Kinect 本身采集范围小的特点, 采集了走路、踏步、跳跃等 8 种常见类型的运动共 8 000 帧数据。将实验数据分为训练集 (4 800 帧) 和测试集 (3 200 帧)。对运动数据人工标定足迹, 定义一个两位的足迹标签, $L_i = [L_i, R_i]^T$, $L_i, R_i \in \{-1, 1\}$, L_i, R_i 分别为左脚和右脚的足迹标签, 1 表示此帧为足迹帧, -1 表示此帧为非足迹帧。

实验结果的评价标准是算法标定足迹相对于人工标定的正确率, 这个正确率通过计算算法得到的分类标签与人工标定的标签相同的数量与总标签数的比值得到, 这个标准在进行足迹检测时被广泛应用, 能够有效衡量算法的优劣。使用本文算法进行足迹检测时, 时间主要用在训练决策函数之上, 训练之后检测每一帧运动数据平均耗时 0.39 ms。

5.1 双边滤波算法有效性的验证

为验证上文中提出的自适应双边滤波去噪算法, 分别对进行双边滤波和不进行双边滤波的数据进行足迹检测并对检测正确率进行比较, 结果如表 2 所示, 可以看出使用本文的滤波算法之后, 能够有效地处理数据中的噪声, 提高了足迹检测的正确率, 正确率平均提高了 6% 左右。

5.2 本文算法与基准线法以及 K 近邻方法的比较

以 Kinect 采集的运动数据作为数据源, 本实验分别比较了基准线算法、K 近邻算法与本文算法的

足迹检测效果。基准线方法又分成了高度基准线、速度基准线以及双基准线法。实验结果在表3和表4中给出。

表2 滤波前后使用 SVM 算法进行足迹检测的正确率比较
Table 2 Comparison of accuracy of foot plant detection with ABF and without

运动类型	滤波前正确率%	滤波后正确率%
正常走路	92.60	97.05
踏步	93.55	98.65
双脚跳跃	96.05	98.25
单脚跳跃	90.40	95.95
踢足球	87.95	97.10
单脚站立	91.10	98.20
醉酒走路	91.95	95.20
踢毽子	89.10	97.20

表3比较了本文算法与基准线方法在足迹检测时的检测正确率。可以看出,基准线方法在足迹检测时,正确率与运动的类型有关,对于踏步和单脚站立检测效果较好,但对于不规则的运动,如踢足球和醉酒走路,算法的正确率出现了下降。实验结果表明,使用本文算法可以很好地对类似现象进行正确的处理,整体的足迹检测正确率比基准线方法提高了10%左右,实验结果稳定可靠,体现出良好的鲁棒性。

表3 本文算法与基准线算法的比较

Table 3 Comparison of accuracy of foot plant detection with baseline method and our method

运动类型	总帧数	正确率/%			
		高度基准线法	速度基准线法	双基准线法	本文算法
正常走路	1 000	84.80	84.50	82.05	97.05
踏步	1 000	98.40	73.80	93.60	98.65
双脚跳跃	1 000	86.45	86.90	79.25	98.25
单脚跳跃	1 000	90.40	85.30	93.00	95.95
踢足球	1 000	90.35	77.75	79.75	97.10
单脚站立	1 000	95.00	62.20	95.10	98.20
醉酒走路	1 000	88.65	87.40	86.90	95.20
踢毽子	1 000	94.35	88.25	90.80	97.20

表4 本文算法与 K 近邻算法的比较

Table 4 Comparison of accuracy of foot plant detection with KNN method and our method

运动类型	总帧数	正确率/%		耗时/(ms/帧)	
		KNN	本文算法	KNN	本文算法
正常走路	1 000	87.55	97.05	2.28	0.30
踏步	1 000	98.05	98.65	2.33	0.42
双脚跳跃	1 000	95.35	98.25	2.32	0.17
单脚跳跃	1 000	80.75	95.95	2.33	0.39
踢足球	1 000	91.75	97.10	2.34	0.62
单脚站立	1 000	98.15	98.20	2.29	0.40
醉酒走路	1 000	83.75	95.20	2.31	0.41
踢毽子	1 000	86.70	97.20	2.36	0.42

表4比较了本文方法与 K 近邻方法在足迹检测时的检测正确率以及标记每帧数据需要的时间。可以看出,对于有些运动(如踏步和单脚站立) K 近邻方法标记足迹的正确率和本文算法相近。但是由于 K 近邻作为一种懒惰的学习算法,每次检测都需要扫描整个训练集数据库,标定每一帧运动的时间会随着训练集的变大而变长,本文采用的 SVM 算法通过训练可以得到一个决策函数,训练之后对于新来的运动数据,只需要根据决策函数进行判断,训练集的大小只会影响训练时间,检测时间不会受到影响。从表中可以看出,本文算法检测一帧运动足迹的速度是 K 近邻方法的 7 倍左右,在时间性能上本文算法有较大的优势。

图5是使用本文算法进行足迹检测时的实验结果图。图5(a)(c)分别是踢球和双脚跳跃的运动序列图,图5(b)(d)是对应的足迹检测的标定结果图。

6 结论

微软 Kinect 可以实时获取到人运动时主要关节的空间坐标信息,而且价格低廉,让普通用户自己采集运动数据并进行编辑制作动画成为可能。足迹检测是运动编辑算法的首要步骤,但是 Kinect 运动数据的噪声使得使用现有运动数据处理算法进行足迹检测变得非常困难。

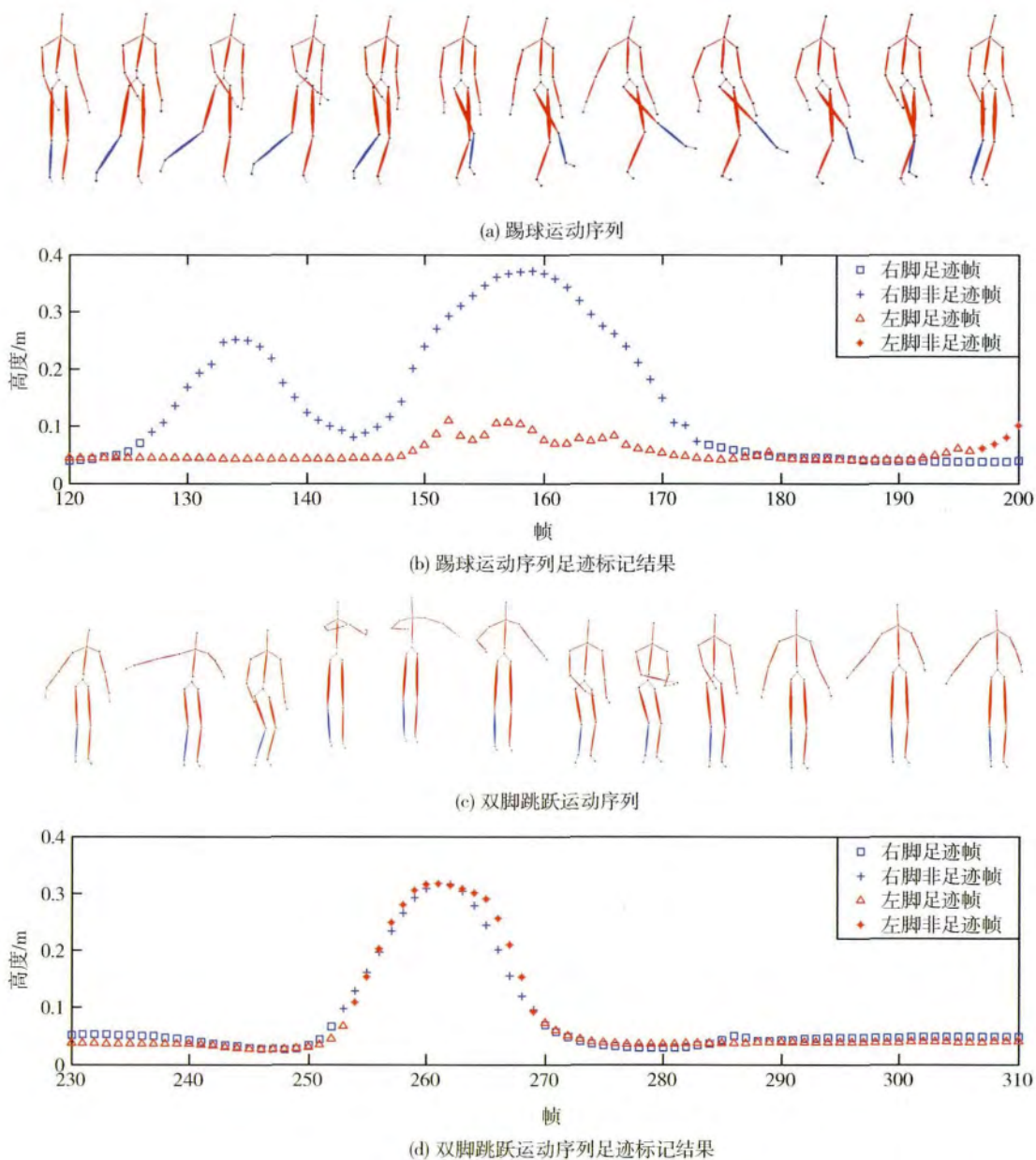


图 5 足迹检测实验结果图

Fig. 5 Results of foot plant detection with our method((a) (c) is screen shot of playing football and two-foot jump , respectively; (b) (d) is the accompany results of foot pant labeling. Horizontal axis denotes number of frame , vertical axis denotes the distance between foot and the ground)

本文提出一种面向 Kinect 运动数据的足迹检测算法: 首先提出了一种自适应双边滤波算法, 可以有效减少 Kinect 运动数据中的噪声, 再进一步选取多种脚部运动特征提高检测足迹检测的正确率, 最后使用 SVM 算法训练决策函数并用于足迹检测。上述特点使得本文算法能够解决 Kinect 运动数据的噪声问题, 提高了算法的鲁棒性, 并且使得算法具

有良好的时间性能。

实验结果表明, 本文算法具有良好的足迹检测正确率以及时间性能, 对于踢足球、醉酒走路等复杂不规则运动仍然具有良好的鲁棒性。由于本文算法作为有监督的学习方法, 下一步将尝试无监督的学习方法进行足迹检测, 减少手工标定的工作。

参考文献(References)

- [1] Kovar L, Schreiner J, Gleicher M. Footskate cleanup for motion capture editing [C] //The 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. New York: ACM, 2002: 97-104. [DOI: 10.1145/545261.545277]
- [2] Yamane K, Ariki Y, Hodgins J. Animating non-humanoid characters with human motion data [C] //The 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. Switzerland: Eurographics Association, 2010: 169-178. [DOI: 10.2312/SCA/SCA10/169-178]
- [3] Guo S, Southern R, Chang J, et al. Adaptive motion synthesis for virtual characters: a survey [J]. The Visual Computer, 2014, 30(12): 1-16. [DOI: 10.1007/s00371-014-0943-4]
- [4] Firouzmanesh A, Cheng I, Basu A. Perceptually guided fast compression of 3-d motion capture data [J]. IEEE Transactions on Multimedia, 2011, 13(4): 829-834. [DOI: 10.1109/TMM.2011.2129497]
- [5] Ménardais S, Kulpa R, Mutton F, et al. Synchronization for dynamic blending of motions [C] //The 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. Switzerland: Eurographics Association, 2004: 325-335. [DOI: 10.1145/1028523.1028567]
- [6] Sung M. Automatic Fixing of Foot Skating of Human Motions from Depth Sensor [M]. Multimedia and Ubiquitous Engineering. Berlin: Springer Netherlands, 2013: 405-412. [DOI: 10.1007/978-94-007-6738-6_50]
- [7] Ikemoto L, Arikian O, Forsyth D. Knowing when to put your foot down [C] //Symposium on Interactive 3D Graphics and Games. New York: ACM, 2006: 49-53. [DOI: 10.1145/11778765.1778809]
- [8] Wu X, Kumar V, Quinlan J R, et al. Top 10 algorithms in data mining [J]. Knowledge and Information Systems, 2008, 14(1): 1-37. [DOI: 10.1007/s10115-007-0114-2]
- [9] Nguyen C V, Izadi S, Lovell D. Modeling kinect sensor noise for improved 3d reconstruction and tracking [C] //Proceedings of the 2nd International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission. Zurich: IEEE, 2012: 524-530. [DOI: 10.1109/3DIMPVT.2012.84]
- [10] Tomasi C, Manduchi R. Bilateral filtering for gray and color images [C] //Proceedings of the 8th International Conference on Bombay: IEEE, 1998: 839-846. [DOI: 10.1109/ICCV.1998.710815]
- [11] Zhang Z Q, Wang W Y. A modified bilateral filtering algorithm [J]. Journal of Image and Graphics, 2009, 14(3): 443-447. [张志强, 王王玉. 一种改进的双边滤波算法 [J]. 中国图象图形学报, 2009, 14(3): 443-447.] [DOI: 10.11834/jig.200903/10]
- [12] Bosco A, Bruna A, Messina G, et al. Fast method for noise level estimation and integrated noise reduction [J]. IEEE Transactions on Consumer Electronics, 2005, 51(3): 1028-1033. [DOI: 10.1109/TCE.2005.1510518]
- [13] Saha P K, Udapa J K, Odhner D. Scale-based fuzzy connected image segmentation: theory, algorithms, and validation [J]. Computer Vision and Image Understanding, 2000, 77(2): 145-174. [DOI: 10.1006/cviu.1999.0813]
- [14] Chen K. Adaptive smoothing via contextual and local discontinuities [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1552-1567. [DOI: 10.1109/TPAMI.2005.190]
- [15] Guyon I, Elisseeff A. An introduction to variable and feature selection [J]. The Journal of Machine Learning Research, 2003, 2003(3): 1157-1182.
- [16] Chen S L, Sun Z X, Zhang Y, et al. An example-based method for Chinese dance motion retrieval [J]. Journal of Computer-Aided Design & Computer Graphics, 2014, 26(2): 198-210. [陈松乐, 孙正兴, 张岩, 等. 民族舞蹈运动数据的实例检索方法 [J]. 计算机辅助设计与图形学学报, 2014, 26(2): 198-210.]
- [17] Cortes C, Vapnik V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297. [DOI: 10.1007/BF00994018]
- [18] Chen Q, Tian J, Huang H N, et al. Study on SAS image segmentation using SVM based on statistical and texture features [J]. Chinese Journal of Scientific Instrument, 2013, 34(6): 1413-1420. [陈强, 田杰, 黄海宁, 等. 基于统计和纹理特征的 SAS 图像 SVM 分割研究 [J]. 仪器仪表学报, 2013, 34(6): 1413-1420.]
- [19] Chang C C, Lin C J. Training v-support vector classifiers: theory and algorithms [J]. Neural Computation, 2001, 13(9): 2119-2147. [DOI: 10.1162/089976601750399335]
- [20] Schölkopf B, Smola A J, Williamson R C, et al. New support vector algorithms [J]. Neural Computation, 2000, 12(5): 1207-1245. [DOI: 10.1162/089976600300015565]
- [21] King D E. Dlib-ml: A machine learning toolkit [J]. The Journal of Machine Learning Research, 2009, 2009(10): 1755-1758.