

基于 Kinect 手势识别的应用与研究

于泽升¹ 崔文华¹ 史添玮²

(辽宁科技大学电子与信息工程学院 鞍山 114051)¹ (辽宁科技大学国际金融与银行学院 鞍山 114051)²

摘 要 为解决当前智能家居系统操作繁琐的问题,同时为获得更简单的控制方式,并增加用户的体验感受,研究了基于 Kinect 骨骼信息的手势识别技术,并将其融入至智能家居的人机交互系统中。在该系统中,用户可以自定义手势动作或语音实现家居设备的智能控制。使用了一种基于加权动态时间规整的模板匹配手势识别算法。通过 Kinect 的深度摄像头获取手势深度图像和骨骼图像数据,并采用加权动态时间规整算法进行识别。实验表明使用该算法实现手势识别是可行且有效的,且其最佳识别位置是在 Kinect 的正前方 2~2.5m 处,识别准确率达到 96% 左右。

关键词 Kinect, 手势识别, 加权动态时间规整算法, 智能家居

中图分类号 TP399 文献标识码 A

Application and Research on Gesture Recognition by Kinect Sensors

YU Ze-sheng¹ CUI Wen-hua¹ SHI Tian-wei²

(Department of Electronics and Information Engineering, University of Science and Technology, Anshan 114051, China)¹

(College of International Finance and Bank, University of Science and Technology, Anshan 114051, China)²

Abstract In order to solve the problems of complex operations in the current smart home system and acquire a simple control method and increase feeling experience for users, the gesture recognition methods based on Kinect were re-researched and integrated into the human-computer interaction (HCI) system in smart home. In this system, users can customize the gestures to realize the intelligent control of household equipment. The utilized template matching gesture recognition method is accomplished based on the dynamic time warping (DTW) algorithm. It employs the Kinect depth camera to obtain the skeleton depth images and gestures. The results of actual gesture recognition experiments show that the gesture recognition based on DTW algorithm is feasible and effective. The best identification distance is in range of 2-2.5 m in front of Kinect and the highest recognition accuracy can reach 96%.

Keywords Kinect, Gesture recognition, Weighted DTW algorithm, Smart home

1 介绍

智能家居是人们的一种居住环境,它是利用先进的计算机技术、网络通讯技术等融合个性需求,将与家居生活有关的各个子系统通过网络化综合智能控制和管理,实现家庭生活更加便利和舒适。随着人机交互技术的发展,人与机器设备之间的交互体验变得非常简单、自然。现今,人们完全可以脱离开关等有线控制方式,通过语音识别、手势识别与人脸识别等技术来控制家庭里的设备^[1],简化了用户的操作过程,增强了用户的体验感受。随着社会的发展,人们对舒适、自然的居住环境的需求更加强烈,因此设计人机交互环境友好的智能家居系统成为一种势在必行的趋势。当前,采用较广泛的技术是在智能家居设备中嵌入无线模块,通过无线感知、远程控制等为家庭用户提供操作功能,而这仍需要家庭用户通过各种终端进行操作。为解决当前智能家居系统存在的操作繁琐问题,传统手势识别技术易受光照、背景等外部因素影响,手

势识别准确率低的缺点^[2],本文研究了一种基于 Kinect 传感器的智能家居手势控制系统,通过 Kinect 采集图像深度数据和骨骼数据,采用 DTW 算法识别手势动作,完成家居设备的控制功能。本文第 2 节介绍了骨骼数据采集的过程;第 3 节介绍了手势识别的技术方法;第 4 节和第 5 节分别介绍了实验结果和结论。

2 骨骼数据的获取

基于骨骼信息的手势识别数据是通过 Kinect 获取的, Kinect 是微软发布的一款体感外设设备,它拥有 3 个摄像头,可同时获取 RGB 彩色图像、深度图像和人体骨骼图像,是研究基于计算机视觉的手势识别技术的理想设备^[3,4]。其外观如图 1 所示,中间是用来捕捉图像的 RGB 摄像头,其采集速度可达 30f/s(帧/秒);两边分别是用于深度图像获取的红外线发射器和红外线接收器,其由红外线投影机加单色 CMOS 传感器组成,使得设备能识别 3D 空间;底部为多点阵列麦克

本文受辽宁省教育厅基金项目(2016HZZD05)资助。

于泽升 男,硕士生,主要研究方向为人工智能, E-mail: 525358121@qq.com; 崔文华(1968—),女,博士,教授,主要研究方向为人工智能物联网; 史添玮(1968—),男,博士,讲师,主要研究方向为机械电子。

风:主要永用于声音输入及定位,并可以帮助过滤环境噪声。

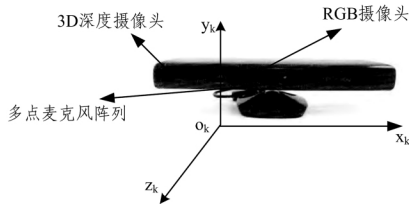


图1 观测坐标系与对象坐标系之间的关系

2.1 骨骼追踪技术

骨骼跟踪技术是在实现深度信息提取的前提下,通过处理深度数据来计算每个关节点的位置信息^[5-7],并用坐标 (x, y, z) 表示。不同的开发平台能定位的关节点个数自然也不同,本研究使用的 Kinect SDK 是微软发布的 v1.8 版本,能够捕获 20 个关节点的三维坐标,其中任何一个关节点都可以用 Tracked, NotTracked 和 Inferred 3 种状态描述,这 20 个关节点组成一幅人体骨架图,如图 2 所示。

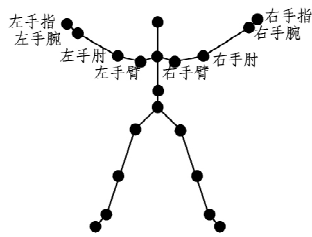


图2 人体骨骼点分布图

2.2 手势特征提取

手势特征的提取是手势识别技术研究中极其重要的一步^[8],能否提取有效的手势特征将直接决定识别算法的准确性和有效性。Kinect 骨骼坐标的原点被定位于摄像头的所在位置,摄像头向前是 Z 轴正向,向上和向右方向分别为 Y 轴正向和 X 轴正向。因为使用者的体态以及所处位置的不同,使关节点的坐标也产生不小的变化,所以不能直接使用节点坐标作为手势识别的依据。对此,本文通过关节点之间的相对位置来表示手势特征,成功地消除了这种影响。选取 20 个骨骼点中的 8 个骨骼点(见图 2),提取动态特征,即左、右手节点,左、右手腕节点,左、右手肘节点和左、右手臂节点。假设在某一帧骨骼图中,一个关节点的三维坐标为 (x_0, y_0, z_0) ,节点 n 的三维坐标为 (x_n, y_n, z_n) ,二者之间的距离为:

$$D_n = \sqrt{(x_n - x_0)^2 + (y_n - y_0)^2 + (z_n - z_0)^2} \quad (1)$$

其中, $n(n=1,2,3)$ 表示手势特征个数。

3 手势识别的技术方法

手势一般可分为静态手势和动态手势,静态手势指在手不动的情形下,手指与手掌做出不同的动作和形状;动态手势则指随着时间的变化,手的位置和形状都发生改变的手势,而动态手势能更丰富地表达操作者意图^[9,10]。手势识别与语音识别的算法基本相同,主要有 3 种:动态规划的方法、矢量量化(Vector Quantization)的方法和隐马尔柯夫模型(HMM)的方法。表 1 对各种识别方法进行了比较。

表1 识别技术方法的比较

方法	优点	缺点
矢量量化	容易理解	识别率低
隐马尔柯夫	识别率高	计算过程复杂
动态时间规整	简单、高效	易受环境因素影响

综上所述,动态手势识别主要使用 HMM 和 DTW 算法,针对简单孤立的识别,两种方法的效果差别不大,但由于 HMM 算法计算极其复杂,结合应用需求,本文重点研究基于 DTW 的动态手势识别。该方法是一种基于模板匹配的方法,需要建立手势模板库,然后利于 DTW 算法对控制手势与模板库进行匹配,识别出手势。基于 DTW 的识别流程如图 3 所示。

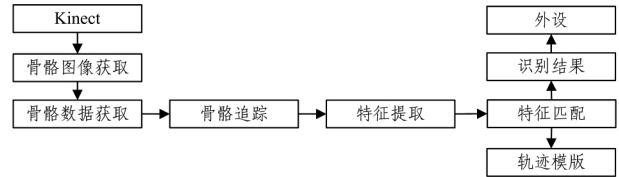


图3 基于 DTW 算法的手势识别流程图

3.1 DTW 算法

DTW 是一种模板匹配算法,即找到一个测试模式与参考模式之间的最佳匹配,其中参考模式被表示为一个特征时间序列。图 4 展示了两个序列的匹配过程。

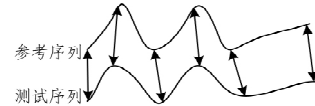


图4 测试序列与参考序列的匹配过程

设参考模式与测试模式的时间序列分别为 $R = \{R_1, R_2, \dots, R_N\}$ 和 $T = \{T_1, T_2, \dots, T_N\}$,使用非线性映射对齐时间轴上的两个序列,这样的规整路径可以被表述为如下给出的一个有序的点集: $P = (P_1, P_2, \dots, P_L)$, $P_l = (n_l, m_l)$,其中 $P_l = (n_l, m_l)$ 表示时间轴 r_{n_l} 到时间轴 t_{m_l} 的映射。 $P_l \in [1:N] \times [1:M]$, $l \in [1:L]$,其中 L 代表映射的次数。设参考序列 R 与测试序列 T 之间相对于距离函数 $d(r_i, t_j)$ 的一条规整路径 P 的总代价为 D ,即所有映射序列元素之间的距离和,其中 $i \in [1:N]$, $j \in [1:M]$ 。

$$D_p = \sum_{l=1}^L d(r_{n_l}, t_{m_l}) \quad (2)$$

其中, D_p 是路径 p 的总代价, $d(r_i, t_j)$ 代表元素 r_i 和 t_j 之间的距离。对于手势识别,距离则被选做参考手势 R 与测试手势 T 的相应关节点位置之间的距离。映射可以看作是在二维网格中的一条路径,即代价矩阵,其大小为 $N \times M$ (见图 5),其中网格节点 (r_i, t_j) 表示 r_i 和 t_j 之间的距离。节点 (r_1, t_1) 通常放置在网格的左下角,其通过匹配第一个序列元素开始对齐过程。二维网格中的每个路径 p 均与使用式(2)计算得到的总代价 D 相关。值得注意的是,在所有可能的路径中,最想得到的是累计总代价最小而期望满足约束的路径,因此,最优路径 P^* 是指具有最小总成本的路径,使用最优路径来表示两个序列之间的 DTW 距离为:

$$DTW(R, T) = D_{P^*}(R, T) \quad (3)$$

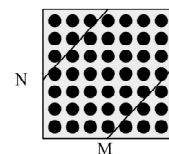


图5 序列 R 和 T 的代价矩阵

在规整路径上,已经提出一些著名的以约束消除序列之间不切实际的对应关系。最基本的约束不仅应用于手势识别还应用于其他领域,其约束如下^[11]:

(i) 边界条件: $p_1 = (1, 1)$, $p_L = (N, M)$ 。

(ii) 步长条件: $p_{l+1} - p_l \in \{(0, 1), (1, 0), (1, 1)\}$ for $l \in [1; L-1]$ 。

边界条件需要整个参考序列映射到整个测试序列, 并且若非严格需要, 其是可以修改的。步长条件要求只有两个序列中的一个元素可以在 Bellman 准则^[12]的每次代价计算过程中被跳过, 因此最优路径可以从一组受约束的前个节点集合中得到改进(见图 6)。

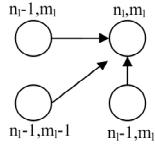


图 6 动态规划中的节点分布

由于所有元素在时间轴上是有序的, 因此前一个节点在当前节点的左下方, 首先定义 $C(n_l, m_l)$ 为:

$$C(n_l, m_l) = DTW(R(1:n_l), T(1:m_l)) \quad (4)$$

其中, $C(N, M)$ 等同于 $DTW(R, T)$ 。假设已经计算得到之前 3 个节点的最优路径总代价 $((n_l-1, m_l), (n_l, m_l-1))$ 和 (n_l-1, m_l-1) 。由于路径 (n_l-1, m_l-1) 的第 $l-1$ 个位置被限制为 2D 网格中 3 个节点之中的一个, 因此根据 Bellman 准则得:

$$C(n_l, m_l) = \min\{C(n_l, m_l-1), C(n_l-1, m_l), C(n_l-1, m_l-1)\} + d(r_{nl}, t_{ml}) \quad (5)$$

最后, 最小代价路径对齐了两个时间序列, 且具有代价 $C(N, M) = DTW(R, T)$ 。此外, 测试序列从参考序列中匹配到具有最小代价函数的参考序列。尽管式(5)得到了两个序列之间的最小代价函数, 但其没有输出最优路径。为了找到用于测试序列元素与参考序列元素之间映射的最优路径, 可以由路径的最后一个节点作为起始点回溯最优路径。值得注意的是, 如果满足边界条件(整个测试序列完全映射到参考序列), 则 $(n_L, m_L) = (N, M)$, $(n_1, m_1) = (1, 1)$ 。

3.2 加权动态时间规整算法

传统的 DTW 算法通过基于样本间距离来对齐序列的方法计算两个时间序列之间的差异, 如式(5)所示。若序列样本是多维的, 则使用欧式距离来描述所有维度之间的同等重要性。基于身体关节点与某个特定的手势分类之间是如何相关的, 本文提出了在代价计算中使用加权距离。所谓关节点与手势的相关性是指它对该手势动作的完成所做贡献的大小。本文衡量关节点对形成指定手势所做出的贡献是通过算出关节点总的偏移量 C_j^g 来实现的。

$$C_j^g = \sum_{n=2}^N Dist^j(f_{n-1}^g, f_n^g) \quad (6)$$

其中, g 为手势指数, j 为关节点指数, n 为骨架标号。 $Dist^j()$ 表示第 j 个关节点的偏移, 它通过两个连续的特征矢量 f_{n-1}^g 和 f_n^g 计算得出, 而通过这些连续的偏移可以计算出所选参考序列关节点的总偏移。当计算出总的偏移量后, 如式(7)所示, 本文需要限定偏移的范围以及去噪处理, 避免输出过高或过低的参量权重值。

$$C_j^g = \begin{cases} C_a, & 0 \leq C_j^g < T_1 \\ \frac{C_j^g - T_1}{T_2 - T_1}(C_b - C_a) + C_a, & T_1 \leq C_j^g < T_2 \\ C_b, & \text{其他} \end{cases} \quad (7)$$

其中, C_a 和 C_b 为设置的阈值, T_1 和 T_2 是决定阈值的边界

值。根据得出关节点的总偏移可以评估关节点对其手势的贡献, 关节点的权重可由式(8)获得:

$$W_j^g = \frac{1 - e^{-\beta C_j^g}}{\sum_k (1 - e^{-\beta C_k^g})} \quad (8)$$

其中, β 是一个参数变量。式(8)中一个关节点的加权值可以依赖于手势分类而变化。使用这些权重计算 DTW 代价函数, 距离函数 $d(r_n, t_m)$ 成为相邻两个关节点距离的加权平均值。

$$d(r_n, t_m) = \sum_j Dist^j(r_n, t_m) w_j^g \quad (9)$$

同一种手势类中不同样本之间的代价函数大小称为类内差距。类间差距则表示不同手势样本之间的代价函数大小。对于关节点权重的参数模型的获得与一个参数变量 β 有关, 本文的目的是选择一个满足类内差距 D_w 最小而类间差距 D_B 最大的 β 值。通过使相关节点的贡献小于代价函数的贡献来实现类内差异最小化, 不减少或增加关节点的权重可以帮助区分不同的手势。假设 $D_{g,h}(\beta)$ 表示所有手势类 g 的样本及手势类 h 的样本间的平均加权 DTW 代价, 则类间差距 D_B 代表所有 $D_{g,h}(\beta)$ 的平均值:

$$D_B(\beta) = \sum_g \sum_{h \neq g} D_{g,h}(\beta) \quad (10)$$

这可计算出手势类之间的平均距离的总和, 给出一个 β 值就能够得出一个手势和其余手势之间的平均距离。对于所有手势类, 类内差异是类内距离 $D_{g,g}(\beta)$ 的总和。

$$D_w(\beta) = \sum_g D_{g,g}(\beta) \quad (11)$$

通过式(11)可以计算出所有手势类 g 的样本之间的平均距离 $D_{g,g}(\beta)$ 。 $R(\beta)$ 是 β 的判别率, 可以通过式(12)得出。

$$R(\beta) = \frac{D_B(\beta)}{D_w(\beta)} \quad (12)$$

选择一个能使 R 达到最大值的 β , 通过式(13)即可得出最优的 β^* :

$$\beta^* = \arg \max_{\beta} R(\beta) \quad (13)$$

DTW 的手势识别是一种基于模板匹配的方法, 需要建立手势模板库。然后利于 DTW 算法对控制手势与模板库匹配, 识别出手势。基于 DTW 的识别流程如图 3 所示。

4 系统实现与实验结果

4.1 基于 Kinect 的手势识别系统

本文提出的手势识别系统的界面共分为 5 部分: 主显示界面、用户注册/识别界面、手势录入/识别界面、TCP/UDP 数据通讯界面, 如图 7 所示。



图 7 系统主界面

使用该系统进行手势识别的过程如图 8 所示。首先, 使用者应该站在传感器前方 1.5~2.5m 处; 其次, 点击输入动作按钮开始 5s 的倒计时, 之后动作录入提示被展示在主界面上, 接着人体骨骼框架变为红色。当使用者摆出动作的起始姿态时, 展示界面的背景变为绿色, 当识别这个动作之后, 动作识别结果栏显示数字“1”。

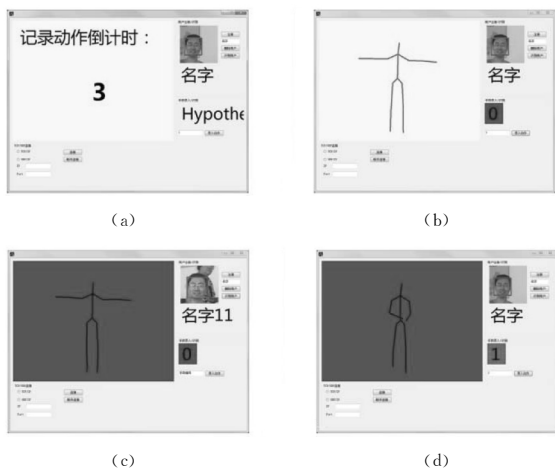


图 8 程序实现的结果

4.2 实验结果

为验证使用方法与提出系统的可行性与有效性,并找到完成手势识别的最佳有效距离,本文做了一个系统手势识别成功率验证实验。选取 5 名被试者,其中 3 男 2 女,平均年龄为 31 ± 8 岁,平均身高为 170 ± 5 cm。每个被试者使用系统完成指定的动作,被试者分别在设备的有效识别距离内完成实验。本文指定每名被试者在同一段距离内做出 5 种不同的手势动作,如图 9 所示。分别使用该系统识别 10 次,计算每种手势 10 次识别的成功率、5 种手势的识别成功率、5 组成功率的均值与方差,5 名被试者在不同距离段的所有数据结果如表 2 和图 10 所示,图 11 为被试者完成实验的示意图。



图 9 手势动作

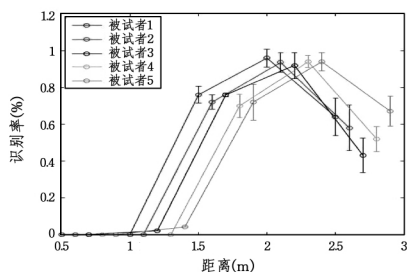


图 10 识别率

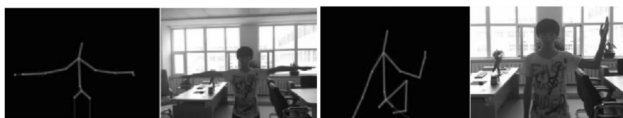


图 11 识别结果

表 2 识别率(%)

被试者	距离(m)				
	0—1	1—1.5	1.5—2	2—2.5	2.5—3
1	0	0	76±4.6	96±4.9	64±10.2
2	0	0	72±4	93.6±5.3	58±12.5
3	0	2	76±10.2	92±7.4	43±9.4
4	0	0	70±6.3	94±3.27	52±6.8
5	0	4	72±9.8	94±4.9	67±8.7

结果显示,使用者在使用系统时需要站在 Kinect 传感器前 1.5~2.5 m 处,其中 2~2.5 m 处效果最佳,超过 2.5 m 会出现用户识别不准确的情况,低于 1.5 m 则无法识别手势,且可能会出现内存不足的现象。如果用户想要识别坐姿的动作,应注意坐姿状态下双手很难达到腹部以下位置,传感器会无法识别。在坐姿录入动作时应尽量采用双手不低于腹部的动作。

结束语 本文采用加权动态时间规整的模板匹配算法实现了一个基于 Kinect 的手势识别系统,实验证明使用 Kinect 手势控制家居设备是可行且有效的;识别距离在 Kinect 正前方 2~2.5 m 处效果最佳。

参 考 文 献

- [1] Yang B L. The new trend of intelligent home appliances human-computer interaction [J]. Application of Integrated Circuits, 2013,18(3):32-35
- [2] Chen Y L, Jin J D. Human-computer interaction of hand posture and wave speed recognition[J]. Journal of Electronic Measurement and Instrument, 2014,24(1):56-61
- [3] 方华,刘诗雄,田敬北. 基于 Kinect 骨骼系统的手势识别研究[J]. 计算机光盘软件与应用, 2014,46(2):65-68
- [4] 樊景超,周国民. 基于 Kinect 骨骼跟踪技术的手势识别研究[J]. 安徽农业科学, 2014(11):3444-3446
- [5] SkeletonTracking[OL]. <http://msdn.microsoft.com/en-us/library/hh855347.aspx>
- [6] 刘阳,尚赵伟. 基于 Kinect 骨架信息的交通警察手势识别[J]. 计算机工程与应用, 2015,51(3):157-161
- [7] 赵飞飞,吕玉祥. 基于 Kinect 骨骼信息的手势识别[J]. 中国科技论文, 2015,37(20):2412-2415
- [8] Tang M. Recognition hand gestures with Microsoft's Kinect [J]. Stanford University, 2011,14(3):42-51
- [9] Oka K, sato Y, Koike H. Real-time fingertip tracking and gesture recognition [J]. IEEE Computer Graphics and Applications, 2002,22(6):64-71
- [10] Su C J, Chiang C Y, Huang J Y. Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic [J]. Applied Soft Computing, 2014,25(22):652-666
- [11] 邹洪. 实时动态手势识别关键技术研究[D]. 广州:华南理工大学
- [12] Bellman R. The theory of dynamic programming[J]. Bull. Amer. Math. Soc, 1954,60(6):51-503