DOI:10.19479/j.2095-719x.2402133

基于 Kinect 动态手势识别的改进 DTW 算法研究与实现

洪 姣,张雪娇,刘国瑞,刘 琦
(天津城建大学 计算机与信息工程学院,天津 300384)

摘 要: 为了解决传统算法在动态手势识别中存在的准确率不高、难以适应复杂背景等问题,本文提出了一种基于 Kinect 传感器的改进 DTW 算法动态手势识别方法.使用 Kinect Studio 和 Visual Gesture Builder 训练建立手势模板库,对传统的 DTW 算法进行限制搜索范围和设置失真度阈值两点改进,利用改进后的 DTW 算法对提取出来的人物指定关节信息和模板信息进行匹配,从而识别出手势.实验结果表明,改进 DTW 算法具有较高的动态手势识别准确率,并且能够更好地适应复杂环境.

关键词: 动态; 手势识别; Kinect; 改进 DTW 算法
中图分类号: TP391.4 **文献标志码:** A **文章编号:** 2095-719X(2024)02-0133-05

Improvement and Implementation of DTW Algorithm on Dynamic Gesture Recognition based on Kinect

HONG Jiao, ZHANG Xuejiao, LIU Guorui, LIU Qi
(School of Computer and Information Engineering, TCU, Tianjin 300384, China)

Abstract: In order to solve the problems of low accuracy and difficult adaptation to complex background in traditional dynamic gesture recognition, this paper proposes an improved DTW algorithm based on Kinect sensor. It creates the Gesture template library by using Kinect Studio and Visual Gesture Builder, and improves the traditional DTW algorithm by limiting the search range and setting the distortion threshold. In order to recognize hand gestures, the improved DTW algorithm is used to match the joint angle information and the template information. The experimental results show that the improved DTW algorithm has high accuracy of dynamic gesture recognition and can better adapt to the complex environment.

Key words: dynamic; gesture recognition; Kinect; improved DTW algorithm

随着科技的进步,人工智能技术的发展也日趋成熟,计算机的各种算法能够模仿人类大脑,经过学习与训练,变成智能化的工具,解放人类劳动力,促进了社会发展, 也为人们的生活出行等带来了巨大便捷.人机交互渐渐融入了人们的日常生活,成为人们生活不可或缺的一部分.由于传统的利用硬件设备进行人机交互缺乏使用的自然性与用户友好性,并且在特定场景下具有很大局限性,因此,建立一个和谐自然、符合人类交流习惯的人机交互方式成为当下关注的重点[1-2].

目前,一些智能、简单的人机交互方式正在慢慢兴起.例如语音助手、手势识别、人脸识别等新的交互方式逐步发展起来.手是人类肢体动作中最常用和灵活的一个部位,能够传达丰富的信息,符合人们日常交流习惯,而手势作为日常生活中使用频率最高的一种方式,受到了众多研究者的广泛关注,也成为了研究的重点[3-5].

在计算机视觉、深度学习等技术的飞速发展和深入研究下,开发并实现了许多基于手势的人机交互应用技术,融入到人们的日常生活中,例如智慧家居、智慧城市等[6-7].目前,最常用的手势识别方法是基于接触式设备的识别和基于计算机视觉的手势识别.基于计算机视觉的手势识别主要是利用普通二维摄像机和深度摄像机采集动作图像, 在识别手势动作时,主要是利用传统识别算法或基于深度学习的识别算法.

随着科学技术的飞速发展,手势交互作为一种新颖的交互方式,在体感游戏、手语识别、辅助驾驶、医疗器械以及智能家电控制等领域得到了广泛的应用[8-10],为了精准捕获用户手势动作以得到良好的用户交互

收稿日期: 2022-10-07; **修订日期:** 2022-11-1
基金项目: 天津市教委科研计划项目(2019KJ094);天津市科技计划项目(22YDTPJC00670)
作者简介: 洪 姣(1984—),女,河北保定人,讲师,博士.

体验，自然手势识别技术在其中起着至关重要的作用，因此被科研单位高度关注并成为社会研究的热点问题.

本实验以手势交互为研究目标，建立一种新的人机交互模式，将基于 Kinect 手势识别技术融入到交互系统中，通过对传统动态时间规整（dynamic time warping，DTW）算法进行改进，提高动态手势识别的准确性，为用户提供自然流畅的交互体验，具有较高的科研价值和良好的市场应用前景.

1　手势特征提取

1.1　Kinect 数据获取

手部特征提取是手势识别过程中的关键步骤之一，因此在手势识别之前获取有效手势特征至关重要. 获取的手势特征既要能够突出手部特点，又要保障其能够代表手势的规律性特征，同时计算量要尽可能低，复杂度也要尽可能低.

手势运动主要包括手腕、手肘以及肩部节点的运动. Kinect 是微软推出的一款基于平台的运动感知输入设备，是一种基于全新空间定位技术的体感外设，能够实时进行动作捕捉，提供稳定、准确的场景与用户信息，并且能够在各种复杂背景环境中，更准确地检测到人体，同时能够实时处理数据. Kinect v2 较 Kinect v1 增加了 5 个可识别的关节点，能够获取更多的手部信息，因此本实验使用 Kinect v2 获取人体骨架信息.

如果把 Kinect v2 获取的 25 个骨骼节点空间坐标数据全部提取并作为人体的运动特征，会大大增加系统的计算量和复杂度，降低训练效率. 由于本实验研究的动态手势动作主要由手臂来完成，因此只需要选取 4 组手臂部分的关节点相对距离系数用来作为手势特征.

1.2　矢量特征构造

1.2.1　获取关节点间欧式距离

在使用 Kinect v2 获取人体骨骼信息时发现，当人体距离 Kinect 传感器的远近不同时，所获取到的关节点位置信息也有一定的差异，为了避免这种差异对实验结果的影响，在计算节点间距离时需要进行中心化处理. 由于中肩节点在人体手势运动过程中位置相对稳定，因此选取中肩节点作为坐标基准点.

根据人们日常手部运动的习惯可知，手部运动主要涉及的是手腕、手肘和肩部的运动. 因此，实验中选取 4 组关节点的距离作为描述手部运动的特征之一，分别是左手到肩部、左肘到肩部、右手到肩部以及右肘到肩部的欧式距离，如图 1 所示.

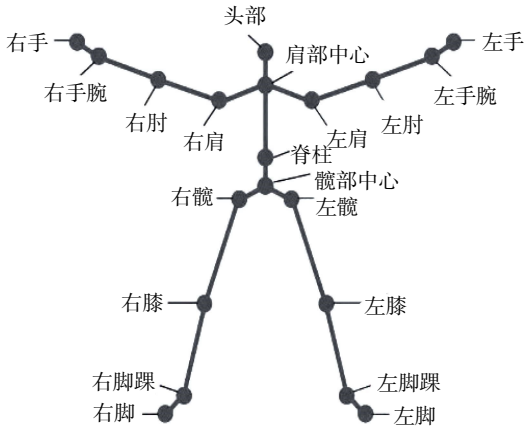


图 1　关节点结构向量

假设左手、右手、左肘、右肘和中肩 5 个节点的三维坐标分别为 $A(x_a,y_a,z_a)$、$B(x_b,y_b,z_b)$、$C(x_c,y_c,z_c)$、$D(x_d,y_d,z_d)$、$S(x_s,y_s,z_s)$，则 4 组欧式距离可以分别用公式(1)–(4)表示.

$$|\overrightarrow{AS}| = \sqrt{(x_s-x_a)^2+(y_s-y_a)^2+(z_s-z_a)^2} \tag{1}$$

$$|\overrightarrow{BS}| = \sqrt{(x_s-x_b)^2+(y_s-y_b)^2+(z_s-z_b)^2} \tag{2}$$

$$|\overrightarrow{CS}| = \sqrt{(x_s-x_c)^2+(y_s-y_c)^2+(z_s-z_c)^2} \tag{3}$$

$$|\overrightarrow{DS}| = \sqrt{(x_s-x_d)^2+(y_s-y_d)^2+(z_s-z_d)^2} \tag{4}$$

1.2.2　骨骼之间夹角特征提取

首先获取人体的骨骼三维坐标. 以右手为例，以右手手肘为坐标原点，手肘与右肩节点连接的延长线为 x 轴，手肘与手腕连接的延长线为 y 轴，定义 O、A、B 3 个点，SR 代表右肘节点，SS 代表右肩节点，WR 代表右手手腕节点，获取它们的三维坐标后，利用两两向量间的余弦值代表两者间的夹角，如图 2 所示.

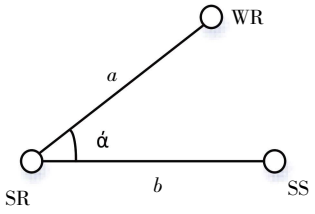


图 2　关节点夹角

假设 O、A、B 3 个关节点的三维空间坐标分别为 (x_1,y_1,z_1)、(x_2,y_2,z_2)、(x_3,y_3,z_3)，根据两点间向量的表达方式，有向量 $a=(x_2-x_1,y_2-y_1,z_2-z_1)$，向量 $b=(x_3-x_2,y_3-y_2,z_3-z_2)$，$a$ 来表示$<a,b>$的夹角，由公式(5)–(8)可以计算出关节点的夹角值.

$$|a| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{5}$$

$$|b| = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2 + (z_3 - z_2)^2} \tag{6}$$

$$a \cdot b = ((x_2 - x_1)(x_3 - x_2) + (y_2 - y_1)(y_3 - y_2) + (z_2 - z_1)(z_3 - z_2)) \tag{7}$$

$$\cos \alpha = \frac{a \cdot b}{|a||b|} \tag{8}$$

以挥手动作为例，这个手势需要跟踪记录中肩、右手和右肘 3 个关节点的运动信息,然后计算它们之间的夹角来进行判断. 记录骨骼点间的欧式距离变化以及它们之间的夹角变化作为特征向量,对挥手动作进行描述,如图 3 所示.

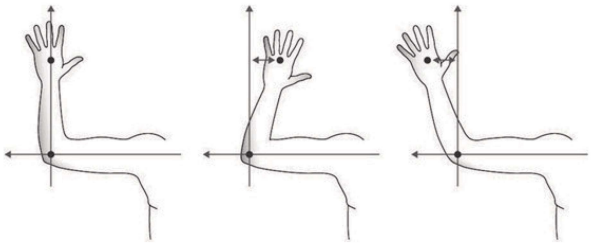


图 3　挥手手势

2　动态手势识别算法

2.1　DTW 算法原理

动态时间规整(DTW)算法是计算 2 个时间长度不同的动作序列间的相似度的一种动态规划算法. 目标是寻找一条最优路径,也就是 2 个序列间的累计距离最小，并计算出 2 个时间序列的相似度. 此算法已经成为了语音识别领域的经典算法,可以解决非等长时间序列数据的相似性计算问题. 随着研究的不断拓展,DTW 算法逐步应用于各个领域. 手势与语音有一定的相似性,手势识别中,不同的人有不同的运动速度但有相似的运动轨迹,对应于语音识别中不同人有不同的语音速度却有相似的语调,因此,这种算法也被广泛研究并应用于动态手势识别[11].

假定有 2 个动作序列 X 和 Y，长度分别为 m 和 n,如公式(9)和公式(10)所示.

$$X = \{x_1, x_2, \cdots, x_m\} \tag{9}$$

$$Y = \{y_1, y_2, \cdots, y_n\} \tag{10}$$

式中:X 为参考序列;Y 为测试序列. 通过计算两者之间的距离来判定两者间的相似度,距离越小相似度越高. 以欧几里得距离作为距离函数，从 X 和 Y 中的各个对应点之间的距离开始累计计算. 本实验引用一个 $m \times n$ 的矩阵 $\boldsymbol{M}$,该矩阵里的每个元素(i,j),表示 2 个序列两点之间的距离 $d(X_i, Y_j)$,距离函数常采用欧式距离来表示,如图 4 所示. 若要对齐 2 个序列,就要找到 2 个序列间的最优路径.

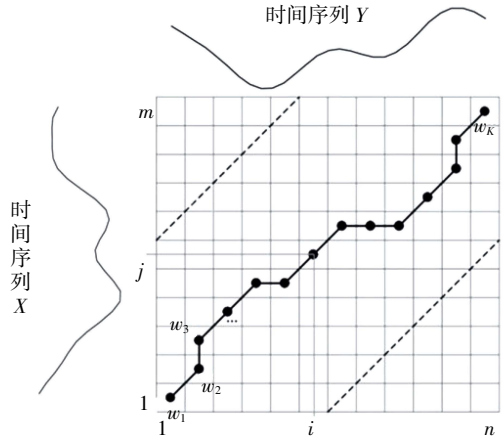


图 4　DTW 算法示意图

规整路径的表达式如公式(11)所示. 由公式可以看出，最优路径 W 可以理解为在矩阵 $\boldsymbol{M}$ 中从起始坐标(1,1)向(n,m)寻找对应点间的最短距离,使得 2 个序列的累计距离 Dis 最小. 可以把 X 与 Y 的规整过程看成是经过 $\boldsymbol{M}$ 中的点寻找最优路径.

$$W = \{w_1, w_2, \cdots, w_T\}, \max(m,n) \leqslant T \leqslant m + n - 1 \tag{11}$$

求最优路径的问题可以转换为在满足约束条件的情况下,求解最优路径函数,使得从起点到终点的路径的累计距离最小. 实验设定了以下 3 个约束条件:

(1)边界性:规整路径的起点为 $w_1 = (1,1)$,终点是 $w_T = (m,n)$.

(2)连续性:若 $W_t = (a,b)$,$W_{t'} = (a',b')$则必须满足 $a - a' \leqslant 1, b - b' \leqslant 1$，这样保证了 X 和 Y 中的每个坐标都会在规整路径中出现.

(3)单调性:若 $W_t = (a,b)$,$W_{t'} = (a',b')$,则必须满足 $a - a' \geqslant 0, b - b' \geqslant 0$,表示规整路径是随时间单调递增的.

根据上述约束条件可以发现,矩阵 $\boldsymbol{M}$ 中的每个元素在寻找下一个像素点的位置时,只能从 3 个方向进行前进,也就是以当前元素为原点,它的 x 轴、y 轴以及对角线方向. 例如从点(i,j)开始,下一步的走向有 3 个点:$(i+1,j)$,$(i,j+1)$或者$(i+1,j+1)$.

定义一个累计距离 Dis_{ij},表示当前元素的值 d_{ij} 和前一个元素的最短距离的累计距离之和,如公式(12)所示. 根据公式计算得到的路径可以认为是最优的规整路径,$DTW(\alpha,\beta)$为 α 和 β 的最短匹配距离,如公式(13)所示.

$$Dis_{ij} = d_{ij} + \min\{Dis_{(i-1)j}, Dis_{(i-1)(j-1)}, Dis_{i(j-1)}\} \tag{12}$$

$$DTW(\alpha,\beta) = \min\left\{\frac{1}{T}\sqrt{\sum_{t=1}^{T} W_t}\right\} \tag{13}$$

由上述公式计算得到的累计距离就是衡量 2 个动作是否相似的标准. 累计距离和相似度成反比,累计距离 Dis_{ij} 越小,相似度越高;反之,相似度越低.

以左手右滑手势为例,这个手势需要左肩的上半肘和下半肘的夹角信息来判断. 由于人的身高体型各不相同, 会导致手势的运动速度和轨迹有所差异,但是运动轨迹整体形状大致相同, 在匹配的过程中, *DTW* 算法会对 2 个序列进行逐点对齐,得到两个序列间的最优路径.

2.2 DTW 算法改进

传统的 DTW 算法简单、高效、识别率高,但是其算法的复杂度和计算成本高. 因此, 本实验在传统 DTW 算法基础上进行改进.

从图 4 可以看出,当 2 个序列完全对应时,就是最理想的情况,这时规整路径是起点到终点的对角线. 路径越接近对角线, 相似度越高. 但是随着动作的复杂以及样本数量的增加,且序列之间存在一对多的映射关系, 使得规整路径无法逼近对角线. 针对上述问题,结合实验中的手势特征,对传统 DTW 算法进行改进,提高识别精度和识别效率.

2.2.1 限定搜索范围

对上节提出的规整路径的计算方式进行改进,判断 2 个时间序列的相似度. 优化后的公式如(14)所示.

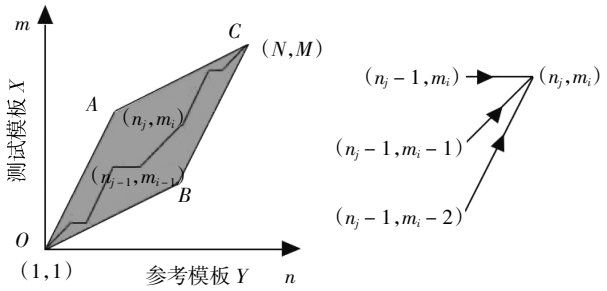
$$Dis_{ij} = d_{ij} + \min\{\alpha\cdot Dis_{(i-1)j}, \beta\cdot Dis_{(i-1)(j-1)}, \gamma\cdot Dis_{i(j-1)}\} \tag{14}$$

式中:α、β、γ 均为优化系数;d_{ij} 为该点距离;Dis_{ij} 为累计距离之和.

优化系数可以有效缩短搜索范围和减少搜索时间,同时约束路径的走势,使得规整路径无限逼近对角线. 在 $Dis_{(i-1)j}$ 和 $Dis_{i(j-1)}$都等于 1 的情况下,$Dis_{(i-1)(j-1)}$刚好是对角线的距离$\sqrt{2}$,若要使路径尽可能地逼近对角线,应使得 $Dis_{(i-1)(j-1)}$大于 $Dis_{(i-1)j}$ 和 $Dis_{i(j-1)}$,则需要使 $Dis_{(i-1)j}$ 和 $Dis_{i(j-1)}$中的最大值大于$\sqrt{2}\ Dis_{(i-1)(j-1)}$. 如图 5 所示,*OA* 和 *BC* 的斜率为 2,*AC* 和 *OB* 的斜率为 0.5,将路径的搜索范围约束在阴影部分,这样计算相似度时明显减少了计算量,提高了效率.

2.2.2 失真度阈值

考虑到实际的手势运动过程中,有的手势是负样本,是不需要进行模板匹配的,因此这种类型的手势可以直接舍弃, 以减少工作量. 在上述限定搜索范围



(a) 约束路径　　(b) 局部路径

图 5　限定搜索范围

的基础上,设置一个边界阈值 *T*,当模板之间的匹配距离 *DTW*(α,β)超过此阈值时直接跳过,不需要再进行计算. 根据实验数据分析和研究发现,2 个动作之间的相似度累积值范围为 40~120. 本实验设置阈值 *T* 为 100,若 *DTW*(α,β)的值低于 100,则表明手势是手势样本库中的手势动作,进行相似度匹配. 若 *DTW*(α,β)的值高于 100,则表明不属于手势样本库里的手势动作,不需要再做匹配计算.

3 实验结果分析

本实验对上滑、下滑、左滑、右滑、左挥手和右挥手 6 种手势动作进行了预定义,使用 Kinect 采集手势动作并利用改进 DTW 算法对采集到的手势动作与训练样本里的手势模板进行匹配, 通过观察 Visual Gesture Builder 中的显示波浪范围来判断手势动作是否被正确识别.

3.1 准确性验证

在相同环境下, 分别对 6 种预定义手势进行 80 次识别, 得出如图 6 所示的结果. 通过对比改进前后 DTW 算法的动态手势识别准确率可以看出改进后的优化参数方法较传统的 DTW 算法 6 种预定义的动态手势识别准确率均有提高, 改进后的 DTW 算法平均识别率为 98.75%, 较传统 DTW 算法动态手势识别准确率提高约 3%,说明改进后的 DTW 算法在相同环境情况下具有更好的识别准确性.

3.2 环境复杂性验证

在不同的光照强度及复杂环境下,通过使用改进 DTW 算法对预定义的 6 种手势进行 80 次动态识别,实验结果如表 1 所示. 由表 1 可以看出,改进后的DTW 算法在强光照、弱光照及复杂背景条件下的动态手势识别准确率平均值分别为 98.33%、97.92%、98.96%,由图 6 数据可以得出传统 DTW 算法在正常情况下的平均识别准确率为 95.62%, 因此改进 DTW 算法能够

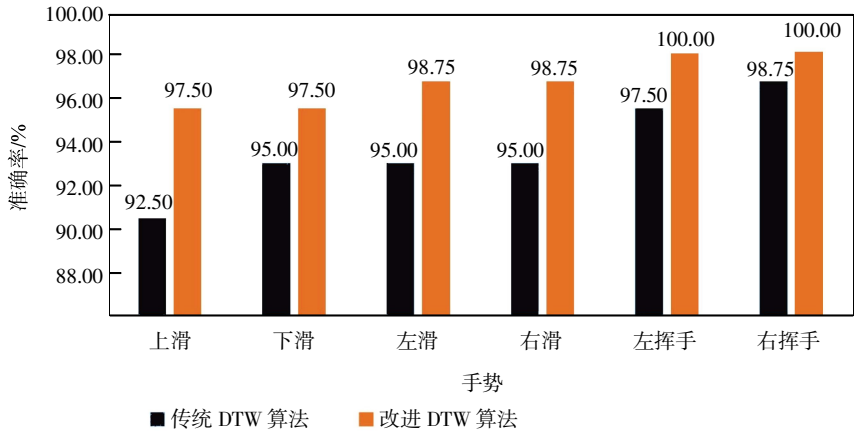


图 6　改进 DTW 优化参数测试实验结果

更好地适应复杂环境,可以有效避免复杂环境对动态手势识别准确性的影响.

表 1　不同环境条件下的动态识别准确率　%

手势	强光照	弱光照	复杂背景
上滑	97.50	96.25	98.75
下滑	98.75	97.50	100.00
左滑	100.00	100.00	100.00
右滑	96.25	98.75	97.50
左挥手	100.00	98.75	100.00
右挥手	97.5	96.25	97.50

4　结　语

本文充分利用 Kinect 传感器特点，通过对传统 DTW 算法进行改进,提出了一种新的动态手势识别方法. 利用 Kinect v2 获取深度图像，对动态手势进行分割,对传统 DTW 算法进行限制 DTW 搜索范围和设置失真度阈值两点改进. 利用改进后的 DTW 算法对提取出的手部关节信息与模板信息进行匹配,进而识别出手势. 从实验结果可以看出,改进 DTW 算法具有较高的动态手势识别效果,可以有效避免复杂环境对手势识别的影响.

参考文献:

[1] 魏秋月,刘雨帆. 基于 Kinect 和改进 DTW 算法的动态手势识别[J]. 传感器与微系统,2021,40(11):127-130.

[2] 王　兵,董洪伟,张明敏,等. 基于 Kinect 的动态手势识别[J]. 传感器与微系统,2018,37(2):143-146.

[3] 陈嘉伟,韩　晶,郝瑞玲,等. 基于改进 KNN 算法的动态手势识别研究[J]. 中北大学学报,2020,41(3):232-237.

[4] 邵天培,蒋　刚,留沧海. 基于 Kinect 的动态手势识别研究[J]. 计算机测量与控制,2021,29(2):161-165.

[5] GUO X L,YANG T T. Gesture recognition based on HMM-FNN model using a Kinect [J].Journal on Multimodal User Interfaces,2016,11(1):1-7.

[6] 林清宇. 基于 Kinect 的手势检测与追踪研究[D]. 南京:南京邮电大学,2020.

[7] 黄东方,杨晶东. 基于改进 ND-DTW 算法的动态手势识别[J]. 电子科技,2017,30(3):37-40.

[8] 李国友,孟　岩,闫春玮,等. 基于 Kinect 的动态手势识别算法改进与实现[J]. 高技术通讯,2019,29(9):841-851.

[9] 张莹莹，郭　星. 基于 Kinect 动态手势识别算法的研究与实现[J]. 计算机技术与发展,2017,27(12):11-15.

[10] SHEIKHAN M,GHARAVIAN D,ASHOFTEDEL F. Using DTW neural-based MFCC warping to improve emotional speech recognition[J]. Neural Computing and Applications, 2012, 21(7):1765-1773.

[11] 袁　菲. 基于 Kinect 的手势识别系统设计与应用[D]. 西安:西安科技大学, 2018.