

改进的 LB 算法在动态手势识别中的应用

贺霄琛, 韩 燮, 李顺增

(中北大学 计算机与控制工程学院, 山西 太原 030051)

摘 要: 为了提高动态手势识别中的匹配效率, 提出一种改进的 LB 算法. 首先, 对采集手势深度信息进行实时分割, 并使用一种新的特征提取方法对手势进行描述. 然后, 引入锚定距离的概念, 并将其应用到 LB 算法当中, 同时进行分类和识别. 最后, 分别对比其他两种相关算法的执行时间, 以及改进 LB 算法的识别率. 实验结果表明, 提出的方法在不影响识别率的情况下, 对算法执行效率有了很大的提升.

关键词: 动态手势识别; 手势分割; 锚定距离; DTW; LB

中图分类号: TP391.9

文献标识码: A

文章编号: 1000-7180(2016)04-0055-05

The Application of Improving LB Algorithm in Dynamic Gesture Recognition

HE Xiao-chen, HAN Xie, LI Shun-zeng

(College of Computer and Control Engineering, North University of China, Taiyuan 030051, China)

Abstract: In order to improve the matching efficiency of dynamic gesture recognition, we have put forward a modified LB algorithm. First of all, we segment the collecting depth information of gesture in real time, and use a new feature extraction method to describe gesture. After that, we introduce the concept of anchor distance, and apply it to the LB algorithm, for simultaneous classification and recognition. Finally, comparing execution time of other two kinds of algorithms respectively and recognition rate of modified LB algorithm. The experiment results show that the method proposed in this article has improved algorithm performance efficiency in the case of having no influence on recognition rate.

Key words: dynamic gesture recognition; gesture segmentation; anchor distance; DTW; LB

1 引言

由于手势具有直观、自然等优点, 被越来越多地应用于人机交互领域. 手势分割和特征提取是进行手势识别的第一步, 手势分割的结果往往直接影响后续的分类效率. 文献[1]提取图像中的特征点并建立模型, 同时设计一种级联的分类器, 识别实现动作识别. 文献[2]根据手势的轮廓及手型的凸起作为手势特征, 通过检测手势的指尖数来识别手势. 该方法计算量较小, 但仅适合对手势静态特征的描述. 本文受其思路启发, 提出一种新的特征提取方法, 相比较有更好的实时性并成功应用于动态手势的识别中.

DTW 算法^[3]具有简洁、识别效率高等优点, 被广泛应用于动态手势的识别. 文献[4]采用传统的 DTW 算法实现了连续的动态手势识别, 但该方法同时进行分割和识别的时本较大. 文献[5]提出一种基于 Kinect 骨骼模型的改进 DTW 算法, 但仅将手部当作一个点, 没有记录手形信息, 实用性有一定限制.

本文融合文献[6]提出的锚定距离的思想和文献[7]提出的基于 DTW 的语音连接词 LB 识别算法, 结合手势特征, 提出一种新的距离模式并将其应用于动态手势的分类识别. 在不影响手势识别率的情况下, 有效提高了算法的性能. 同时, 本文的算法基于图像的深度信息, 对复杂背景及不同的光照强

度具有很强的鲁棒性. 算法流程图如图 1 所示.

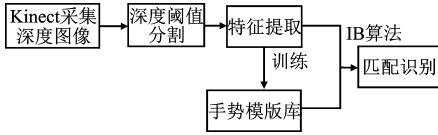


图 1 本文算法流程图

2 手势分割

2.1 深度阈值预分割

介于深度图像可以提供空间的三维信息, 而且对环境中的颜色、纹理及光照等因素不敏感, 因此可以很好地应对复杂背景. 本文中采用微软公司发布的 Kinect 相机来提取手势图像的的深度信息.

将手势区域从背景信息中分割出来, 是任何方式进行手势识别的最基础也是最关键的步骤, 其结果的好坏往往直接影响后面手势识别的效果. 在人机交互过程中, 手势往往处于交互场景的最前方, 因此可以按手势和背景的不同深度值将手势分割为背景区域和手势区域.

为了使算法具有较好的执行效率, 手部区域包含较多的特征点, 提出如下约定: 假设人手的位置是距离相机最近的, 且与 Kinect 相机之间的距离在 1 m 以内. 此外, 规定手势的指尖向上, 手心朝向摄像头, 手部区域平面与相机平面的夹角小于 45° . 基于之前的约定, 通过深度阈值算法对深度信息进行处理.

深度阈值分割步骤如下:

- (1) 令 $X_{u,v}$ 是获取的深度图像任意一点;
- (2) 扫描深度图像, 得到其最小的深度值 $D_{\min}$;
- (3) 以 $T_{\text{depth}} = 20 \text{ cm}$ 为阈值, 进行分割, 得到手势区域 H (该区域包含部分手臂, 后续会进行处理) 可以表示为

$$H = \{X_{u,v} \mid D(u,v) < D_{\min} + T_{\text{depth}}\} \quad (1)$$

(4) 将步骤(3)中得到的手势区域 H 二值化, 得到掩模 $B(u,v)$:

$$B(u,v) = \begin{cases} 1, & x \in H \\ 0, & \text{其他} \end{cases} \quad (2)$$

对深度图像进行预分割, 并得到二值化图像, 如图 2 所示.

2.2 手势区域细分

上一节中分割出的手势区域包括部分手腕和手臂信息, 此部分对手势识别不起作用, 且会影响后续的特征提取, 属于冗余信息, 需要剔除. 在剔除之前,



图 2 深度阈值预分割

首先需要找到掌心的位置. 由于手掌区域一般情况下比手指和手掌区域的面积要大, 所以 B 区域具有最大的点密度.

根据文献[8], 采用高斯滤波来过滤掩模 $B(u,v)$, 并为高斯滤波设置一个非常大的标准差.

$$\sigma = 150 \cdot \frac{1}{D_{\min}} \quad (3)$$

这可以保证过滤区域是足够大的, 以捕捉到手腕和关节区域.

令高斯滤波输出的区域为 $B^f(u,v)$, 取该区域内具有最大灰度值对应的点 C_g (即 B 中密度最大的点), 则认为该点就是要寻找的掌心位置 C , $C = C_g$. 当然, 某些情况下, 由于手部倾斜角度太大会导致找到的掌心位置向手腕方向偏移, 这也是在上一节中提出约定的原因. 图 3 所示为高斯滤波的结果.



图 3 高斯滤波结果

本文中特征提取的方法需要分离出手指部分, 在确定掌心位置后, 通过椭圆拟合将得到的手势区域进一步细分. 椭圆拟合步骤如下:

- (1) 以掌心 C 为手掌区域内切圆的圆心, 令半径 $R = 1 \text{ pxl}$ 为初始值;
- (2) 依次增长 R 值;
- (3) 直到属于区域 H 的点 95% 都包含在该内切圆之内.

至此, 可以将上文获取的手势 (包含手腕) 区域 H 进一步分割为三部分: 手指区域 F 、手掌区域 P , 和手腕区域 W (如图 4 所示). 其中, 手掌区域 P 是手掌内切圆内的部分; 手腕区域 W 为 H 区域的子集, 其中 $x \leq -R$; 手指区域为 $F = H - W - P$.



图4 椭圆拟合分割

3 特征提取

对于任意一点 $X_i \in F$, 计算对应点到手掌区域的标准距离:

$$d_{xi} = ||X_i - C|| - R \quad (4)$$

以手中指方向为 X 轴正向, 记为 i_x . 手指集合上的点 X_i 与中指的夹角就是 $X_i - C$ 与轴线 i_x 的夹角, 记为 θ_i . 对于指尖点其距离表示如下:

$$L(\theta) = \max d_{xi} \quad (5)$$

式中, 考虑误差因素, 手指尖对应的角度区域为 θ , 并设误差标准为 $\Delta = 2^\circ$.

$$\theta_i - \frac{\Delta}{2} \leq \theta \leq \theta_i + \frac{\Delta}{2} \quad (6)$$

不同的手势是由不同手指的伸展与弯曲决定的, 实验过程中的误差主要来源于实验对象个体的体型差异. 为了消除这些误差, 对实验数据进行标准化处理. 得到编号为 i 的手指在 t 时刻的特征分量为

$$f_{i,t} = \frac{L(\theta_{i,t})}{L_{\max}} \quad (7)$$

式中, i 表示手指集合 $i \in [1, 5]$, $L_{\max}$ 为中指的长度.

最后, 将所有特征分量统一收入 F_t 中作为 t 时刻的特征向量.

4 动态手势识别

4.1 DTW 算法描述

动态时间规整 (Dynamic Time Warping, DTW) 算法, 最早被应用于语音识别领域, 算法思想是在测试序列和参考序列通过时间弯折函数建立一条非线性的弯折路径, 沿着该路径上的两个序列之间的距离越小, 则相似度越高^[9].

设测试序列为 $T = \{t_1, t_2, \dots, t_n\}$, 参考序列为 $R = \{r_1, r_2, \dots, r_m\}$, 其中 n, m 分别为对应序列手势帧的时序标号, 用 D 来表示序列 R 和 T 来表示的累积距离矩阵:

$$D = \sum_{k=1}^f d(i_k, j_k) \quad (8)$$

式中, $d(i, j)$ 表示测试序列 T 第 i 帧的特征向量与参考序列 R 第 j 帧的特征向量之间的距离, f 为距离映射的个数. 则每条测试序列到参考序列的路径对应上式的距离矩阵 D , 开销最小的距离矩阵就是两序列之间的最佳匹配. 图 5 所示为 DTW 算法原理图.

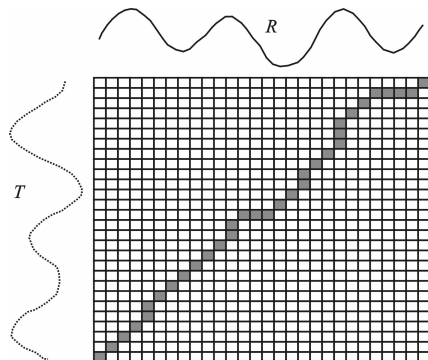


图5 DTW 算法原理

4.2 提出一种新的距离模式

DTW 为了寻找两序列之间最佳的匹配路径, 会寻找所有可能的路径, 其计算开销很大. 此外 DTW 在处理多个连接手势组成的序列时其分类表现较差. 考虑到上述因素, 提出一种新的距离函数, 提高算法的运行效率, 同时, 新的距离函数在分类识别时也表现出较好的性能.

令模板库 $M: \{X_1, X_2, \dots, X_n\}$, M 中的每个元素代表一个特定的手势类. 对于给定的手势序列, 定义一种新的逐帧距离表示方法. 在执行动态时间规整过程中, 将 D 中的距离平均到一个时间窗口 R 中, 计算平均值只捕获的一对帧之间的最大距离, 之后再进行归一化处理, 得到

$$D(k, l) = 1 - e^{-\sum_{k=R}^{k+R+l} ||f_{X_i}(k) - f_{X_j}(l)||^2} \quad (9)$$

式中, $l = \{1, \dots, (K/R)\}$, $k = \{1, \dots, (L/R)\}$, $f_{X_i}(k)$ 表示的是序列 X_i 中对应第 k 帧的特征向量. 这样, D 在的距离被切分成大小相等的 $R \times R$ 块.

用 $w(X_i, X_j)$ 表示对 X_i 和 X_j 进行第一次动态时间规整得到的距离矩阵, 称之为序列 X_i 与 X_j 的偏差向量. 下面, 引入锚定距离的概念.

锚定距离使用第三个序列来计算两序列之间距离. 假设 X_i 为模板库中的序列, Q 为查询序列, X_j 为锚定序列, 在确定查询序列 Q 与模板序列 X_i 的距离时, 以 X_j 为锚定. $w(X_i, X_j)$ 是序列 X_i 与 X_j 的偏差向量, $w(Q, X_j)$ 是序列 Q 与 X_j 的偏差向量.

$w(w(X_i, X_j), w(Q, X_j))$ 是对两偏差向量再一次进行动态时间规整, 实际上就是对两个距离矩阵进行比较. 锚定距离可以表示为

$$s(X_i, Q | X_j) = w^T 1 \tag{10}$$

式中, 向量 1 表示将 w 中的所有的值加起来, 锚定距离 $s(X_i, Q | X_j)$ 的值越小, X_i 与 Q 的相似度越高.

我们的目标是计算查询序列 Q 与每个模板序列的距离 $d(Q, X_i)$. 距离 d 即在模板库 M 中找到所有锚定距离最小的那个, 如下式:

$$d(X_i, Q | \{X_1, X_2, \dots, X_n\}) = \min_{j \neq i} (X_i, Q | X_j) \tag{11}$$

为了减少分类识别过程中的运算成本, 提出一些对锚定序列预选择的策略:

(1) 对于给定的模板库 $M: \{X_1, X_2, \dots, X_n\}$, 需要找到对于特定模型 X_i 最合适的锚定序列 X_j , 模板库本身作为查询序列进行测试.

(2) 在选择锚定序列时, 如果某一个序列被选为锚的次数超过第二位 10% 以上, 则可以将该序列预选定为锚定序列. 如果没有这样的序列, 则随机选择一个作为锚定序列.

4.3 改进的 LB 算法

一个查询序列可以看做由数个标签组成, 这些标签对应于模板库中的手势类. 因此, 动态手势识别需要对查询序列进行动态、精确的分割. LB(Level-Building)算法是一种融合分割和识别匹配有效的方法, 其基本思想是: 将参考序列与查询序列进行分层匹配, 首先对未知的查询序列进行逐点匹配, 求出当前层内最可能的路径, 然后再逐层匹配求出整个过程中的最优路径, 直到动态路径搜索完毕之后, 再对路径进行回溯确定切分点得到识别结果^[10]. LB 算法每一步的匹配计算中采用 DTW 方法, 本文结合锚定距离, 提出了改进的 LB 算法.

设待识别的查询序列中最大的手势个数为 $L_{\max}$, 算法中可以分为 $L_{\max}$ 层进行匹配. 算法步骤如下:

(1) 第一层 LB 算法

每一层对应于手势在查询序列中可能的顺序, 因此, 第一层是查询序列中第一个可能的标签. 当 $l = 1$ 时, 有如下等式:

$$A(i, m, l) = d_{i \neq y}(X_i, Q(1:m) | X_y) \tag{12}$$

式中, 矩阵 $A(i, m, l)$ 表示对计算成本的累积和, M^* 为查询序列总帧长, $1 \leq l \leq L_{\max}, 1 \leq m \leq M^*, d$ 为锚定距离.

为了对路径回溯, 对应矩阵 A 定义索引矩阵

$\varphi(i, m, l)$ 来记录对应模板序列的索引号.

$$\varphi(i, m, l) = \operatorname{argmin}_i (d_{i \neq y}(X_i, Q(1:m) | X_y)) \tag{13}$$

(2) 2 ~ L 层的迭代计算

第 l 层对应查询序列中的第 l 个标签, 该标签的起点对应上一层次中标签的终点. 同样, 可以得到累积矩阵 $A(X_i, Q, l)$ 和索引矩阵 $\varphi(X_i, Q, l)$, 如下:

$$A(i, m, l) = \min_{k, j} A(k, j, l-1) + d_{i \neq y}(X_i, Q(j+1:m) | X_y) \tag{14}$$

$$\varphi(i, m, l) = \operatorname{argmin}_k (A(k, j, l-1) + d_{i \neq y}(X_i, Q(j+1:m) | X_y)) \tag{15}$$

最后, 通过回溯索引矩阵 φ 得到识别结果.

5 实验结果分析

本文定义数字 1 到 9 共 9 个手势类模板, 一个动态手势命令由若干个手势类组成, 令 $L_{\max} = 5$, 即每种手势命令中包含的手势类不超过 5 个, 我们定义 4 种不同的手势命令 (A, B, C, D). 邀请 10 名志愿者分别进行测试, 每个手势命令测试 20 次, 每个手势命令获取获取 200 个样本, 其中 50 次作为训练样本, 其余 150 次作为测试样本. 分别采用文献[4]中传统的 DTW 算法、文献[7]中的 LB 算法和本文提出的改进 LB 算法对手势命令进行分类识别, 比较三者的处理时间和识别率, 结果如表 1 所示.

表 1 算法处理时间对比

手势命令	平均处理时间/ms		
	文献[4]中 DTW 方法	文献[7]中 LB 算法	本文改进的 LB 算法
A	438	319	253
B	523	387	297
C	481	354	265
D	497	361	270

从表 1 中可以看出, LB 算法在处理连接手势时相对于 DTW 时间成本更低, 本文提出的改进 LB 算法使用锚定距离对匹配过程进行优化, 进一步降低了处理时间.

从表 2 可以看出, 改进 LB 算法对手势命令有很高的识别率.

表 2 改进 LB 算法手势识别率

手势命令	测试个数	识别个数	识别率/%
A	150	141	94.0
B	150	129	86.0
C	150	137	91.3
D	150	132	88.0

6 结束语

本文提出了基于深度图像的动态手势识别方法,该方法基于深度图像,对光照强度和复杂背景有较强的鲁棒性,且在不影响识别率的情况下,相对于文献[4]中的 DTW 算法和文献[7]中的 LB 算法在处理时间上有了明显的提升.本文的手势特征有很好的识别效率,但特征提取过程相对复杂,如何找到一种简化的手势特征使得算法效率进一步提高,是下一步的研究方向.

参考文献:

[1] Trelea I C. The particle swarm optimization algorithm; Convergence analysis and parameter selection[J]. Information processing letters, 2003, 85(6): 317-325.

[2] 翁汉良, 战荫伟. 基于视觉的多特征手势识别[J]. 计算机工程与科学, 2012, 34(2): 123-127.

[3] Guyon I, Athitsos V, Jangyodsuk P, et al. Results and analysis of the chlearn gesture challenge 2012 [M]//Depth Image Analysis and Applications. Berlin, Heidelberg: Springer , 2013: 186-204.

[4] Ney H, Ortmanns S. Dynamic programming search for continuous speech recognition[J]. Signal Processing Magazine, IEEE, 1999, 16(5): 64-83.

[5] 何超, 胡章芳, 王艳, 等. 一种基于改进 DTW 算法的

动态手势识别方法[J]. 数字通信, 2013, 40(3): 21-25.

[6] Krishnan R, Sarkar S. Similarity measure between two gestures using triplets[C]//Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on. Oregon:IEEE, 2013: 506-513.

[7] Myers C, Rabiner L. A level building dynamic time warping algorithm for connected word recognition[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1981, 29(2): 284-297.

[8] 曹雏清, 李瑞峰, 赵立军. 基于深度图像技术的手势识别方法[J]. 计算机工程, 2012, 38(8): 16-18.

[9] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1978, 26(1): 43-49.

[10] 徐华, 俞一彪. 基于 MIM_LB 算法的连续数字语音识别[J]. 微电子学与计算机, 2004, 21(5): 33-36.

作者简介:

贺霄琛 男,(1991-),硕士研究生. 研究方向为计算机仿真、虚拟现实. E-mail:76156860@qq.com.

韩 雯 女,(1964-),教授,博士生导师. 研究方向为仿真与可视化、虚拟现实.

李顺增 男,(1966-),副教授. 研究方向为计算机硬件、嵌入式技术.

(上接第 54 页)

[12] Lu H , Kot A C , Shi Y Q. Distance-reciprocal distortion measure for binary document Images[J]. IEEE Signal Processing Letters, 2004, 11,(2):228-231.

作者简介:

冯 炎 男,(1979-),硕士,讲师. 研究方向为信号与信息处理、图像处理. E-mail:fy4528@163.com.