

基于线性判别分析和自适应 K 近邻法的手势识别

温俊芹¹ 王修晖²

(1. 浙江经济职业技术学院数字信息技术学院, 杭州, 310018; 2. 中国计量大学信息工程学院, 杭州, 310018)

摘要: 基于小容量数据集的手势识别是人机交互技术研究中的一个重要课题。本文提出了一种基于线性判别分析和自适应 K 近邻法的手势识别方法。首先, 应用高斯背景建模方法从包含目标交互者的训练视频集中提取各类手型图像, 并调整到相同尺度来构建手势训练集。然后, 通过改进的线性判别分析对训练数据进行特征提取。最后提出一种自适应 K 近邻法对实时交互过程中得到的手型信息进行分类和识别。应用上述方法自建小型手势库进行实验和比较分析, 结果显示与现有的手势识别算法相比, 本文方法具有更高的识别率。

关键词: 线性判别分析; 手势识别; 自适应 K 近邻法; 人机交互

中图分类号: TP301 **文献标志码:** A

Hand Gesture Recognition with Linear Discriminant Analysis and Adaptive K-Nearest Neighbor Algorithms

Wen Junqin¹, Wang Xiuhui²

(1. Digital Information Technology Institute, Zhejiang Technical Institute of Economics, Hangzhou, 310018, China; 2. School of Information Engineering, China Jiliang University, Hangzhou, 310018, China)

Abstract: Gesture recognition based on small samples is one of the main trends in the advanced human-computer interaction research. A novel gesture recognition method based on adaptive K-nearest neighbor (A-KNN) and linear discriminant analysis (LDA) is presented. First, hand-shape images are segmented from the given interaction videos, and scaled to the same size to construct the training set. Then an optimized LDA algorithm is designed to extract gesture features. Finally, an improved KNN algorithm is introduced with adaptive K value to classify the real-time gesture information. Test results show that the correct recognition rate of the proposed approach is higher than most existing methods.

Key words: linear discriminant analysis (LDA); hand gesture recognition; adaptive K-nearest neighbor; human computer interaction

引 言

手势交互是一种更加自然的非接触式人机交互方式, 它通过将人手的各种姿态作为控制语义输入到计算机中实现对特定目标或者场景的控制。手势是通过人手的连续动作产生的各种状态或者状态的组合, 通常包括静态的手势和动态的手势(泛指一组静态手势的组合)^[1]。随着高级人机交互技术的广

泛应用,如何实现更加自然的人机交互成为相关领域的关注焦点,包括人体姿态识别、手势识别、步态识别以及自然语言识别等^[2]。其中手势识别作为一种非接触式的关键输入技术,在机器人、虚拟现实游戏和工业控制等领域应用广泛。根据传感器类型不同,现有关于手势识别的研究成果大致可以分为基于物理传感器和基于视频两类方法。基于物理传感器的手势识别方法包括使用数据手套检测手指变化情况的方法^[3]和通过在手指关节佩戴感应装置来判断手指状态的方法^[4]。这类方法的一个共同点是依赖于交互者手上佩戴的某种机械或者电子感知设备,在很大程度上影响了用户的方便性。此外,由于感知装置本身灵敏度对人手尺度的固有敏感性和装置的易磨损性,进一步限制了该类手势识别方法的使用空间。基于视频的手势识别主要通过计算机视觉和模式识别技术对交互者手部轮廓进行建模和特征提取,从而抽象出不同的交互语义。文献[5]提出了一种简单的骨骼模型来描述人的手部特征。在此基础上,融合遗传算法中迭代优选的思想,实现当前交互者的手型轮廓与模式库中手形轮廓的匹配。文献[6]提出了一种静态的手势辨识方法,该方法利用手势空间分布特征来提升手势辨识的正确率,但是该方法难以解决遮挡带来的问题。文献[7]提出了一种使用 SVM 作为分类器的手势识别方法。文献[8]提出了一种基于无线信号的手势识别系统,该系统提供基于家庭的手势识别解决方案。总之,基于视频的手势识别和交互技术能够实现更加自然的非接触式交互。同时手势正确识别率也有所提高。但是由于人手非刚体的本质,手势识别在误判率和识别效率等方面都存在很大的困难。为了解决当前基于视频的手势识别方法难以满足实时交互要求的难题,本文在文献[5]基础上,设计了一种无模型的手势识别解决方案。该方案分为离线手势分类器训练和实时手势识别两个步骤。在离线处理阶段,应用高斯背景建模方法从包含目标交互者的训练视频集中提取各类手型图像来建立训练集,并通过改进的线性判别分析对训练数据进行特征提取和压缩。在实时手势识别阶段,采用一种自适应 K 近邻法(Adaptive K-nearest neighbor, A-KNN)对实时交互过程中得到的手型信息进行分类和识别。

1 手势分类器训练

1.1 手型图像提取和归一化

手型图像提取主要研究如何从连续的手势交互图像中分离出完整的手型部分。本文根据实际交互中的运动连续性、人体肤色特征以及腕部环状标识信息,使用渐进性方法完成手型部分的分割^[5]。然后,对得到的各种手型模板进行旋转和尺寸归一化(全部展开的手指中分线朝上;尺寸 200 像素×220 像素)。本文选取了 18 种差异较大的静态手势建立训练集,图 1 显示了 18 种手势之一的“两手指分开伸展手势”归一化处理后得到的 10 个模板。

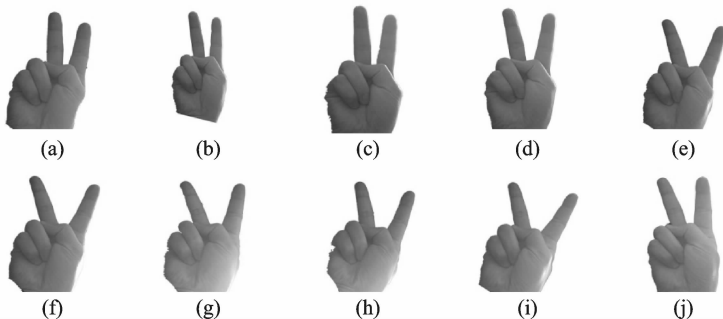


图 1 两手指分开伸展手势的归一化模板

Fig. 1 Normalized templates for hand gesture with two-finger separate extension

1.2 基于改进 LDA 算法的手势分类器训练

假设手势训练数据集 $X = \{X_1, X_2, \dots, X_c\}$, 其中 $X_i (i \in [1, C])$ 为第 i 类手势数据集。令 $n_i (i \in [1,$

C])为先验概率,样本 $x_1, x_2, \dots, x_n \in \mathbf{R}^D$, 则可以得到标准 Fisher 法则, 则

$$J(\boldsymbol{\varphi}) = \boldsymbol{\varphi}^T \mathbf{S}_b \boldsymbol{\varphi} / \boldsymbol{\varphi}^T \mathbf{S}_w \boldsymbol{\varphi} \quad (1)$$

$$\mathbf{S}_b = \sum_{i=1}^c n_i (\mathbf{m}_i - \mathbf{m}) (\mathbf{m}_i - \mathbf{m})^T \quad (2)$$

$$\mathbf{S}_w = \sum_{i=1}^c \sum_{x \in X_i} (\mathbf{x} - \mathbf{m}_i) (\mathbf{x} - \mathbf{m}_i)^T \quad (3)$$

式中: \mathbf{S}_w 为同一类内部个体之间的离散度矩阵; \mathbf{S}_b 为不同类之间的离散度矩阵; $\mathbf{m} = \frac{1}{n} \sum_{i=1}^n x_i$ 为样本均值; $\mathbf{m}_i = \frac{1}{n_i} \sum_{x_i \in X_i} x_i$ 为当前类型手势的样本均值。通过求取式(1)的最优解获得用于训练手势分类器的最优投影向量。但在拉格朗日乘数法求解中

$$\mathbf{S}_b \mathbf{W} = \lambda \mathbf{S}_w \quad (4)$$

式中: λ 为拉格朗日因子; \mathbf{S}_w 可能奇异, 此时式(4)无法直接求解。针对小样本手势样本库的特点, 使用类似文献[9-11]的思路对 LDA 算法进行改进, 并使用改进后算法来训练手势分类器。改进算法的基本思想是在 \mathbf{S}_w 内求解 1 个线性变换矩阵, 使得在 \mathbf{S}_w 上 \mathbf{S}_b 最大, 从而将 LDA 求解问题转换为一个最优化问题, 则

$$\mathbf{M} = \operatorname{argmax}(\mathbf{M}^T \mathbf{S}_b \mathbf{M}) \quad (5)$$

使得 $\mathbf{M}^T \mathbf{S}_w \mathbf{M} = 0$, $\mathbf{M} \in \mathbf{R}^D$ 。式(5)已经在文献[9]中进行了收敛性证明, 并给出了具体解决方法, 这里不在赘述。

2 基于自适应 KNN 算法的手势识别

K 近邻(K-nearest neighbor, KNN)算法的基本思想是在某一特征空间中, 挑选出与待分类的样本 X 距离最近的 k 个已知样本构成目标集, 如果目标集中大多数样本属于某个类别 T , 则 X 属于 T 。该方法在分类过程中只使用最接近的 k 个样本的已知类别来判定未知样本的最终归属。KNN 法根据某种距离度量对未知样本进行分类, 简单直观, 而且在多数情况下分类效果较好。然而, KNN 的分类正确率严重依赖于所选取的 k 值: 当 k 值选择偏小时, 所获取的紧邻样本集合过小, 会对偶然误差和噪声过于敏感, 损失分类的精度和鲁棒性。反之如果 k 值选取偏大, 同时未知样本所归属的类在训练样本集中个体较少, 则在选择其 k 个近邻时与未知样本距离较大的样本也会被选中, 从而可能造成误判。针对上述 KNN 的缺陷, 本文提出了 A-KNN 来降低手势识别的误判率。假设手势样本集 $R = \{Y_j^{<i> \dots </i>} \mid i = 1, 2, \dots, c; j = 1, 2, \dots, N_i\}$, 样本数 $N = \sum_{i=1}^c N_i$, N_i 为当前手势类型的样本数。通过自适应处理获得优化的 K 值 \hat{k} , 统计待分类手势样本的 \hat{k} 个最近邻中大多数样本所属的类, 并将 x 判归该类。定义 $\{\omega_i \mid i = 1, 2, \dots, c\}$ 类的判决函数为

$$\{D_i(x) = n_i \mid i = 1, 2, \dots, c\} \quad (6)$$

判决条件为

$$x \in \omega_t \quad t = \max_{i=1, 2, \dots, c} (D_i(x)) \quad (7)$$

训练样本数量较少时, KNN 算法的误判率对距离度量方法比较敏感。A-KNN 算法使用的欧氏距离为

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (8)$$

式中: $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$; $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$; n 为输入特征的维数。对应的置信度函数为

$$\eta_i = \frac{\sum_{j=1}^k d_{ij}^y}{\sum_{j=1}^k d_{ij}^{-y}} \quad i = 1, 2, \dots, n; j \neq i \quad (9)$$

本文提出的 A-KNN 算法流程如下:

(1) 数据预处理,使用与离线处理阶段类似的方法将待分类手势样本 x 从输入视频流中分割出来,并调整尺寸大小实现归一化。

(2) 获得自适应 k 值,设定 k 初始值为 $0.5 \times N$, N 为训练集容量。

(3) 创建一个长度为 k 的队列保存最相似的训练样本集,并按照式(8)定义的距离从大到小排序。首先从训练样本集中随机地选择 k 个样本构成最相似的训练样本集,然后计算这些样本与待分类样本之间的距离,并将对应序号存入队列。

(4) 依次计算训练集中所有个体与未知手势样本的距离 d ,并将其与队列中的最大距离 D 比较:如果 $d \geq D$,则舍弃该样本集;如果 $d < D$,则删除队列中 D 对应的样本集,并将当前样本集存入队列。

(5) 完成遍历之后,统计队列中多数样本所属分类,并将其作为标准测试集的分类。

(6) 计算测试结果的置信度,然后选取 k 值继续训练。最终取置信度最大的 k 值作为输出。

3 算法性能实验

3.1 手势识别测试系统实现

在多台投影仪拼接实现的大屏幕环境下,搭建了一个沉浸式手势识别系统来测试本文提出算法的性能。该系统的运行包括手势分类器训练和手势识别两个阶段,如图 2 所示。其中手势轮廓提取和手部运动目标跟踪虽然是该系统的重要模块之一,相关内容已经在相关工作^[5]中进行了详细描述,不属于本文研究内容,这里不再详述。从图 2 可以看出,该系统的输入包括训练图像序列和手部实时运动视频两类,其中训练图像序列用于训练离线阶段的手势分类器,而实时视频则用于测试手势识别算法。在离线的预处理阶段,输入的训练图像序列包含当前交互者的不同手势。因此,在对手势分类器进行训练前,需要先提取手势轮廓。本文实验采用了高斯混合背景建模方法建立手势模型,根据人体肤色、手指和手掌的基本关系等进行手势轮廓检测,并进行图像分割和尺度调整,建立规格化的手势训练集。在手势分类器训练过程中,采用本文提出的改进线性判别分析方法提取手势特征并进行降维操作,以提高分类器训练的效率。实时手势识别阶段包括以下几个处理过程:(1)手部运动目标跟踪。与离线阶段的手势轮廓提取不同,实时视频流中的背景比较复杂。为了方便下一步进行轮廓提取,需要从来自于监控摄像头的实时视频数据中检测出交互者的手部,并实现可靠的跟踪。(2)手势轮廓提取。这一步与离线处理阶段的手势轮廓提取类似,需要采用混合高斯模型建模和肤色检测方法检测出手部的位置并调整到统一的规格。但是该阶段的手部移动更加显著,为了提高手势识别的效率,需要利用运动的连贯性。(3)手势识别。该步骤使用本文提出的 A-KNN 算法,利用训练好的手势分类器,对步骤 2 提取的手势轮廓数据进行实时分类,从而实现手势识别。在该手势识别系统中,离线阶段完成计算相对复杂的手势分类器训练,能够较好地提高手势识别的实时性;同时,手势识别中采用 A-KNN 算法,相比采用固定 k 值的 KNN 算法,较大幅度地提高了识别率。

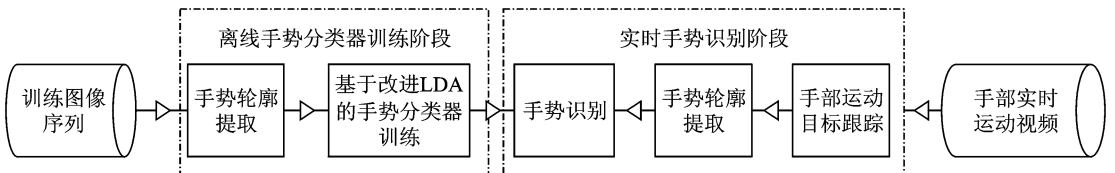


图 2 手势识别测试系统

Fig. 2 Hand gesture recognition system

3.2 实验结果与分析

本文设计了两个实验,分别将本文算法与通过手机、手套硬件实现的手势识别方法,基于序贯相似检测算法(Sequential similarity detection algorithm,SSDA)的识别算法和基于自适应遗传算法的识别算

法^[5]进行了比较,并对结果进行了分析。测试中的用例为对虚拟常见中的物体进行“选中”、“移动”和“旋转”等3种基本操作。实验中所用的数据手套采用了5DT Data虚拟现实体感手套,每个节点14个传感器。用于视频采集的摄像头采用了巨峰科技的高清网络摄像头,焦距为6 mm,清晰度1 080 P。测试系统的运行环境为HP Pro 3380 MT(4GB RAM, I5-3470 CPU)。实验1比较了本文提出的手势识别方法和基于硬件传感器实现的手势识别方法,实验结果如图3所示。实验中“选中”、“移动”和“旋转”3种测试用例各执行了200次。可以看出借助硬件传感器实现的手势识别方法的误判率比本文方法高,尤其是对“平移”和“旋转”测试用例中一些相互差异较小的手势。这主要由硬件手势识别装置中的传感器对不同的人手部尺度的依赖性和机电装置内在感应度局限性造成。另一方面,从交互的自然和谐方面来说,本文的手势识别方案无需交互者穿戴特殊装置,更符合人类日常习惯。然而尽管本文手势识别算法的平均识别速度已经能够满足一般的实时交互需求,但是在整体识别速度方面,仍然无法和硬件实现方法相比拟,因此这里只比较误判率。实验2比较了本文手势识别算法与其他两种手势识别算法,结果如图4,5所示。实验过程中,3种测试用例(“选中”、“移动”和“旋转”)各执行了200次。其中的KNN算法,其前期处理仍然采用改进的LDA算法,在手势识别中采用了KNN算法。图4的结果显示,在误判率方面本文算法显著优于基于SSDA的方法和基于自适应遗传算法的手势识别方法^[5],在一些存在区分度较小手势的组合操作中差别更加明显;与传统的KNN算法比较也有明显的提高。这是由于本文提出的基于自适应KNN的手势识别算法自适应选择置信度最高的K值,降低了传统KNN算法由于K值偏大或者偏小造成的识别精度损失。此外在平均辨识速度^[5]方面,本文提出的方法与SSDA算法和同样采用改进LDA进行前期处理的KNN算法类似,但是和文献^[5]相比有了较大改善。

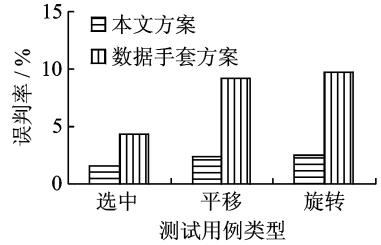


图3 本文识别方案与数据手套识别方案的比较
Fig. 3 Comparison between proposed method and data glove scheme

但是和文献^[5]相比有了较大改善。

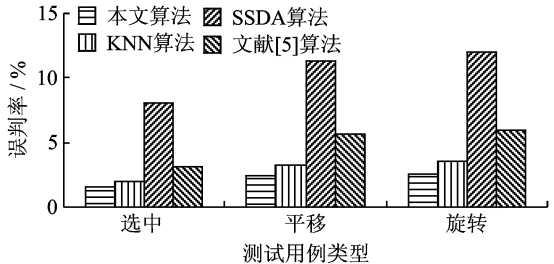


图4 本文手势识别算法与其他手势识别算法误判率比较
Fig. 4 Comparison of proposed method and other recognition algorithms in misjudgment rate

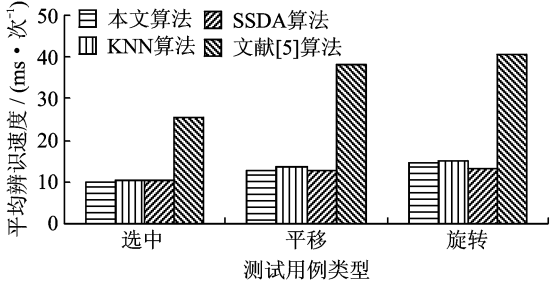


图5 本文手势识别算法与其他手势识别算法平均识别时间比较
Fig. 5 Comparison of proposed method and other recognition algorithms in average recognition time

4 结束语

手势识别的正确率和识别速度是评价一个手势识别算法的主要标准,但是通常情况下手势识别率和识别速度是难以兼顾的。如何在提高手势识别率的同时,改善识别速度是提升手势识别算法实用性的关键。本文提出一种基于小容量手势集的视频手势分类和识别方法。该方法通过在离线阶段实现复杂度较高的手势分类器训练,并在实时手势识别中采用 A-KNN 算法提高手势识别率。实验结果显示,本文提出的算法在很大程度上提高了实时手势识别的正确率,并改善了交互阶段的识别效率。

参考文献:

- [1] Kirishima T, Sato K, Chihara K. Real-time gesture recognition by learning and selective control of visual interest points[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(3): 351-364.
- [2] Gonzalez-Sanchez T, Puig D. Real-time body gesture recognition using depth camera[J]. *Electronics Letters*, 2011, 47(12): 697-698.
- [3] 李琪,白英彩,曾芬芳. 一个虚拟手原型的实现[J]. *上海交通大学学报*, 2009, 33(11): 1377-1380.
Li Qi, Bai Yingcai, Zeng Fenfang. Implementation of a virtual hand prototype[J]. *Journal of Shanghai Jiaotong University*, 2009, 33(11): 1377-1380.
- [4] Haritaoglu I, Harwood D, Davis L S. W4: Real-time surveillance of people and their activities[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 809-830.
- [5] 王修晖,鲍虎军. 基于自适应遗传算法的手势识别[J]. *计算机辅助设计与图形学学报*, 2007, 19(8): 1056-1063.
Wang Xiuhui, Bao Hujun. Gesture recognition based on adaptive genetic algorithm[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2007, 19(8): 1056-1062.
- [6] 杨波,宋晓娜,冯志全,等. 复杂背景下基于空间分布特征的手势识别算法[J]. *计算机辅助设计与图形学学报*, 2010, 22(10): 1841-1848.
Yang Bo, Song Xiaona, Feng Zhiquan, et al. Gesture recognition in complex background based on distribution features of hand[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2010, 22(10): 1841-1848.
- [7] Wu Chihhung, Chen Weilun, Lin Changhong. Depth-based hand gesture recognition[J]. *Multimedia Tools and Applications*, 2016, 75(12): 7065-7086.
- [8] Pu Qifan, Gupta S, Gollakota S. Shwetak patel: Gesture recognition using wireless signals[J]. *ACM GetMobile*, 2015, 18(4): 15-18.
- [9] Zhang Ya, Chen Weiyuan, Zha Hongyuan, et al. A time-topic coupled LDA model for IPTV user behaviors[J]. *IEEE Transactions on Broadcasting*, 2015, 61(1): 56-65.
- [10] Yang Feng, Mao K Z, Lee G K K, et al. Emphasizing minority class in LDA for feature subset selection on high-dimensional small-sized problems[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(1): 88-101.
- [11] 刘绍毓,周杰,李弼程,等. 基于多分类 SVM-KNN 的实体关系抽取方法[J]. *数据采集与处理*, 2015, 30(1): 202-210.
Liu Shaoyu, Zhou Jie, Li Bicheng, et al. Entity relation extraction method based on multi-SVM-KNN classifier[J]. *Journal of Data Acquisition and Processing*, 2015, 30(1): 202-210.

作者简介:



温俊芹 (1977-), 女, 副教授, 研究方向: 数字图像处理、模式识别, E-mail: wenjunqin@zjtie. edu. cn.



王修晖 (1978-), 男, 博士, 副教授, 研究方向: 计算机图形学、计算机视觉和模式识别。