

# 基于马尔可夫模型的手势识别算法

齐邦强 齐苏敏

(曲阜师范大学 计算机科学学院, 山东 曲阜 273165)

**摘 要** 同时应用手形及其运动轨迹两大特征实现动态手势识别。在轮廓跟踪过程中, 获得手部轮廓, 同时用轮廓质心的坐标表示手的位置获取手势的运动轨迹, 则可得动态手势的特征向量, 即观察值序列; 然后采用左右结构的带有 4 个状态的离散马尔可夫模型实现手势识别。试验结果表明算法提高了手势识别率, 也达到了实时性效果。

**关键词** 手形; 运动轨迹; 手势识别; 马尔可夫模型

## 1 引言

基于隐马尔可夫模型的动态手势识别一般是应用手势的时间特性进行识别<sup>[1]</sup>。单一的手势可被认为是不同手形序列组成的, 多个手势可由手形及其运动轨迹加以区分。本文同时应用手形及其运动轨迹两大特征进行手势识别, 提高了手势识别率, 也达到了实时性的效果。在轮廓跟踪过程中, 获得当前的手部轮廓, 同时用轮廓质心的坐标代表当前手的位置获取手势的运动轨迹, 这样就可从图像序列中得到动态手势的特征向量, 即观察值序列, 记作  $O_l, l=1, \dots, L$ , 其中  $L$  为序列总数。本文采用左右结构的带有 4 个状态的离散 HMM 模型用于手势识别, 其输入为  $O_l, l=1, \dots, L$ 。

## 2 特征向量的提取

在本文算法中, 手势特征向量分为手形与手势运动轨迹两类, 下面我们分别介绍这两类特征向量的提取方法。

手形特征主要为手部轮廓特征, 可在轮廓跟踪过程中<sup>[2]</sup>获得。但作为离散 HMM 的观察值, 手形需分类量化。我们首先定义几种手形, 记为  $Y^k = \{y_1^k, y_2^k, \dots, y_{m_k}^k\}, k=1, \dots, K$ ,  $K$  为手形总数,  $m_k$  为表示轮廓  $Y^k$  的像素总数, 图 1 表示算法中的两种手形样本: 开手形(图 1(a))与闭手形(图 1(b))。设当前提取手形为  $X^l = \{x_1^l, x_2^l, \dots, x_{m_l}^l\}$ , 则可采用 Chamfer 距离模板匹配方法加以分类识别。手形  $X^l = \{x_1^l, x_2^l, \dots, x_{m_l}^l\}$  与  $Y^k = \{y_1^k, y_2^k, \dots, y_{m_k}^k\}$  的 Chamfer 距离如式(1)所示。

$$\rho(Y^k, X^l) = \frac{1}{m^l} \sum_i d_{y^k}(x_i^l) + \frac{1}{m^k} \sum_i d_{x^l}(y_i^k) \quad (1)$$

其中测度  $d_F(y)$  表示点  $y$  到轮廓  $F$  的最近距离。

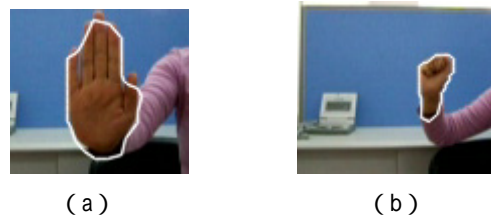


图 1 手形图

手势轨迹的基本特征是位置、速度和角度。由于同一个用户, 手势速率的变化也是比较大的, 因此, 本文采用位置和角度相结合的方法获取运动轨迹。通过基于卡尔曼滤波器<sup>[2]</sup>与自适应形状模型的轮廓跟踪算法<sup>[3]</sup>, 我们可得滤波器最优估计值  $\hat{x}_k = (x_c^t, y_c^t, x_c^t - x_c^{t-1}, y_c^t - y_c^{t-1})^T$ , 其中  $(x_c^t, y_c^t)$  为  $t$  时刻的轮廓质心,  $(x_c^{t-1}, y_c^{t-1})$  为  $t-1$  时刻的轮廓质心, 由此可得手势轨迹的角度, 如图 2(a) 所示。

$$\phi^t = \tan^{-1} \left( \frac{y_c^t - y_c^{t-1}}{x_c^t - x_c^{t-1}} \right) \quad (2)$$

根据所得  $\phi^t$ , 我们采用 12 方向链码进行离散化, 如图 2(b) 所示<sup>[4]</sup>。

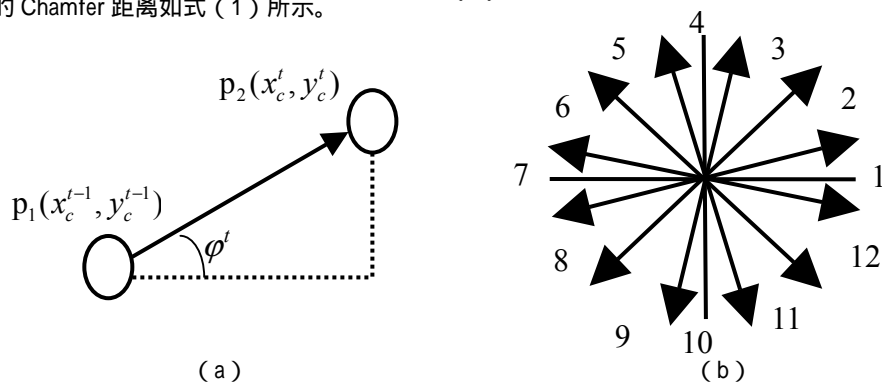


图 2 12 方向链码提取特征向量

在轮廓跟踪过程中,跟踪器根据上述方法按固定时间间隔从图像序列中提取手势的特征向量,可得到手势的特征观察值序列  $O_l, l=1, \dots, L$ 。

### 3 训练手势的隐马尔可夫模型

在获取手势的特征观察值序列后,则可对 HMM 模型进行训练。首先要对 HMM 模型  $\lambda = (A, B, \pi)$  初始化,由于我们采用左右结构的离散 HMM 模型,因此我们设定状态  $i$  只能返回到本身或转移到  $i+1$  状态,状态转移矩阵  $A$  可初始化为<sup>[4]</sup>:

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 & \dots & 0 & 0 \\ 0 & a_{22} & a_{23} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & a_{N-1, N-1} & a_{N-1, N} \\ 0 & 0 & \dots & \dots & 0 & a_{NN} \end{pmatrix};$$

对于初始概率分布,设定:

$$\begin{aligned} \pi_1 &= 1, \\ \pi_i &= 0 (i \neq 1); \end{aligned}$$

即假设 HMM 是从第一个状态开始的。对于观察概率矩阵  $B$ , 设

$$b_{ik} = \frac{1}{M}, 1 \leq i \leq N, 1 \leq k \leq M。$$

初始模型确定以后,利用前后向递推法计算观察值序列  $O$  在这个模型下的  $P(O/\lambda)$ ; 然后利用 Baum-Welch 算法对初始 HMM 模型  $\lambda = (A, B, \pi)$  进行重新计算, HMM 的各个参数在这一过程得到重新估计,即求得新的 HMM 模型  $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ ; 再利用 Viterbi 算法计算出观察序列  $O$  在这个模型下的  $P(O/\hat{\lambda})$ 。为了估计出最接近观察值序列  $O$  的 HMM 模型,我们设定阈值  $\varepsilon$ , 当  $|P(O/\hat{\lambda}) - P(O/\lambda)| < \varepsilon$  时即得到训练出的 HMM 模型; 否则令  $\lambda = \hat{\lambda}$ 。重复此步骤,直到  $P(O/\lambda)$  收敛,得到最接近观察值的 HMM 模型。

### 4 基于隐马尔可夫模型的识别过程

基于 HMM 模型进行手势识别研究,大多采用计算某测试样本与参考模型集的似然度,通过判断最大值以确定测试样本的类别<sup>[5]</sup>。基于这一原理,采用 HMM 实现动态手势识别, HMM 的输入为跟踪器提取的特征观察值序列  $O_l, l=1, \dots, L$ 。图像序列中手势识别过程可归纳为(流程图如图 3 所示):

(1) 检测手部位置初始化跟踪器,跟踪算法如文献[2]所述。

(2) 在执行手势的图像序列中提取手部轮廓并按固定时间间隔获得手形与手势运动轨迹两类特征向量。

(3) 由(2)可得特征观察值序列  $O_l$ , 将其作为训练所得 HMM 模型的输入值,并应用前后向递推算法求其似然值,判断出最大可能手势。

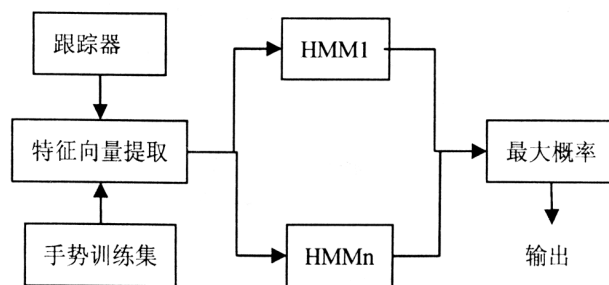


图 3 HMM 手势识别流程图

### 5 实验结果及讨论

在实验中,我们选择了六种不同的动态手势,这些手势用于实现远程控制的六种不同功能,可用指令描述。手势指令为人和机器之间的一种自然接口<sup>[4]</sup>,当用户做出手势动作后,计算机将识别结果发送到远程接口,控制系统作出相应的反应。这六种手势指令分别为<sup>[5]</sup>:

- (1) 向前
- (2) 向后
- (3) 向前然后向右
- (4) 向前然后向左
- (5) 向后然后向右
- (6) 向后然后向左

这六种指令对应的手势动作分别为:

(1) 闭手形向前至开手形:起始时为闭手形,向前靠近摄像机,然后停止,手形由闭至开。

(2) 开手形向后至闭手形:起始时为开手形,向后远离摄像机,然后停止,手形由开至闭。

(3) 闭手形向右至开手形:起始时为闭手形,然后向右移动、停止,手形由闭至开。

(4) 闭手形向左至开手形:起始时为闭手形,然后向左移动、停止,手形由闭至开。

(5) 开手形向右至闭手形:起始时为开手形,然后向右移动、停止,手形由开至闭。

(6) 开手形向左至闭手形:起始时为开手形,然后向左移动、停止,手形由开至闭。

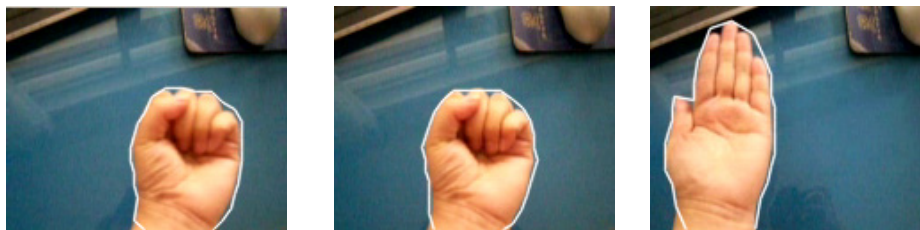
从上述可知,这六种手势具有相似性,我们选取三种手势分析它们的执行过程。图 4、图 5 与图 6 显示了含有手势指令(1)(4)与(6)图像序列,分别选取了三帧查看其执行过程。在含有手势动作的序列图像中,跟踪与识别系统需要解决两个问题:

- (1) 确定何时手势开始。
- (2) 确定何时手势结束。

对于第一个问题,我们的解决方法为:若在跟踪窗口内有超过 75% 的肤色区域,则表示手势的开始,否则跟踪器忽略该帧图像。对于第二个问题,以 10 帧为一个周期,手部区域变化保持为  $\pm 10\%$ , 其变化大于 20% 则表示手势结束<sup>[6]</sup>。

实验中我们选取四个用户执行上述六种手势动作, 摄取序列图像用于手势跟踪与识别。我们分别就手部轮廓跟踪、静态手形识别与动态手势识别三类实验结果做出分析与讨论。

### 1) 跟踪结果

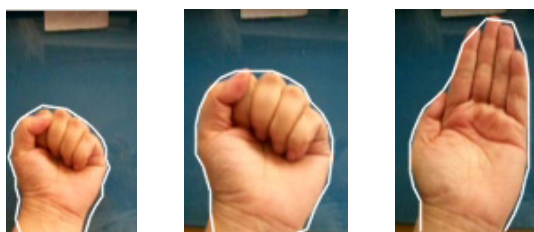


(a) 第 10 帧

(b) 第 30 帧

(c) 第 80 帧

图 4 手势指令向前然后向左的跟踪结果

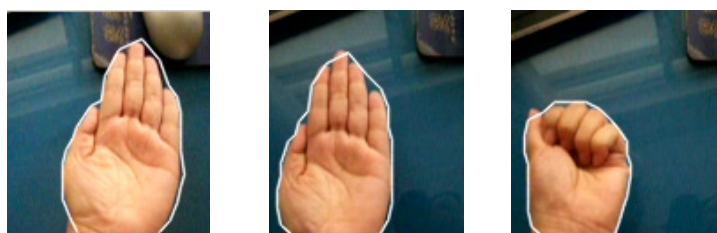


(a) 第 1 帧

(b) 第 40 帧

(c) 第 74 帧

图 5 手势指令向前的跟踪结果



(a) 第 1 帧

(b) 第 40 帧

(c) 第 65 帧

图 6 手势指令向后然后向左的跟踪结果

### 2) 静态手形识别结果

算法实现两种静态手势: 开手形(图 1(a)), 闭手形(图 1(b))的识别。从多个序列中进行测试, 其中开手形的识别率能够达到 97%以上, 闭手形的识别达到 90%以上。从识别效果上看, 当手腕的露出部分较多时, 跟踪器将提取不到准确手形, 从而导致识别率降低。因此, 去除手部之处的肤色区域或类肤色区域是识别算法的一个研究重点。

### 3) 基于隐马尔科夫模型的手势识别结果

对于手势的图像序列, 算法采用每 10 帧提取特征向量(静态手形, 运动轨迹), 得到特征观察值序列, 采用左右结构的带有 4 个状态的离散 HMM 模型。实验中选取四个用户分别执行六种手势动作, 摄取相应的图像序列。对所得的图像序列基于本章算法训练与测试 HMM 模型, 表 1 列出了实验训练次数与测试次数, 表 2 列出了手势识别的正确率, 表中数据分母数值表示测试次数, 分子数值代表了识别正确的

图 4、图 5 与图 6 分别显示了手势指令向前然后向左、向前与向后然后向左的图像序列跟踪手形结果。我们分别从这三个序列选取三帧查看跟踪效果, 从图可以看出算法能够较好地实现手部轮廓跟踪。

次数。从表中可以看出, 手势识别率较高, 能达到 80%以上, 基本满足人机交互的需要。

表 1 手势识别实验训练次数及测试

手势命令	训练次数	测试次数
命令 1	100	40
命令 2	120	40
命令 3	130	60
命令 4	140	60
命令 5	130	60
命令 6	140	60

表2 手势识别实验识别率

手势 用户	命令 1	命令 2	命令 3	命令 4	命令 5	命令 6
1	8/10	7/10	18/20	17/20	17/20	19/20
2	9/10	9/10	17/20	19/20	18/20	17/20
3	9/10	8/10	9/10	6/10	7/10	8/10
4	7/10	8/10	8/10	8/10	9/10	9/10

## 6 结束语

本文同时应用手形及手部运动轨迹两大特征作为 HMM 特征观察值实现了动态手势识别。试验中实现了用于远程控制的六种手势识别,结果表明算法提高了手势识别率,同时也达到了实时性效果。

## 参考文献

[1]段洪伟. 基于视觉的手势识别及其应用 [D]. 上海:上海大学, 2004

[2]NUMMIARO K, KOLLER - MEIER E, VAN GOOL L. An adaptive color-based particle filter [J]. Image and Vision Computing, 2002, 21 ( 1 ): 99 - 110

[3]齐苏敏, 黄贤武, 孟静. 基于自适应活动轮廓模型的实时手势跟踪 [J]. 计算机科学, 2006, 33 ( 11 ): 192 - 194.

[4]孔晓明. 人机交互系统中动态手势识别的研究 [D]. 上海:上海大学, 2005

[5]FREEMAN W, WEISSMAN C. Television control by hand gestures [C]. Proc IEEE Int'l Conf. on Automated Face and Gesture Recognition ( FG'95 ), 1995 : 179 - 183

[6]RAMAMOORTHY A, VASWANI N, etl.all. Recognition of dynamic hand gestures [J]. Pattern Recognition, 36 ( 9 ): 2069 - 2081

收稿日期: 3 月 20 日 修改日期: 4 月 14 日

作者简介: 齐邦强 ( 1970 - ), 男, 山东曲阜人, 硕士, 副教授, 主要研究方向: 单片机与机器视觉; 齐苏敏 ( 1976 - ), 女, 山东曲阜, 博士, 副教授, 主要研究方向: 视频跟踪与机器视觉。

(上接第 41 页)

就是大量动态链接库的维护问题。如何保证可读性、互斥性、版本的可控性等是至关重要的。这些维护开销都需要考虑到系统开发过程中。

### 1) 可读性

制定统一的命名规则,使每一个动态链接库都有唯一的标识,命名时尽量选用能说明该动态链接库主要用途的词汇。

### 2) 互斥性

应该尽量保证各个动态链接库的功能没有交叉,即库功能的原子性,这样可以避免上层程序使用时不够明确,带来维护上的困难。另外如果功能交叉,会存在相同功能代码的不一致性问题。比如,两个库函数有相似功能的函数,一方面很难确定哪一个库函数功能更完备,另一方面很难确定升级时应该使用哪个库作为升级对象。

### 3) 版本可控性

严格管理各个动态链接库的版本。按照设计,动态链接库的数量将非常庞大,明确每个动态链接库的版本是非常困难的,尤其是在升级不受限制的情况下。所以应该配置相应的数据库,以精确管理动态链接库的各种版本状态,以保证信息系统中各个模块功能的正确性。

## 7 总结

将复杂信息系统分层设计有利于分解任务,降低整个系统的复杂度,提高资源的利用率。使用动态链接库封装信息系统各层各模块功能,有助于数据共享,代码共享,提高了系统的运行效率,增强了系统性能。而且,各个动态链接库可独立开发,有效促进了分工合作,大大提高了信息系统开发的效率和准确性。总之,基于动态链接设计开发复杂信息系统,充分发挥了动态链接库的优势。即便如此,对于大型的复杂信息系统的开发,应用这种方式还需做更多的实践尝试,以求更好的完成任务目标。

## 参考文献

[1] Charles Petzold. Windows Programming Fifth Edition[M]. 北京博彦科技发展有限公司译. 北京:北京大学出版社, 1999

[2] 文欣秀. 基于动态链接库实现软件界面组件化方法研究. 计算机应用与软件. 2007(7):18 - 20

收稿日期: 2 月 26 日 修改日期: 3 月 26 日

作者简介: 刘立辉 ( 1981 - ), 男, 硕士, 助理工程师, 从事系统软件设计与开发。孟庆鑫 ( 1980 - ), 男, 硕士, 助理工程师, 从事系统软件设计与开发。