

# 连续隐马尔科夫的静态手势识别法<sup>①</sup>

吴彩芳<sup>1</sup>, 谢 钧<sup>1</sup>, 俞 璐<sup>2</sup>, 周开店<sup>1</sup>

<sup>1</sup>(解放军理工大学 指挥信息系统学院, 南京 210007)

<sup>2</sup>(解放军理工大学 通信工程学院, 南京 210007)

**摘要:** 首次将连续型隐马尔科夫模型应用于静态手势识别中, 根据该方法的特点, 选用手型轮廓像素点坐标值序列作为静态手势的数据特征. 采用微软公司的 Kinect 体感设备提取并追踪手势, 为几种常用的静态手势训练 HMM 模型库, 并使用该模型库进行静态手势识别实验. 实验将该方法与使用 SVM 方法进行对比, 结果表明这种方法的识别率高, 训练模型所需样本少, 简单灵活.

**关键词:** 静态手势识别; 连续; 隐马尔科夫模型; Kinect; 手型轮廓点

## Static Gesture Recognition Arithmetic Based on CHMM

WU Cai-Fang<sup>1</sup>, XIE Jun<sup>1</sup>, YU Lu<sup>2</sup>, ZHOU Kai-Dian<sup>1</sup>

<sup>1</sup>(School of Command Information System, PLA University of Science and Technology, Nanjing 210007, China)

<sup>2</sup>(School of Communication Engineering, PLA University of Science and Technology, Nanjing 210007, China)

**Abstract:** In the paper, the continuous hidden Markov model is applied in static gesture recognition for the first time. According to the characteristics of CHMM, the pixel coordinates sequence of hand contour is chosen as a learning sample. It extracts and tracks the hand gesture by Kinect, and trains CHMM model library which is used for static gesture recognition. At last, the method is compared with the SVM method in an experiment. Experimental result shows the method is efficient, flexible and need minority samples.

**Key words:** static gesture recognition; continuous; Hidden Markov Model; Kinect; the pixel coordinates sequence of and contour

近年来, 基于计算机视觉的手势识别的人机交互技术正在成为模式识别和人工智能领域研究的热点. 这种交互方式通过摄像机捕获静态手型或者动态手势动作, 并由计算机进行分析理解, 然后完成交互任务. 与传统的人机交互设备——鼠标和键盘相比, 手势识别的人机交互方式使人与计算机之间的信息传递方式更加符合人自身的习惯, 摆脱了这些设备对人的束缚, 从而使人机交互的舒适度与效率均得到提高.

手势识别可以分为静态手势识别和动态手势识别两种, 静态手势是指手摆出的某一个手的形态, 现实生活中人们通常根据伸出的手指数、手指名称、手指朝向来判断手势含义, 并且大部分的手势的含义不受手的翻转、平移影响; 动态手势是由手移动、翻转或

手型变化形成的, 人们通常根据手势移动的轨迹、速度、方向或手型变化前后的状态进行识别. 使用动态手势进行人机交互需要用户手的位置和形状进行有规律的变化, 适用于大型界面演示的交互场景; 与动态手势相比, 使用静态手势进行人机交互只需用户在摄像头前用手摆出一个姿势即可, 手无需移动, 简单方便, 用户体验的舒适度较高. 因此, 本文主要针对静态手势识别进行研究.

传统的手势识别获取数据的方法有基于数据手套和基于计算机视觉信息两种. 使用数据手套的方法虽然精度高, 但是设备昂贵, 使用舒适感欠佳, 难以推广, 因此现在一般都采用基于计算机视觉的方法获取数据<sup>[1]</sup>. 常用的静态手势识别的方法主要有模板匹配

① 基金项目: 国家高技术研究发展计划(863)(2012aa01a509, 2012aa01a510); 国家自然科学基金(61101202)

收稿时间: 2015-10-30; 收到修改稿时间: 2016-01-11 [doi:10.15888/j.cnki.csa.005266]

法、构造分类器或使用支持向量机法等<sup>[2]</sup>。隐马尔科夫模型法是一种很成熟的匹配时变数据的数学方法,适用于特征数据是一组序列的情况,该方法包括训练隐马尔科夫模型和分类识别两个阶段,过去常将其用于动态手势识别中<sup>[3]</sup>,例如文献[4]中以动态手势轨迹的方向角序列作为动态手势的特征,应用隐马尔科夫模型对动态手势进行识别。使用隐马尔科夫进行动态手势识别有几个难点不易解决:如何分割连续的几个动态手势、如何区别有意识的动态手势和无意识的动态手势,例如连续做几个向右挥动的手势,其中手返回的动作是无意识的动作,易被识别成有意识的向左挥动手势,从而产生误差。而静态手势识别不涉及这些问题。本文首次将隐马尔科夫模型方法用于静态手势识别,以静态手型轮廓像素点的坐标值序列作为手势特征,训练HMM模型,并用训练好的HMM模型识别不同的静态手势。最后,对五种常用手势分别使用隐马尔科夫模型法和SVM的方法进行识别,结果表明,使用隐马尔科夫法进行静态手势识别训练模型需要的样本数少,识别率高。

## 1 数据获取及预处理

本文使用微软的Kinect体感设备进行静态手势识别的研究。Kinect传感器有两个摄像头和一个投影机,分别是彩色摄像头、深度摄像头和红外投影机<sup>[5]</sup>。红外投影机发出激光,遇到物体反射回来被深度摄像头接收,生成景深图像,并且追踪和提取出了视野范围内人的骨骼信息<sup>[1]</sup>。实验所使用的Kinect2.0是微软的第二代Kinect体感设备,深度摄像头的可视范围是距离摄像头平面的0.5m到8m之间的范围,能够检测到的人体骨骼节点有25个,其中与实验相关的节点有腕关节、掌指关节、食指指尖点和拇指指尖点。深度摄像头每秒生成30帧景深图像,景深图像中每个像素点上的数值代表物体与Kinect摄像头的距离,如果值为0,则表示该点处在视野范围内没有发现任何物体。Kinect追踪并提取的人体骨骼关节点坐标信息中的Z坐标值是节点距离摄像头的距离。通过Kinect获取的人体骨骼关节点信息可以定位到手的位置,根据骨骼关节点之间深度值的关系设定深度阈值,可以提取出手部像素点。

静态手势数据获取及预处理包括提取手部像素点、二值图去噪、提取手型轮廓点并归一化处理几个

部分。具体过程如下:

1) 提取手部像素点:实验者位于Kinect体感设备正前方的可视范围内,面向摄像头伸出手,结合Kinect获取的深度数据和骨骼数据,提取镜头中手部关节点附近、深度值小于实验者伸出手的腕关节深度值的所有像素点,这些像素点即为手部像素点。生成一个二值图,图上手部像素点的值为1、其他点值为0。

2) 二值图去噪:提取二值图中的连通分量,使像素点个数小于100的连通分量被周围像素点多的大连通分量吞并,即像素点个数小于100的小连通分量像素点的值置为包围它的大连通分量像素点的值,返回生成的二值图。

3) 提取手型轮廓点并归一化处理:去掉手部像素点上下左右的多余的空白区域,使原二值图剩下仅包含手型的二值图(手型四周留一圈像素的空间以保证轮廓曲线能够闭合)。使用文献[6]中的手部轮廓跟踪算法找到手型轮廓。设手型轮廓上的像素点横坐标为 $X1$ ,纵坐标为 $Y1$ ,其中最左侧点横坐标值为 $minX$ ,最右侧点横坐标值为 $maxX$ ,最上侧纵坐标值为 $maxY$ ,最下侧纵坐标值为 $minY$ ,手型轮廓像素点坐标值通过下面所示的公式进行尺度归一化,以消除图像大小对识别结果的影响。

$$X2 = \frac{X1 - minX}{maxX - minX}$$

$$Y2 = \frac{Y1 - minY}{maxY - minY}$$

4) 获取手势特征:手型轮廓像素点坐标值归一化后,以手型轮廓左上方为起点,顺时针的顺序依次获取手型轮廓上像素点的坐标值,形成序列,该序列就是手势的特征数据。

## 2 基于连续隐马尔科夫的手势识别

### 2.1 CHMM 概述

隐马尔科夫模型是一个统计模型,它可同时对空间特性和时间特性及其相关关系建立模型。HMM的基本思想是用双重随机过程来描述一个模式,一个随机过程是Markov链,描述模式内部状态变化的序列;另一个随机过程是描述状态和观测值之间的关系。一个隐马尔科夫模型可以定义做:

$$\lambda = (N, M, \pi, A, B)$$

其中:  $N$  为模型中 Markov 链的状态数目,  $M$  为各状态对应的观测值的数目,  $\pi$  为起始状态概率分布矢量,  $A$  为状态转移概率矩阵,  $B$  为某个时刻因隐藏状态而可观察的状态的概率, 即输出概率. 根据其输出观测值的不同, 分为离散隐马尔科夫模型(DHMM)和连续隐马尔科夫模型(CHMM), DHMM 模型的观测变量为离散型, 称为观测矢量, 其输出概率参数  $B$  是一个概率矩阵; CHMM 模型的观测变量为连续型变量, 其输出概率参数  $B$  是一个概率密度函数<sup>[7]</sup>. 考虑到手势特征选取的是顺时针取手型一圈轮廓的像素点坐标序列, 具有连续性, 所以实验中选取了连续隐马尔科夫模型.

## 2.2 训练连续隐马尔科夫模型

隐马尔科夫模型的训练是通过给定的观察值序列  $O$  调整模型  $\lambda = (\pi, A, B)$  的参数, 获取使得条件概率  $P(O | \lambda)$  最大的模型  $\lambda$ . 解决这种参数估计问题最经典方法的是 Baum-Welch 算法. 连续隐马尔科夫模型的输出概率  $B$  是一个概率密度函数. 由概率论可知, 如果高斯概率密度函数足够多, 则混合高斯密度可以逼近任意的概率分布函数, 因此使用几个高斯概率密度函数的线性组合可以模拟观测序列的产生, 连续隐马尔科夫模型的输出概率密度函数  $B$  可用下面的公式表示:

$$b_j(O) = \sum_{l=1}^R c_{jl} G(O; \mu_{jl}, U_{jl}), 1 \leq j \leq N$$

其中,  $O$  是观测值向量,  $G$  表示正态高斯概率密度函数,  $\mu_{jl}$ ,  $U_{jl}$  分别为第  $j$  个状态第  $l$  个混合高斯元均值和方差,  $c_{jl}$  为第  $j$  个状态第  $l$  个混合高斯元的权值(混合系数),  $R$  为每个状态包含的混合高斯元个数,  $N$  是 HMM 的状态数<sup>[8]</sup>. 因此, 训练连续隐马尔科夫模型的目的是获取以下参数的值: 初始概率矩阵  $\pi$ , 状态转移矩阵  $A$ , 混合高斯元的权值矩阵  $C$ 、均值矩阵  $\mu$ 、方差矩阵  $U$ . Baum-Welch 算法的具体过程如下所述.

### 2.2.1 初始化连续隐马尔科夫的参数

在进行隐马尔科夫模型训练之前, 需要对模型参数赋初值, 不同的参数初值会训练出不同的隐马尔科夫模型, 因此参数初值的选取直接影响手势识别结果. 本文实验采用下述的方法初始化模型参数.

设手势  $X$  的隐状态数为  $a$ , 每个隐状态的高斯元的个数为  $b$ :

初始化混合高斯元的权值矩阵、均值矩阵、方差矩阵: 将手势  $X$  的所有样本的像素点根据坐标进行  $k$

均值聚类, 将它们聚类成  $a$  个类, 记录所有像素点的类标号和每类质心点坐标, 每一类的像素点对应一个隐状态; 再将每类像素点聚类成  $b$  个部分, 每个部分的像素点对应一个高斯元. 根据样本中属于隐状态  $i$  的第  $j$  部分像素点的个数占属于隐状态  $i$  的像素点的个数的百分比估计隐状态  $i$  第  $j$  个高斯元的权值. 用  $a \times b$  的矩阵记录估计的权值矩阵. 将所有属于隐状态  $i$  第  $j$  个部分的像素点的质心坐标估计为隐状态  $i$  第  $j$  个高斯元的均值, 用  $a \times b$  的矩阵记录均值矩阵. 同理用统计样本初始状况的方法统计方差矩阵.

初始化初始概率矩阵: 取手势  $X$  的所有样本数据的第一个像素点, 根据这些像素点分布在隐状态  $i$  中的个数占手势  $X$  的样本总数的百分比估计初始概率. 用  $a \times 1$  的矩阵记录估计的初始概率矩阵.

初始化状态转移矩阵: 根据统计手势  $X$  的所有样本中由属于隐状态  $i$  的像素点转移到属于隐状态  $t$  的像素点的次数占样本中属于隐状态  $i$  的像素点个数的百分比, 来估计状态转移的概率. 用  $a \times a$  的矩阵记录估计的转移矩阵.

### 2.2.2 重估模型参数

这一部分的核心思想是通过递归方式更新权重, 以得到能够更好的解释训练样本序列的模型参数. 定义前向  $\alpha_i(t)$  为使系统在  $t$  时刻位于状态  $i$ , 并且已经产生了到  $t$  时刻为止的目标序列的概率:

$$\alpha_i(t) = \begin{cases} 0 & t=0 \text{ 且 } j \neq \text{初始状态} \\ 1 & t=0 \text{ 且 } j = \text{初始状态} \\ \sum_{i'} \alpha_{i'}(t-1) a_{ij} b_j(o_t) & \text{其他} \end{cases} \quad (1)$$

定义后向变量  $\beta_i(t)$  为  $t$  时刻位于状态  $i$ , 并且将产生  $t$  时刻之后的目标序列(时间范围为从  $t+1$  到  $T$ )的概率:

$$\beta_i(t) = \begin{cases} 0 & t=T \text{ 且 } i \neq \text{初始状态} \\ 1 & t=T \text{ 且 } i = \text{初始状态} \\ \sum_{i'} \beta_{i'}(t+1) a_{ij} b_j(o_{t+1}) & \text{其他} \end{cases} \quad (2)$$

根据前向变量和后向变量的定义, 显然可以得到下面这个等式:

$$P(O|\lambda) = \sum_{i=1}^a \alpha_i(1) a_{ij} b_j(o_{t+1}) \beta_j(t+1) \quad (1 \leq t \leq T-1) \quad (3)$$

定义变量  $\gamma_i(i)$  为在给定观察序列  $O$  和 HMM 模

型参数  $\lambda$  的条件下, 在  $t$  时刻处于状态  $i$  的概率, 并且满足  $\sum_{i=1}^N \gamma_t(i) = 1$ ,  $\gamma_t(i)$  用前向变量和后向变量表示为:

$$\gamma_t(i) = \frac{\alpha_t(i) \cdot \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \cdot \beta_t(i)}, (1 \leq i \leq N) \quad (4)$$

定义变量  $\xi_t(i, j)$  表示在给定观察序列  $O$  和 HMM 模型参数  $\lambda$  的条件下, 在  $t$  时刻处于状态  $i$  并且  $t+1$  时刻处于状态  $j$  的概率,  $\xi_t(i, j)$  用前向变量和后向变量表示为:

$$\xi_t(i, j) = \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_t(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_t(j)} \quad (5)$$

根据  $\gamma_t(i)$  和  $\xi_t(i, j)$  所代表的意义, 得出二者的关系为:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (6)$$

根据上述几个式子可以推导出参数重估公式如下:

$$\begin{aligned} \bar{\pi}_i &= \gamma_1(i) \\ \bar{a}_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{i=1}^{T-1} \gamma_t(i)} \\ \bar{b}_j(k) &= \frac{\sum_{t=1}^{T-1} \gamma_t(j)}{\sum_{i=1}^{T-1} \gamma_t(j)} \end{aligned} \quad (7)$$

### 2.2.3 概率收敛, 得到估计结果

使用上述方法计算观察序列  $O$  在模型  $\lambda$  下的概率  $P(O|\lambda)$  以及在  $\bar{\lambda}$  下的概率  $P(O|\bar{\lambda})$ , 并计算  $|\log P(O|\bar{\lambda}) - \log P(O|\lambda)|$ , 若满足  $|\log P(O|\bar{\lambda}) - \log P(O|\lambda)| < \varepsilon$  ( $\varepsilon$  为设定的收敛门限值), 则  $P(O|\lambda)$  收敛, 此时的  $\bar{\lambda}$  即为训练得到的最接近手势样本的 HMM. 否则令  $\lambda = \bar{\lambda}$ , 继续执行 2.2.2 中步骤, 直至  $P(O|\lambda)$  收敛.

### 2.3 手势识别

手势识别是一个隐马尔科夫模型的估值问题. 估值问题是在已知确定的隐马尔科夫模型  $\lambda = (\pi, A, B)$  和给定观测序列  $O$  的前提下, 求该 HMM 模型产生出观测序列  $O$  的概率  $P(O|\lambda)$ . 解码问题常用的方法是前向算法. 具体的应用原理是: 在已知多个确定的隐马尔科夫模型  $\lambda_1, \lambda_2, \dots$  的条件下, 计算给定的观测序列  $O$  的条件概率  $P(O|\lambda_i)$  ( $i=1, 2, \dots$ ) 的值. 取  $\lambda^* = \arg \max_i P(O|\lambda_i)$  为识别结果, 即为最符合该观测序列的隐马尔科夫模型<sup>[9]</sup>.

手势识别时首先获取手型数据特征序列, 分别求该手势数据特征序列在手势模型库中的每个 HMM 模型下的条件概率  $P(O|\lambda)$ , 使该条件概率最大的 HMM 模型对应的手势就是识别的结果.

### 3 实验过程及结果

本文对常见的十种手势进行识别实验. 每种手势由随机选取的 10 个人每人做 26 次, 即每个手势有 260 个数据. 实验随机取每种手势的 10 个数据作训练样本, 剩下的数据作为测试样本用于手势识别, 记录识别率. 再依次随机取每个手势的 20 个、30 个、40 个、100 个数据作为训练样本, 剩下的数据作为测试样本做手势识别, 记录识别率, 比较识别效果. 上述实验分别用以下三种方法做, 比较识别率:

- 1) 手型轮廓点坐标序列作为数据特征, 训练 HMM 模型用于手势识别;
- 2) 手势轮廓的 Hu 矩作为数据特征, 训练 SVM 模型用于手势识别;
- 3) 手势轮廓点坐标序列作为数据特征, 训练 SVM 模型用于手势识别.

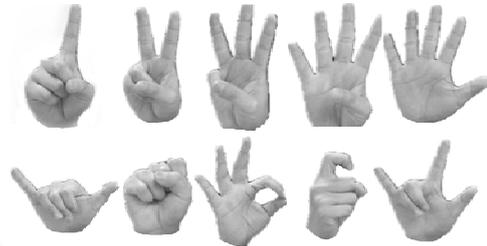


图 1 要识别的几种手势

上述实验的程序使用 C# 语言编程实现, 硬件使用 Kinect2.0 采集数据, 手势识别程序运行在一台拥有物理双核速度达 3.1GHz、64 位处理器以及有专用 USB3.0 总线的 PC 机上, PC 机的操作系统为 windows8.1. 实验结果如下表所示.

每种手势训练样本数

表 1 静态手势识别实验结果(%)

每种手势训练 样本数	10	20	30	40	100
方法					
方法(1)	94.8	96.2	96.9	96.5	99.7
方法(2)	84.8	86.2	87.8	89.9	98.23
方法(3)	59.5	62.3	40.9	46.8	67.2

从实验结果可以看出,使用方法(3)的识别效果最差,几乎不能用于工程中.方法(1)和方法(2)识别率均较高,两者相比而言,方法(1)的识别效果要比方法(2)好.方法(1)和方法(2)的识别率均会随着训练样本数的增多而升高,但是使用方法(1)进行静态手势识别时,仅用 10 个训练样本训练 HMM 模型的识别率就可高达 94.89%,方法(2)在训练样本数比较少的情况下,识别效果不太理想,要想达到较高的识别率需要更多的训练样本.综上所述,基于手型轮廓像素点坐标序列和连续隐马尔科夫模型的静态手势识别方法最优.

SVM 的特征数据的特点是特征维度低,信息量密度大, Hu 矩正符合上述条件;相比而言,手型轮廓像素点坐标序列的数据特征的维度高,信息量分散,不符合训练 SVM 的条件.因此使用轮廓像素点坐标序列训练 SVM 模型进行手势识别的效果明显低于使用 Hu 矩训练 SVM 模型的方法.由于 Hu 矩不具备序列特征,不符合隐马尔科夫模型的使用条件,实验并未尝试使用 Hu 矩作为数据特征训练隐马尔科夫模型的识别实验.

#### 4 结语

隐马尔科夫模型法适用于特征数据是一组序列的情况,本文利用隐马尔科夫模型法这一特点,使用手型轮廓像素点坐标值序列作为手势特征进行静态手势识别.实验表明,这种方法的识别率高,训练模板所用的样本数少,使用起来灵活,方便用户自己设定交互手势及训练模板.在下一步的工作中,我们将结合多种数据特征,使用特征选择的方法选出使得手势区分度最高的几个数据特征,用于训练 HMM 模型和识别手势.将静态手势识别和动态手势识别结合起来的

复杂的手势识别也是一个新的识别方向,要实现复杂手势识别系统,在本文研究的基础上,还需要解决孤立动态手势的识别问题、连续动态手势的分割问题以及静态手势和动态手势组合等问题.

#### 参考文献

- 1 Mitra S, Acharya T. Gesture Recognition: A Survey. IEEE Trans. on Systems, Man, and Cybernetics-Part C: Applications and Reviews, 2007, 37(3): 311-324.
- 2 任海兵,祝远新,徐光祐,林学嵩,张晓平.基于视觉手势识别的研究综述.电子学报,2000(2):118-121.
- 3 Starner TE, Pentland A. Visual recognition of American Sign Language using hidden markov models. Proc. of the Intl. Workshop on Automatic Face- and Gesture-Recognition. Zurich. 1995.
- 4 尹彦敏.基于时空轨迹特征的 HMM 动态手势识别研究[硕士学位论文].青岛:青岛科技大学,2013.
- 5 余淘.Kinect 应用开发实战:用最自然的方式与机器对话.北京:机械工业出版社,2012.
- 6 Li Y. Hand gesture recognition using Kinect. Proc. of 2012 IEEE 3rd International Conference on Software Engineering and Service Science. 2012. 196-199.
- 7 Rabiner LR, Juang BH. An introduction to hidden markov models. IEEE ASSP Magazine. January 1986. 4-16.
- 8 洪文,黄凤岗,苏茜.基于连续隐马尔科夫模型的步态识别.应用科技,2005,32(2).
- 9 Yamato J, Ohya J, Ishii K. Recognizing human action in time-sequential images using hidden markov models. Proc. 1992 IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Press. 1992. 379-385