

# 基于不变矩和支持向量机的手势识别

董立峰, 阮军, 马秋实, 汪雷

(武汉理工大学 信息工程学院, 湖北 武汉 430070)

**摘要:** 选取 Hu 不变矩、手势轮廓的凹陷个数及其周长与面积比为手势识别的主要特征, 采用了基于径向基核的 SVM 分类器进行 0~9 十种手势的识别。实验结果表明, 在背景单一、光照情况良好条件下, 该方法具有很高的识别率, 并且简单快速。

**关键词:** 手势识别; Hu 矩; 支持向量机

中图分类号: TP391

文献标识码: B

文章编号: 1674-7720(2012)06-0032-04

## The gesture identification based on invariant moments and SVM

Dong Lifeng, Ruan Jun, Ma Qiushi, Wang Lei

(College of Information Engineering, Wuhan University of Technology, Wuhan 430070, China)

**Abstract:** This paper selects Hu moment invariants, the sag number of gestures outline and the ratio of its perimeter and area as the main characteristics of gesture recognition, adopts SVM classifier based on the RBF kernel to recognize ten kinds gestures of 0~9. The result of the experiment shows that this method has a high recognition rate under a single background and good illumination condition, and it is more simple and faster.

**Key words:** gesture identification; Hu moments; support vector machine

手势识别<sup>[1]</sup>、人脸识别、指纹识别和唇读识别等都是符合人际交流习惯的备受关注的人机交互技术, 在计算机辅助哑语教学、电影特技处理、医疗研究和智能机器人等各个领域均有广泛的应用。其中, 基于计算机视觉的手势识别因其设备成本低廉、操作自然方便等特点越来越受到人们的追捧, 是手势识别未来的发展趋势。然而, 人手的自由度和弹性、背景、光照以及噪声等因素给手势识别带来了很大困难, 如何克服这些困难以获得更加自然、和谐、完善的手势识别系统是当今人们研究的目标。

手势特征选取和分类器设计一直是影响手势识别效果及速度的两个关键问题。良好的特征应具有可靠性、可区别性、数量少以及独立性强等特点, 比较常用的特征提取方法有 Hu 矩<sup>[2]</sup>、Zernike 矩、傅里叶轮廓矩和小波矩等。其中, Zernike 矩的优势是能够构造任意的高阶矩, 对噪声不敏感, 主要用于图像恢复; 傅里叶描述子具有较好的轮廓描述能力, 但对细节特征非常敏感, 误识率比较高; 小波矩虽然能达到很高的识别率但需首先对图像进行归一化处理, 计算比较复杂。本文的仿真实验在背景单一、光照情况良好的条件下进行, 选择了能同时满足平移、旋转和比例不变性的 Hu 不变矩、手势轮廓的凹陷个数以及其周长与面积比作为主要的特征参数。

目前, 基于视觉的手势识别算法研究主要有通过计算模板的相关系数来进行模板匹配的手势识别、用神经网络方法进行手势识别以及基于 Euclidean 距离的手势识别等。支持向量机<sup>[3]</sup>是基于结构风险最小化的一种新的模式识别方法, 在解决小本、非线性及高维模式识别问题中具有许多特有的优势。本文在对手势轮廓进行 Hu 不变矩特征提取的基础上设计了支持向量机分类器, 对分类器进行训练和测试, 实验结果证明, 系统具有很高的识别率。

### 1 特征的选择

#### 1.1 不变矩理论及参数选择

矩的定义源于概率理论, 是一种重要的数字特征, 同时也具有直观的物理意义。手势图像函数  $f(x, y)$  的  $(p+q)$  阶原点矩 (也称普通矩) 定义为:

$$m_{pq} = \sum_x \sum_y f(x, y) x^p y^q \quad (1)$$

其中,  $x^p$  和  $y^q$  为变换核,  $p, q = 0, 1, 2, \dots$ 。

如果手势图像函数是分段连续且在  $xy$  平面的有限部分中有非零值, 即函数的各阶矩存在, 其  $(p+q)$  阶中心矩可表示为:

$$\mu_{pq} = \sum_x \sum_y ((x-\bar{x})^p (y-\bar{y})^q) f(x,y) \quad (2)$$

其中,  $(\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}})$  表示手势图像的重心坐标,  $p, q = 0, 1, 2, \dots$ 。

为了保持几何矩的尺度不变性, 归一化后的几何矩为:

$$\eta_{pq} = \frac{\mu_{pq}}{(\mu_{00})^\gamma} \quad (3)$$

其中,  $\gamma = \frac{p+q+1}{2}$ ,  $q+p=2, 3, \dots$ 。

根据 Hu 氏理论, 通过各种不同级别的几何矩的数学组合, 可以得到 7 个特征量。当图像移动、旋转和比例大小改变时, 它们具有保持其数值不变的特性, 因此称这 7 个特征量为不变矩特征量。实验中对采集的视频按帧读取静态图像并计算 Hu 矩, 最后取平均值作为标准值。可以看出, 高阶矩不仅计算量大, 而且更容易受到噪声的影响, 因此本文取前 4 个 Hu 矩作为前 4 个特征参数  $M1 \sim M4$ :

$$M1: \phi_1 = \eta_{20} + \eta_{02} \quad (4)$$

$$M2: \phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \quad (5)$$

$$M3: \phi_3 = (\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2 \quad (6)$$

$$M4: \phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \quad (7)$$

手势轮廓的凸包和凹陷可用于描述手势的形状, 由于文中自定义的 10 个手势各具形态, 手指的数目不尽相同, 手势的凸包和凹陷也是不同的, 它们能很好地体现手以及手状态的特征, 因此考虑提取手势的凹陷点数目作为手势的特征向量。其计算方法是首先计算轮廓是否为凸, 再计算其连接凸包的外接多边形并提取出来, 根据外接多边形计算凹陷点个数, 如图 1 所示。本文选取手势轮廓的凹陷点个数作为第 5 个特征参数  $N5$ 。

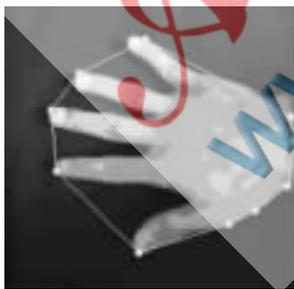


图 1 手势外接多边形与凹陷个数

手势轮廓的周长定义为手势轮廓在矩形区域内包含的像素总数, 通过扫描手势轮廓图中的手势矩形区域, 计算黑色点个数即可算出周长。手势轮廓的面积定义为轮廓部分和起始点连线构成的封闭部分的面积, 其值受轮廓方向的影响有可能为负值。本文构造轮廓的周长与面积之比作为第 6 个特征参数  $N6$ 。

## 1.2 特征值提取

在本文设计的基于计算机视觉的手势识别系统中, 首先以帧为单位对摄像头采集到的手势视频采用了图像平滑、色彩空间转换、图像二值化、边缘检测、轮廓提取以及形态学处理等算法进行图像预处理以获得手势的二值化图像、边缘图像以及轮廓矩阵, 然后分别对每帧手势轮廓进行 Hu 矩、轮廓周长面积比和凹陷个数等的运算, 得到各个手势的特征值。图 2 所示为样本视频进行训练时提取到的 0~9 手势轮廓样本。



图 2 0~9 手势的轮廓图像

表 1 列出了手势轮廓的 Hu 矩前 4 个矩  $M1 \sim M4$ 、轮廓周长面积比  $N5$  和凹陷个数  $N6$  的特征值, 它们都是对学习视频中所有帧数的手势轮廓分别求特征值并计算平均值得到的。这里的识别率为 0~9 每一种手势于视频模式下分别对 5 组测试视频进行检测, 然后对分别获得的识别率求平均值获得的。其中, 9 的识别率偏低是由于其手势轮廓与 3 相近, 容易造成分类时的判断错误。

表 1 0~9 手势的轮廓特征值及视频模式识别率

|   | $M1(10^{-3})$ | $M2(10^{-7})$ | $M3(10^{-8})$ | $M4(10^{-9})$ | $N5$     | $N6$      | 识别率/% |
|---|---------------|---------------|---------------|---------------|----------|-----------|-------|
| 0 | 8.858 24      | 3.441 14      | 1.745 17      | 2.744 34      | 0.000 00 | -1.459 32 | 99.7  |
| 1 | 1.030 03      | 3.589 14      | 1.059 15      | 2.366 06      | 0.926 83 | -2.220 14 | 97.5  |
| 2 | 8.667 07      | 2.613 04      | 2.294 34      | 4.508 27      | 1.500 00 | -3.264 36 | 94.8  |
| 3 | 6.645 62      | 9.241 64      | 2.602 45      | 4.119 83      | 2.239 98 | -4.261 57 | 92.1  |
| 4 | 5.683 14      | 5.149 89      | 3.492 58      | 3.280 92      | 3.029 41 | -5.074 98 | 94.5  |
| 5 | 5.091 29      | 1.060 78      | 1.480 16      | 1.229 60      | 3.945 95 | -5.927 42 | 99.9  |
| 6 | 7.600 06      | 3.021 83      | 1.462 82      | 1.605 16      | 1.068 97 | -2.788 03 | 98.5  |
| 7 | 7.448 05      | 8.132 33      | 4.278 79      | 4.792 51      | 2.000 00 | -2.263 72 | 96.3  |
| 8 | 7.689 81      | 3.805 18      | 8.583 76      | 4.749 34      | 2.356 53 | -3.070 07 | 95.2  |
| 9 | 6.393 92      | 5.286 72      | 2.467 13      | 1.627 29      | 2.956 52 | -3.996 09 | 85.6  |

## 2 支持向量机

### 2.1 基本原理

SVM 是统计学习理论中最年轻的部分, 主要用于解决有限样本情况下的模式识别问题。它的基本思想是在样本空间或特征空间构造出最优超平面, 使超平面与小同类样本集之间的距离最大, 从而达到最大的泛化能力。算法思路如下:

(1) 设训练集为  $\{x_i, y_i\}$ , 其中,  $\{x_i\} \in R^n, y_i \in \{1, -1\}, i = 1, 2, \dots, n$ 。非线性函数  $\phi$  实现从输入空间  $X$  到特征空间  $Z(Z \in R^m)$  的映射, 核函数为:

$$K(x_i, x_j) = (\phi(x_i) \cdot \phi(x_j)) \quad (8)$$

(2) 在特征空间构造最优分类超平面:

$$f(x) = (w \cdot \phi(x)) + b \quad (9)$$

## 图形、图像与多媒体

其中,  $w$  是权重向量, 约束条件为:

$$\max \phi(\alpha) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j a_i a_j (x_i \cdot x_j) - \sum_{j=1}^n a_j \quad (10)$$

$$\sum_{i=1}^n y_i a_i = 0 \quad (11)$$

其中,  $a_i \geq 0, i=1, 2, \dots, n$ 。

得到最优解:

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T \quad (12)$$

(3) 计算:

$$\omega^* = \sum_{i=1}^n y_i a_i^* x_i \quad (13)$$

$$b^* = y_i - \sum_{i=1}^n y_i a_i^* (x_i \cdot x_j) \quad (14)$$

(4) 求得决策函数:

$$f(x) = \text{sgn} \left( \sum_{i=1}^n y_i a_i^* K(x_i, x_j) + b^* \right) \quad (15)$$

## 2.2 核函数选择

根据泛函数的相关理论, 只要一种函数满足 Mercer 条件, 它就对应某一变换空间的内积, 此函数称为内积函数或核函数。核函数的种类有很多, 常用的有以下几种:

(1) 样条核函数

$$K(x_i, x_j) = B_{2n+1}(x_i - x_j) \quad (16)$$

(2) 多项式核函数

$$K(x_i, x_j) = (s(x_i g x_j) + c)^d \quad (17)$$

其中,  $s, c, d$  为参数。

(3) 径向基(RBF)核函数

$$K(x_i, x_j) = \exp(-\gamma \|x - y\|^2) \quad (18)$$

其中,  $\gamma = \frac{1}{2\sigma^2}$  为参数。

(4) Sigmoid 核函数

$$K(x_i, x_j) = \tanh(\rho(x_i g x_j) + b) \quad (19)$$

其中,  $\rho, b$  为参数。

在这 4 种核函数中, 本文选取了 RBF 核函数, 因为在低维、高维、小样本和大样本等情况下, RBF 核函数均适用, 具有较宽的收敛域, 是较为理想的分类依据函数。

## 2.3 参数最优化选择

核函数的参数和误差惩罚因子  $C$  是影响支持向量机性能的关键因素。 $C$  用于控制模型复杂度和逼近误差的折中,  $C$  越大则对数据的逼近误差越小, 同时模型也会越复杂, 支持向量机的推广能力也变得越差。RBF 中的  $\sigma$  用于控制回归逼近误差的大小, 从而控制支持向量机的个数和泛化能力, 其值越大则支持向量数目变少, 同时精度变差; 相反, 则支持向量数目变多, 精度变高。

寻找最佳的  $(C, \gamma)$  参数组使该 SVM 的性能最好, 最简单的方法是分别选取不同的参数组合, 得出不同的错误率, 选取其中错误率最小的参数组合作为最优化选择, 这种方法也叫做“穷举法”。本文采用了台湾林智仁

教授开发的 libsvm 支持向量机库中的工具“grid.py”进行了参数的优选, 该工具是采用网格遍历法, 由 gnuplot 生成的搜索结果显示出最小错误率时的参数, 即最优参数为  $(8.0, 0.5)$ 。

## 3 仿真实验

本文设计的手势识别系统采用了 C 语言、Code::Blocks 开发平台、OpenCV2.3.1<sup>[4]</sup> 计算机视觉开源库和 35 万像素摄像头。该系统主要包括手势跟踪、手势图像预处理、特征提取和手势识别 4 个部分。系统框图如图 3 所示。

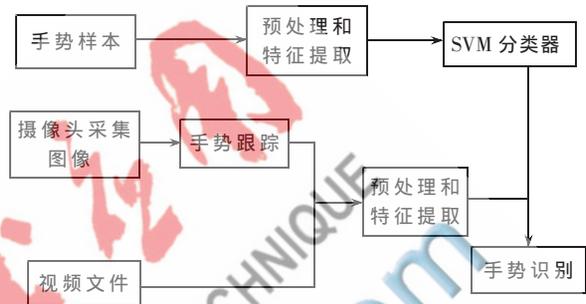


图 3 手势识别系统框图

本系统的程序整体框架分为学习模式和识别模式, 识别模式包括视频文件识别模式和摄像头识别模式。处于学习模式时, 通过读取配置的手势视频文件获取学习模板, 对每一帧图片进行滤波、阈值分割和轮廓提取等预处理得到手势轮廓图像, 获取手势轮廓的 Hu 矩和其他特征参数, 训练特征参数后并保存, 生成相应的判别界面函数。视频文件识别模式通过对每一帧图片进行预处理和特征参数提取后调用 svmPredict 函数进行分类预测, 结果显示在屏幕上。处于摄像头模式时, 首先读取摄像头采集的图像, 设置鼠标事件和界面窗口, 选定跟踪目标的区域和大小, 计算目标区域的色彩直方图和反射投影。本文选取了传统的 Camshift 跟踪算法用于跟踪目标并提取跟踪窗口的目标, 对跟踪目标图片进行预处理并计算特征参数, 最后进行手势识别并显示识别结果。程序运行的最后均要释放资源。

图 4 为摄像头模式下手势识别的截图, 图中实时显



图 4 摄像头模式下对手势的实时识别

示出了手势的外接矩形、外接多边形、手势轮廓、直方图和识别率等信息。

本文开发的基于 Hu 矩和 SVM 的手势识别系统能够在视频模式和摄像头模式下对 0~9 十种手势的实时识别,但光照条件要求相对比较高,人手与摄像头需要保持在一定的距离范围内。目前只选择了 Hu 矩中的  $M1\sim M4$ 、手势轮廓的凹陷个数  $N5$  以及周长与面积的比  $N6$  作为特征参数,如何选取更好的特征参数以提高系统识别性能需要进一步研究。

#### 参考文献

- [1] 任雅祥.基于手势识别的人机交互发展研究[J].计算机工程与设计,2006,27(7):1201-1204.
- [2] FLUSSER J, SUK T, ZITOWA B. Moments and moment invariants in pattern recognitong[M]. Chichester: John Wiley

& Sons Ltd, 2009.

- [3] 罗瑜.支持向量机在机器学习中的应用研究[D].成都:西南交通大学,2007.
- [4] LAGANIÈRE R. OpenCV 2 Computer vision application programming cookbook[M]. Birmingham: Packt Publishing Ltd, 2011.

(收稿日期:2012-01-31)

#### 作者简介:

董立峰,男,1980年生,硕士研究生,主要研究方向:图像处理与智能识别。

阮军,男,1976年生,博士,副教授,主要研究方向:智能系统、信息传输与处理。

