

多尺度特征和神经网络相融合的手写体数字识别

赵元庆¹ 吴 华²

(安阳师范学院计算机与信息工程学院 安阳 455000)¹ (安阳师范学院公共计算机教学部 安阳 455000)²

摘 要 针对传统特征提取方法无法有效解决书写随意性的干扰问题,提出了一种多尺度特征和神经网络相融合的手写体数字识别方法。首先提取手写体数字二值图像的轮廓、笔画次序等结构特征,并旋转坐标轴,提取多角度结构特征;然后将字符从中心点到外边框划分为 K 层矩形子层,提取每层图像的灰度特征,最后以两种多尺度特征构建神经网络模型,并预测测试集合样本。将该算法实际用于以 MNIST 字体库构建的两个数据集识别,其精度高达 99.8%,并能有效降低倾斜等手写字体的随意性影响。

关键词 多尺度,手写体数字识别,多角度结构特征,多层次灰度特征

中图法分类号 TP391 **文献标识码** A

Handwritten Numeral Recognition Based on Multi-scale Features and Neural Network

ZHAO Yuan-qing¹ WU hua²

(School of Computer and Information Engineering, Anyang Normal University, Anyang 455000, China)¹

(Computer Teaching Department of Anyang Normal University, Anyang 455000, China)²

Abstract Aiming at the problem that tradition handwritten numeral recognition method can not solve the interference from writing arbitrary, a new handwritten numeral recognition method was proposed based on multi-scale features and neural network. Firstly, two structural features of outline and strokes were extracted, and multi-angle structural features were extracted by rotating the datum line. Second, Multi-level grayscale pixel features were extracted by dividing the image to K sub-layer from the inside out. Thirdly, BP neural network model was build based on the two features. Lastly, new method was used for The MNIST font library, and the prediction precision reached 99.8%. The result shows that new algorithm can effectively reduce the impact of tilt.

Keywords Multi-scale, Handwritten numeral recognition, Multi-angle structural features, Multi-level grayscale pixel features

1 引言

手写体数字识别是光学字符识别技术(Optical Character Recognition, OCR)的主要研究内容之一。联机的手写数字识别已取得较好的效果^[1],而脱机的手写数字识别精度还不够理想。手写体数字识别系统一般包括图像获取、图像预处理、特征提取、手写体数字识别4个主要部分^[2],其中特征提取更是其核心部分,提取能完整反映字符属性的特征可有效提高识别系统的识别精度。传统基于数字图像结构特征、统计特征的特征提取方法取得了一定的效果^[3-6],但非特定人的手写数字随意性较高,没有固定的属性,并存在倾斜、断笔等问题,当识别数字字符规范性较差时,其识别精度比较低^[7,8]。

针对上述手写体数字字符随意性高的问题,提出了一种基于多角度结构特征与多层次灰度特征的手写体数字识别算法。将该算法应用于 MNIST 字体库手写数字的识别,其精

度达到了 99.8%,且该算法不受手写字体随意性的影响。

2 多尺度和神经网络相融合的手写体数字识别算法

基于多角度结构特征与多层次灰度特征的多尺度手写体数字识别系统主要包括图像预处理、多角度结构特征提取、多尺度结构特征 PCA 整合、多层次灰度特征提取、BP 神经网络建模5个主要部分。该方法首先提取手写体数字的轮廓、笔划等结构特征,然后将坐标轴旋转多个角度,提取其多角度结构特征,并以主成分分析(Principal Component Analysis, PCA)融合多组结构特征,以消除多个角度不匹配而导致特征可能错位的问题。然后提取其多层次灰度特征,将图形从中间到外边框分为 K 层子图,提取各层的灰度特征。最后以两组多尺度特征构建 BP 神经网络预测模型。该方法基于多尺度的思想,提取手写体数字的多角度结构特征与多层次灰度特征,可以有效解决字体倾斜等随意性导致的特征改变问题。其过程见图1。

到稿日期:2012-10-25 返修日期:2013-01-13 本文受国家自然科学基金青年基金项目(41001251)资助。

赵元庆(1976—),男,讲师,主要研究方向为图形图像处理与网络, E-mail: Zyzq7643@163.com; 吴 华(1970—),男,副教授,主要研究方向为多媒体与网络。

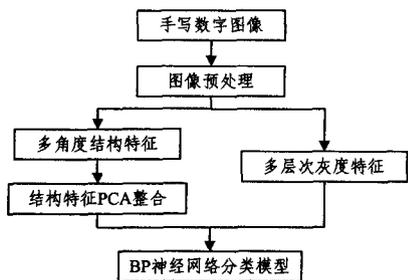


图1 多尺度算法流程图

2.1 图像预处理

小波去噪法能较好地保留图形细节,且更具灵活性。所以采用小波去噪法去除手写数字字符图像中的噪音影响;然后将字符图像二值化,使得后续计算更加方便,即将其变为仅有黑白两色的灰度模式。通过直方图变换法可以将图像快速、高质量地转为二值图像;最后确定字符图像的最小外边框,即图像上、下、左、右最外层的像素点所构成的矩形框。在确定外边框后,将所有字符都归一化为统一的大小,以解决字符大小不一的问题。

2.2 多层次灰度特征提取

手写体数字的灰度图像保留了初始字符的完整信息,提取其灰度特征进行识别,作为结构特征的补充,能有效提高字符的识别准确度。

在灰度图像经过去噪处理后,还需进一步进行字符区域紧致处理,消除图像中存在的杂斑等污染杂质。首先将灰度图片进行二值化,以最大类间方法计算阈值^[9],对图像进行字符区域紧致。

(1)二值字符图像对应的灰度图像作为输入;

(2)寻找二值字符图像中的连通区域,规定最大的连通区域为保留字符区域;

(3)二值字符图像确定保留字符区域后,将对应的灰度图像中非保留区域的全部像素点清除。

对灰度字符图像进行紧致处理后,将图像由内到外分割成 L 个子区块,然后提取各子区块的灰度特征组成多层次灰度特征。字符层次划分见图2。



图2 字符6的多层次划分

2.3 手写体数字结构特征

字符的结构特征包括轮廓特征、笔划特征等,具有提取简单、识别效果优异的优点,一般都是从字符的二值图像中提取得到。

2.3.1 轮廓特征提取

二值图像的轮廓特征能较好地描述字符的边框信息,可分为内、外轮廓特征两种^[10]。内轮廓特征(INL)定义为二值字符内部黑色像素点数,即从第一个黑白交界像素点到第一个黑白交界点之间的像素点数;外轮廓特征(OUL)定义为从图像外周到第一个白色像素点之间的像素数目。在计算二值图像的轮廓特征时,首先将字符图像从行方向分割为 n_H 个子图像,在左、右两个方向都可以得到 n_H 个内轮廓特征与 n_H 个

外轮廓特征,共得到 $4 * n_H$ 个轮廓特征;然后将整个二值图像从列方向分割为 n_K 个子图像,同样,在上下两个方向都可以提取 n_K 个内轮廓特征与 n_K 个外轮廓特征,共得到 $4 * n_H$ 个轮廓特征。则整个字符图像可以提取 $4 * (n_H + n_K)$ 个轮廓特征。

2.3.2 笔划结构特征

轮廓特征仅能表征字符的边框信息,不能反映字符的笔划次序的变化信息。笔划特征的提取方法如下:

(1)对一幅 $N_1 \times N_2$ 的二值字符图片,先从行的方向将整个图片等分为 n_h 个子区间: J_1, J_2, \dots, J_m , 保证每个子区间包含有 N_1/n_h 行像素点。

(2)计算上述平均划分的行区域中黑白或者白黑交界点出现的次数 $j_i (i=1, 2, \dots, n_h)$, 可获得 n_h 维行方向笔画结构特征。

(3)将整个图片从纵向方向均等分割为 n_k 块,与横方向一样,在纵向方向可以计算获得 n_k 维笔画结构特征。那么,对整幅图片,在行列两个方向共可获得 $n_h + n_k$ 维笔画结构特征。

显然,对一个二值字符图像,可以提取内、外轮廓特征和笔画特征共 $4 * (n_H + n_K) + n_h + n_k$ 维的结构特征 \vec{F} 。

2.4 多角度结构特征

当数字字符出现倾斜时,其结构特征差异会随着倾斜角度变大而变大。将坐标轴按每5度旋转,计算一组当前结构特征,直到旋转360度,共可得到72组结构特征。该72组结构特征表征了各角度下的字符结构特征情况,从而有效解决了字符倾斜而导致单组结构特征失效的问题。

2.5 结构特征 PCA 整合

上述多组结构特征表征了各角度下的字符结构特征,但不能确定各组特征的前后顺序,当字符倾斜时,导致其与非倾斜字符的特征无法匹配。通过对多组特征进行PCA整合,以保留主成分代表72组结构特征,解决了特征可能错位的情况,并且能有效降低特征维数,简化模型。

PCA是一种高效的用于特征降维的统计分析方法。它找出一个包含少量互不相关的综合因子的最佳特征子集,用来代替初始的多个特征因子,并使其最大程度地保留初始特征集的信息,从而达到简化初始特征集合的目的,并可以去掉初始特征之间的冗余信息。对于一样本数为 N 、描述特征为 $\vec{x} (i=1, 2, \dots, N)$ 的数据集,其主成分的获得描述如下。

首先获得该数据集各特征的样本均值 m 。

$$\vec{m} = \frac{1}{N} \sum_{i=1}^N \vec{x}_i \quad (1)$$

求得各特征样本均值后,再产生数据集的协方差矩阵。

$$R = \frac{1}{N} \sum_{i=1}^N (\vec{x}_i - \vec{m})(\vec{x}_i - \vec{m})^T \quad (2)$$

然后用雅克比方法求解特征方程 $|R - \lambda I| = 0$ 的 P 个大于0的特征值 $\lambda_1 > \lambda_2 > \dots > \lambda_p$ (经过排序处理),各特征值 λ_j 对应的特征向量为:

$$C^{(j)} = (C_1^{(j)}, C_2^{(j)}, \dots, C_p^{(j)}), j=1, 2, \dots, p \quad (3)$$

各特征向量要满足以下条件:

$$C^{(j)} C^{(k)} = \sum_{q=1}^p C_q^{(j)} C_q^{(k)} = \begin{cases} 1, & j=k \\ 0, & j \neq k \end{cases} \quad (4)$$

将初始特征经特征向量投影后即可得到 p 个主成分 Z_1, Z_2, \dots, Z_p , 当前面 m 个主成分的方差之和占全部总方差的比

例 $a = (\sum_{j=1}^m \lambda_j) / (\sum_{j=1}^p \lambda_j)$ 比较大时,即选择前 m 个主成分为保留主成分,用于后续分析。当 a 大于 0.85 时,前 m 个主成分基本保留了初始特征的信息量,因此可以以 0.85 作为阈值确定 m 值的大小。

对 2.4 节提取的多维结构特征 \vec{F} 进行 PCA 降维处理后,选择方差总和大于 85% 的前 m 个主成分为保留主成分。

2.6 BP 神经网络建模

将上述提取的多层次灰度特征与整合的多角度结构特征同时作为 BP 神经网络输入,构建多尺度手写数字预测模型,并对测试集字符进行预测。

3 结果与分析

3.1 数据来源

用于仿真实验的手写数字来源于 MNIST 字体库。该库收藏了 60000 个训练样本与 10000 个测试样本,各图像全部为 28×28 的图片。本研究基于该库构建了两个仿真数据集。数据集一:从该库中选择各类型字符 1000 个用于训练,500 个用于测试;数据集二:从中选择 1000 个正立的数字作为训练集,250 个正立、250 个倾斜共 500 个数字作为测试。

3.2 参比模型

为检验新方法的预测性能,实验过程中同时构建了 6 个参比模型,分别为:仅基于结构特征与灰度特征的模型;基于结构特征与灰度特征组合的模型;仅基于多角度结构特征的模型;仅基于多层次灰度特征的模型;本研究提出的多尺度预测模型。

3.3 实验结果

各模型对两个手写体数字数据集的预测结果见表 1。表 2 列出了现有部分方法在 MNIST 字体库上的识别精度^[11]。

表 1 手写体数字各模型预测精度

识别率(%)	数据一	数据二	
		正立样本	倾斜样本
单角度结构特征	83.2	96.4	68.0
多角度结构特征	98.6	99.2	98.4
单层灰度特征	79.2	94.4	63.20
多层灰度特征	97.4	96.8	95.4
单尺度组合特征	86.0	99.6	72.4
多尺度组合特征	99.80	100	99.4

表 2 现有部分方法识别精度

识别方法	识别精度
Linear classifier (1-layer NN)	91.60
K-nearest-neighbors, Euclidean (L2)	98.20
Support vector machine	98.60

由以上结果可知,在两个数据集中,基于多尺度组合特征的手写体数字识别算法在所有参比模型中精度最高,在正立样本中达到了 100% 的识别率。对比各模型可以发现:

(1) 对比灰度特征模型与结构特征模型,不论是单尺度结构还是多尺度情况,结构特征模型预测精度都要高于灰度特征模型的预测精度。可见,对于手写数字的识别,结构特征要优于灰度特征。

(2) 比较单尺度模型与对应的多尺度模型,对于第一个数

据集,多角度结构特征模型识别精度达到 98.6%,高出单角度结构特征模型 15.4 个百分点,多层次灰度特征模型识别率同样高于单层次灰度特征模型;在第二个数据集上,对于正立测试样本,单尺度与多尺度差别并不是特别明显,但在预测倾斜测试样本时,两种特征的多尺度模型识别精度均高出单尺度模型 30 个百分点。由此可见,多尺度更能描述数字字符的多种倾斜情况,降低手写随意性的影响。

(3) 不管是单尺度还是多尺度模型,两种特征的组合模型识别性能均高于单一的特征模型。单一的特征并不能完整地描述字符的属性,以两种特征组合可以弥补仅以单一特征表述字符时存在的信息缺失,进一步提高模型识别性能。

(4) 基于多尺度的组合特征模型在所有参比模型中识别精度最高,并优于当前主流方法的识别精度。当仅以正立样本训练,测试样本为正立字符时,识别精度达到 100%,当测试样本为倾斜字符时,识别精度同样达到 99.2%,显然多尺度组合模型能较好地处理手写数字的随意性问题。

结束语 针对手写体数字识别随意性高的问题,基于多尺度思想,提出了一种新的基于多角度结构特征与多层次灰度特征的组合识别算法,其以两种特征组合表征字符图像,有效降低了单一特征可能造成的信息缺失;通过多角度与多层次的多尺度思想,描述了字符各种情况下的特征属性,并以 PCA 分析解决了多角度结构特征因倾斜导致的不匹配的问题。新算法仅以正立样本作为训练集,在正立和倾斜测试样本中都能取得较好的识别性能,在实际工作中具有较好的应用前景。

参考文献

- [1] 任丹,陈学峰. 手写数字识别的原理及应用[J]. 计算机时代, 2007(3):7-21
- [2] 孙征,李宁. 用专家乘积系统实现手写体数字识别[J]. 计算机仿真, 2006, 23(5):197-120
- [3] 刘春丽,吕淑静. 基于混合特征的孟加拉手写体数字识别[J]. 计算机工程与应用, 2007, 43(20):214
- [4] 张伟,王克俭,秦臻. 基于神经网络的数字识别的研究[J]. 微电子学与计算机, 2006, 23(8):206-208
- [5] 陈艳平,赵磊. 基于 8 邻域编码逐像素跟踪法的指纹细节特征提取算法[J]. 电脑知识与技术, 2008, 4(31):932-934
- [6] 耿西伟,张猛,沈建京. 基于结构特征分类 BP 网络的手写数字识别[J]. 计算机技术与发展, 2007, 17(1):130-132
- [7] 罗鑫,吴伟,杨晓敏,等. 一种基于 PCA 的多模板字符识别[J]. 电子测量技术, 2007, 30(1):138-142
- [8] 于真. 基于支持向量机的人脸识别技术研究[J]. 计算机仿真, 2011, 28(12):296-299
- [9] Automatica O N. Threshold Selection Method from Gray-level Histograms [J]. IEEE Transactions on Systems, Man and Cybernetics, 1979, 9(1):62-66
- [10] 朱峰. 手写体数字汉字识别技术的研究与实现 [D]. 镇江:江苏大学, 2006:18-19
- [11] 双小川,张克. 基于统计和结构特征的手写数字识别研究[J]. 计算机工程与设计, 2012, 33(4):1533-1537