

“有效行”特征对手写体字符的识别

王贵新*

(重庆后勤工程学院 重庆 400014)

刘建胜 居琰 汪同庆 彭健 杨波

(重庆大学光电工程学院 重庆 400000)

【摘要】从无限制的手写体数字的结构出发,提出“有效行”特征的概念及其提取算法。该特征具有维数小、平移不变、字符小角度旋转不变等特点。建立了相应的字符特征库,利用BP网络对样本字符进行研究,通过大量的手写体数字识别测试表明:该方法在识别速度上优于利用单一的矩特征、小波特征等传统识别方法;在误识率方面也优于一般的单一识别方法。

关键词 特征提取; 识别; 神经网络; 手写体

中图分类号 TP393

字符的识别已是一个比较成熟的技术,从比较简单的印刷体识别转向比较困难的脱机手写体识别^[1,2]。不管是采用人机结合的集成方式或多分类器集成^[3],都存在一个有效特征的选取问题。特征的提取和选择是字符识别的关键环节^[4],主要从下面三个方面考虑:1)特征的抽取速度要快;2)特征的稳定性好,即同一字符发生微小变化时,特征的变化要小;3)特征的分类能力强。从已有的文献来看,特征的种类和抽取的方法多种多样,比如中线特征、凹特征、方向特征、特征曲线,但选取的特征总存在稳定性与分类能力之间的矛盾,特别是有些特征的选取由于维数过大而影响了分类速度。

本文提出一种新的“有效行”特征和对应的提取方法,在一定程度上解决稳定性与分类能力之间的矛盾。在抽取特征和形成特征库的过程中,考虑了相同字符的不同模式的特征情况,对笔画的粗细利用阈值进行自适应调整,对一定的断笔画已可进行自适应的修补,结果特征的维数大大降低。

本文假设我们已对扫描的字符进行了分割和去噪等预处理,每个字符形成32×24点阵的二值图像。全文采用MATLAB5.3仿真。

1 有关定义和描述

1.1 有关定义

- 1) 字符行的起、止点:字符点阵中,某行第一次出现字符的像素点 (x_1, y_1) 称为该行的起点;该行最后一次出现字符的像素点 (x_2, y_1) 称为该行的止点。
- 2) 字符行的长度:行的止点横坐标减去起点横坐标,再加1(即 $x_2 - x_1 + 1$)。
- 3) 连续行:某行的起、止点之间的所有点都是字符的像素点。
- 4) 间断行:某行的起、止点之间至少有一个点不是字符的像素点。
- 5) 间断行的连续段和间断段:间断行上的连续点形成字符像素点的某一段,称为间断行的连续段;间断行上的连续点不形成字符像素点的某一段,称为间断行的间断段。
- 6) 空白行:某行的所有点都不是字符的像素点。
- 7) 有效行:某行是间断行或连续行。

2000年11月1日收稿

* 男 32岁 博士生

8) 间断行的补段: 对于间断行的间断段(开始点 $B(x_1, y_1)$, 结束点 $E(x_2, y_1)$)若满足: (1) 在该行的两端有连续段存在; (2) 若相邻的两行中, 一行是连续行, 另一行是空行; 或另一行纵坐标满足 $<=x_1$ (或 $>=x_2$)的所有点都不是字符的像素点, 此时称该间断段是间断行的补段。

例如, 图1所示的数字7中, 第3行是间断行, 其间断段也是补断, 同样数字5的第8行的间断段也是补断。

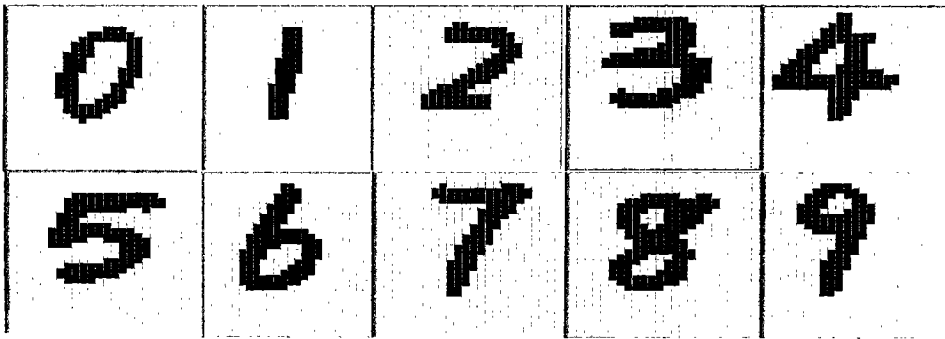


图1 经分割和去噪等预处理后的二值图像

1.2 有关描述

“间断行”用数字0表示; “连续行”中, 若字符行的长度 \leq 阈值(笔画的宽度)用数字1表示, 否则用数字2表示。由于数字字符笔画的连续封闭性, 则认为“连续行”、“间断行”才是描述字符的“有效行”。若在“有效行”间出现了“空白行”, 则以“有效行”看待, 其他情况的“空白行”不予考虑。

2 字符的“有效行”特征及其提取和形成

2.1 字符的特征

本文提取的字符“有效行”特征是: 有效行描述形成的方向向量 $A=\{a_1, a_2, \dots, a_k\}$, 称为“有效行”的主特征; 以及向量 A 的每一分量 a_i 在字符点阵中连续出现的数目 b_i 形成的对应向量 $B=\{b_1, b_2, \dots, b_k\}$, 称为“有效行”的从特征($1 \leq i, k \leq 32$)。显然 a_i 是为0、1或2的值, 根据下面的特征提取规则, A 中的相邻两分量不会相同, i 会远远小于32。

2.2 “有效行”特征的提取

提取的顺序从字符点阵的左上角开始, 从上往下进行; 特征的提取过程中, 阈值 bhk (笔画的宽度)、阈值 mcd (毛刺的大小)应根据学习或与特征库的对照情况自动调整, 使主特征 A 中分量是1的个数 ≤ 2 ; 间断行的补段对应的点应视为字符的像素点; 有效行间出现了“空白行”, 以相邻(或最接近)的有效行看待; 若 $a_i=0$, 则 $a_{i-1}=2$; $a_{i+1}=2$ 。此时如果 $a_{i-2}=1$ (或 $a_{i+2}=1$)且 b_{i-2} (或 b_{i+2}) \leq 阈值 mcd (毛刺的大小)时, a_{i-2} (或 a_{i+2})不予考虑在内; 间断行的长度 \leq 阈值 bhk (笔画的宽度)时, 应视为连续行; 若 $a_i=1$ 且 $b_i \leq$ 阈值 mcd (毛刺的大小)时, 此时 a_i 不予考虑在内。

根据上面的规则, 对图1的10个数字字符和一些变化形式进行了特征抽取编码, 得到本文的结果如表1所示(其中特征 $A=\{a_1, a_2, \dots, a_k\}$, 特征 $B=\{b_1, b_2, \dots, b_k\}$)。

图1的阈值 bhk (笔画的宽度) $=6$, 阈值 mcd (毛刺的大小) $=1$; 每字符的特征 A 和特征 B 的第1种情况是图1的特征, 其他是手写字符的可能变化情况。比如“0”在实心的情况下, $A=\{2\}$; 特征 A 的编码在一般情况下可固定不变, 特征 B 的编码随着字符点阵的大小而改变。为了得到比较稳定的特征, 归一化处理后的具体算法如下: $b_i' = b_i / \sum b_k$, $B=\{b_1, b_2, \dots, b_k\}$ 。

表1 数字字符的“有效行”特征

字符	特征A	特征B	字符	特征A	特征B
0	{2,0,2}{2}	{1,12,2}{15}	2	{2,1,2}{2,0,2,1,2}{1,2,0,2}	{3,10,3}{3,2,0,8,3}{8,2,6,1}
1	{1}	{18}	3	{2,0,2,1,2,0,2}{2,1,2,1}	{3,2,4,3,0,2,3}{3,4,4,10}
4	{1,2,0,2,1}	{4,2,6,3,6}	6	{1,2,0,2}{1,2}	{9,3,6,2}{9,8}
5	{2,1,2,0,2,1,2}	{4,2,3,2,0,3,3}	8	{2,0,2,0,2}{2,0,2}{2,0,2}	{3,4,5,4,2}{12,4,2}{3,4,11}
7	{2,1}	{4,18}	9	{1,2,0,2,1}{1,2,1}	{2,0,4,4,12}{2,8,12}

从上面的特征可知：其特征的维数小于8，一般情况在识别时可根据“有效行”的主特征A就可完成；如果主特征存在重复的情况，可将特征A和特征B结合，形成一个维数为14的特征向量。该特征具有极好的稳定性，由特征的提取算法可知：特征具有维数小、平移不变、字符小角度旋转不变等特点，同时不需要字符图像的细化工作，对字符笔画的粗细和起笔位置无关。

为了能让A、B两特征形成特征库，本文采取了树形结构的存储方式，这样在访问特征库时能大大提高速度，具体安排为：根结点代表某标准字符，根结点的分支代表该字符可能出现的“有效行”特征情况，第n层、第s个分支的叶结点代表该字符的第s个特征的主特征A和从特征B的第n个分量构成的数组。例如对表2的字符“0”，其树形结构为：根结点是“0”，第1分支有3层叶结点，每层也结点存储信息为(2,0.067)、(0,0.8)、(2,0.13)；第2分支只有1层叶结点，结点存储信息为(2,1)。

3 实验设计和结果分析

人工神经网络(ANN)由于具有很强的分类能力和容错性、鲁棒性，并且可在特征空间内形成任意复杂的决策区域等优点。所以ANN的运用取得了极大的进展，特别是在比较成熟的BP算法方面，ANN为OCR的研究提供了新的手段。本文的“有效行”特征具有维数较少和特征较稳定的特点，故大大地减少了网络的复杂度和难度。在实验中，采用输入层神经元为14，输出神经元数为4，1个隐层的方式。实验设计如下：首先对特征A和特征B进行规范化处理，把分量数不足7的特征扩展为7维，办法是在原来的基础上添加0。然后A、B合成一个14维的向量。即 $X_\mu=(a_1, a_2, \dots, a_7, b_1, b_2, \dots, b_7)$ ，输出神经元的状态为0或1，因此对0~9这10个数字，其二进制编码从0000到1001，共需输出神经元数为4，神经元的其他输出状态表示待识别字符的拒识情况。BP网络的拓扑结构如图2所示。以 (X_μ, Y_μ) 作为训练样本对(假设有N对)，误差函数为 $A_p = \frac{1}{2} \sum_{k=1}^s (y_{pk} - \hat{y}_{pk})^2$ 是第p表征矢量的误差， y_{pk} 是第k个输出神经元的实际值， \hat{y}_{pk} 是第k个输出神经元的期望值，利用Sigmoid函数，

方向传播训练算法所需的微分方程为： $\Delta w_{ji} = \eta \delta_{pj} \hat{y}_{pi}$ (其中 η 表示学习效率， δ_{pj} 表示L层神经元j的误差信号， \hat{y}_{pi} 表示L-1层神经元i的输出)， δ_{pj} 可以表示为：对于输出神经元有 $\delta_{pj} = (y_{pj} - \hat{y}_{pj}) \hat{y}_{pj} (1 - \hat{y}_{pj})$ ，对隐含层神经元有 $\delta_{pj} = y_{pj} (1 - \hat{y}_{pj}) \sum \delta_{pk} w_{kj}$ 。

按上述权值调整规则，应用迭代算法，每类字符用10个样本进行了学习，然后对每类字符200个样本进行了测试，实验取得的结果如表2所示。本文将学习样本的数量按每次增加10的方式重新测试表明，其识别正确率并没有显著提高，而拒识数有一定的增加。但是，对同样的样本按照文献[5]的不变特征进行识别，其误识率比本文的方法大8.4%，拒识数由于样本数的增加而显著的增加；同样的样本按照文献[6]的特征测试，尽管识别效果较好，但速度很慢。对上述现象分析如下：

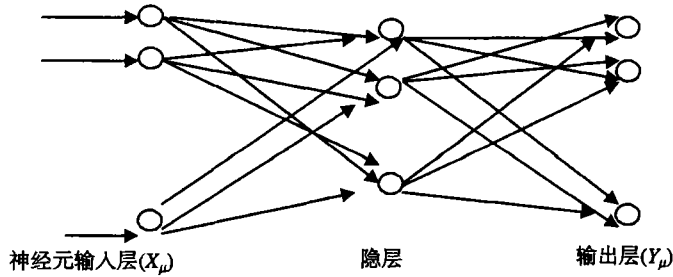


图2 神经网络的拓扑结构

1) 本文方法由于特征提取算法简单, 特征较稳定, 因此网络的学习效果在少量的样本下就能取得很好识别的效果, 这样大大提高了学习效率, 克服了BP网络学习效率低下的毛病。

2) 虽然矩特征的不变性好, 但由于公认的计算量大, 且不能达到其完备性; 象字符“6”、“9”等出现误识的情况也较多, 因此识别速度和精度大受影响。

3) 小波变换的方法尽管能处理字符的局部噪声, 其提取的特征有较好的抗干扰能力, 但要求其样本点需尽可能多, 因此影响识别速度。

4) 从表2可知, 实验取得了预期的结果, 但是由于采用的分类只用到字符的一种特征, 效果还很容易受到一些特殊情况下字符识别的影响。利用字符的多种特征, 采用多层分类器的识别方式, 是取得良好识别结果的有途径。

表2 BP网络的识别情况

字符	0	1	2	3	4	5	6	7	8	9
正确数	196	197	190	191	189	190	192	194	193	195
误识数	3	2	6	7	5	8	7	2	5	4

参 考 文 献

- 1 张炳中. 汉字识别技术, 北京: 清华大学出版社, 1992
- 2 张炳中. 印刷体汉字文本的微型计算机自动识别. 中国科学(A), 1990, 1: 97~104
- 3 郝红卫, 戴汝为. 人机结合的集成方法及其在字符识别中的应用. 模式识别与人工智能, 1996, 9(1): 10~19
- 4 Terier O D, Jain A K, Txi T. Feature extraction methods for character recognition—a survey. Pattern Recognition, 1996, 29(4): 641~662
- 5 Perantonis S J, P J Lisboa. Translation, rotation and scale invariant pattern recognition by higher-order neural networks and moment classifiers. IEEE Trans. Neural Networks 1992, 3: 241~251
- 6 Fayolle Jacques, Riou Laurence, Ducottet Christophe. Robustness of a multiscale scheme of feature points detection. Pattern Recognition, 2000, 33:1 437~1 453

Handwritten Numerals Recognition Based on "Effective-rows" Feature

Wang Guixin

(Institute of Logistic and Engineering of Chongqing Chongqing 400014)

Liu Jiansheng Ju Yan Wang Tongqing Peng Jian Yang Bo

(Chongqing University Chongqing 400000)

Abstract The concept and extracting method of the effective-rows feature, which possesses low dimension, shift-invariance and rotation-invariance under little angle is proposed on the basis of the structure of unconstraint and handwritten numerals in this paper. Its features is also constructed. According to the result of widely verifying recognition BP network, it is demonstrated that its recognition-speed is superior to the traditional recognition methods only based on the feature of moment or wavelet and its recognition-extract is higher than some conventional methods.

Key words feature extraction; recognition; neural network; handwritten

(上接第283页)

Application of SMIL on Distance Education

Peng Dezhong Huang Diming

(Dept. Applied Mathematics, College of Computer Science and Engineering, UEST of China Chengdu 610054)

Abstract Synchronized multimedia integration language(SMIL) is an important application of the new generation web language--XML, and can synchronize variant media, save network bandwidth and support super-link. The development and integration system of network-multimedia courseware is a tool of creating and editing network teaching courseware that is based on synchronized multimedia integration language(SMIL). It takes the advantage of SMIL's good on creating network multimedia, and introduces modeling conception, which provides teachers and students a function-strong and using-simple network distance education environment. It has a lot of advantages of saving bandwidth and synchronizing easily as well.

Key words synchronized multimedia integration language; distance education; extensible markup language; multimedia