

基于矩和小波变换的数字、字母字符识别研究

沈会良 李志能

(浙江大学信息与电子工程学系, 杭州 310027)

摘要 欲实现汽车监控和管理智能化, 必须正确识别牌照字符. 在识别过程中, 关键是特征向量的提取. 小波变换能有效地提取字符的结构特征, 而矩能够很好地对其进行描述. 该文提出了一种用线性矩和小波变换提取数字、字母字符特征的方法, 实验证明该方法有很高的识别率, 达到 97% 以上, 能够有效地进行字符的分类, 可满足实际应用.

关键词 字符识别 特征向量 矩 小波变换

中图分类号: TP391.43 文献标识码: B 文章编号: 1006-8961(2000)01-0249-04

A Study of Number and Letter Character Recognition Based on Moments and Wavelet Transform

SHEN Hui-liang, LI Zhi-neng

(Department of Information and Electronic Engineering, Zhejiang University, Hangzhou 310027)

Abstract For the sake of vehicle identification and intelligent management, it's necessary to recognize the characters on the license in high accuracy. Selection of a feature extraction method is probably the most important factor in recognition process. The feature of number and letter character construction can be extracted by wavelet transform effectively and described by moments. A method of character recognition based on linearity moments and wavelet transform is presented in this paper. The result of experiment shows high recognition rate above 97%, which indicates that the method can be put into practical use.

Keywords Character recognition, Feature vector, Moments, Wavelet transform

0 引言

在字符识别过程中, 特征量的选取至关重要. 有效的特征描述量有: 矩^[1]、边界投影^[2]、样条曲线拟合^[3]、傅立叶描绘子^[4]、模板匹配等, 针对不同的应用, 特征量的选择也不尽相同. 当然, 最终的效果还有赖于模式分类器的性能, 一个好的分类器应具有很低的误识率和拒识率. 一般来讲, 一个字符识别系统的步骤如图 1 所示:

矩是一个能很好地描述模式特征的量, 在模式识别中有较广泛的应用, 但表达方式繁琐. 本文提出一种能有效地表述字符的线性矩, 它形式简洁, 具有平移和伸缩不变性.

小波变换能给出信号完全而稳定的表示, 能够得到不同尺度下信号的形式. 从特征提取的角度来分析, 其变换后的信号正包含了信号的主要特征, 如大致形状、变化位置及幅度等.

本文结合矩与小波变换来提取有效的字符特征, 组成一个特征向量, 送入 BP 神经网络进行识别.

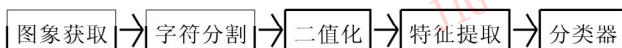


图 1 字符识别系统框图

1 用矩和小波变换提取特征

1.1 线性矩的描述

一幅包含有物体的二值图象可以表述如下:

$$f(i, j) = \begin{cases} 0, & (i, j) \in \text{物体} \\ 1, & (i, j) \notin \text{物体} \end{cases}$$

在图象上的初始参考点 $C_0 = (i_0, j_0)$ 定义如下:

$$\begin{cases} i_0 = \min(i: \exists j, \text{使得} f(i, j) = 0, j = 0, 1, \dots, \\ \quad W - 1; i = 0, 1, \dots, D - 1) \\ j_0 = \min(j: \exists i, \text{使得} f(i, j) = 0, i = 0, 1, \dots, \\ \quad D - 1; j = 0, 1, \dots, W - 1) \end{cases}$$

其中 W, D 分别为图象的宽度和高度. 在图象上定义行、列上的线性矩向量 M_{row} 和 M_{col} , 为叙述方便, 仅以 M_{row} 为例说明. 设 M_{row} 矩向量的元素个数为 N , 第 k 个元素记为 $M_{\text{row}}[k]$ ($N \leq D$, 第 k 个矩元素的行扫描线是图象的第 $\lfloor \frac{D}{N}k + 0.5 \rfloor$ 行, $\lfloor \cdot \rfloor$ 表示取整.) 在扫描线上, 有黑线段 B_{kl} 和白线段 W_{kl} , 其中 B_{kl} 表示物体的一个分段, W_{kl} 表示背景的一个分段, 记 $1 \leq l \leq N_k, N_k$ 为第 k 个元素的扫描线上黑线段的数目. 同时记 L_{kl} 为 B_{kl} 的长度, C_{kl} 为 B_{kl} 的中心点. 有

$$M_{\text{row}}[k] = \sum_{l=1}^{N_k} (C_{kl} - C_0) \cdot L_{kl}$$

可以看到, 矩的大小与黑线段的数目、长度、位置间隔成正比关系. 记行向的伸缩因子为 S , 平移后的中心点为 \tilde{C}_{kl} , 图象中物体的初始点为 \tilde{C}_0 , 则

$$\begin{aligned} \tilde{C}_{kl} - \tilde{C}_0 &= S \cdot (C_{kl} - C_0) \\ \tilde{L}_{kl} &= S \cdot L_{kl} \end{aligned}$$

从而

$$\begin{aligned} \tilde{M}_{\text{row}}[k] &= \sum_{l=1}^{N_k} (\tilde{C}_{kl} - \tilde{C}_0) \cdot \tilde{L}_{kl} \\ &= S^2 \cdot \sum_{l=1}^{N_k} (C_{kl} - C_0) \cdot L_{kl} \\ &= S^2 \cdot M_{\text{row}}[k] \end{aligned}$$

定义线性矩为

$$m_{\text{row}}[k] = \frac{M_{\text{row}}[k]}{\sum_{k=1}^{N_k} M_{\text{row}}[k]}$$

则

$$\tilde{m}_{\text{row}}[k] = \frac{\tilde{M}_{\text{row}}[k]}{\sum_{k=1}^{N_k} \tilde{M}_{\text{row}}[k]} = \frac{S^2 \cdot M_{\text{row}}[k]}{\sum_{k=1}^{N_k} S^2 \cdot M_{\text{row}}[k]} = m_{\text{row}}[k]$$

即 $m_{\text{row}}[k]$ 是平移、伸缩不变的.

对于灰度图象, 同样定义矩, 只不过在形式上稍有差异, 可写为

$$M_{\text{row}}[k] = \sum_{l=1}^{N_k} (C_{kl} - C_0) \cdot L_{kl} \cdot \bar{f}_{kl}$$

其中, \bar{f}_{kl} 为第 l 个灰线段的平均灰度值. 同样可以定义线性矩 m_{row} , 它具有平移、伸缩不变性, 在此不再赘述.

1.2 小波变换描述

信号分析的方法一般有 Fourier 变换、窗口 Fourier 变换 (Gabor 变换)、小波变换 3 种. 我们要求在分析信号时, 在低空间频率处有较高的频谱分辨率, 同时在高空间频率处有较高的空间分辨率, 即要求变换的时-频窗口是根据信号的变化而自适应调节的. 很显然, Fourier 变换、Gabor 变换都不能满足这个要求, 只有小波变换具有这种性质.

假设 $\varphi(x), \psi(x)$ 分别是尺度函数和小波函数,

定义 $\varphi_j(x) = \frac{1}{2^{j/2}} \varphi\left(\frac{x}{2^j}\right), \psi_j(x) = \frac{1}{2^{j/2}} \psi\left(\frac{x}{2^j}\right)$ 则空间函数 $f(x)$ 在尺度 2^j 上位置 x 处的小波变换为

$$\begin{aligned} S_{2^j}^d f &= f(x) * \varphi_j(x) = \int_{-\infty}^{\infty} f(u) \varphi_j(x - u) du \\ W_{2^j}^d f &= f(x) * \psi_j(x) = \int_{-\infty}^{\infty} f(u) \psi_j(x - u) du \end{aligned}$$

推导得, 在各个尺度上有

$$\begin{aligned} S_{2^j}^d f &= S_{2^{j-1}}^d f * H_{j-1} \\ W_{2^j}^d f &= S_{2^{j-1}}^d f * G_{j-1} \end{aligned}$$

式中 H_j 和 G_j 为低通和高通滤波器^[5].

人眼对字符的识别在很大程度上是基于字符结构的, 即是依赖于字符的笔划特征, 如横向、竖向和斜向的笔划. 由此, 为提取出图象 4 个方向上的边缘, 对字符图象作小波分解.

在图 2 中, 用两个分离的一维小波的乘积定义二维小波, 使图象的分解可先按行, 再按列进行, 其中 $A_{j-1}^d f$ 是图象在尺度 2^j 上的轮廓, $D_{j-1}^1 f$ 是横向笔划, $D_{j-1}^2 f$ 是竖向笔划, $D_{j-1}^3 f$ 是斜向笔划, 分解过程采用 Mallat 算法^[6], 结果见图 3(c).

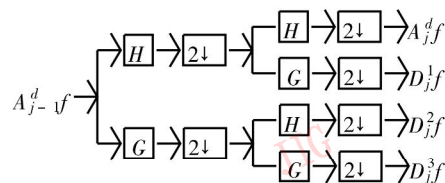


图 2 图象的小波分解示意图

$\boxed{2\downarrow}$ 表示行或列系数取偶数位置的值

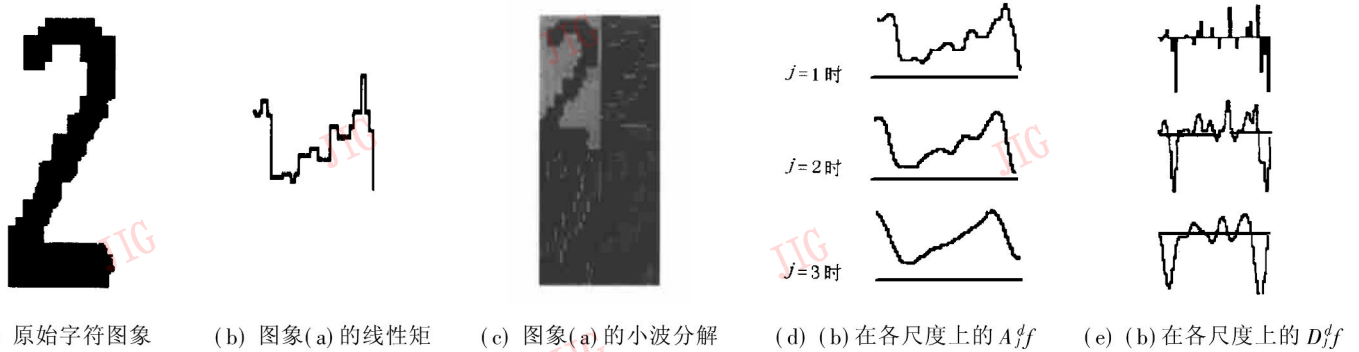


图 3 用矩和小波变换提取特征

我们提取两类图象特征, 具体描述如下.

(1) 对二值图象进行线性矩变换, 得到矩向量信号, 对其作一维小波变换, 得到各个尺度上的信号如图 3(d)、(e) 所示. 可以看到, $A_j^d f$ 是信号的平滑, 随着尺度的增加, 平滑效果更加明显; 而 $D_j^d f$ 反映了信号奇异点的位置, 在高尺度上, 反映了主要的奇异点. 对较小的字符图象, 适宜在 $j=1$ 时的信号上取特征, 从 $A_j^d f$ 中抽样出 $N/2$ (N 为总的离散点) 个点作为特征, 在 $D_j^d f$ 上定义 4 个统计变量作为特征:

$$\text{平均值: } m = \frac{1}{N} \sum_{k=1}^N f(k) \quad (1)$$

$$\text{能量: } e = \frac{1}{N} \sum_{k=1}^N f^2(k) \quad (2)$$

$$\text{标准方差: } s = \frac{1}{N} \sum_{k=1}^N [f(k) - m]^2 \quad (3)$$

$$\text{平均残差: } a = \frac{1}{N} \sum_{k=1}^N |f(k) - m| \quad (4)$$

(2) 对图象作小波分解, 得到 3 个高频图象, 如前所述. 为叙述方便, 仅以反映横向笔划的图象 $f'(i, j)$ 说明处理过程(其它图象同样处理). 设定阈值 B (B 大于图象的平均灰度), 作新的图象 $g(i, j)$, 有

$$g(i, j) = \begin{cases} 0, & f'(i, j) \leq B \\ f'(i, j) - B, & f'(i, j) > B \end{cases}$$

为了减少数据量, 再在 $g(i, j)$ 上粗化灰度为 M (M 在本算法中取 4) 级, 得到图象 $G(i, j)$, $G(i, j) = [g(i, j) \cdot M / f_{\max}]$ ($[]$ 表示取整, f_{\max} 为最大灰度值), 在图象 G 上进行灰度线性矩的提取(斜向笔划图象中矩的扫描按 45° 、 135° 进行), 同样计算前面的 4 个统计量(式(1)~(4))作为特征向量.

将特征向量归一化后作为 BP 神经网络的输入量, 进行字符的识别.

2 实验及结果分析

采用从汽车牌照上提取的数字、字母字符作为识别对象. 由于牌照挂法的不规范、摄像机角度、光照等的影响, 字符有某种程度上的扭曲、平移、退化. 首先对数字的识别作一描述. 将 500 个字符图象(平均每个数字 50 个)做样本训练神经网络, 将另外 500 个图象作为测试集, 得到的结果如表 1 所示.

表 1 神经网络对字符的分类结果

数字	0	1	2	3	4	5	6	7	8	9
识别率(%)	96	100	100	100	96	100	98	100	96	100

总的识别率是 98.6%. 字母字符的识别结果与数字接近, 为 97.3%. 汽车牌照上的汉字图象严重退化, 二值化效果很差, 不能采用前面提出的特征分析法, 而应该直接在汉字灰度图象上进行特征提取, 这方面的研究工作正在进行之中.

本文中识别的字符图象较小 (32×16), 特征不宜在太大的尺度上提取, 因为大尺度意味着大的数据压缩量, 会造成不同图象间平滑分量的区分错误. 而如果原始数据量很大, 则应该采用较大尺度上的信号.

3 结 论

小波分析是近年来信号处理中研究的热点, 各种新算法层出不穷, 但用于字符识别的文献仍不多见. 本文提出了基于矩和小波变换的特征向量提取应用于识别字符图象, 实验表明, 平均字符识别率达 97% 以上, 此方法可以推广到实际工程应用中.

参 考 文 献

- 1 Reiss T H. The revised fundamental theorem of moment invariants. IEEE Trans Pattern Anal Mach Intell, 1991, 13: 830~834.
- 2 Kimura F, Shridhar M. Handwritten numerical recognition based on multiple algorithms. Pattern Recognition, 1991, 24(10): 969~983.
- 3 Sekita I, Toraichi K, Mori R *et al.* Feature extraction of handwritten Japanese characters by spline functions for relaxation matching. Pattern Recognition, 1988, 21(1): 9~17.
- 4 Kuhl F P, Giardina C R. Ellipse Fourier features of a closed contour. Comput Vis Graphics Image Process, 1982, 18: 236~258.
- 5 秦前清, 杨宗凯. 实用小波分析. 西安: 西安电子科技大学出版社, 1988, 25~27.
- 6 Mallat S. Theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans. Pattern Recognition Mach Intell, 1989, 11(7): 674~693.



沈会良 1974年生, 1996年毕业于浙江大学信息与电子工程学系, 博士生. 主要研究方向为图象处理、模式识别.



李志能 现为浙江大学信息与电子工程学系教授, 物理电子专业博士生导师. 主要从事光电信息处理、模式识别、神经网络等方面的科学研究与教学工作.

新 书 推 介

《计算智能信息处理技术及其应用》

由湖南大学电气工程系王耀南教授编著的《计算智能信息处理技术及其应用》一书已由湖南大学出版社于1999年6月出版. 该书从信息科学的角度出发, 系统地介绍了计算智能信息处理技术的基础理论及其各种新的处理方法与技术. 全书共分十章, 内容包括计算智能信息处理技术综述、模糊计算、模糊信息处理、神经计算、模糊神经网络信息处理、进化计算的基本方法、进化计算的信息处理、混沌与分形计算、混沌与分形信息处理、计算智能信息处理技术在图象识别、压缩, 机器人控制, 复杂工业过程控制、规划, 故障诊断等方面的应用.

该书涉及了目前国内外计算智能信息处理的技术研究成果, 综合了作者近年来的教学心得与科研成果, 取材新颖、内容丰富、注重了理论与实践相结合, 论述深入浅出, 力求使读者较快掌握和应用这门高新技术, 是一本很好的“软计算”专著. 该书可作为自动化、计算机应用、人工智能、图象处理与模式识别、智能控制与信息处理、电子工程、机械工程、系统工程等专业研究生或高年级本科学生的教材和参考书, 也可供有关工程技术人员和科研工作者参考.