

## 基于压缩传感的手写字符识别方法

刘长红<sup>1,2</sup>, 杨 扬<sup>1</sup>, 陈 勇<sup>3</sup>

(1. 北京科技大学 信息工程学院, 北京 100083; 2. 江西师范大学 计算机信息工程学院, 南昌 330022;

3. 南昌工程学院 管理工程系, 南昌 330099)

(ahong286@126.com)

**摘 要:** 基于新出现的压缩传感理论, 提出了一种鲁棒的手写字符识别方法, 能很好地对含有噪声的字符进行识别。该方法通过对测试字符进行稀疏表示, 采用  $l_1$  范数最小化算法求得最稀疏的系数解, 所获得的系数具有明显的类别信息, 从而易于对测试字符进行分类。实验结果表明, 该方法具有很好的噪声鲁棒性。

**关键词:** 手写字符识别; 压缩传感; 稀疏表示;  $l_1$  最小化

**中图分类号:** TP391.4 **文献标志码:** A

## Handwritten character recognition based on compressive sensing

LIU Chang-hong<sup>1,2</sup>, YANG Yang<sup>1</sup>, CHEN Yong<sup>3</sup>

(1. School of Information Engineering, Beijing University of Science and Technology, Beijing 100083, China;

2. School of Computer Information and Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China;

3. Department of Management Engineering, Nanchang Institute of Technology, Nanchang Jiangxi 330099, China)

**Abstract:** A robust method for handwritten character recognition with noises based on compressive sensing was presented. The sparsest representation of the test character computed by  $l_1$ -minimization had distinct class information; therefore, it is easy to classify the characters. The experimental results show that the proposed method is a noise robust technique.

**Key words:** handwritten character recognition; compressive sensing; sparse representation;  $l_1$ -minimization

### 0 引言

虽然字符识别问题的研究已经历了多年, 也取得了很多成果, 但是在实际应用中, 字符识别性能的改善仍然是一个主要的问题。对于完全没有约束的手写西文、数字与汉字, 没有一种简单的方案能达到很高的识别率和识别精度, 最近这方面的努力向着更为成熟、复杂、综合的方向发展。一方面, 研究工作者努力把新的知识运用到预处理、特征提取和分类当中, 如神经网络、隐马尔可夫、支持向量机 (Support Vector Machine, SVM)、数学形态学等<sup>[1-2]</sup>; 另一方面, 采用信息集成的方式从多角度对手写字符进行综合分析<sup>[3-5]</sup>。

近年来, 视觉感知的研究表明, 在中低层人体视觉中, 许多神经元对各种外部刺激具有选择性。在视觉系统 V1 区神经细胞的处理过程中, 一个重要的约束就是编码的稀疏性, 利用较少的资源尽可能有效地编码更多的信息; 而自然图像经过稀疏编码后得到的基函数类似于 V1 区简单细胞感受野的反应特性<sup>[6]</sup>。考虑到基函数的超完备性, Olshausen 和 Field<sup>[7]</sup>又提出了一种超完备基的稀疏编码算法, 利用基函数和系数的概率密度模型成功地建模了 V1 区简单细胞感受野。

而在信号处理领域, 关于超完备基的稀疏线性表示计算问题, 即压缩传感理论 (Compressive Sensing, CS), 引起了广大研究者的兴趣<sup>[8-11]</sup>。该理论表明, 如果信号表示是足够稀疏的, 就可通过凸优化的方法来计算<sup>[12]</sup>。压缩传感研究的初衷是用于信号的压缩和表示, 每个基函数并没有任何的语义, 采用标准基 (如: 傅里叶、小波、Gabor 等) 或随机测量矩阵, 但其最稀疏的表示具有很好的判别性, 因为它选择最能紧致表

示信号的基。文献[13]利用稀疏表示的判别性有效地实现了人脸识别, 但不是使用以上所述的超完备基, 而是使用训练样本作为超完备基。因为如果每类样本足够充分的话, 那么测试样本可表示为同类样本的线性组合, 这对于整个样本集来说, 其表示自然是非常稀疏的, 那么分类问题即可通过压缩传感的方法来实现。本文利用其思想提出了一种新的简单的手写字符识别方法, 使用稀疏性识别相关的模式或特征。首先对要识别的字符寻求其稀疏表示, 然后基于压缩传感理论求解其最稀疏的解, 根据所得的解直接进行分类。因为该方法选择最能紧致表示的基, 所以得到的最稀疏解具有显著的类别信息, 可直接进行分类。该分类器可认为是流行的最近邻 (Nearest Neighbor, NN) 和 K-最近邻 (K-NN) 分类器的推广, 最近邻法基于单个训练样本的最好表示来进行分类, 本文的方法则考虑所有可能的样本, 在每类或多类中自适应地选择能最好表示测试样本的最少训练样本。

### 1 手写字符的稀疏表示

目标识别的基本问题就是根据  $K$  类训练样本确定测试样本属于哪一类。假设手写字符训练样本种类为  $K$  类, 同一类训练样本位于一个线性子空间中, 那么新测试样本可看成是同类样本的线性组合。每一类的样本数分别为  $N_1, N_2, \dots, N_k$ , 每个训练样本字符为  $w * h$  大小的灰度图片, 按列排列为  $m$  维的向量  $V, V \in \mathbf{R}^m (m = w * h)$ , 第  $i$  类所有的  $N_i$  个样本作为矩阵  $A_i$  的列,  $A_i = [V_{i,1}, V_{i,2}, \dots, V_{i,N_i}] \in \mathbf{R}^{m \times N_i}$ , 则给定第  $i$  类的新样本  $y \in \mathbf{R}^m$ , 它可近似表示为该训练样本的线性组

收稿日期: 2009-02-26; 修回日期: 2009-05-01。 基金项目: 江西省教育厅科技项目 (GJJ09143)。

作者简介: 刘长红 (1977-), 女, 江西南丰人, 讲师, 博士研究生, 主要研究方向: 模式识别、图像处理; 杨扬 (1955-), 男, 河北承德人, 教授, 博士生导师, 主要研究方向: 图像处理、计算机网络; 陈勇 (1973-), 男, 江西南昌人, 讲师, 硕士, 主要研究方向: 电子商务、图像处理。

合,即:  $y = a_{i,1}V_{i,1} + a_{i,2}V_{i,2} + \dots + a_{i,N_i}V_{i,N_i}, a_{i,j} \in \mathbf{R}, j = 1, 2, \dots, N_i$  为常量。

由于事先不知道  $y$  属于哪个样本, 所以将  $y$  表示为所有样本的线性组合:

$$y = Ax_0 \in \mathbf{R}^m \quad (1)$$

其中系数  $x_0 = [0, \dots, 0, a_{i,1}, a_{i,2}, \dots, a_{i,N_i}, 0, \dots, 0]^T$ ,  $T$  表示转置, 除了与第  $i$  类有关的系数, 其他的系数都为零。  $A = [A_1, A_2, \dots, A_i, \dots, A_k] = [V_{1,1}, V_{1,2}, \dots, V_{i,1}, \dots, V_{k,N_k}]$ , 共有  $N = N_1 + N_2 + \dots + N_k$  个样本, 为所有  $K$  类训练样本组成的矩阵。

为了用稀疏向量  $x_0$  表示字符样本  $y$ , 需要求解方程(1); 典型情况下  $m \ll N$ , 方程(1)是欠定的, 因此它的解不是唯一的。传统的求解方法一般采用最小  $l_2$  范数来进行求解:

$$\hat{x}_2 = \operatorname{argmin} \|x\|_2 \text{ 约束条件为:}$$

$$Ax = y \quad (2)$$

但此时求解出来的系数  $x_0$  中非零项元素非常密集, 几乎没有零。而根据系数  $x_0$  的稀疏性条件, 应通过最小  $l_0$  范数(即计算向量中非零元素的个数)来求解, 但此时在欠定线性方程组中求最稀疏解是 NP 难问题。压缩传感理论中<sup>[12]</sup>证明了如果信号足够稀疏, 那么  $l_0$  最小化问题的解等同于  $l_1$  最小化问题的解, 通过最小  $l_1$  范数可求出唯一的最稀疏解; 也就是说, 如果系数  $x_0$  足够稀疏, 那么可通过  $x_0$  的最小  $l_1$  范数来求解:

$$\hat{x}_1 = \operatorname{argmin} \|x\|_1 \text{ 约束条件为:}$$

$$Ax = y \quad (3)$$

$\|\cdot\|_1$  表示向量中各元素绝对值之和。图 1(a) 是对测试字符“7”通过最小  $l_1$  范数求解所得的系数  $x_0$ , 从图中可以看出非零元素非常稀疏, 大多数几乎都近似为 0, 而且非零项并没有分布在各类上, 主要集中在字符“7”这类上, 最大的两个系数值都对应训练字符“7”。

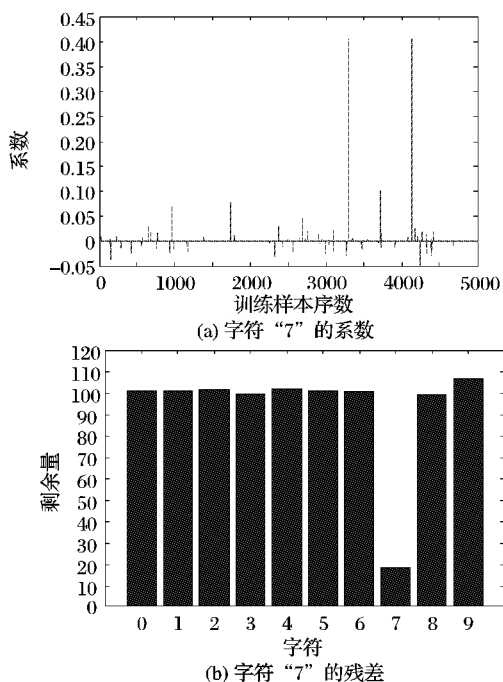


图1 采用最小  $l_1$  范数的方法识别字符“7”得到的系数和残差  $r_i(y)$

由于实际字符数据中一般都含有噪声, 也就是说,  $y = Ax_0 + e$ , 其中噪声满足  $\|e\|_2 < \varepsilon$ , 那么最稀疏解  $x_0$  可通过解下面稳定的  $l_1$  最小化问题来求得:

$$\hat{x}_1 = \operatorname{argmin} \|x\| \text{ 约束条件为:}$$

$$\|Ax - y\|_2 \leq \varepsilon \quad (4)$$

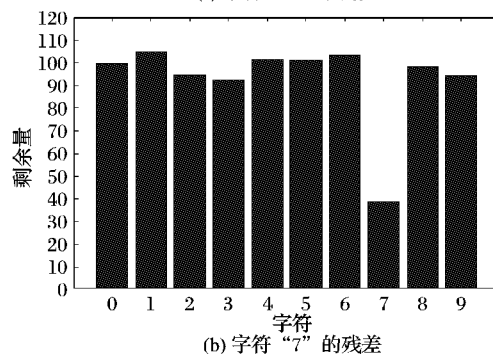
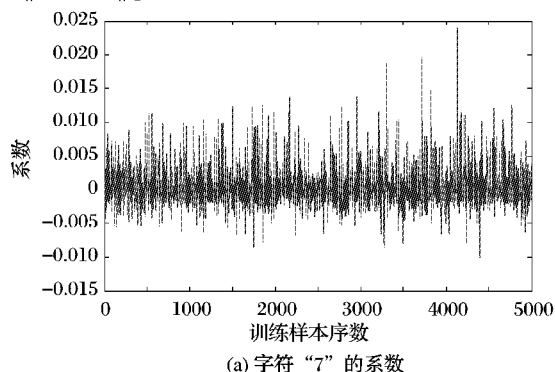


图2 采用最小  $l_2$  范数的方法识别字符“7”得到的系数和残差  $r_i(y)$

## 2 字符分类

给定新测试字符  $y$ , 通过式(3)或(4)计算其稀疏表示  $\hat{x}_1$ , 理想情况下得到的  $\hat{x}_1$  中非零项应该都是与第  $i$  类有关的项, 但是由于噪声或模型错误等的影响, 导致  $\hat{x}_1$  中有一些与其他类有关的小的非零项(如图1中左图所示), 所以需要通过一些简单方式来对其进行分类。本文采用文献[6]中的方式, 对于每一类  $i$ , 利用一个函数  $T_i(\hat{x}_1)$  在系数  $\hat{x}_1$  中选取与第  $i$  类有关的元素项, 其中除了与第  $i$  类有关的元素为非零外, 其他的都为零; 重新生成一个近似的  $\hat{y}_i, \hat{y}_i = AT_i(\hat{x}_1)$ , 然后求最小残差:

$$\min r_i(y) = \|y - AT_i(\hat{x}_1)\| \quad (5)$$

残差最小的  $i$  即为  $y$  的类标志。分别根据最小  $l_1$  范数和最小  $l_2$  范数求解出的系数产生的各类残差  $r_i(y)$ , 如图1(b)和图2(b)所示。从图可以看出, 采用最小  $l_1$  范数求解的系数产生的残差中, 最小剩余量与次小剩余量相差更大, 其比值为 1:5.6, 而采用最小  $l_2$  范数所得的比值只有 1:2.4, 所以采用最小  $l_1$  范数求解的系数  $\hat{x}_1$  中包含了更多测试字符的信息, 更容易进行分类识别。

## 3 干扰字符识别

在实际应用中, 可能经常会出现提取出来的字符受到干扰, 而且由于字符中受干扰的位置不确定, 所以不能仅仅加上一个小小的噪声。为了使本文的方法能够对干扰的字符进行识别, 可将式(1)改为:  $y = y_0 + e_0 = Ax_0 + e_0$ , 其中  $e_0 \in \mathbf{R}^m$  是错误向量, 即:

$$y = [A \ I] \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = Bw_0 \quad (6)$$

其中  $B = [A \ I] \in \mathbf{R}^{m \times (n+m)}, w_0 = [x_0 \ e_0]^T$ 。

对于一般的欠定系统  $y = Bw$ , 如果非零项个数小于等于  $m/2$  的话, 那么  $\hat{w}$  就有唯一的最稀疏解, 由于  $x_0$  是稀疏的, 所以只要  $e_0$  中干扰像素小于一定覆盖量, 就可通过下面  $l_1$  最小

化问题来求解:

$$\hat{w}_1 = \operatorname{argmin} \|w\|$$

约束条件为:

$$Bw = y \quad (7)$$

其中  $B = [A \ I]$ ,  $w = [x \ e]$ 。求出系数  $\hat{w}_1 = [\hat{x}_1 \ \hat{e}_1]$  以后, 那么就可获得去干扰的清晰字符  $y_{\text{raw}}$ ,  $y_{\text{raw}} = y - \hat{e}_1$ , 然后再对去除干扰后的字符  $y_{\text{raw}}$  根据式(5)进行识别。

#### 4 实验结果与分析

本文在基准手写字符数据库 Mnist 上进行实验, 每个字符为  $28 \times 28$  的灰度图像, 训练样本字符为 60 000 个, 测试字符为 10 000 个。为了便于测试, 本文随机抽取了 5 000 个训练样本字符和 1 000 个测试字符进行实验, 目前没有利用任何特征提取算法, 只将训练样本字符和测试字符都下采样为  $14 \times 14$  的灰度图像, 分别采用本文的方法(基于 CS)与 NN、K-NN、SVM 的方法进行分类, 表 1 给出了实验结果。实验结果表明, 本文方法的识别率高于最近邻和 SVM 方法。

表 1 各种比例噪声字符的识别率

方法	识别率/%	方法	识别率/%
NN	91.9	SVM	94.5
KNN	92.4	CS	96.8

受干扰字符的实验也是在 Mnist 手写字符数据库上进行, 训练样本字符取 3 000 个, 训练字符采用原字符样本, 即每个字符为  $28 \times 28$  的灰度图像。测试字符随机取 100 个, 分别进行了含 10%、25%、50%、60% 和 75% 噪声的测试, 图 3(a~c) 从上到下分别为字符“5”含 25% 和 50% 的噪声图像、去噪后的字符以及提取出的噪声, 图 3(d) 为含 50% 噪声的字符去噪后对应的系数, 从图中可以看出去噪后的系数仍然非常稀疏, 且主要集中在训练字符为“5”的这类上, 具有明显的类别信息。

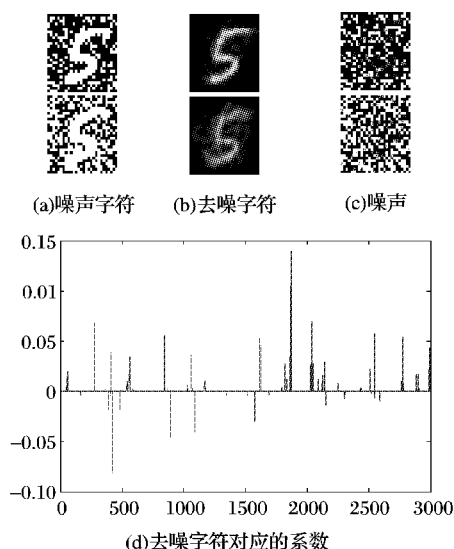


图 3 字符去噪以及去噪后的字符对应的系数

图 4 给出了本文的方法(基于 CS)与 NN、K-NN、SVM 在含各种比例噪声下的手写字符的识别率, 实验结果表明, SVM 对噪声很敏感, 最近邻法在噪声比例为 50% 时识别率迅速下降; 而本文的方法对于含有 25% 噪声的识别率可达到 96%, 接近于无噪声时的识别率, 即使含有 50% 噪声的字符识别率

也达到 89%。因此可以看出本文的方法优于其他方法, 能够有效地对字符去噪, 对噪声字符具有很高的识别率。

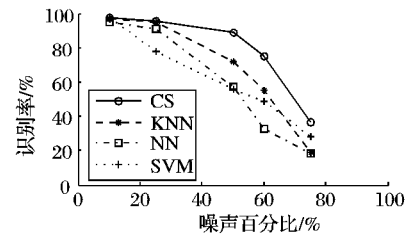


图 4 各种比例噪声字符的识别率比较

#### 5 结语

本文基于压缩传感理论提出了一种噪声鲁棒的手写字符识别方法, 能够对含有很大噪声干扰的字符进行有效的识别, 并在 Mnist 手写字符数据库上对该方法进行了验证。实验结果表明, 在相同噪声情况下, 本文的方法得到的识别率均比最近邻、SVM 的结果具有更好的性能。

由于字符识别过程中对测试字符和训练样本进行了下采样, 整体识别率仅达到 96.8%, 结合其他的特征提取方法以及弹性变形等技术提高识别率将是下一步的工作。

#### 参考文献:

- [1] ARICA N, YARMAN-VURAL F T. An overview of character recognition focused on off-line handwriting[J]. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 2001, 31(2): 216 - 233.
- [2] 王先梅, 王宏, 颜斌. 一种基于距离相似性度量和 HMMs 的字符识别方法[J]. 光电子. 激光, 2008, 19(8): 1100 - 1103.
- [3] SUEN C Y, LAM L. Multiple classifier combination methodologies for different output levels[C]// Proceedings of the First International Workshop on Multiple Classifier Systems, LNCS 1857. Berlin: Springer, 2000: 52 - 66.
- [4] LIU C L. Classifier combination based on confidence transformation[J]. Pattern Recognition, 2005, 38(1): 11 - 28.
- [5] 付强, 丁晓青, 刘长松. 用于手写汉字识别的级联 MQDF 分类器[J]. 清华大学学报: 自然科学版, 2008, 48(10): 1605 - 1608.
- [6] OLSHAUSEN B, FIELD D. Emergence of simple - cell receptive field properties by learning a sparse code for natural image[J]. Nature, 1996, 381(6583): 607 - 609.
- [7] OLSHAUSEN B, FIELD D. Sparse coding with an overcomplete basis set: A strategy employed by V1?[J]. Vision Research, 1997, 37(23): 3311 - 3325.
- [8] CANDEÈS E. Compressive sampling[C/OL]// Proceedings of the International Congress of Mathematicians, 2006. [2009 - 02 - 01]. <http://www.acm.caltech.edu/~emmanuel/papers/Compressive-Sampling.pdf>.
- [9] CANDÈS E, WAKIN M. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine, 2008, 25(2): 21 - 30.
- [10] DONOHO D. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289 - 1306.
- [11] BARANIUK R. Compressive sensing[J]. IEEE Signal Processing Magazine, 2007, 24(4): 118 - 121.
- [12] DONOHO D. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution[J]. Communications on Pure and Applied Mathematics, 2006, 59(6): 797 - 829.
- [13] WRIGHT J, YANG A, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210 - 227.