

基于最佳鉴别变换的 HMM 手写数字字符识别

芮挺^{1,2)} 沈春林¹⁾ 丁健²⁾

¹⁾(南京航空航天大学自动化学院, 南京 210016) ²⁾(解放军理工大学工程兵工程学院, 南京 210007)

Qi TIAN

(德克萨斯大学计算机科学系圣安东尼奥 TX78249)

摘要 基于隐马尔可夫模型(HMM)的手写字符识别方法是近年来的一个研究热点,针对HMM编码稳定性和建模过程复杂的问题,提出了一种新方法,即采用统计不相关最佳鉴别变换对模式进行特征抽取和降维,获得最佳鉴别特征向量,并在此基础上对各最佳鉴别方向的投影结果进行编码,作为HMM的观测值序列,由于统计不相关最佳鉴别变换保证了变换特征向量集类内散布最小,类间散布最大的条件,使HMM编码的稳定性和模式的可分性得到明显改善,通过对美国国家邮政局USPS手写字库的识别实验证实了该算法的准确性和鲁棒性。

关键词 模式识别 手写字符识别 最佳鉴别变换 编码 隐马尔可夫模型

中图法分类号: TP391.43 **文献标识码**: A **文章编号**: 1006-8961(2004)08-1008-06

Handwritten Character Recognition Using HMM Based on Optimal Discriminant Transformation

RUI Ting^{1,2)}, SEHN Chun-lin¹⁾, DING Jian²⁾, Qi TIAN³⁾

¹⁾(College of Automation Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016)

²⁾(Engineering Institute of Engineering Corps, PLA University of Science & Technology, Nanjing 210007)

³⁾(Department of Computer Science, University of Texas at San Antonio, San Antonio, TX 78249)

Abstract Handwritten character recognition using the hidden Markov model (HMM) has been an active research topic for the past decade. One of the major problems, however, is that the handwritten characters may not exhibit consistent patterns due to different people's different writing styles. To enhance HMM's encoding stability and to reduce its modeling complexity, we propose a new approach in this paper. Specifically, we first obtain a set of uncorrelated optimal discriminant vectors by conducting feature extraction and dimension reduction using the uncorrelated Foley-Sammon transformation. Next, using a new feature space spanned by the optimal discriminant vectors, we obtain the projection coefficients of the raw data onto this new feature space. We then use these coefficients to form the observation sequence of the HMM. Because the uncorrelated Foley-Sammon transformation ensures minimum intra-class distance and maximum inter-class distance, it significantly improves HMM's encoding stability and difference classes' separability. In fact, the transformation allows different characters to be separable in many projection directions. To validate the accuracy and robustness of the proposed approach, we conduct experiments on the widely used US Postal Service (USPS) data set. Experiments show that the integration of the uncorrelated Foley-Sammon transformation and the HMM performs very well, achieving a recognition rate of 92%. It not only is better than regular HMM, but also is superior to the widely used neural network based approaches.

Keywords handwritten character recognition, optimal discriminant transformation, encode, hidden Markov models

1 引言

手写体字符识别是 OCR (optical character recognition) 的一个重要组成部分,尤其是手写阿拉伯

数字,由于在邮政编码、统计报表、财务报表、银行票据等方面的广泛用途,在过去的30多年里,其一直是图像处理、模式识别领域的研究热点。但是脱机手写数字字符的识别问题至今还是没有得到圆满的解决,主要原因在于书写风格和书写习惯的不同造成了各种字符

变形,使字符的模式特征变得不稳定和极其复杂。

隐马尔可夫模型(hidden Markov models, HMM)在许多处理序列化动态非平稳信号的问题中表现出优越的性能。Rabiner 等人首先将 HMM 作为语音信号的线性预测编码系数用于语音识别技术^[1],并取得了巨大成功;之后 Schuster 等人将 HMM 用于在线序列图像处理,识别手势变换^[2];Samaria 等人采用 HMM 进行人脸特征提取和分类识别^[3];Povlow 等人把 HMM 用于图像纹理分析^[4]。从 20 世纪末,HMM 方法开始应用到手写字符识别领域^[5],并显示出很大的潜力。

在以往的研究中,字符特征的编码方式一般是基于字符局部形态结构如边缘骨架^[6]或是局部像素特征^[7]。但由于手写字符本身结构有较大的随意性,使得在提取字符模式特征的稳定性和 HMM 模型码本(code book)的生成等方面还存在不少问题。

在模式识别领域,基于 Fisher 鉴别准则函数的最佳鉴别变换有着重要的地位和影响^[8]。其基本思想是在 Fisher 鉴别准则函数最大的条件下,求得一个最佳鉴别方向,然后将模式高维特征投影到该方向上,构成一维鉴别特征;Foley-Sammon 在此基础上通过求解与 Fisher 最佳鉴别方向正交的第 2 最佳鉴别方向提出了对两类问题的最佳鉴别平面^[9];Dochne 和 Leclercq 进一步完善了对多类问题的 Foley-Sammon 最佳鉴别矢量集计算方法^[10]。通过最佳鉴别变换,可以保证在各最佳鉴别投影方向上获得的各模式特征向量的类间散布最大,类内散布最小。同时由于抽取了最有效的模式特征,在很大程度上消除了模式特征的不稳定性。

鉴于此,提出针对手写字符结构不稳定的特点,不采用提取字符局部特征的方法,而是考虑字符的整体特征,首先通过在最佳鉴别方向上投影获得最佳鉴别矢量集,并在各最佳鉴别方向上对样本进行编码,生成码本;然后利用 Baum-Welch 算法建立各字符的 HMM 模型;最后通过 Forward-Backward 算法计算在测试样本观测序列下各字符 HMM 的条件概率,实现字符识别,其系统框图如图 1 所示。

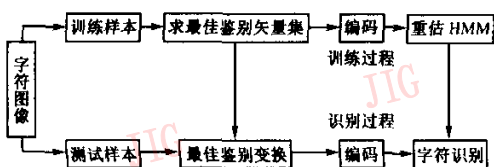


图 1 字符识别系统框图

2 最佳鉴别变换

2.1 Foley-Sammon 变换

设模式特征为 $X = \{x_i\}$ 的 N 维矢量,共有 m 个模式类别 w_1, w_2, \dots, w_m , 模式类内散布矩阵为 S_w , 类间散布矩阵为 S_b , 总体散布矩阵为 S_t 。

$$S_b = \sum_{i=1}^m P(w_i) (\mu_i - \mu_0) (\mu_i - \mu_0)^T \quad (1)$$

$$S_w = \sum_{i=1}^m P(w_i) E\{(x_i - \mu_i)(x_i - \mu_i)^T | w_i\} \quad (2)$$

$$S_t = S_b + S_w \quad (3)$$

式中, $P(w_i)$ 、 $E\{\}$ 、 μ_0 、 μ_i 分别为先验概率,协方差,全体平均矢量和类内平均矢量。

Fisher 鉴别准则函数定义如下:

$$F(\varphi) = \frac{\varphi^T S_b \varphi}{\varphi^T S_w \varphi} \quad (4)$$

其中, φ 为 N 维列向量。

使 Fisher 鉴别准则函数式(4)取得最大值的向量 φ_1 即为 Fisher 最佳投影方向,模式样本集在该方向上的投影保证具有最大的类间散布和最小的类内散布, φ_1 是广义特征方程式

$$S_b \varphi_1 = \lambda S_w \varphi_1 \quad (5)$$

中最大特征值所对应的特征向量。

在求出 $r(r \geq 1)$ 个最佳鉴别方向 $\varphi_1, \varphi_2, \dots, \varphi_r$ 后,第 $r+1$ 个最佳鉴别方向是在满足下式正交条件下的使式(4)取得最大值的向量 φ_{r+1}

$$\varphi_{r+1}^T \varphi_i = 0 \quad (i = 1, 2, \dots, r) \quad (6)$$

在模式识别理论中,特征抽取的一般原则是所抽取特征间的统计相关性越小越好,最好是抽取不相关的特征,在 Foley-Sammon 最佳鉴别变换中,任意两个最佳鉴别方向是正交的,但投影所得到两个鉴别特征却可能是统计相关的。为了获得具有统计不相关的最佳鉴别矢量集,文献[11]提出一种新的方法。

2.2 统计不相关的最佳鉴别矢量集

在求出广义特征方程式(5)的最大特征值对应的特征向量作为第 1 最佳鉴别方向 φ_1 后,为了得到统计不相关的鉴别特征,第 $r+1(r \geq 1)$ 个最佳鉴别方向可取在满足下式共扼正交条件下的使式(4)取得最大值的向量 φ_{r+1}

$$\varphi_{r+1}^T S_t \varphi_i = 0 \quad (i = 1, 2, \dots, r) \quad (7)$$

由于有共扼正交条件式(7),使投影后的最佳鉴别矢量集中任意两个鉴别特征向量统计不相关。

统计不相关最佳鉴别方向 φ_{r+1} 的解是广义特征方程式

$$PS_b\varphi_{r+1} = \lambda S_w\varphi_{r+1} \quad (8)$$

中最大特征值对应的特征向量。
式(8)中

$$P = I - S_r D^T (D S_r S_w^{-1} S_r D^T)^{-1} D S_r S_w^{-1} \quad (9)$$

这里, I 为单位矩阵

$$D = [\varphi_1 \quad \varphi_2 \quad \dots \quad \varphi_r]^T \quad (10)$$

对于一个 N 维向量 x , 通过下式在 $k(k < N)$ 个最佳鉴别方向上投影可以获得 k 维特征向量 γ 。

$$\gamma = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} \varphi_1^T \\ \varphi_2^T \\ \vdots \\ \varphi_k^T \end{bmatrix} x \quad (11)$$

通过以上的统计不相关最佳鉴别变换, 就可以得到降维后的鉴别特征向量。

3 识别模型的建立

在得到鉴别特征向量后, 通常的选择是用神经网络建立识别模型。手写字符模式特征的稳定性在投影变换后虽有改善, 但神经网络仍难以描述其复杂的变化。因此, 采用对动态不平稳模式具有良好描述性能的 HMM 建立识别模型。

3.1 HMM 字符识别方法

隐马尔可夫过程的基本思想是用双随机过程来描述一个模式^[1], 一是 Markov 链, 这是基本随机过程, 它描述模式内部状态序列; 二是描述状态值和观测值之间关系的随机过程。因此它对复杂系统的描述能力比单纯的 Markov 模型要强大得多^[1]。从观察者的角度只能看到观测值, 而看不到内部状态, 只能通过反映观测值和内部状态之间的随机过程去感知状态的存在及其特性。

一阶隐马尔可夫过程可用隐马尔可夫模型(HMM)描述, 隐马尔可夫模型由 5 个参数描述并确定: (1) 模型中 Markov 链的状态数目 N ; (2) 各状态对应观测值的数目 M ; (3) 起始状态分布概率 Π ; (4) 状态转移概率矩阵 A ; (5) 观测值概率矩阵 B 。因此一个 HMM 可以记为: $\lambda = (N, M, \Pi, A, B)$ 或简写为 $\lambda = (\Pi, A, B)$ 。

在 HMM 的应用中主要涉及以下 3 个关键问题:

(1) 估值问题, 即已知模型和一个观测序列, 计算由该模型产生出该观测序列的概率;

(2) 揭示模型内部状态, 即通过观测序列估计模型内部状态序列;

(3) 模型训练问题, 即通过观测序列样本建立最佳模型。这 3 个问题可分别由 Forward-Backward 算法、Viterbi 算法和 Baum-Welch 算法解出^[1]。

本文算法涉及问题 1 和问题 3, 其具体步骤如下:

(1) 按码本对经过统计不相关最佳鉴别变换的样本进行编码, 得到各样本的观测值序列;

(2) 根据多个训练样本的观测值序列, 通过问题 3 建立 10 个字符的 HMM;

(3) 对于一个测试样本输入, 通过问题 1 分别求得对于各 HMM 的条件概率 $P(O|\lambda_i)$, 其中最大值对应的模型即为识别结果。

因此, 该算法中的关键问题就是如何生成码本。

3.2 生成码本

模式识别理论认为模式特征都存在一个最佳维数, 当模式特征的维数等于最佳维数时, 模式可以获得最佳的可行性; 在模式特征的维数较低时, 增加特征有利于提高识别率; 在模式特征的维数较高时, 增加特征反而使识别率降低。根据 Fukunaga 提出的维数定理^[12], 对于 m 个类别的模式识别问题, $m-1$ 个后验概率函数是理想的分类特征集, 同时实验表明广义特征方程式(8)由大到小的前 $m-1$ 个特征值对应的特征向量均为实数解。因此将对于 10 个手写字符统计不相关最佳鉴别变换定义为 $10-1 (= 9)$ 个投影方向, 而这 9 个方向的投影结果可作为 HMM 的观测值序列。

由于满足 Fisher 最佳鉴别准则, 在同一投影方向的尺度空间, 各类模式均满足类间散布距离最大, 类内散布距离最小的条件。基于此, 提出在 9 个投影方向上分别编码生成码本的编码方法。例如在 φ_i 方向上, 设样本均值投影距离最大的两类模式为 M_{max} 和 M_{min} , 则以 M_{max} 和 M_{min} 为边界将投影线段均分为 14 段, 并以 2~15 编码, 样本投影所在编码段即为样本在 φ_i 方向上的编码, 若样本投影小于 M_{min} 或大于 M_{max} , 则其编码为 1 或 16。这样在每个方向上都可以获得长度为 16 位的码本, 如图 2 所示。选择距离最大的两类样本均值的投影作为边界是为了消除个别奇异样本可能造成编码段过大的问题。

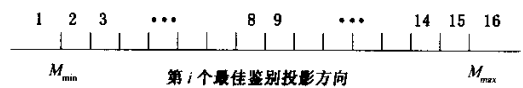


图 2 在 φ_i 方向上生成码本

例如字符“0”的 5 个样本在 ϕ_1 上的投影结果分别为 0.044 6, 0.045 5, 0.044 6, 0.044 2, 0.043 6; 该方向上 $M_{min}=0.030 6, M_{max}=0.045 8$, 则 5 个样本在 ϕ_1 上的编码分别为 14, 15, 14, 14, 14, 编码结果非常稳定。

可以看出, 在统计不相关最佳鉴别方向上利用投影结果进行编码的方式可以保证同类模式样本编码的稳定性和不同类样本编码的最大差异。

4 实验结果与分析

实验使用美国国家邮政局数据库 (US postal service database) 收集的手写数字字符。该字符数据库中包括训练样本集 (USPS training data) 和测试样本集 (USPS testing data) 两部分, 其中每个样本都只经过简单的预处理并归一化为 $16 \times 16 (=256)$ 像素的灰度图。图 3 中是该字库 3 组字符的图像, 可以看出该字库中各字符笔画的形态、粗细和灰度等级的差别都很显著。

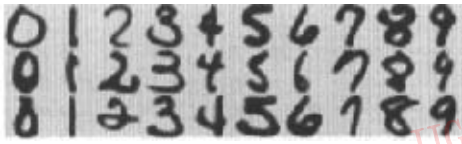


图 3 USPS 手写字库样本

为了验证算法的鲁棒性, 实验中未对 USPS 字符集做进一步预处理。实验分两部分, 一是检验经过不相关最佳鉴别变换后抽取特征向量的有效性; 二是检验基于最佳鉴别矢量的 HMM 方法的有效性, 并对比 HMM 方法和神经网络方法的识别率。

4.1 最佳鉴别变换的有效性

首先将样本像素点的灰度值进行了归一化, 这样可以在一定程度上改善字符灰度的一致性。

根据前面的分析, 实验中将 USPS training data 中的全部样本作为训练样本得到 9 个统计不相关最佳鉴别方向。在 USPS testing data 中, 从每类字符中随机选取 30 个样本, 共 $30 \times 10 (=300)$ 个测试样本作投影。图 4 是字符“1”的 30 个测试样本分别在 9 个最佳鉴别方向上的投影结果, 图中横坐标表示 9 个最佳鉴别方向, 纵坐标表示 30 个 256 维向量投影后的数值分布。从图中可以看出, 30 个 256 维向量投影后在各投影方向上都密集地聚集在一起, 表明具有良好的 consistency。

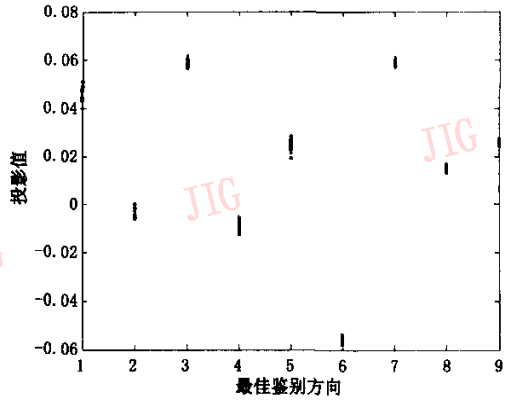
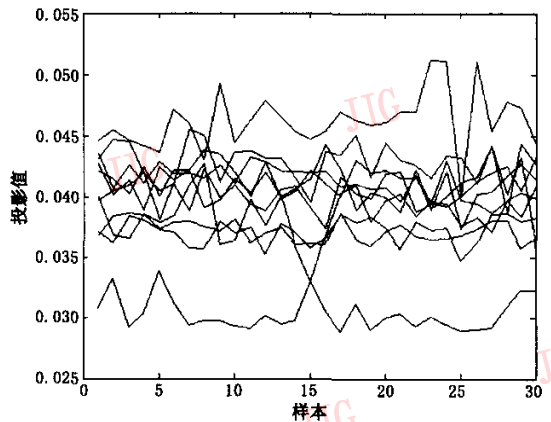
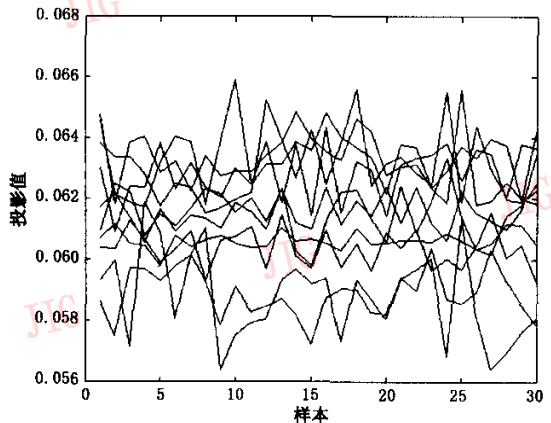


图 4 30 个同类字符“1”的投影结果

图 5 显示的是 0~9 每组 30 个的测试样本在不同最佳鉴别方向上投影后的分布, 图中横坐标表示 30 个测试样本, 纵坐标表示样本在最佳鉴别方向上的投影结果。由于篇幅所限, 在图 5 中只显示了在 ϕ_1, ϕ_2 两个最佳鉴别方向上的投影分布。



(a) 样本在 ϕ_1 上的投影分布



(b) 样本在 ϕ_2 上的投影分布

图 5 样本在最佳鉴别方向上的投影分布

从图5中可以看出在最佳鉴别方向上各类样本的分布都是在其样本均值投影的附近做窄幅震荡,并且在其中的任意一个方向上有2类以上的模式是完全可分的。

实验证明测试样本通过最佳鉴别方向投影后得到的9维特征向量使模式的稳定性和可分性得到了很大的改善。

4.2 HMM 识别方法的有效性

在手写数字字符识别实验中,训练样本集是从USPS training data中每类选取的150个共 $150 \times 10 (= 1500)$ 个样本;测试样本集为USPS testing data中每类选取的50个共 $50 \times 10 (= 500)$ 个样本;9个最佳鉴别方向由USPS training data全部样本训练获得。

在9个最佳鉴别方向上按上节中提出的方法进行16位码本的编码,并分别按Baum-Welch算法建立0~9字符的隐马尔可夫模型。

实验中HMM状态数 N 取6,实验结果共有462个字符得到正确的识别。表1中是 $N=6$ 时部分测试样本在各模型下按Forward-Backward算法得到的 $\log likelihood$,其中 $likelihood$ 为似然值。

从表1中结果可以看出在大多数情况下,字符样本在其所属HMM下的 $\log likelihood$ (表1中黑体部分)比在非其所属HMM下的 $\log likelihood$ 要大得多。表明样本模式在HMM中具有良好的可分性。在分析识别错误的字符时发现,出错的字符如: $\Psi(4)$, $\Psi(4)$, $\Psi(8)$ 字型都和其自身所属类的字型差别较大,同时又和某类其他字符相近,导致识别错误。

表2所示是用传统的神经网络法,以相同的训练样本和测试样本进行识别测试的实验结果对比。实验中使用3层前向神经网络,输入层节点为9,隐含层节点为19,输出层节点为10,采用加入动量项的学习速率渐小算法,其中动量项 $\alpha=0.7$,初始学习率 $\eta_0=0.01$,网络收敛误差 $\varepsilon=0.001$ 。

实验对比说明HMM对手写字符这类不稳定模式的描述能力明显好于神经网络。这主要是因为HMM用双随机过程来描述一个模式,并通过状态转移概率矩阵和观测值概率矩阵将双随机过程联系起来,通过多样本序列的训练,其内部状态和观测值丰富的变化使其有能力对不平稳的复杂系统进行细致的描述。

因此对于手写字符这类不稳定模式,通过抽取最佳鉴别特征后,用HMM达到了比较理想的识别效果。

表1 实验中部分样本的测试结果

类别	样本	Model0	Model1	Model2	Model3	Model4	...
字符0	1	-21.564	-29.429	-32.714	-28.101	-33.858	...
	2	-21.769	-31.851	-31.804	-33.686	-33.317	...
	3	-22.648	-33.502	-27.247	-32.584	-31.317	...
	4	-21.551	-37.495	32.179	-28.932	-30.283	...

字符1	1	-29.434	-15.515	-30.160	-32.367	-30.123	...
	2	-26.253	-16.839	-29.595	-30.134	-27.988	...
	3	-33.058	-23.616	-27.414	-36.292	-34.462	...
	4	-26.042	-21.783	-28.224	-35.975	-32.311	...

字符2	1	-27.944	-42.352	-24.678	-27.980	-36.499	...
	2	-25.772	-42.742	-22.833	-26.865	-38.073	...
	3	-33.348	-35.125	-26.205	-28.629	-30.758	...
	4	-30.955	-41.310	-25.293	-30.586	-39.611	...

字符3	1	-32.920	-39.053	-27.524	-22.938	-25.049	...
	2	-31.542	-36.321	-34.720	-26.624	-34.791	...
	3	-32.535	-38.465	-25.652	-23.120	-35.623	...
	4	-30.797	-39.425	-29.002	-18.802	-30.661	...

字符4	1	-28.548	-30.021	-31.440	-35.785	-17.307	...
	2	-27.677	-29.483	-26.015	-35.075	-15.027	...
	3	-33.634	-42.403	-24.651	-29.297	-22.081	...
	4	-34.745	-35.672	-32.755	-30.900	-21.790	...

:	:	:	:	:	:	:	

表2 实验结果对比

	测试集	正确识别字符数	识别率(%)
HMM	500	462	92
BPN	500	421	84

5 结论

本文针对手写字符书写随意,模式特征不稳定的特点,提出将最佳鉴别变换和HMM相结合解决识别问题的新方法,并提出了一种简单而又稳定的码本生成方式。

首先从字符整体特征出发,通过最佳鉴别变换降维和抽取特征,获得相对稳定和具有较好可分性的特征向量;然后在各最佳鉴别方向上分别编码;最后采用对不平稳序列具有很好描述能力的HMM进行字符识别。实验证实本文算法具有良好的准确性和鲁棒性。本文算法不仅可以用于手写数字字符

识别,也可以推广到其他手写字符识别中。

本文中字符图像预处理方法和 HMM 算法未作完善的优化。进一步的工作主要在两方面,一是研究改善字符形态、笔画粗细一致性的图像预处理方法,以期获得更稳定的模式特征;二是研究对 HMM 初始模型的选择和 HMM 相关算法优化的方法。

参 考 文 献

- 1 Rabiner L R. A tutorial on hidden Markov models and selected application in speech recognition[J]. Proceedings of IEEE, 1989,77(2) :257~286.
- 2 Schuster M, Rigoll G. Fast on-line video image sequence recognition with statistical methods[A]. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing[C]. New York, 1996:450~454.
- 3 Samaria F, Fallside F. Face identification and feature extraction using hidden Markov models [A]. In: Image Processing Theory and Applications[C]. Amsterdam; Elsevier Science Publishers, 1993:295~302.
- 4 Povlow B. R, Dunn S M. Texture classification using noncausal Hidden Markov Models [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1995,17(10):295~302.
- 5 Agazzi, Kuo O E, Levin S S, et al. Connected and degraded text recognition using planar hidden Markov models [A]. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing[C], Minneapolis, USA 1993, 5: 113~116.
- 6 张引, 潘云鹤. 工程图纸自动输入字符识别的二维隐马尔可夫模型方法[J]. 计算机辅助设计与图形学学报, 1999,11(5): 403~406.
- 7 冯兵, 丁晓青, 吴佑寿. HMM 方法识别脱机手写汉字[J]. 模式识别与人工智能, 2002,15(1):84~88.
- 8 Duda R O, Hart P E, Stork D G. Pattern Classification (second edition)[M], New Jersey: John Wiley & Sons, Inc., 2001.
- 9 Foley D H, Sammon J W. An optimal set of discriminant vectors [J]. IEEE Transactions on Computers, 1975,24(3): 281~289.

- 10 Duchene J, Lecereq S. An optimal transformation for discriminant and principal component analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1988,10(6) :978~983.
- 11 金忠, 杨静宇, 胡钟山. 有效最佳鉴别特征的抽取与维数问题 [J]. 计算机学报. 2000,23(1):108~110.
- 12 Fukunaga K. Introduction to Statistical Pattern Recognition [M]. New York: Academic Press, 1990.

芮 挺 1972 年生。2001 年于解放军理工大学获机械工程博士学位,现于南京航空航天大学自动化学院博士后流动站从事研究工作。研究方向为图像处理和模式识别。

E-mail: rting@public1.ptt.js.cn



沈春林 1935 年生。教授、博士生导师,美国纽约科学院院士,美国 IEEE 学会高级会员。研究方向为超低空飞行及其低空突防技术、精确制导与打击技术、智能控制技术、网络与控制技术。



丁 健 1973 年生。1998 年于解放军理工大学获硕士学位,现于中国科学院地理湖泊研究所攻读博士学位。研究方向为地理信息系统、图像处理。



Qi TIAN 1971 年生。2002 年于伊利诺大学-香槟分校 (University of Illinois at Urbana-Champaign) 获博士学位,现为德克萨斯大学 (University of Texas at San Antonio) 计算机科学系助理教授,美国 IEEE 学会高级会员。研究方向为多媒体信息检索、数字信号/图像处理与计算机视觉。

