

Style Context with Second-Order Statistics

Sriharsha Veeramachaneni and George Nagy, *Life Fellow, IEEE*

Abstract—Patterns often occur as homogeneous groups or fields generated by the same source. In multisource recognition problems, such isogeny induces statistical dependencies between patterns (termed *style context*). We model these dependencies by second-order statistics and formulate the optimal classifier for normally distributed styles. We show that model parameters estimated only from pairs of classes suffice to train classifiers for any test field length. Although computationally expensive, the style-conscious classifier reduces the field error rate by up to 20 percent on quadruples of handwritten digits from standard NIST data sets.

Index Terms—Interpattern feature dependence, writer consistency, continuous styles, quadratic discriminant classifier.

1 INTRODUCTION

WHEN patterns are presented as groups (or *fields*) to a classifier, statistical dependencies between the features of different patterns can be exploited to improve recognition accuracy over that of a *singleton* classifier that operates on the patterns one at a time. Linguistic context (interpattern class dependence) has been widely used to improve classification accuracy. Feature dependence between adjacent patterns is another commonly occurring context which can be caused by ligatures in cursive handwriting and by coarticulation in speech and is dependent on the relative position of the patterns. Another kind of interpattern feature dependence is present in multisource recognition problems owing to the commonality of origin (*isogeny*) of test fields. Such dependence is usually independent of the order of the patterns in the field and is called *style context*. Whereas feature dependence between adjacent patterns occurs even in single-source recognition problems, style context arises only when multiple sources are present. This communication presents new models and algorithms that exploit style context for field classification.

The single-source interpattern feature context has been widely studied and utilized. Such a context can usually be modeled as Markov dependence of pattern features on the features of the previous pattern [1].

When interpattern feature dependence is due to ligatures, coarticulation, etc., the order in which the patterns occur is important. However, when the feature dependencies arise only because of the isogeny of the patterns, the relative order of the patterns in the field should not affect the classification decision. This condition disallows the modeling of style context by Markov processes.

It is clear that, even in the absence of any linguistic context, spatial context, i.e., the stationary nature of typeset, typeface and shape deformations over a long sequence of symbols, can be used to improve classification accuracy [2].

Human readers achieve higher accuracy on word groups from single writers than on word groups from different writers [3]. Hong and Hull use visual relations between images, i.e., similarities between words or word-parts, to increase word recognition accuracy [4].

Sarkar and Nagy model styles by multimodal distributions with two layers of hidden mixtures, corresponding to styles and variants [5], [6], [7]. The parameters of the mixture distributions are estimated using the EM algorithm. Although, under the assumption that there are only a few styles with Gaussian style-conditional class distributions, this approach is optimal, the complex EM estimation stage is prone to small-sample errors. Furthermore, in some problems, such as handwritten text recognition, the assumption of a limited number of discrete styles is questionable.

Under essentially the same assumptions as ours, Kawatani attempts to utilize style consistency by identifying “unnaturalness” in test patterns with respect to other patterns by the same writer [8]. His criteria are based on distances to other patterns classified into the same class or correlations with patterns classified into different classes. This method is similar in principle to our proposed method. However, our method replaces the above heuristic criteria by a unified method of classification using all the information available. Koshinaka et al. model style context as the dependence between subcategory labels of adjacent patterns [9].

Some multisource recognition methods attempt to improve accuracy by extracting features that are invariant to source-specific peculiarities such as size, slant, skew, etc., [10]. Dehkordi et al. extract principal components based only on class means to determine directions which have the most interclass variability [11]. Although such strategies are suitable for pattern classifiers operating on one pattern at a time, correlations between intraclass variations contain useful information for style-constrained word recognition.

Another approach to multifont recognition is first recognizing the font of a document and performing font-specific classification. The font recognizer developed by Zramdini and Ingold uses typographical features such as weight, size and slope of the text [12]. Shi and Pavlidis propose extracting information from two sources—global page properties (which help to distinguish fixed-pitch from variable-pitch fonts) and from short-word recognition (that helps to distinguish between serif and sans-serif fonts). The

• S. Veeramachaneni is with the Automated Reasoning Division, Istituto per la Ricerca Scientifica e Tecnologica, Via Sommarive 18, Povo, Trento 38050, Italy. E-mail: sriharsha@itc.it.

• G. Nagy is with the Department of Electrical, Computer, and Systems Engineering, 6020 Johnsson Engineering Center, Rensselaer Polytechnic Institute, 110 8th street, Troy, NY 12180. E-mail: nagy@ecse.rpi.edu.

Manuscript received 25 Oct. 2002; accepted 21 May 2004.

Recommended for acceptance by L. Vincent.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 117670.

knowledge of the font family is then used to guide the text recognition [13].

Bazzi et al. observe that, because of the assumption that successive observations are statistically independent, HMM methods have higher error rate for rare styles (because, due to the independence assumption, the infrequent styles are exponentially penalized) [14]. They propose modifying the mixture of styles in the training data to overcome this effect. In contrast, our method uses the a priori style probabilities during classification.

There are, of course, many factors that impact text recognition accuracy: segmentation errors, variability in character shape, image quality, and the extent of linguistic context. For classification, the resulting distribution of features has been modeled most successfully by means of Gaussian and composite Gaussian multivariate density functions [15], [16]. The discriminant functions are then completely determined by the first and second-order statistics (mean vectors and covariance matrices). While higher-order statistical dependences may occur, estimating even second-order dependences among features requires large training sets for the high-dimensional feature spaces customary in character recognition. Language models that govern the sequence of pattern classes can be readily incorporated into the Gaussian statistical framework, typically through Hidden Markov Methods. Although we have not applied a language model, their use is entirely compatible with style-constrained classification.

In OCR, touching character sequences are usually processed by integrating segmentation with recognition [17], [18]. The classifier determines the most probable path through a trellis of oversegmented character images. Therefore, the classification itself is still based on isolated patterns and style constraints will raise the accuracy of the combined scheme.

In Section 2, we introduce precise notation for field classification and extend the Gaussian quadratic discriminant function to fields of isogenous patterns generated by discrete sources. In Section 3, we develop a model for sources with normally distributed means that appears appropriate for handwriting. Experimental results on fields of NIST data sets are reported in Section 4. Our conclusions and future work are in Section 5.

2 STYLE-CONSCIOUS QUADRATIC FIELD CLASSIFIER—DISCRETE SOURCES

In the first two sections, we formally define the problem of style-conscious field classification when homogeneous test fields are generated by one of several discrete sources. In Sections 2.3, 2.4, and 2.5, we describe the proposed style-conscious quadratic discriminant classifier. The characteristics of various field classifiers for this problem are explored through an extended example.

2.1 Problem Statement

We consider the problem of classifying an isogenous input field-feature vector $\mathbf{y} = (\mathbf{x}_1^T, \dots, \mathbf{x}_L^T)^T$ (each \mathbf{x}_i represents d feature measurements for one of L patterns in the field) produced by one of S sources s_1, s_2, \dots, s_S (writers, fonts, etc.). More than one source can be from the same style. Let $\mathcal{C} = \{c_1, \dots, c_N\}$ be the set of singlet-class labels. Let c^i represent the class of the i th pattern

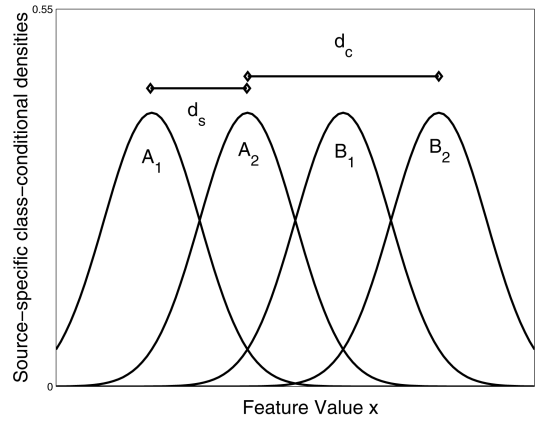


Fig. 1. The source-and-class-conditional feature distributions—discrete sources.

of the field.¹ We make the following assumptions on the class and feature distributions:

1. $p(s_k | \mathbf{c}^1, \mathbf{c}^2, \dots, \mathbf{c}^L) = p(s_k) \forall k = 1, \dots, S$. That is, any linguistic context (interpattern class dependence) is source independent. For multiwriter word recognition, this assumption implies that the handwriting style of a writer does not depend on her vocabulary.
2. $p(\mathbf{y} | \mathbf{c}^1, \mathbf{c}^2, \dots, \mathbf{c}^L, s_k) = p(\mathbf{x}_1 | \mathbf{c}^1, s_k) p(\mathbf{x}_2 | \mathbf{c}^2, s_k) \dots p(\mathbf{x}_L | \mathbf{c}^L, s_k) \forall k = 1, \dots, S$. The features of each pattern in the field are class-conditionally independent of the features of every other pattern in the same field. For multifont zip code recognition, this assumption implies that, for the zip code “12180” in a particular font, the noise in the “2” is independent of the noise in the “0.”

Under the above assumptions, for the field-feature vector $(\mathbf{x}_1^T, \dots, \mathbf{x}_L^T)^T$ and field-class $(c_{i_1}, c_{i_2}, \dots, c_{i_L})$, it can be shown that, for any function $f(\cdot)$ and $l = 1, \dots, L$,

$$E\{f(\mathbf{x}_l) | c_{i_1}, c_{i_2}, \dots, c_{i_L}\} = E\{f(\mathbf{x}_l) | c_{i_l}\}. \quad (1)$$

This result will be useful for deriving the formulae for the field-class-conditional means and covariance matrices.

To illustrate the problem of style-constrained classification and contrast various classifiers we reexamine the example studied by Sarkar in [5].

Example 1. Consider the problem of classifying test fields of length L generated by one of two sources $\{s_1, s_2\}$. The possible singlet-class labels are $\mathcal{C} = \{A, B\}$. For simplicity, we assume that the classes are equally likely (i.e., $p(A) = p(B) = 1/2$) and no linguistic dependence is present (i.e., $p(AA) = p(A)p(A)$ etc.). Additionally, we assume that the sources are equally likely (i.e., $p(s_1) = p(s_2) = 1/2$).

We ensure that the style-specific-class-conditional feature distributions are unidimensional Gaussian and are configured, as shown in Fig. 1. $\{A_1, B_1\}$ represent the

1. This egregious notation avoids nested subscripts. If the second pattern of the field belongs to the fifth class, then it is denoted $c^2 = c_5$.

class-conditional densities for A and B from source s_1 , and $\{A_2, B_2\}$ represent those from source s_2 . The feature distributions are

$$(x|A, s_1) \sim N(\mu_{A_1}, \sigma^2), (x|A, s_2) \sim N(\mu_{A_2}, \sigma^2),$$

$$(x|B, s_1) \sim N(\mu_{B_1}, \sigma^2) \text{ and } (x|B, s_2) \sim N(\mu_{B_2}, \sigma^2).$$

The parameters d_c and d_s control the interclass and interstyle distances. Since only the relative positions of the means matter, the means were fixed as follows:

$$\mu_{A_1} = 0, \quad \mu_{A_2} = d_s, \quad \mu_{B_1} = d_c \text{ and } \mu_{B_2} = d_c + d_s.$$

For $L = 2$, the field-class-conditional field-feature distributions have two components—one for each source.

$$p(x_1, x_2|c_1, c_2) = (1/2)\{p(x_1|c_1, s_1)p(x_2|c_2, s_1) + p(x_1|c_1, s_2)p(x_2|c_2, s_2)\} \text{ for } c_1, c_2 \in \{A, B\}.$$

Similarly, for any field length, the field-class-conditional feature distributions are bimodal composite Gaussian.

2.2 The Discrete-Style (DS) and the Style-First (SF) Field Classifiers

If the source-and-class-conditional feature distributions are known, the *optimal* field classification strategy is to assign the label (c_1^*, \dots, c_L^*) to the test field feature vector \mathbf{y} , where

$$\begin{aligned} (c_1^*, \dots, c_L^*) &= \underset{(c_1, \dots, c_L) \in \mathcal{C}^L}{\operatorname{argmax}} \{p(\mathbf{y}|\mathbf{c}^1 = c_1, \dots, \mathbf{c}^L = c_L) \\ &\quad p(\mathbf{c}^1 = c_1, \dots, \mathbf{c}^L = c_L)\} \\ &= \underset{(c_1, \dots, c_L) \in \mathcal{C}^L}{\operatorname{argmax}} p(c_1, \dots, c_L) \\ &\quad \sum_{m=1}^S p(\mathbf{x}_1|c_1, s_m) \dots p(\mathbf{x}_L|c_L, s_m)p(s_m). \end{aligned} \quad (2)$$

Such a classifier, being optimal for discrete styles, will henceforth be referred to as the *discrete-style* (DS) field classifier. In reality, the feature distributions are seldom known and have to be estimated from training data, implying the need for sufficient samples per class for each style. Sarkar attempts to alleviate this requirement by modeling the data as being drawn from much fewer styles [5].

Another straightforward approach for exploiting the isogeny of pattern fields is to *first* recognize the source label from the field and use the particular source-specific decision boundaries to classify the patterns in the field. This classifier will hence be referred to as the *style-first* (SF) field classifier.² The SF classifier first identifies the style according to

$$\begin{aligned} s^* &= \underset{s_k}{\operatorname{argmax}} p(s_k|\mathbf{y}) = \underset{s_k}{\operatorname{argmax}} p(s_k|\mathbf{x}_1, \dots, \mathbf{x}_L) \\ &= \underset{s_k}{\operatorname{argmax}} \{p(\mathbf{x}_1, \dots, \mathbf{x}_L|s_k)p(s_k)\} \end{aligned} \quad (3)$$

2. The style-first classifier is different from Sarkar's *label-style* (LS) classifier [5], which essentially classifies each field by S style-specific classifiers and chooses the label assigned by the style with the maximum field-feature likelihood (weighted by the a priori style probability). The LS classification rule is

$$(c_1^*, \dots, c_L^*) = \underset{(c_1, \dots, c_L) \in \mathcal{C}^L}{\operatorname{argmax}} p(c_1, \dots, c_L) \max_{s_m} p(\mathbf{x}_1|c_1, s_m) \dots p(\mathbf{x}_L|c_L, s_m)p(s_m).$$

and then classifies the test field according to

$$(c_1^*, \dots, c_L^*) = \underset{(c_1, \dots, c_L) \in \mathcal{C}^L}{\operatorname{argmax}} p(c_1, \dots, c_L) p(\mathbf{x}_1|c_1, s^*) \dots p(\mathbf{x}_L|c_L, s^*). \quad (4)$$

In general, the style-first classifier has a higher field error rate than the discrete-style classifier because of errors in style identification.

2.3 Quadratic Discriminant Field Classifier

In contrast to the above classifiers, we propose a quadratic discriminant field classifier that models style context in a field as correlations between normally distributed singlet feature vectors. This classifier, which is a natural extension of the singlet Gaussian quadratic discriminant classifier, will henceforth be referred to as the *style-conscious quadratic discriminant field* (SQDF) classifier.

We will first derive the expressions for field-class-conditional means and covariance matrices. For now, let $L = 2$ and $\mathbf{c} = (c_i, c_j)^T$ be a field-class label.

The mean vector for the field-class \mathbf{c} is given by

$$\begin{aligned} \mu_{ij} &= E\{\mathbf{y}|\mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\} = \begin{pmatrix} E\{\mathbf{x}_1|\mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\} \\ E\{\mathbf{x}_2|\mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\} \end{pmatrix} \\ &= \begin{pmatrix} E\{\mathbf{x}_1|c_i\} \\ E\{\mathbf{x}_2|c_j\} \end{pmatrix} \text{ from (1)} \\ &\triangleq \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}. \end{aligned} \quad (5)$$

Therefore, the field-class-conditional mean vector can be constructed by concatenating the component singlet-class-conditional mean vectors.

Let us now compute the field-class-conditional covariance matrix for the class $\mathbf{c} = (c_i, c_j)^T$, which we will denote by \mathbf{K}_{ij} .

$$\begin{aligned} \mathbf{K}_{ij} &= E\{(\mathbf{y} - E\{\mathbf{y}|\mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\}) \\ &\quad (\mathbf{y} - E\{\mathbf{y}|\mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\})^T\} \\ &= E\{(\mathbf{x}_1^T, \mathbf{x}_2^T)^T (\mathbf{x}_1^T, \mathbf{x}_2^T) | \mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\} \\ &\quad - E\{(\mathbf{x}_1^T, \mathbf{x}_2^T)^T | \mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\} \\ &\quad E\{(\mathbf{x}_1^T, \mathbf{x}_2^T) | \mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\} \\ &\triangleq \begin{pmatrix} C_i & C_{ij} \\ C_{ji} & C_j \end{pmatrix}, \end{aligned} \quad (6)$$

where

$$\begin{aligned} C_i &= E\{\mathbf{x}_1 \mathbf{x}_1^T | c_i\} - E\{\mathbf{x}_1 | c_i\} E\{\mathbf{x}_1^T | c_i\}, \\ C_{ij} &= E\{\mathbf{x}_1 \mathbf{x}_2^T | \mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\} - E\{\mathbf{x}_1 | c_i\} E\{\mathbf{x}_2^T | c_j\}, \end{aligned} \quad (7)$$

and, similarly,

$$C_{ii} = E\{\mathbf{x}_1 \mathbf{x}_2^T | \mathbf{c}^1 = c_i, \mathbf{c}^2 = c_i\} - E\{\mathbf{x}_1 | c_i\} E\{\mathbf{x}_2^T | c_i\}.$$

Thus, \mathbf{K}_{ij} can be written as an $L \times L$ block matrix with $d \times d$ blocks (d is the singlet-pattern feature dimensionality), where the diagonal blocks are just the class-conditional singlet covariance matrices. It can be shown that the above derivations generalize to longer fields. In the general case,

the mean $\mu_{i,j,\dots,k}$ and covariance matrix $\mathbf{K}_{i,j,\dots,k}$ for the field-class $\mathbf{c} = (c_i, c_j, \dots, c_k)^T$ are given by

$$\mu_{i,j,\dots,k} = \begin{pmatrix} \mu_i \\ \mu_j \\ \vdots \\ \mu_k \end{pmatrix}; \mathbf{K}_{i,j,\dots,k} = \begin{pmatrix} C_i & C_{ij} & \dots & C_{ik} \\ C_{ji} & C_j & \dots & C_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ C_{ki} & C_{kj} & \dots & C_k \end{pmatrix}.$$

Hence, the means and covariance matrices for field-classes of arbitrary length can be constructed from the singlet-class-conditional means, the N singlet-class-conditional covariance matrices C_1, C_2, \dots, C_N , and the $N(N+1)/2$ "cross-covariance" matrices (since $C_{ij} = C_{ji}^T$) $C_{11}, C_{12}, \dots, C_{NN}$. This means that the estimation of field-class-conditional covariance matrices for long fields is no more demanding than for $L = 2$.

2.4 Estimation of the Classifier Parameters

Here, we present expressions from which the estimators for the field-class-conditional means and covariance matrices, given class and source labeled data, can be derived by replacing all expectations by the corresponding sample averages.

The field-class-conditional means as shown in (5) are formed just by concatenating the appropriate singlet-class means estimated from data (neglecting source labels). As mentioned earlier, the diagonal blocks of the field-class-conditional covariance matrices, C_i , are the class-conditional singlet-class covariance matrices. They will be estimated from the weighted sum of the source-and-class-conditional "power" matrices.

$$C_i = E\{\mathbf{x}_1 \mathbf{x}_1^T | c_i\} - \mu_i \mu_i^T = \sum_{k=1}^S p(s_k) E\{\mathbf{x}_1 \mathbf{x}_1^T | c_i, s_k\} - \mu_i \mu_i^T. \quad (8)$$

In general, the accurate estimation of parameters that describe style context (here, C_{ij} , the off-diagonal "cross-covariance" matrices) requires a large number of field-samples for each field-class. We will show that the assumptions made in Section 2.1 simplify the estimation of the cross-covariance matrices. From (7),

$$\begin{aligned} C_{ij} &= E\{\mathbf{x}_1 \mathbf{x}_2^T | \mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j\} - E\{\mathbf{x}_1 | c_i\} E\{\mathbf{x}_2^T | c_j\} \\ &= \sum_{k=1}^S p(s_k) E\{\mathbf{x}_1 \mathbf{x}_2^T | \mathbf{c}^1 = c_i, \mathbf{c}^2 = c_j, s_k\} - \mu_i \mu_j^T \\ &= \sum_{k=1}^S p(s_k) E\{\mathbf{x}_1 | c_i, s_k\} E\{\mathbf{x}_2 | c_j, s_k\}^T - \mu_i \mu_j^T \\ &= \sum_{k=1}^S p(s_k) \mu_i^{(k)} \mu_j^{(k)T} - \mu_i \mu_j^T. \end{aligned} \quad (9)$$

Thus, the cross-covariance matrices can be computed from the estimates of the source-specific singlet-class means ($\mu_i^{(k)} = E\{\mathbf{x} | c_i, s_k\}$). Since the cross-covariance matrices encapsulate the style information, it is intuitively appealing that the existence of any style context is determined only by the variation of class means across sources and not by variation within a source. Furthermore, the cross-covariance matrices can be estimated more accurately from source-specific sample means than from individual patterns from each source.

2.5 The SQDF Field Classification Rule

SQDF field classification is a straightforward application of Gaussian quadratic discrimination, albeit for field-classes. The field quadratic discriminant function for field-class (c_i, c_j, \dots, c_k) is given by

$$g_{i,j,\dots,k}(\mathbf{y}) = (\mathbf{y} - \mu_{i,j,\dots,k})^T \mathbf{K}_{i,j,\dots,k}^{-1} (\mathbf{y} - \mu_{i,j,\dots,k}) + \log |\mathbf{K}_{i,j,\dots,k}| - 2 \log p(c_i, c_j, \dots, c_k). \quad (10)$$

The test field pattern \mathbf{y} is assigned the field label that yields the minimum discriminant value. We note that the SQDF classifier is computationally expensive because of the exponential increase in the number of field-classes with field length. In addition, the quadratic discriminant function involves the computation of quadratic forms over high-dimensional spaces. To classify a field with L singlet patterns, d singlet features, and N singlet classes, the SQDF classifier must compute N^L quadratic forms over an Ld -dimensional space. Furthermore, to avoid inverting covariance matrices during runtime, N^L inverse covariance matrices of size $Ld \times Ld$ must be stored.

We now cite from [19] some interesting and useful properties of the SQDF classifier and the field-class-conditional covariance matrices. 1) With identical class means across sources, the SQDF classifier reduces to the singlet quadratic discriminant classifier. 2) The discriminant computation for field-classes that are permutations of one another can be performed using the means and covariance matrices of only one of those classes. This reduces the storage complexity of the classifier. 3) The relationship between the discriminant values computed by the singlet classifier and the field classifier can be used to derive lower bounds on the field discriminant values. These bounds can be exploited to develop a computationally less expensive branch and bound algorithm for field classification.

Example 1 (continued). For the classification problem in Example 1, the discrimination rules for various classifiers are shown below, followed by the corresponding error rates and decision boundaries in the field feature space.

- DS: From (2), the (optimal) discrete-style (DS) classification rule for $L = 2$ is

$$(c_1^*, c_2^*) = \operatorname{argmax}_{(c_1, \dots, c_j) \in \mathcal{C}^2} \sum_{m=1}^2 p(x_1 | c_i, s_m) p(x_2 | c_j, s_m)$$

since the sources and classes are equally likely and no linguistic dependence exists.

- SQDF: For the SQDF classifier, we require the field-class-conditional means and covariance matrices. For $L = 2$, the means are

$$\begin{aligned} E\{\mathbf{y} | AA\} &= \mu_{AA} = \begin{pmatrix} d_s/2 \\ d_s/2 \end{pmatrix}, \\ E\{\mathbf{y} | AB\} &= \mu_{AB} = \begin{pmatrix} d_s/2 \\ d_c + d_s/2 \end{pmatrix}, \\ E\{\mathbf{y} | BA\} &= \mu_{BA} = \begin{pmatrix} d_c + d_s/2 \\ d_s/2 \end{pmatrix}, \quad \text{and} \\ E\{\mathbf{y} | BB\} &= \mu_{BB} = \begin{pmatrix} d_c + d_s/2 \\ d_c + d_s/2 \end{pmatrix}, \end{aligned}$$

TABLE 1
Field Error Rates in % for the DS Classifier
for Different Values of d_c and d_s

d_s	d_c						
	0	σ	2σ	3σ	4σ	5σ	6σ
0	74.9	52.3	29.5	12.8	4.5	1.3	0.3
	74.9	52.3	29.5	12.8	4.5	1.3	0.3
σ	74.9	55.0	34.1	17.3	6.9	2.4	0.6
	74.7	54.6	33.1	16.7	6.7	2.1	0.6
2σ	74.8	61.0	45.0	29.4	15.1	6.7	2.2
	75.2	56.3	38.4	21.7	10.2	4.4	1.4
3σ	74.8	60.5	50.3	43.9	28.4	15.2	6.4
	74.9	54.1	35.3	25.1	13.7	5.9	2.1
4σ	75.2	56.2	41.6	38.5	43.5	28.4	15.4
	75.0	52.5	31.6	18.6	17.2	8.3	3.3

The top number in each cell is the field error rate for the singlet classifier and the bottom number is the field error rate for the field classifier.

and the covariance matrices are equal for all field-classes, given by

$$\mathbf{K} = \begin{pmatrix} \sigma^2 + d_s^2/4 & d_s^2/4 \\ d_s^2/4 & \sigma^2 + d_s^2/4 \end{pmatrix}.$$

As indicated in Section 2.3, the field-class-conditional means and covariance matrices can be constructed for any field length from the above expressions. Note that the SQDF singlet classifier approximates the class-conditional bimodal densities with a single Gaussian. Thus, even for $L = 1$, the SQDF and the DS classifiers are different.

- SF: Finally, the style-first (SF) classifier first identifies the style from the test samples and classifies them independently using the particular style-specific quadratic singlet classifier. The SF classification rule for a field of length $L = 2$ is

$$(c_1^*, c_2^*) = (\operatorname{argmax}_{c \in \mathcal{C}} p(x_1|c, s^*), \operatorname{argmax}_{c \in \mathcal{C}} p(x_2|c, s^*))$$

where

$$s^* = \operatorname{argmax}_{s \in \{s_1, s_2\}} p(x_1, x_2|s).$$

2.5.1 Error Rates of the Classifiers

For every set of parameters (interclass distance d_c and interstyle distance d_s), we generated 30,000 random test fields of length two, satisfying the distributions described in Example 1. No training was required because we used the true parameters for classification. The resulting field error rates are summarized in Tables 1 and 2.

When $d_c = 0$, the classifiers have no means of distinguishing between A s and B s, leading to a field error rate of 75 percent (since all four field classes have the same posterior probability at every point in the field-feature space). When $d_s = 0$, there is no style variation and neither the DS nor the SQDF field classifier proffer an improvement over their corresponding singlet classifier. As expected, when the interstyle distance (d_s) is large compared to the interclass distance (d_c), the single Gaussian approximation is poor, resulting in a higher error rate for the SQDF

TABLE 2
Field Error Rates in % for the SQDF Classifier
for Different Values of d_c and d_s

d_s	d_c						
	0	σ	2σ	3σ	4σ	5σ	6σ
0	74.9	52.3	29.5	12.8	4.5	1.3	0.3
	74.9	52.3	29.5	12.8	4.5	1.3	0.3
σ	75.0	55.2	34.3	17.4	7.2	2.5	0.7
	75.1	54.8	33.2	16.8	6.8	2.3	0.5
2σ	75.2	61.2	45.2	29.1	15.5	6.9	2.2
	75.0	57.4	40.1	22.7	10.9	4.3	1.4
3σ	74.8	67.8	57.6	43.7	28.6	15.5	6.6
	75.1	59.1	44.9	29.0	15.9	6.7	2.4
4σ	75.4	71.4	66.9	57.1	43.7	28.8	15.1
	74.6	59.9	46.8	33.9	21.7	10.9	4.2

The top number in each cell is the field error rate for the singlet classifier and the bottom number is the field error rate for the field classifier.

TABLE 3
Character Error Rates in % as a Function of Field Length
for Discrete Sources for Two Different Configurations

L	$d_c = 4, d_s = 2 (2.3\%)$			$d_c = 6, d_s = 2 (0.14\%)$		
	DS	SQDF	SF	DS	SQDF	SF
1	8.0	8.1	8.1	1.12	1.14	1.14
2	6.0	6.1	6.2	0.67	0.72	0.72
3	4.7	5.2	4.8	0.43	0.50	0.48
4	4.1	4.6	4.0	0.34	0.42	0.37
5	3.5	4.2	3.5	0.25	0.33	0.25
6	3.2	3.9	3.3	0.22	0.32	0.18

The single source error rate for each configuration is indicated.

classifier over the DS classifier. This effect is evident even for the singlet classifiers.

Both the DS and SQDF field classifiers are more accurate than the corresponding singlet classifier. Even after accounting for statistical variation, the observed improvement in accuracy of the field classifiers is significant. This improvement in accuracy is most striking when both d_c and d_s are large. This can be attributed to the advantage in recognizing the style identity when the interstyle distance is large. Additionally, only when the classes are well separated can we reliably recognize the style identity from only one pair of isogenous patterns. Detailed analysis showed that the field classifier decreases the error rate on different-class pairs more than on same-class pairs. From the results, we note that the SQDF field classifier is a good approximation to the optimal DS pair classifier when the interstyle distance is small compared to the interclass distance.

Table 3 shows the character error rates of the DS, SQDF, and SF classifiers with increasing field length for two different parameter sets. The top row shows the minimum achievable error rate, i.e., the single-source error rate that can be attained when the source labels for the test data are known. All three classifiers improve with increasing field length. The single Gaussian assumption of the SQDF classifier causes its error rate to flatten out at a higher value than the single-source error rate. For this simple example, the SF classifier is a good approximation to the optimal DS classifier.

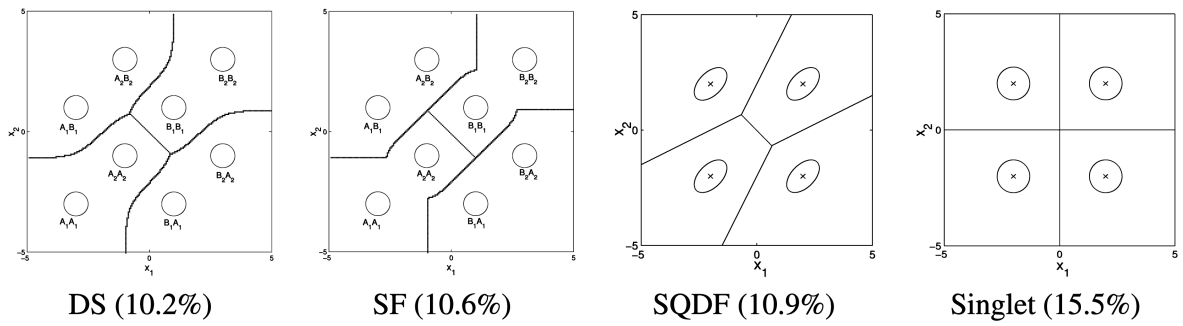


Fig. 2. Discrete styles—field classification boundaries and field error rates for different classifiers, with $d_c = 4$, $d_s = 2$.

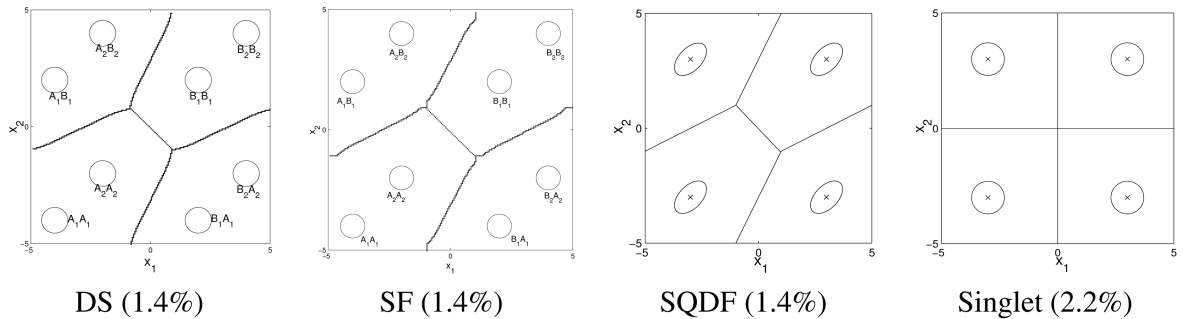


Fig. 3. Discrete styles—field classification boundaries and field error rates for different classifiers, with $d_c = 6$, $d_s = 2$.

2.5.2 Decision Boundaries

In this section, we will highlight the differences between the various classifiers by studying the differences in their decision boundaries in the field-feature space (for $L = 2$). Figs. 2 and 3 show the decision regions for the various classifiers. The labels for the decision regions are obvious from the position of the style-specific field-class means (shown in the subfigures for the DS and SF classifiers). The assumed single normal distribution for the SQDF pair and quadratic singlet classifier are also shown.

The singlet classifier operates on each pattern in the field independently: The classification boundaries are parallel to the coordinate axes. The style-conscious classifiers (DS, SQDF, and SF) utilize the dependence between co-occurring singlets to improve accuracy. In Fig. 2 ($d_c = 4$ and $d_s = 2$), we notice that, although the quadratic field classifier's boundaries differ considerably from those of the DS classifier, its field-error rate is nearly optimal. When $d_c = 6$ and $d_s = 2$ (Fig. 3), the suboptimal quadratic classifier decision boundaries are almost identical to those of the optimal field classifier. This is due to the large class separation compared to the style separation, which is generally the case in character recognition applications. Hence, we believe that the SQDF classifier can reasonably approximate the optimal DS field classifier even though the patterns originate from only two discrete sources. The boundaries for the SQDF field classifier

are piecewise linear because all four field-classes have equal covariance matrices.

3 SQDF FIELD CLASSIFIER—CONTINUOUSLY DISTRIBUTED SOURCES

It is conceivable that there are classification problems where the possible styles are not discrete but are drawn from a continuous distribution. For example, there is almost a continuous variability in handwriting styles. We present a mathematical model for continuously distributed sources and show that, under the assumption of normality of feature and style distributions, the SQDF classifier is optimal for field classification.

3.1 Model for Continuous Styles

We consider again the problem of classifying a field $\mathbf{y} = (\mathbf{x}_1^T, \dots, \mathbf{x}_L^T)^T$ (each $\mathbf{x}_i \in \mathbb{R}^d$) into one of the field-classes in \mathcal{C}^L , where $\mathcal{C} = \{c_1, c_2, \dots, c_N\}$. We assume the existence of a "hidden" Nd -dimensional random vector $\mathbf{s} = (\mathbf{m}_1^T, \dots, \mathbf{m}_N^T)^T$, where each \mathbf{m}_i is a d -dimensional random vector, $\forall i = 1, \dots, N$. Here, the random vector \mathbf{s} represents the style whose identity is entirely determined by its class-conditional means.³ We make the following assumptions on the feature distributions:

3. A style is generated by selecting the class means. The mean for a particular class is a random translation from the overall mean (μ_i) of that class (grand class mean over all styles). The translation vector for each class is normally distributed with zero mean. Furthermore, the translations for different classes are correlated (as given by Σ_s). After the style is chosen, the features are generated for each class independently, according to the style-specific means and a covariance matrix (Σ_i) that depends on the class but not on the style.

1. $\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$, where

$$\boldsymbol{\mu}_s = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_s = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1N} \\ C_{21} & C_{22} & \dots & C_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \dots & C_{NN} \end{pmatrix},$$

where $C_{ij} = E\{(\mathbf{m}_i - \mu_i)(\mathbf{m}_j - \mu_j)^T\}$.

2. $(\mathbf{y}|\mathbf{c}^1 = c_i, \dots, \mathbf{c}^L = c_k, \mathbf{s}) \sim \mathcal{N}((\mathbf{m}_i^T, \dots, \mathbf{m}_k^T)^T, \boldsymbol{\Sigma}_{i,\dots,k})$, where

$$\boldsymbol{\Sigma}_{i,\dots,k} = \begin{pmatrix} \Sigma_i & \dots & 0_{d \times d} \\ \vdots & \ddots & \vdots \\ 0_{d \times d} & \dots & \Sigma_k \end{pmatrix}.$$

The covariance matrix $\boldsymbol{\Sigma}_s$ specifies the style variability, while the matrices $\{\Sigma_1, \dots, \Sigma_N\}$ specify the within-style variance in the patterns.

Let us now define

$$\mathbf{z} = (\mathbf{y}|\mathbf{c}^1 = c_i, \dots, \mathbf{c}^L = c_k, \mathbf{s} = (\mathbf{m}_1^T, \dots, \mathbf{m}_N^T)^T) - \mathbf{m},$$

where

$\mathbf{m} = (\mathbf{m}_i^T, \dots, \mathbf{m}_k^T)^T$. Clearly, \mathbf{z} is Gaussian ($\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{i,\dots,k})$) and independent of \mathbf{s} .

The field-class-conditional density is given by

$$\begin{aligned} p(\mathbf{y}|c_i, \dots, c_k) &= \int_{\mathbf{s}} p(\mathbf{y} = \mathbf{y}|c_i, \dots, c_k, \mathbf{s})p(\mathbf{s})d\mathbf{s} \\ &= \int_{\mathbf{m}} p(\mathbf{y} = \mathbf{y}|c_i, \dots, c_k, \mathbf{m})p(\mathbf{m})d\mathbf{m} \\ &= \int_{\mathbf{m}} p(\mathbf{z} = \mathbf{y} - \mathbf{m}|c_i, \dots, c_k, \mathbf{m})p(\mathbf{m})d\mathbf{m} \\ &= p(\mathbf{z}) \star p(\mathbf{m}) \end{aligned} \quad (11)$$

(where \star is the convolution operator).

Since both \mathbf{z} and \mathbf{m} are normally distributed, so are the field-class-conditional feature distributions $(\mathbf{y}|c_i, \dots, c_k)$. Therefore, the SQDF classifier yields the minimum field classification error rate (the Bayes error).

The field-class conditional means and covariance matrices can be shown to be

$$\mu_{i,j,\dots,k} = E\{\mathbf{y}|c_i, c_j, \dots, c_k\} = \begin{pmatrix} \mu_i \\ \mu_j \\ \vdots \\ \mu_k \end{pmatrix} \quad (12)$$

$$\begin{aligned} \mathbf{K}_{i,j,\dots,k} &= E\{(\mathbf{y} - \mu_{i,j,\dots,k})(\mathbf{y} - \mu_{i,j,\dots,k})^T | c_i, c_j, \dots, c_k\} \\ &= \begin{pmatrix} \Sigma_i + C_{ii} & C_{ij} & \dots & C_{ik} \\ C_{ji} & \Sigma_j + C_{jj} & \dots & C_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ C_{ki} & C_{kj} & \dots & \Sigma_k + C_{kk} \end{pmatrix}. \end{aligned} \quad (13)$$

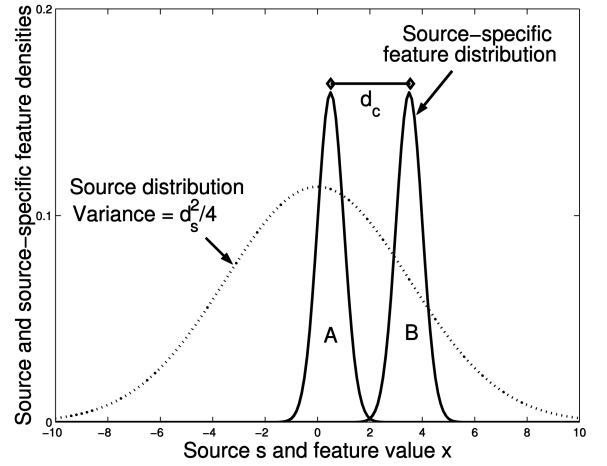


Fig. 4. The source-and-class-conditional feature distributions—continuously distributed sources.

Example 2. We modify Example 1 from the previous section to illustrate the model for continuous styles. There are still only two singlet classes $\{A, B\}$ that are a priori equally likely and there is no linguistic dependence.

The class-and-source conditional singlet-feature distributions are

$$(x|A, s = s) \sim N(s, \sigma^2) \quad (x|B, s = s) \sim N(d_c + s, \sigma^2)$$

and the sources are distributed according to

$$s \sim N(0, d_s^2/4). \quad (14)$$

The distribution of sources and the feature distributions are shown in Fig. 4. Here, the source is identified by only one number (instead of the two class-conditional means) because, given the mean of A (denoted m_A) of a particular source, we can obtain the source-specific mean of B (denoted m_B) by $m_B = m_A + d_c$. That is, the mean of A and the mean of B are maximally correlated. So, the single style variable just determines the “shift” of the mean of A from the origin.

For any field length, the features are field-class-conditionally normally distributed. For $L = 2$, the field-class means are

$$\begin{aligned} E\{\mathbf{y}|AA\} &= \mu_{AA} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad E\{\mathbf{y}|AB\} = \mu_{AB} = \begin{pmatrix} 0 \\ d_c \end{pmatrix}, \\ E\{\mathbf{y}|BA\} &= \mu_{BA} = \begin{pmatrix} d_c \\ 0 \end{pmatrix}, \text{ and } E\{\mathbf{y}|BB\} = \mu_{BB} = \begin{pmatrix} d_c \\ d_c \end{pmatrix} \end{aligned}$$

and the covariance matrices are equal for all field-classes, given by

$$\mathbf{K} = \begin{pmatrix} \sigma^2 + d_s^2/4 & d_s^2/4 \\ d_s^2/4 & \sigma^2 + d_s^2/4 \end{pmatrix}.$$

Again, the means and covariance matrices for any field length can be constructed from those for $L = 2$. These parameters are used by the SQDF classifier for field classification. Note that the means, covariance matrices, and decision boundaries are identical to those of Example 1 where there were only two styles.

TABLE 4
Handwritten Numeral Data Sets

	Writers	Number of samples
SD3-Train	0-399 (395)	13481
SD7-Train	2100-2199 (99)	3615
SD3-Test	400-799 (399)	13491
SD7-Test	2200-2299 (100)	3704

4 EXPERIMENTAL RESULTS

4.1 Description of the Data

We experimented with databases SD3 and SD7, which are contained in the NIST Special Database SD19 [20]. The database contains samples of handwritten numerals labeled by writer and class. SD3 was the training data released for the First Census OCR Systems Conference and SD7 was the test data. SD3 and SD7 were obtained from different populations and SD7 is considered to be much more difficult to recognize. There are approximately 10 samples per class per writer.

We constructed four data sets, two from each of SD3 and SD7, as shown in Table 4. The writers in the Train and Test sets are disjoint, which allows us to verify our hypothesis that broad styles can be gleaned from a sufficiently large sample of writers. We experimented with only a subset of three classes (1, 2, 7) that contributed most to the error for singlet classification.

Since we compute the field class-conditional covariance matrices from source-specific class-conditional matrices, we require that each writer have at least two samples for each singlet class. We therefore deleted all writers not satisfying this criterion from the training sets (but not from the test sets). In Table 4, the numbers in the parentheses indicate the total number of writers from each set that remain after the deletion.

We extracted 100 blurred directional (chain-code) features from each sample [21]. The top 25 principal component features were picked for the following experiments in order to avoid complicated covariance regularization schemes as well as to reduce the computational cost for classification. The samples of each writer in the test sets were randomly permuted and L patterns were chosen at a time to simulate fields of length L . All the algorithms were implemented in MATLAB; therefore, we don't consider it appropriate to report running times. Since the error rates reported below are the field error rates for a subset of classes, they are not directly comparable with the error rates reported on the NIST benchmark. For comparison, the best character error rate of the baseline singlet classifier (with covariance regularization) on all 10 classes of SD3-Test + SD7-Test, with all 100 features, was 1.4 percent.

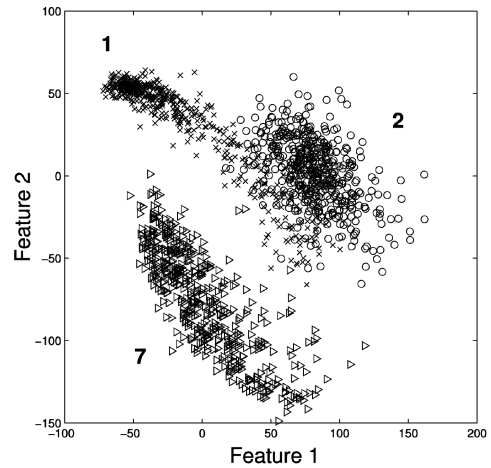


Fig. 5. Scatter plot of the top two principal component features of writer-specific class means.

Fig. 5 shows the scatter plot of the top two principal component features of the writer-specific class means of the writers in the training sets. The writer-specific class means seem to vary in a continuous fashion.

4.2 Classification Results

We present below the recognition results for the SQDF field classifier and compare them with those of the quadratic discriminant singlet classifier on the handwritten digits. Table 5 lists the field error rates for the singlet and SQDF field classifiers on fields of length two through five for various test sets.

Fig. 6 shows some of the field patterns classified incorrectly by the singlet classifier but correctly by the SQDF field classifier.

The results in Table 5 show the advantage of exploiting style context. The field error rate (3.8 percent) of the style-conscious field classifier is significantly lower on NIST hand-printed digits than that of the singlet classifier (4.6 percent) for fields of length 4. For $L = 4$, the results are over 4,495 test fields.

5 DISCUSSION AND FUTURE WORK

We have formulated the style-conscious quadratic discriminant function (SQDF) classifier as a natural extension of the widely used Gaussian quadratic discriminant classifier. We have shown that the SQDF classifier can be trained with information only from pairs of same-source characters to classify test fields of arbitrary length. We proposed a model for the generation of style consistent fields under which the SQDF classifier is optimal for field error rate.

TABLE 5

Field Error Rates for Fields in % of Length $L = 2$ and $L = 3$ on Handwritten Data for Singlet and SQDF Field Classification

Test set	$L = 2$		$L = 3$		$L = 4$		$L = 5$	
	Singlet	Field	Singlet	Field	Singlet	Field	Singlet	Field
SD3-Test	1.4	1.3	2.0	1.7	2.5	2.0	3.0	2.5
SD7-Test	7.4	6.7	10.1	9.1	12.5	10.4	13.8	11.7
SD3-Test + SD7-Test	2.7	2.4	3.7	3.3	4.6	3.8	5.3	4.5

The training set is SD3-Train+SD7-Train.

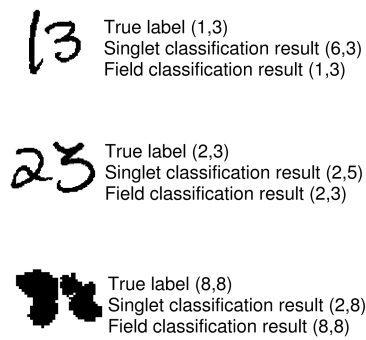


Fig. 6. Some test fields with recognition results.

The simulations indicate under what circumstances style-conscious classification is most advantageous and reveal the nature of the resulting pair-classification boundaries. On real data (Table 5), as field length increases, the field classification accuracy for the singlet classifier as well as the SQDF field classifier approaches zero, albeit at different rates. The decrease in relative gain by the field classifier from $L = 4$ to $L = 5$ can be attributed to this. For long fields, it may be more advisable to optimize classifiers for character error rate than for field error rate. In any case, its excessive computational and storage complexity disqualifies the SQDF classifier for longer fields. We propose investigating methods to further exploit the special structure of the field-class-conditional covariance matrices to reduce the computational complexity of the classifier.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Hiromichi Fujisawa, Dr. Cheng-Lin Liu, and Dr. Prateek Sarkar for their valuable comments and suggestions on both the content and presentation of this work.

REFERENCES

- [1] M. Gilloux, M. Leroux, and J.M. Bertille, "Strategies for Handwritten Words Recognition Using Hidden Markov Models," *Proc. Second Int'l Conf. Document Analysis and Recognition*, pp. 299-304, 1993.
- [2] G. Nagy, "Teaching a Computer to Read," *Proc. 11th Int'l Conf. Pattern Recognition*, vol. 2, pp. 225-229, 1992.
- [3] R. Plamondon, D.P. Lopresti, L.R.B. Schomaker, and R. Srihari, "Online Handwriting Recognition," *Wiley Encyclopedia of Electrical and Electronics Eng.*, pp. 123-146, New York: John Wiley and Sons, 1999.
- [4] T. Hong and J.J. Hull, "Visual Inter-Word Relations and Their Use in OCR Postprocessing," *Proc. Third Int'l Conf. Document Analysis and Recognition*, vol. 1, pp. 442-445, 1995.
- [5] P. Sarkar, "Style Consistency in Pattern Fields," PhD thesis, Rensselaer Polytechnic Inst., Troy, N.Y., 2000.
- [6] P. Sarkar and G. Nagy, "Style Consistency in Isogenous Patterns," *Proc. Sixth Int'l Conf. Document Analysis and Recognition*, pp. 1169-1174, 2001.
- [7] P. Sarkar and G. Nagy, "Style Consistent Classification of Isogenous Patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, Jan. 2005.
- [8] T. Kawatani, "Character Recognition Performance Improvement Using Personal Handwriting Characteristics," *Proc. Third Int'l Conf. Document Analysis and Recognition*, vol. 1, pp. 98-103, 1995.
- [9] T. Koshinaka, D. Nishikawa, and K. Yamada, "A Stochastic Model for Handwritten Word Recognition Using Context Dependency between Character Patterns," *Proc. Sixth Int'l Conf. Document Analysis and Recognition*, pp. 154-158, 2001.

- [10] M. Shridhar and F. Kimura, "Character Recognition Performance Improvement Using Personal Handwriting Characteristics," *Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics*, vol. 3, pp. 2341-2346, 1995.
- [11] M.E. Dehkordi, N. Sherkat, and R.J. Withrow, "A Principal Component Approach to Classification of Handwritten Words," *Proc. Fifth Int'l Conf. Document Analysis and Recognition*, pp. 781-784, 1999.
- [12] A. Zramdini and R. Ingold, "Optical Font Recognition Using Typographical Features," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 877-882, Aug. 1998.
- [13] H. Shi and T. Pavlidis, "Font Recognition and Contextual Processing for More Accurate Text Recognition," *Proc. Fourth Int'l Conf. Document Analysis and Recognition*, vol. 1, pp. 39-44, 1997.
- [14] I. Bazzi, R. Schwartz, and J. Makhoul, "An Omnifont Open-Vocabulary OCR System for English and Arabic," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 495-504, June 1999.
- [15] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. New York: John Wiley and Sons 1973.
- [16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic Press 1972.
- [17] H. Fujisawa, Y. Nakano, and K. Kurino, "Segmentation Methods for Character Recognition: From Segmentation to Document Structure Analysis," *Proc. IEEE*, vol. 80, no. 7, pp. 1079-1092, July 1992.
- [18] R.G. Casey and E. Lecolinet, "A Survey of Methods and Strategies in Character Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 690-706, July 1996.
- [19] S. Veeramachaneni, "Style Constrained Quadratic Field Classifiers," PhD thesis, Rensselaer Polytechnic Inst., Troy, N.Y., 2002.
- [20] P. Grother, "Handprinted Forms and Character Database, NIST Special Database 19," technical report, Mar. 1995.
- [21] C.L. Liu, H. Sako, and H. Fujisawa, "Performance Evaluation of Pattern Classifiers for Handwritten Character Recognition," *Int'l J. Document Analysis and Recognition*, vol. 4, no. 3, pp. 191-204, 2002.



Sriharsha Veeramachaneni received the PhD degree in computer engineering from Rensselaer Polytechnic Institute, Troy, New York in 2002. He is currently a researcher at the Istituto per la Ricerca Scientifica e Tecnologica, Trento, Italy. His research interests include statistical pattern recognition, machine learning, and information theory.



George Nagy received the BEng and MEng degrees from McGill University and the PhD degree in electrical engineering from Cornell University in 1962 (on neural networks). For the next 10 years, he studied pattern recognition at the IBM T.J. Watson Research Center in Yorktown Heights, New York. From 1972 to 1985, he was a professor of computer science at the University of Nebraska-Lincoln (nine years as chair) and worked on geographic information

systems, remote sensing applications, and human-computer interfaces. Since 1985, he has been a professor of computer engineering at Rensselaer Polytechnic Institute, where he established the ECSE DocLab. In addition to document image analysis, OCR, geographic information systems, and computational geometry, his students have engaged in solid modeling, finite-precision spatial computation, and interactive computer vision, often with a focus on systems that improve with use. He has benefited from visiting appointments at the Stanford Research Institute, Cornell, the University of Montreal, the National Scientific Research Institute of Quebec, the University of Genoa and the Italian National Research Council in Naples and Genoa, AT&T and Lucent Bell Laboratories, IBM Almaden, McGill University, the Institute for Information Science Research at the University of Nevada, and the Center for Image Analysis in Uppsala. He is a life fellow of the IEEE and the IEEE Computer Society.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.