ELECTRONIC PUBLISHING, VOL. 6(3), 249–260 (SEPTEMBER 1993)

Optical font recognition from projection profiles

ABDELWAHAB ZRAMDINI AND ROLF INGOLD

Computer Science Institute
University of Fribourg
Pérolles – Chemin du Musée 3
CH-1700 Fribourg, Switzerland

email: zramdini@cfruni52.bitnet

SUMMARY

This paper presents a statistical approach for font attribute recognition based on features extracted from projection profiles of text lines and using a Bayesian classifier. The presented features allow the discrimination of the font weight, slope and size.

KEY WORDS Font recognition Projection profiles Discrimination power Bayesian classifier

1 INTRODUCTION

Optical Font Recognition (OFR) is an important but often neglected problem in optical reading. In fact, OFR is useful and necessary in different domains:

- Recognition of logical document structures [1], where knowledge of the font used in a word, line, or text block may be useful for defining its logical label (chapter title, section title or paragraph).
- Document reproduction, where knowledge of the font is necessary in order to reproduce (reprint) the document.
- Document indexing and information retrieval, where word indexes are generally printed in fonts different from those of the running text.
- Text font knowledge may also improve the recognition rate of OCR (Optical Character Recognition) systems, because we believe that mono-font OCR (OCR with an assumed known font) may give better results than omni-font OCR (OCR capable of recognizing characters of any font and size).

One uses different fonts in a document in order to emphasize some parts of the text in such a way that the reader notices them easily. In a document, font changes may occur at particular points (titles, indexes, references, etc.). They may be done by choosing another typeface, or changing the style or the size of the same typeface (normal typeface for the running text, bold for titles, italic for references, monospaced typeface for program illustrations, etc.). In reality, we do few and systematic font changes in a single structured document.

1.1 Font recognition approaches

There are two possible approaches for font recognition:

- Global feature extraction from text entities (word, line, paragraph). These features are generally detected by non-experts in typography (text density, size, orientation

CCC 0894–3982/93/030249–12 *Received 1 September 1993*
Revised 3 December 1993

and spacing of the letters, serifs, etc.). This approach is suitable for *a priori font recognition*, where the font is recognized without any knowledge of the letter classes.
- Local feature extraction from individual letters. The features are based on letter particularities like the shapes of serifs (coved, squared, triangular, etc.) and the representation of particular letters like g and *g*, a and *a*. This kind of approach may derive substantial benefit from knowledge of the letter classes. In this case we talk about *a posteriori font recognition*.

1.2 Goals of our work

In our work we are interested in an *a priori font recognition system* which is integrated into a structured document recognition system in which the analysis is based, among other factors, on font identification. The OCR comes at a second level and uses the knowledge of the font to perform character recognition.

The OFR system uses a global feature extraction approach on a limited and known set of fonts. This set belongs to the system knowledge base, which may contain many fonts. This approach is justified by the fact that in a structured document we generally use a limited number of known fonts. The OFR system is able to work on different text entities (words, lines, paragraphs). It is also independent of document content and language. The knowledge base is generated by a process of learning selected features for the different fonts used in the system.

The following information could be estimated from the knowledge base:

- the power of a feature to discriminate fonts. We talk about *discrimination power*. In fact, it is important to know the risk of confusion between some fonts according to one feature or a whole vector of features;
- global recogniton rate for each font in the system or for a particular set of fonts.

The following section presents some related work on digital font recognition. Section 3 defines the notion of *font* considered in our system. Section 4 discusses the methodology we adopted. In Section 5 we present some of the selected features used in the classification process. Some statistical evaluations applied to the features, and practical results, are presented respectively in Sections 7 and 8.

2 RELATED WORK

A small amount of work has been done on typeface classification and recognition. Typeface classification has mainly been done manually, and operated according to historical considerations and serif styles [2–5]. Nowadays, with the expansion of computers, we are witnesses to a rapid proliferation of digital fonts, but their analysis for document recognition has not yet been considered.

Morris [6] based his study of the problem of digital font recognition on the analysis of Fourier amplitude spectra extracted from word images. The study was mainly done in order to examine the applicability of human vision models to typeface discrimination and to investigate whether spectral features might be useful in typeface production. He applied a Fourier transform to the word image and then extracted a features vector by applying many filters to the resulting spectra. He used a quadratic Bayesian classifier for

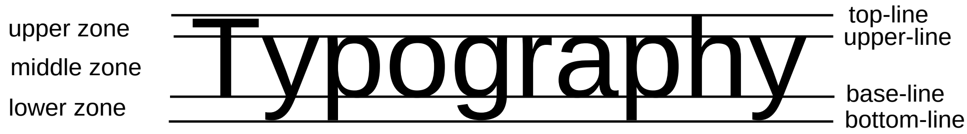

Figure 1. Structure of a text line

font classification and obtained good results, but many important simplifications were done: the images considered were noise-free, because they were created by software instead of being scanned from paper documents. He also considered only one font size for the different samples.

Anigbogu [7], in his multi-font OCR system, defined models for characters and placed them in a tree according to certain attributes (ascenders, descenders, holes, etc.). Some preprocessing was done on the text in order to select one sample for each kind of shape (letters). The selected shapes were placed in the tree according to their attributes. In another operation, a tree was generated for each font in the system (instantiation of a generic tree with the different letters of the given font). The shape tree was then compared with each font tree in order to compute a distance; the smallest distance thus defined the associated font. The font identification was mainly done to improve the performance of the OCR system by limiting the search space (font trees). This approach seems to provide good results if the generated shape tree is complete enough (i.e. it has a sample for each character).

Spitz, in his *Multilingual Document Recognition System* [8], used a function discriminating Roman from Japanese texts for OCR purposes. The discrimination is based on fundamental differences between Roman and Japanese texts such as character spacing, the presence of ascenders and descenders and the homogeneity of text density. It relies on the presence of a statistically significant difference in the distribution of a measure of local optical density.

3 DEFINITIONS

In a global features approach, the features are extracted from zones extending over several characters, words or lines instead of a single character.

3.1 Text line structure

A text line can be considered as being composed of three zones: the *upper zone*, the *middle zone* and the *lower zone* (see Figure 1). These zones are delimited by four virtual lines: the *top-line*, the *upper-line*, the *base-line* and the *bottom-line*. Each text line has at least a *middle zone*; the *upper zone* depends on capital letters and letters with ascenders, like h and k; the *lower zone* depends on letters with descenders, like g, p and y. This structure allows the definition of four kinds of text line:

1. **full line**, with character parts present in all three zones;
2. **ascender line**, with character parts present in the *upper* and *middle* zones;
3. **descender line**, with character parts present in the *lower* and *middle* zones;
4. **short line**, with character parts present in the *middle* zone.

The *middle zone* is the most important part of a text line, where we find the main information. Its height is commonly called the *x-height* because it corresponds to the height of a lowercase x. The proportion of the different zones in the font size differ from one typeface to another.

3.2 Font model

In our OFR system a font is modelled by the following attributes:

1. the font **family**, such as Times, Helvetica or Courier. Commonly this corresponds to the definition of **typeface**;
2. the **size**, expressed in typographic points and with a value range of $[6, \ldots, 96]$;
3. the **weight** of the font, having one of the following values: *light*, *normal* or *bold*;
4. the **slope** indicating the orientation of the letters' main strokes. A font could be *roman*, *slanted* or *italic*;
5. the **spacing mode** specifying the pitch of the characters. A font may have a fixed pitch (mono-spaced) or a proportional one. The latter class may have *condensed*, *normal* or *expanded* spacing mode.

4 METHODOLOGY

We decided to use a Bayesian classifier for our font identification system. The selected features are extracted from projection profiles.

4.1 Bayesian decision theory

A classification process consists of associating a class (from a set of predefined classes) to an observation defined by a features vector $x = (x_1, x_2, \ldots, x_d)$. Bayesian decision theory is based on the assumption that the decision problem is posed in probabilistic terms and that all of the relevant probability values are known [9].

4.1.1 *Bayesian classification*

Let $\Omega = \{w_1, w_2, \ldots, w_n\}$ be the set of classes. The Bayesian classification consists of finding a class ω_i so that $P(\omega_i|x)$ is maximal where $P(\omega_i|x)$ is computed with the Bayes rule:

$$\begin{aligned} P(\omega_i|x) &= \frac{p(x|\omega_i)P(\omega_i)}{P(x)}, \quad \text{where} \\ P(x) &= \sum_{i=1}^{n} p(x|\omega_i)P(\omega_i). \end{aligned}$$

$P(\omega_i)$ expresses the *a priori probability* of the class ω_i, and $p(x|\omega_i)$ represents the *conditional density function* of x, or in other words the probability of obtaining x when its class is ω_i.

A Bayesian classifier can be represented as a set of discriminant functions $g_i(x)$, $i = 1, \ldots, n$. It assigns the class ω_i to the features vector x if $g_i(x) > g_j(x)$ for all $j \neq i$. It is then seen as a machine calculating n discriminant functions and choosing the class

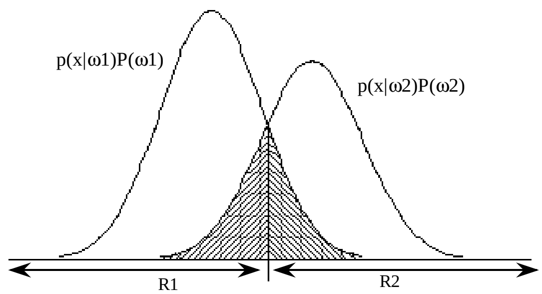

Figure 2. Bayesian error probability

corresponding to the largest discriminant. The selection of the discriminant function is not unique. We can take $g_i(x) = P(\omega_i|x)$, because the largest discriminant corresponds to the highest *a posteriori* probability.

4.1.2 Classification error

The Bayesian classifier splits up the feature space into n decision regions, $R_1 \ldots R_n$. Take the case of two regions; there are two cases that give rise to a classification error: either the observation x falls into R_2 and its true class is ω_1, or x falls into R_1 and its true class is ω_2 (see Figure 2). We have thus

$$\begin{aligned} P(error) &= P(x \in R_2, \omega_1) + P(x \in R_1, \omega_2) \\ &= P(x \in R_2|\omega_1)P(\omega_1) + P(x \in R_1|\omega_2)P(\omega_2) \\ &= \int_{R_2} p(x|\omega_1)P(\omega_1)dx + \int_{R_1} p(x|\omega_2)P(\omega_2)dx. \end{aligned}$$

4.1.3 Feature distribution

As mentioned above, Bayesian decision theory supposes that feature distributions can be expressed by probabilistic parametric functions. In our study we assume, as discussed in Section 7, that feature vectors follow normal lows $\mathcal{N}(\mu,\Sigma)$. This means that their distributions are multivariate normal density functions expressed by:

$$f(x;\mu,\Sigma) = \frac{1}{(2\pi)^{(d/2)}\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right]$$

where x is a d-component features vector, μ is the d-component *mean vector*, Σ is the d-by-d *covariance matrix*, $(x-\mu)^t$ is the transpose of $x-\mu$ and $|\Sigma|$ is the Σ determinant.

4.1.4 Learning

The learning process consists of estimating the parameters of the class conditional density function $p(x|\omega_i)$ so that the Bayes rule can be applied to find the feature discriminant function values. These values are computed on a large-enough sample for each class ω_i. In our case, we used the maximum likelihood estimator for the estimation of the parameters

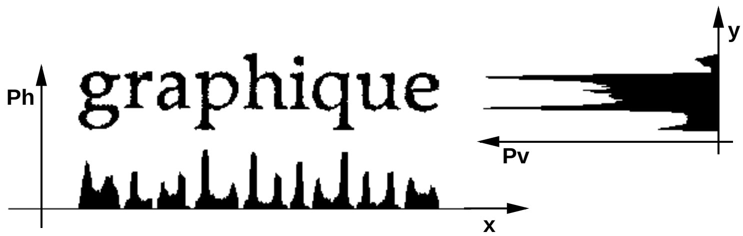

Figure 3. Horizontal and vertical projection profiles

μ and Σ. To be more specific, if x_i is the ith component of features vector x, μ_i is the ith component of μ and σ_{ij} is the i-jth component of Σ, then

$$\mu_i = E(x_i) \quad \text{and}$$
$$\sigma_{ij} = E(xi - \mu_i)(xj - \mu_j), \quad \text{where} \quad E(y) = \frac{1}{n}\sum_{k=1}^{n} y_k.$$

We also assumed that the classes have the same *a priori* probabilities $P(\omega_i)$, so that they became computationally irrelevant.

4.2 Projection profiles

Features are extracted from the projection profiles of text lines. Let $S(N,M)$ be a binary image of N lines and M columns. As is shown by Figure 3, we define:

- *Vertical profile*: sum of black pixels perpendicular to the y axis; this is represented by the vector P_v of size N and defined by:

$$P_v[i] = \sum_{j=1}^{M} S[i,j].$$

- *Horizontal profile*: sum of black pixels perpendicular to the x axis; this is represented by the vector P_h of size M and defined by:

$$P_h[j] = \sum_{i=1}^{N} S[i,j].$$

5 SELECTED FEATURES

We considered five principal features for font discrimination. They have been derived from visual observations of different fonts and their projection profiles.

5.1 Vertical profile heights

The height of the vertical profile (P_v) depends mainly on the font size. As is shown in Figure 4, P_v has four peaks estimating the four virtual lines defined in section 3.1:

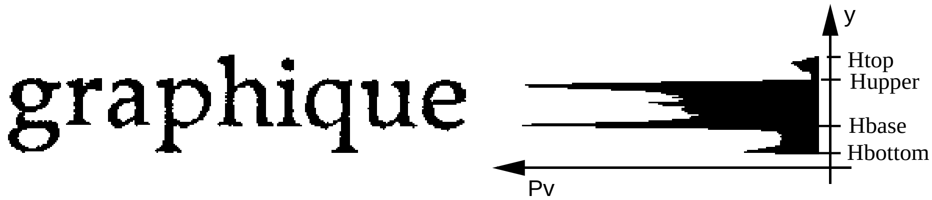

Figure 4. Features extracted from vertical profile

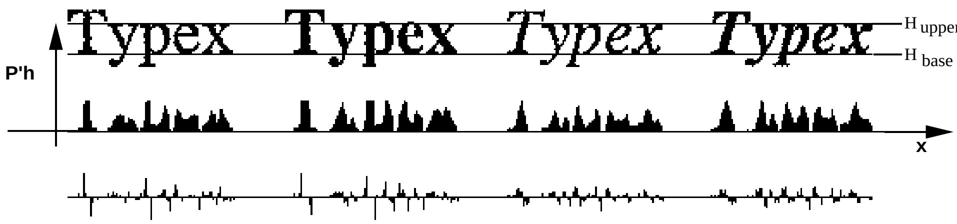

Figure 5. Horizontal projection profile and its first derivative

1. $H_{top} = \max\{i \text{ such that } P_v[i] > 0\}$.
2. $H_{bottom} = \min\{i \text{ such that } P_v[i] > 0\}$.
3. $H_{base} = i$ such that $P_v[i] - P_v[i-1]$ is maximal.
4. $H_{upper} = i$ such that $P_v[i] - P_v[i-1]$ is minimal.

Three features are extracted from P_v:

1. The height of the whole profile, defined by $h_1 = |H_{top} - H_{bottom}|$.
2. The height of the upper part of the profile, defined by $h_2 = |H_{top} - H_{base}|$.
3. The height of the middle part of the profile, defined by $h_3 = |H_{upper} - H_{base}|$.

5.2 Density of black pixels

The weight of a font is reflected by the density of black surfaces on the white background. This density (dn) is extracted from the horizontal profile P'_h (see Figure 5). It is computed on the central part of the line located between H_{upper} and H_{base}, in order to be independent of the text line structure. The feature is thus defined by

$$dn = \frac{1}{n} \sum_{x=1}^{n} P'_h[x].$$

5.3 Variance of horizontal profile derivative

One can observe from the horizontal profile that roman texts are characterised by a set of upright and tall peaks. For italic texts the peaks are less tall, rounded and broader (see Figure 5). This feature is revealed by the first derivative of the horizontal projection profile (dr), defined by

$$dr = \frac{1}{n-1} \sum_{x=1}^{n-1} (P'_h[x+1] - P'_h[x])^2.$$

The latter two features are sensitive to text formatting: in a justified text, word spacing is not fixed and depends on the number of words in the line. In order to be independent of formatting, each space between two words is replaced by a small fixed space.

6 LEARNING

The learning process consisted of estimating, for each font, the mean vector of five features (dr, dn, h_1, h_2, h_3) and the corresponding covariance matrix. This has been done on 100 full English text lines (with ascenders and descenders) of about 6 cm length for each font. The texts have been arbitrarily extracted from existing documents and produced by a laser printer. Binary images have been produced from these texts by scanning at 400 dpi.

Some preprocessing has been done on the image of the text lines, in order

- to detect and correct possible skew;
- to filter some marks such as noise, punctuation and diacritical signs which do not carry any pertinent information.

The knowledge base, used in the theoretical evaluations and in the classification experiments, contains 112 fonts compounded from the combinations of the different font attributes. Seven font families have been considered: four are seriffed (Times, Palatino, Bookman, New Century Schoolbook), two are sanserif (Avant Garde, Helvetica) and one is monospaced (Courier). Four sizes (10, 11, 12, 14) and four styles (normal, italic, bold, bold-italic) often used in documents have been selected for each font family.

7 STATISTICAL EVALUATION

Bayesian decision theory allows the estimation of theoretical classification error rates, which can be computed from the feature distribution functions (see subsection 4.1). Thus, in a first step, we made a statistical evaluation in order

- to evaluate the power of the selected features to discriminate the weight, slope and size attributes;
- to study the influence of the family and size attributes on the weight and slope.

7.1 Evaluation environment

In this evaluation, feature discrimination powers have been analysed individually, where we did not take into account their correlations. Individual features follow univariate normal lows $\mathcal{N}(\mu,\sigma)$ with the distribution function

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

A verification of the normality of marginal distributions has been done by statistical evaluation in which we used the χ^2 test [10]. Nevertheless, we assume without any rigorous verification that feature vectors are normally distributed.

Table 1. Theoretical confusion rates for weight (normal, bold) and slope (roman, italic) with known family and size (12) and with unknown size (all sizes merged)

Font			Decision with *dn* for **normal** weight		Decision with *dr* for **roman** slope	
Family	*Weight*	*Slope*	Size 12	All sizes	Size 12	All sizes
Courier (1)	normal	roman	1.0	0.999	0.998	0.995
	bold	roman	0.0	0.001	0.983	0.964
	normal	italic	1.0	0.999	0.009	0.026
	bold	italic	0.0	0.001	0.006	0.026
Avant Garde (2)	normal	roman	1.0	0.999	1.0	0.999
	bold	roman	0.0	0.001	1.0	1.0
	normal	italic	1.0	0.999	0.0	0.001
	bold	italic	0.0	0.001	0.0	0.0
Times (3)	normal	roman	0.996	0.973	1.0	0.0
	bold	roman	0.002	0.010	1.0	0.0
	normal	italic	0.998	0.989	0.0	1.0
	bold	italic	0.003	0.026	0.0	1.0
(1)+(2)+(3)	normal	roman	0.752	0.709	0.999	0.995
	bold	roman	0.238	0.254	0.999	0.996
	normal	italic	0.755	0.712	0.001	0.003
	bold	italic	0.241	0.251	0.001	0.002

7.2 Evaluation results

Table 1 gives an overview of the power of *dn* (density of black pixels) and *dr* (variance of horizontal projection profile derivative) features to discriminate the weight and slope of a font when its family and size are known, and when its size is unknown. *dn* was used to discriminate the weight and *dr* to discriminate the slope. We can see that:

- these features are pertinent, and have a discrimination power greater than 97%;
- the slope is easier to detect than the weight;
- the font size has a very low influence on the discrimination of weight and slope.

The last line of Table 1 shows that *dr* is still very accurate in discriminating the slope when the font family is unknown, while *dn* is less accurate in discriminating the weight. This may be explained by the fact that fonts do not have homogeneous typographic grey levels.

Other tests with other font families (Helvetica, Palatino, Bookman, New Century Schoolbook) gave similar results.

Table 2 gives an overview of the power of the h_3 feature to discriminate the font size with known family, weight and slope (h_3 is presented because it estimates the main part of a text line). The confusion rates were computed, first for sizes 10, 11 and 12 and second for sizes 10, 12 and 14. The table shows that size discrimination is easy for non-consecutive sizes and is more difficult for successive ones.

Other tests with the h_1 and h_2 features led to the same conclusions. In fact, h_1, h_2 and h_3 depend on the font family and have very low discrimination power for merged families, for example h_3 has the same value for Helvetica-10 and Times-11.

8 EXPERIMENTS

In a second step, a multivariate Bayesian classifier has been designed for the experiments in order to confirm the statistical evaluation results of Section 7 and to see the influence of font families on the classification results.

Table 2. Theoretical confusion rates between font sizes using h_3 with known family, weight and slope

Font				Decision for size with h_3			Decision for size with h_3		
Family	*Weight*	*Slope*	*Size*	10	11	12	10	12	14
Courier	normal	roman	10	**0.798**	0.189	0.013	**0.987**	0.013	0.0
			11	0.312	**0.588**	0.100	-	-	-
			12	0.017	0.093	**0.890**	0.017	**0.960**	0.023
			14	-	-	-	0.0	0.018	**0.982**
Avant Garde	normal	roman	10	**0.987**	0.012	0.001	**0.999**	0.001	0.0
			11	0.013	**0.903**	0.084	-	-	-
			12	0.001	0.115	**0.884**	0.001	**0.998**	0.001
			14	-	-	-	0.0	0.001	**0.999**
Times	normal	roman	10	**0.951**	0.047	0.002	**0.999**	0.001	0.0
			11	0.107	**0.592**	0.301	-	-	-
			12	0.001	0.129	**0.870**	0.001	**0.998**	0.001
			14	-	-	-	0.0	0.001	**0.999**

8.1 Preclassification

The classifier has been configured so that it takes into account the text line structures defined in subsection 3.1. A text line preclassification was done for each text line, and following its estimated structure we use different feature vectors:

- a 5-component vector (dr, dn, h_1, h_2, h_3) is used when the line is estimated to be *full*.
- a 4-component vector (dr, dn, h_2, h_3) is used for *ascender* lines.
- a 3-component vector (dr, dn, h_3) is used for *descender* and *short* lines.

8.2 Classification results

The classification was done on 100 French text lines for each font with almost the same length as those used for learning.

In a first trial, the classification was done with known font families. Global recognition rates for different attributes (weight, slope and size) are shown in Table 3. For each family 16 fonts have been considered (four sizes, two slopes and two weights).

The classification gave better discrimination than the statistical evaluation from the knowledge base. This is due to the property of multivariate classification which implicitly takes into account correlations between features. This confirms the interdependence of the different features used in the classification process.

Another classification trial was done on a set of font families, including the seven learned font families, with a database of 112 fonts. The recognition rates are shown in Table 4. The weight and slope discrimination are still accurate, even when the family and sizes are unknown. The family and size discrimination were surprisingly good, confirming the fact that the features are interdependent.

Finally, confusion between font families was tested; the results are presented in Table 5. If we consider three font family classes (seriffed, sanserif and monospaced), when a font family is misclassified it is mainly confused with a font family within the same class.

9 CONCLUSIONS

We have shown in this paper the importance of font identification and the reliability of an *a priori* identification based on statistical analysis of projection profiles. We have presented some features allowing an accurate discrimination of font weight and slope, but we think

Table 3. Recognition rates of a Bayesian multivariate classifier for each family

Family	**Weight**		**Slope**		**Size**			
	normal	*bold*	*roman*	*italic*	*10*	*11*	*12*	*14*
Courier	1.0	1.0	1.0	0.997	0.974	0.885	0.920	0.995
Helvetica	1.0	1.0	1.0	0.999	1.0	0.998	1.0	1.0
Avant Garde	1.0	1.0	1.0	0.991	0.998	0.987	1.0	1.0
Times	0.992	0.996	1.0	0.991	0.945	0.904	0.981	1.0
Palatino	1.0	0.994	1.0	0.993	1.0	0.966	0.995	1.0
Bookman	1.0	1.0	1.0	0.997	0.997	0.988	0.982	0.994
New Century Schoolbook	1.0	0.995	1.0	0.985	0.928	0.852	0.916	0.994

Table 4. Global recognition rates of a Bayesian multivariate classifier including all families

	Family	**Weight**	**Slope**	**Size**
Courier	0.992	0.995	0.995	0.887
Helvetica	0.844	0.991	0.996	0.978
Avant Garde	0.936	0.998	0.996	0.975
Times	0.895	0.997	0.998	0.958
Palatino	0.924	0.996	0.997	0.950
Bookman	0.922	0.999	0.996	0.976
New Century Schoolbook	0.923	0.999	0.999	0.977

Table 5. Confusion rates between font families including all families

	Courier	Helvetica	Avant G	Times	Palatino	Bookman	New C
Courier	**0.992**	0.002	0.003	0.001	0.0	0.001	0.002
Helvetica	0.005	**0.844**	**0.115**	0.012	0.004	0.010	0.010
Avant Garde	0.002	**0.033**	**0.936**	0.011	0.0	0.011	0.007
Times	**0.015**	0.003	0.002	**0.895**	**0.016**	0.003	**0.066**
Palatino	0.004	0.002	0.001	**0.044**	**0.924**	**0.013**	**0.011**
Bookman	0.003	0.004	0.007	0.004	0.006	**0.922**	**0.054**
New Century Schoolbook	0.002	0.0	0.0	**0.042**	**0.020**	**0.012**	**0.923**

that we do not yet have a classifier for omni-font recognition (discrimination of the weight and slope of any font family and size). Size discrimination was accurate when the font family was known. The correlation between the selected features seems to be powerful for family discrimination on a limited number of fonts but needs a more thorough analysis to be confirmed.

The classifier's theoretical performance has been evaluated, using intuitive methods. A more comprehensive study needs to be done. A standard eigenvalue analysis of variance on the classifier must be done to show the discrimination power of individual features and to find if some linear combination of the basic features can be better than any of them taken separately.

Some other problems, on which we are presently working, still need to be analysed. First, other features are studied in order to discriminate font families or family classes. Second, an evaluation of the approach on a large number of fonts has to be done. Third, the influence of sample length on discrimination power (line length, number of words) has to be analysed in order to evaluate the approach on single words. Finally, an analysis of the influence of document degradation (faxes, photocopies) should also be considered.

REFERENCES

1. Rolf Ingold, *Structures de documents et lecture optique: une nouvelle approche*, collection Méta, Presses Polytechniques Féderales, 1989.
2. J. Dreyfus and F. Richaudeau, *La chose imprimée: histoire, technique esthéthique et réalisation de l'imprimerie*, Editions Retz, 1985.
3. B. Bauermeister, *Manual of Comparative Typography: The PANOSE System*, VNR Company, New York, 1988.
4. Peter Karow, *Digital Formats for Typefaces*, URW Verlag, 1987.
5. R. Rubinstein, *Digital Typography: An introduction to type and composition for computer system design*, Addison-Wesley, 1988.
6. Robert A. Morris, 'Classification of digital typefaces using spectral signatures', *Pattern Recognition*, **25**(8), 869–876, (1988).
7. Julien Ch. Anigbogu, *Reconnaissance de Textes Imprimés Multifontes à l'aide de Modèles Stochastiques et Métriques*, Ph.D. dissertation, Université de Nancy I, 1992.
8. A. Lawrence Spitz, 'Recognition processing for multilingual documents', in *International Conference on Electronic Publishing, Document Manipulation & Typography*, pp. 193–205, Gaithersburg, Maryland, (1990).
9. Richard O. Duda and Peter E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.
10. Harald Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1946.