A Formal Approach to Textons and Its Application to Font Style Detection

A. Schreyer[1], P. Suda[2] and G. Maderlechner[2]

[1]University of Technology Munich, Institute for Human-Machine-Communication
Arcisstr. 21, 80290 Munich, Germany
[2]Siemens AG, Corporate Technology
Otto-Hahn-Ring 6, 81730 Munich, Germany
{Angela.Schreyer,Peter.Suda,Gerd.Maderlechner}@mchp.siemens.de

Abstract. In this paper we present a formal approach to Document Analysis based on the notions of textons and texture. This theoretical framework is then used to define texture-based font style classifiers. Tests on scanned document pages indicate good performance of the classifiers.

1 Introduction

Humans are able to identify regions in a document and their characteristics at first glance, as text is treated as texture by the human visual system. This thesis is generally supported by Watt ([10]) and more specifically by Bloomberg ([1]).

Apparently, text in the image of a document page consists of discrete perceptual units: strokes, characters, words, lines and blocks. Thus, it is reasonable to adopt the basic ideas of a well-known psychological theory of human texture perception: Julesz' Texton Theory ([5]). This theory defines textures as aggregates of textons, where textons are discrete, simply shaped elements of the image with a small number of basic, observable features ("perceptual atoms").

For the use in document analysis we developed a formal theory that fixes the notions of textons and texture in a proper mathematical way. The basic idea underlying this theory is that textures can be described by the textons as building blocks and by two relations: a neighborhood relation and a similarity relation, expressing the spatial distribution and the similarity of textons. Based on the theory, we show how font style classifiers for italics, bold and all-capitals can be defined.

Our special interest in font style detection is due to the fact, that style changes in a document "pop out" ([7]) and guide us in extracting relevant information.

In contrast to Bloomberg ([1]), who constructs morphological classifiers for texture-based font style analysis, we explicitly model elements of font style textures by textons. Automatic detection of italics, bold and all-capitals by explicit features is also described in[2]. The usefulness of font style detection for information retrieval is mentioned, but the paper's aim is improving OCR by previous font style analysis.

S.-W. Lee and Y. Nakano (Eds.): DAS'98, LNCS 1655, pp. 72-83, 1999.

Not only the font style, but the predominant font of running text in an English document is identified in [6] using a database of function words in various fonts.

In [4] the term texton is used in a different meaning, namely for the logical components of text (characters, words, etc.) in a hierarchical manner.

The overall aim of our work is attention-based extraction of relevant information from documents by classifying typography and layout features. Doermann ([3]) also uses these features, in order to determine the role of document parts in the process of transferring information.

2 Formal Definition of Textons

Julesz defines textons to be discrete elements or objects with a set of observable features. This definition can be formalized in a simple way:

Definition 2.1: A *texton structure TS* is a 4-tuple $TS=(D,\Gamma,E,\Phi)$ with the properties

1. $D\in \mathbf{R}^2$ is a non-empty compact subset of $\mathbf{R}^2$.
2. Γ is a finite partition of D and for each $\gamma\in\Gamma$, γ is non-empty and compact.
3. $E=E_1\times\ldots\times E_n$ is a feature space.
4. $\Phi:\Gamma\rightarrow E$ is a function, assigning each γ from Γ a feature vector from E.

Remarks:

1. D is called the *domain,* and a pair $t=(\gamma, \Phi(\gamma))$ is called a *texton* in the texton structure TS. Given a texton $t=(\gamma, \Phi(\gamma))$, then $Dom(t)$ denotes the *carrier* γ and $Prop(t)$ the *feature vector* $\Phi(\gamma)$ of the texton t.
2. Given two textons s and t, then $Dom(s)\cap Dom(t)\neq\emptyset \Leftrightarrow s=t$. This models the assumption that textons are discrete objects.
3. Let $TEX_{D,\Gamma,E,\Phi}$ denote the set of all textons of a given texton structure $TS=(D,\Gamma,E,\Phi)$. If it is not misleading we write TEX instead of $TEX_{D,\Gamma,E,\Phi}$.
4. For n-dimensional texton structures we can replace $\mathbf{R}^2$ by $\mathbf{R}^n$. For discrete texton structures we take the integers $\mathbf{Z}$ instead of the reals $\mathbf{R}$.

For a formal definiton of texture we consider that Julesz defines textures as aggregates of similar textons of constant density. In our framework this means that textons have to be connected spatially and that they should have similar features. To model connectedness and similarity we introduce two relations:

Definition 2.2: Let $TS=(D,\Gamma,E,\Phi)$ be a texton structure and TEX the set of all textons. Then $N\subseteq TEX\times TEX$ is called a *neighborhood relation* in TEX, if

1. $\forall t \in TEX$: $(t, t) \in N$.
2. $\forall s, t \in TEX$: $(s, t) \in N \rightarrow (t, s) \in N$.

Definition 2.3: Let E be a feature space,then $\Sigma\subseteq E\times E$ is a *similarity relation*, if Σ is an equivalence relation.

Remarks:

1. Σ induces a partition of the feature space. Given two textons s and t, and a similarity relation Σ, then s and t have equivalent features in respect of Σ, if $(Prop(s), Prop(t))\in\Sigma$.

2. Each similarity relation Σ can be composed from component similarity relations $\Sigma_i \subseteq E_i \times E_i$. It is also evident that each similarity relation Σ induces component similarities $\Sigma_1, ..., \Sigma_n$.

Definition 2.4: Let $TS=(D,\Gamma,E,\Phi)$ be a texton structure, TEX the set of all textons, $N \subseteq TEX \times TEX$ a neighborhood relation and $\Sigma \subseteq E \times E$ a similarity relation. Then two textons s and t from TEX are *N-Σ-connected,* if there exists a chain $t_1, ..., t_n \in TEX$ such that:

1. $s = t_1$ and $t = t_n$.
2. $\forall i = 1, ..., n-1$: $(t_i, t_{i+1}) \in N$ and $(Prop(t_i),Prop(t_{i+1})) \in \Sigma$.

Remark:

N-Σ-connectedness is the key to the definition of textures, because the underlying idea is that textures consist of chains of similar textons. It is evident that if s is N-Σ-connected with t, then t is also N-Σ-connected with s.

Definition 2.5: Let $TS=(D,\Gamma,E,\Phi)$ be a texton structure, TEX the set of all textons, $N \subseteq TEX \times TEX$ a neighborhood relation and $\Sigma \subseteq E \times E$ a similarity relation. Then a set $T \subseteq TEX$ is a *N-Σ-texture*, if

1. $\forall s, t \in T$: s and t are N-Σ-connected,
2. $\forall s \notin T$: there is no $t \in T$ such that s and t are N-Σ-connected.

Remarks:

1. Textures are sets of maximally N-Σ-connected textons.
2. Let S and T be N-Σ-textures, then $S \cap T \neq \emptyset \Leftrightarrow S=T$. This means that textures do not share textons. In addition, for textures S and T with $S \neq T$ the sets $\bigcup_{s \in S} Dom(s)$ and $\bigcup_{t \in T} Dom(t)$ are disjoint.

Under the definition's viewpoint document segmentation means to find a hierarchical sequence of textures in the document, i.e. to establish *a texton hierarchy*. When performing a bottom-up analysis, this hierarchy is installed by the lemma:

Lemma 2.1: Let $TS=(D,\Gamma,E,\Phi)$ be a texton structure, TEX the set of all textons and $TEXTURES$ the set of all N-Σ-textures, that are formed by a neighborhood relation $N \subseteq TEX \times TEX$ and a similarity relation $\Sigma \subseteq E \times E$.

Furthermore, be $E^*=E^*_1 \times \ldots \times E^*_m$ a feature space, $\Delta: TEXTURES \rightarrow E^*$ and $\Theta: TEXTURES \rightarrow \Gamma^*$, where $\Gamma^*=\{ \bigcup_{t \in T} Dom(t) : T \in TEXTURES\}$ and $\Theta(T)= \bigcup_{t \in T} Dom(t)$.

Then there exists a function $\Phi^*: \Gamma^* \rightarrow E^*$ such that the diagram Fig.1 commutes.

$$\begin{tikzcd} & \text{TEXTURES} \arrow[dl, "\Theta"'] \arrow[dr, "\Delta"] & \\ \Gamma^* \arrow[rr, "\Phi^*"'] & & E^* \end{tikzcd}$$

Fig. 1. $\Phi^*: \Gamma^* \rightarrow E^*$

Remarks:

1. As Γ^* is a finite partition of D, and Φ^* is a feature assignment according to Definition 2.1, $TS^*=(D,\Gamma^*,E^*,\Phi^*)$ is a texton structure.
2. The mapping Δ reflects that the features of a constructed texton at the new level are calculated from features of those textons used in construction.
3. As higher level textons are constructed from textures (= sets of lower level textons), there is a natural inclusion between the textons in a hierarchy. To denote that a lower level texton t contributed to the creation of a higher level texton s we write in the following $t \succ s$.
4. The theoretical framework also allows to assign statistical features to a higher level texton. For example, let $\Xi \subseteq E \times E$ be a similarity relation different from Σ. Then a feature $e \in E$ is dominant for a texture T in respect of Ξ, if
   $\forall e' \in E$: $|\{t: (Prop(t), e) \in \Xi\}| \geq |\{s: (Prop(s), e') \in \Xi\}|$.
   The feature e can be taken as a characterizing feature of the higher level texton. This is exactly what is done for font style detection with more elaborate statistical methods (Chapter 4).
5. So far, we were only concerned with one texton structure over a domain D and the construction of hierarchies. In practice, we have to deal with different interwoven texton structures simultaneously reflecting different textural aspects.

In general, document analysis uses rather simple features to characterize textons. Among them are size and position of a texton as follows:

Let Γ be a partition of D. Then we can define for each $\gamma \in \Gamma$:

$$TOP(\gamma) = \max\{y : (x, y) \in \gamma\}, \qquad BOT(\gamma) = \min\{y : (x, y) \in \gamma\}$$
$$LEFT(\gamma) = \min\{x : (x, y) \in \gamma\}, \qquad RIGHT(\gamma) = \max\{x : (x, y) \in \gamma\}$$
$$HEIGHT(\gamma) = TOP(\gamma) - BOT(\gamma) + 1, \qquad WIDTH(\gamma) = LEFT(\gamma) - RIGHT(\gamma) + 1$$

Subsequently, one can define a mapping $\Phi(\gamma)=(TOP(\gamma), BOT(\gamma), LEFT(\gamma), RIGHT(\gamma), HEIGHT(\gamma), WIDTH(\gamma))$, providing us with a basic feature vector. Apart from these basic geometric features, a texton may have additional features, e.g. area, orientation and, for higher level textons, the amount of lower level textons used for construction, etc..

3 Grouping and Texton Hierarchy

A short explanation of the grouping process is given, because in our approach the hierarchical grouping of strokes (s) to character textures, character to word textures, etc. is a prerequisite for font style detection. Steps upwards in the texton hierarchy are done, when we construct from a character texture the character texton c, from a word texture the word texton w, from a line texture the line texton l and from a block texture the block texton b, such that

$$s \succ c \succ w \succ l \succ b.$$

Although the texton theory is general enough to describe grouping on bit and run level, we start at the level of strokes. The carrier γ of a stroke is an area of 8-connected black runs, such that there is no split- or merge-run in the area. (split-/merge-run: at least two connected runs in the next/previous image line). The strokes are grouped to character textons by the strokes' geometric properties. The neighborhood relation N is the 8-connection relation between strokes. The similarity relation Σ considers width, height and orientation of a stroke.

On the basis of character textons the word textures can be found, for example with relations similar to the following:

Let $c1, c2, c3$ be character textons, then

$(c1, c2) \in N :\Leftrightarrow$

$RIGHT\_OF(c1, c2) \wedge \forall c3{:}DISTANCE(c1, c2) \leq DISTANCE(c1, c3) \wedge$

$DISTANCE(c1, c2) \leq \min\{HEIGHT(c1), HEIGHT(c2) \wedge$

$DISTANCE(c1, c2) < MAXDIST \wedge \ldots$

$(c1,c2) \in \Sigma :\Leftrightarrow |HEIGHT(c1)-HEIGHT(c2)| \leq \min\{HEIGHT(c1), HEIGHT(c2\} \cdot 0.67 \wedge$

$OVERLAP(c1, c2) > \min\{HEIGHT(c1), HEIGHT(c2)\} / 4 \wedge \ldots.$

The distance is given by the difference between the left border of *c2* and the right border of *c1*. *c2* is right of *c1*, if the left border of *c2* is right of the left border of *c1* and *c1* and *c2* have an overlap in vertical direction.

Next, the word textons are grouped to lines by quite similar neighborhood and similarity relations.

During grouping, features of the newly established textons can be calculated, e.g. the word and line features geometrical borders, base line, top line, average character height, area, and the line feature line skew.

The most sophisticated grouping step is the investigation of block textures, as we get confronted with a multitude of different two-dimensional texture models that are based on the multitude of possible block layouts. The basic neighborhood relation N is given by the distance between two vertically adjacent lines. The basic similarity relation Σ is given by the average character height within a line.

In the following, we use *STROKES* to denote the set of all stroke textons, *CHARS* for the character textons, *WORDS* for the word textons, *LINES* for the line textons, and *BLOCKS* for the block textons in a document.

4 Font Style Detection

Based on the results of the grouping process a further texture analysis determines different font styles for the words in the document. The classification process gives a new feature for a word texton based on the features of the textons on lower levels. This new feature can be used with a new similarity relation in a more elaborate texture analysis to find pop outs on a higher level.

4.1 Italics Classification

Italics font style is characterized by the slant of the approximately vertical stroke textons in a word. The slant of a stroke as a stroke texton feature is given by the angle α between the vertical line and the connecting line of the top-most and bottom-most black run of the stroke (Fig. 2).

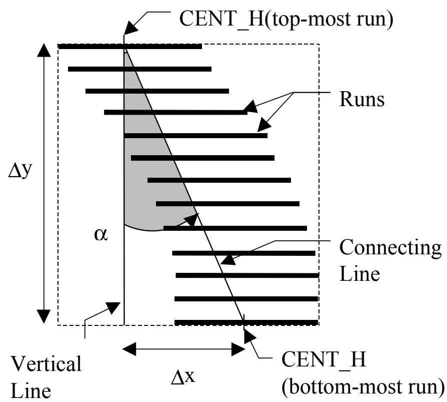

Fig. 2. Slant of a stroke

If $s \in STROKES$, then the slant of s is

$$SLANT(s) = \tan(\alpha) = \frac{CENT\_H(TOP\_MOST\_RUN(s)) - CENT\_H(BOTTOM\_MOST\_RUN(s)) + 1}{HEIGHT(s)},$$

where *CENT_H* denotes the middle of the according top-most and bottom-most run.

To derive a mean slant for a word texton $w \in WORDS$, we investigate the approximately vertical strokes, that contributed to the word texture:

$$A\_VERTICALS(w) = \{s : s \succ w \wedge HEIGHT(s) > WIDTH(s)\}$$

The mean slant of the word w is given by

$$MEAN\_SLANT(w) = \sum_{s \in A\_VERTICALS(w)} SLANT(s) / |A\_VERTICALS(w)|$$

and the mean slant of a text block b∈ BLOCKS by

$$MEAN\_SLANT(b) = \sum_{w \in \{w : w \succ b \wedge w \in WORDS\}} MEAN\_SLANT(w) / \left|\{w : w \succ b \wedge w \in WORDS\}\right| .$$

With these features, the new feature *ITALICS*(w) for the word texton w is given:

Let w be a word texton and b the block texton with $w \succ b$, then

$$ITALICS(w) :\Leftrightarrow MEAN\_SLANT(w) > TH \vee MEAN\_SLANT(b) > TH.$$

According to Table 1, we have set the threshold $TH = 0.12 \approx \tan(7°)$.

Due to the statistical nature of the feature, word slants are only reliable for words that have a significant number of approximately vertical strokes. Experiments have shown nice classifier performance for words with more than three characters.

4.2 Bold Classification

A word in bold font style is characterized by its black intensity, which is caused by the number of black pixels per area in the word core.

The *base line* of a line texton l is the regression line through the bottoms of the characters without descender in the line. The *top line* is the regression line through the tops of the capitals and characters with ascender. As we assume the lines to be deskewed, the regression lines are given by the two y-values $TOP(l)$ and $BASE(l)$.

Given a word texton w in a line l ($w \succ l$), then the *word core* of w is given by:

$$WORD\_CORE(w)= \{(x,y): LEFT(w) \leq x \leq RIGHT(w) \wedge BASE(l) \leq y \leq BASE(l)+h \cdot (TOP(w)-BASE(l)+1)\},$$

with h=0.67. h has been deduced from the difference in height between lowercase and uppercase characters in different fonts. Actually, the ratio is 0.74 (Table 1), but we take a smaller ratio to ensure measurement in the real core area of the word.

Note, that the word core is not a texton. It is only a region in the basic domain D underlying the texton structure.

Given the word core for a word w in line l we can define the *black intensity of the word w* by:

$$BLACK\_INTENS(w)= |WORD\_CORE(w) \cap Dom(w)| / (h(TOP(w)-BASE(l)+1) \cdot WIDTH(w)).$$

As this local intensity cannot determine bold font style, we investigate the intensity distribution across the document by setting up a histogram of the intensity values for the document. From this histogram we derive a threshold for classification.

Let $1 \leq i,j \leq N_{bin}$, then the histogram of black intensity values is given by

$$H(i) = \left|\{w: w \in WORDS \wedge i = \lfloor N_{bin} \cdot BLACK\_INTENS(w) \rfloor\}\right|$$

and the derived *threshold* by

$$TH = (\max\{i : \forall j : H(i) \geq H(j)\} + d) / N_{bin},$$

with N_{bin} =100 and d=8 set empirically.

We also define the *black intensities for blocks*. Let b be a block texton, then

$$BLACK\_INTENS(b)= \sum_{w \in \{w: w \succ b \wedge w \in WORDS\}} BLACK\_INTENS(w) / |\{w: w \succ b \wedge w \in WORDS\}|.$$

Thus, the new feature $BOLD(w)$ for the word texton w is given by:

Let w be a word texton and b the block texton with $w \succ b$, then

$$BOLD(w) :\Leftrightarrow BLACK\_INTENS(w) > TH \vee BLACK\_INTENS(b) > TH.$$

This classifier works nice for words with a reasonably sized word core, i.e. for words with more than 3 characters. For words with less characters we can extend the definition by using the direct predecessor $PRED(w)$ and successor $SUC(w)$:

If w is a word with less than four characters, then

$$BOLD(w) :\Leftrightarrow BLACK\_INTENS(w) > TH \wedge$$
$$(BLACK\_INTENS(PRED(w)) > TH \vee BLACK\_INTENS(SUC(w)) > TH).$$

4.3 All-capitals Classification

All-capitals words are characterized by the heights of the contained characters. For statistic classification, the character heights are compared to an adaptive threshold that is calculated from the mean height of lines in a block.

The mean height of lines in a block b is given by

$$MEAN\_L\_HEIGHT(b) = \sum_{l \in \{l : l \succ b \wedge l \in LINES\}} (TOP(l) - BASE(l) + 1) \,/\, \left|\{l : l \succ b \wedge l \in LINES\}\right|.$$

Given a word w in line l of a block b then

$$ALL\_CAPIT(w) :\Leftrightarrow \forall c \in CHARS, c \succ w : (TOP(c) - BASE(l) + 1 > h \cdot MEAN\_L\_HEIGHT(b)).$$

Given a block b with character c in line l, then

$$ALL\_CAPIT(b) :\Leftrightarrow$$
$$\frac{\left|\{c : c \succ b \wedge TOP(c) - BASE(l) + 1 > h \cdot MEAN\_L\_HEIGHT(b) \wedge c \in CHARS\}\right|}{\left|\{c : c \succ b \wedge c \in CHARS\}\right|} > 0.67$$

Experiments on synthetic documents (Chapter 5.1) have yielded the ratio of 0.74 between the height of lowercase characters without ascender or descender and the height of capital characters. For robust all-capitals classification h=0.85 is chosen.

Essentially, the ALL_CAPIT classifier for blocks counts big characters. If more than 2/3 of a block's characters are big, the whole block is an *all-capitals* block.

We end up in the following classifier for the new feature $ALL\_CAPITAL(w)$:

Let w be a word texton and b the block texton with $w \succ b$, then

$$ALL\_CAPITAL(w) :\Leftrightarrow ALL\_CAPIT(w) \vee ALL\_CAPIT(b).$$

The classifier does not treat words with less than three characters, because with these words there is ambiguity between words completely in all-capitals style and those that contain both capital characters and characters with ascender.

5 Tests and Results

5.1 Tests on Synthetic Documents

Synthetic documents were used to test features for font styles under standardized conditions. There were synthetic documents in four different typefaces and in 3 different font sizes each in italics and bold font style. All synthetic documents were printed on a laser printer at 600dpi and then scanned at 300dpi with fixed threshold.

Each synthetic document contains the alphabet once printed in lowercase and once in uppercase, and the numbers from 0 to 9. The characters are divided into groups of differing basic structure (present or lacking ascender and descender), so that the differently structured characters can be examined separately.

To investigate the threshold *TH* for italics classification, the average of the slants over the whole document was calculated for all synthetic documents, once in italic (A) and once in plain (B) font style (Table 1). A comparison suggests the value *TH*=0.12 at the lower bound of A as an appropriate threshold.

Table 1. Results for slant

Font	A	B
Arial, 8pt	0.13	0.02
Arial, 10pt	0.13	0.02
Arial, 12pt	0.11	0.02
Times New Roman, 8pt	0.17	0.01
Times New Roman, 10pt	0.19	0.01
Times New Roman, 12pt	0.19	0.01
Graphite Light, 8pt	0.19	0.11
Graphite Light, 10pt	0.25	0.12
Graphite Light, 12pt	0.19	0.11
Courier, 8pt	0.13	-0.01
Courier, 10pt	0.14	0.01
Courier, 12pt	0.14	0.02

Table 2 gives the results for the measurement of the ratio of character heights *A* = height(*lowercase without ascender and descender*)/height(*uppercase)*, that is used for defining the word core and classifying all-capitals. Its average is 0.74 for all font families and heights.

Table 2. Results for height ratio

Font	A
Arial, 8pt	0.76
Arial, 10pt	0.76
Arial, 12pt	0.75
Times New Roman, 8pt	0.72
Times New Roman, 10pt	0.72
Times New Roman, 12pt	0.70
Graphite Light, 8pt	0.67
Graphite Light, 10pt	0.73
Graphite Light, 12pt	0.65
Courier, 8pt	0.83
Courier, 10pt	0.77
Courier, 12pt	0.76

5.2 Tests on Scanned Document Pages

We tested our classifiers on document pages from the UW-III Document Image Database ([9]). We chose the 20 pages S000bin.tif to S00kbin.tif, that are scanned directly from original scientific papers at 300dpi. The text on every page (excluding graphics and formulas) was processed by the above classifiers (example: Fig. 3).

Table 3. Results for document pages S000bin.tif to S00kbin.tif

Font style	#(words in respective style)	#(correctly classified words in respective style	#(plain words)	#(falsely classified plain words)
Italics	378	350 (93%)	16338	19 (0.1%)
Bold	568	542 (95%)+11(all-capitals)	16148	9 (0.06%)
All-Capitals	94	89 (95%)	16622	24 (0.1%)+103(numerals)

Table 3 shows, how many words in the respective style were detected correctly and how many words in plain style were marked falsely as being printed in the respective font style. Note, that the bold classifier missed 11 bold words printed in all-capitals. With the present approach all-capitals are a problem, because the greater part of uppercase characters lies outside the defined word core. The all-capitals classifier also suffers from a structural problem, that is given by the fact that the height of numerals is almost the same as the height of uppercase characters.

6 Conclusion and Future Work

We presented a first step towards a texture-based approach to document analysis. We established a formal theory of texture, inspired by Julesz' Texton Theory, and transferred the psychological notion of textons to a mathematical one. Within our formal framework, the paper described classifiers for bold, italics and all-capitals font style. The classifiers can be implemented efficiently, because they use relatively simple features, that can be calculated partially during the preceding grouping process. The good performance of the classifiers on scanned documents showed the practical relevance of our approach.

At present, we are extending our theory, so that the detection of additional layout features in a document can be described consistently in the framework of textons. It is our aim to detect those layout features that would attract a human's attention at first glance, in order to enable attention-based extraction of relevant information from documents ([8]).


case of those who oppose the practice and creates conflicts among potential allies. Are the long-term interests of the scientific community best served by AAU, for example, taking the side of the executive branch versus that of Congress, the side of authorizers versus that of appropriators, and condoning earmarkers on the agriculture appropriations subcommittee? Only the clearest and most inclusive definition of earmarking truly supports the goal of achieving the highest-caliber science, not only within the academy but throughout the federal government. This definition should be adopted by organizations such as the National Academy of Sciences, the American Association for the Advancement of Science, and AAU to help them put a lid on the academic pork barrel.

DON E. DETMER
ELAINE B. STEEN

Patient Records in the Information Age

If you're out of town and need cash, an automatic teller machine can instantly determine if you have money in your account and let you withdraw what you need. If you apply for a loan, the lender can quickly get access to your credit history to help determine your reliability. If the Internal Revenue Service is reviewing your tax return and needs to know how much was withheld from your salary during the year, the information is readily available. But if you need emergency medical care and the physician needs information on your medical history, it is highly unlikely that it can be found and made available promptly. In fact, much of the information necessary to understand the workings of the U.S. health care system is not to be found no matter how long one is willing to wait. And without better access to medical data, the health care system will be hampered in its efforts to provide better care to individuals and to enhance the overall effectiveness of the entire system.

The information-management challenge experienced by health care professionals and institutions is growing daily. At least three factors contribute to this growth. First, health care practitioners must master and track an ever-increasing base of medical knowledge. MEDLINE, the computer data base of biomedical literature, grows by approximately 360,000 new articles per year. Second, patient records include more data as patients live longer, experience more chronic disease, undergo a greater variety of tests, and have more encounters with health care providers. Third, the demand for patient data is increasing. In addition to supporting the diagnostic and therapeutic work of clinicians, patient data are used to document patient risk factors, expectations, and satisfaction with treatment; to perform quality assurance, risk management, cost monitoring, and utilization review; to identify emerging public health problems; to track adverse reactions to pharmaceuticals; to document

Broad cooperation is needed to help institute a national computer-based system of medical data.

Don E. Detmer, vice president and provost for health sciences at the University of Virginia, was chairman of the Institute of Medicine's Patient Record Committee. Elaine B. Steen, assistant to the vice president and provost for health sciences, was staff officer for the IOM committee.

24 ISSUES IN SCIENCE AND TECHNOLOGY

Fig. 3. Grouping and font style detection on s00mbin.tif (dark grey: bold, medium grey: italics, light grey: all-capitals)

References

1. D. S. Bloomberg: Multiresolution Morphological Approach to Document Image Analysis, Proc. ICDAR, Saint-Malo, 1991, pp. 963 – 971
2. B. B. Chaudhuri, U. Garain: Automatic Detection of Italic, Bold an All-Capital Words in Document Images, Proc. ICPR , Brisbane, 1998, pp. 610 – 612
3. D. Doermann, A. Rosenfeld, E. Revlin: The Function of Documents, Proc. ICDAR, Ulm, 1997, pp. 1077 - 108
4. D. Dori, D. Doermann, C. Shin, R. Haralick, I. Phillips, M. Buchmann, D. Ross: The Representation of Document Structure: A Generic Object-Process Analysis in P. S. P. Wang and H. Bunke (editors): Handbook on Optical Character Recognition and Document Image Analysis, 1996
5. B. Julesz, J. R. Bergen: Textons: The Fundamental Elements in Preattentive Vision and Perception of Textures, The Bell System Technical Journal, Vol. 62, No. 6, 1983, pp. 1619 - 1645
6. S. Khoubyari, J. J. Hull: Font and Function Word Identification in Document Recognition, Computer Vision and Image Understanding, Vol. 63, No. 1, 1996, pp. 66 - 74
7. T. S. Klitz, J. S. Mansfield, G. E. Legge: Font "Pop Out" in Text Images, OSA Annual Meeting Technical Digest, 23, 1992, pp. 170
8. A. Schreyer, P. Suda, G. Maderlechner: The Idea of Attention-Based Document Analysis, Proc. DAS, Nagano, 1998
9. UW-III English/Technical Document Image Database, CD-ROM, University of Washington, 1996
10. R. Watt: The Visual Analysis of Pages of Text in R. Sassoon: Computers and Typography, Intellect Books, Oxford, 1993, pp. 179 – 201