# Reading Systems: An introduction to Digital Document Processing

Lambert Schomaker

Artificial Intelligence Dept.
Groningen University
The Netherlands
schomaker@ai.rug.nl
http://www.ai.rug.nl **

**Abstract.** As an introduction to the area of digital document processing we first take a few steps back and take a look at the purpose of digital document processing. Subsequently a detailed comparison between the human and the artificial reading system is made. Finally, the chapter provides an overview on the book as a whole.

Methods for the *creation and persistent storage* of text [10] have existed since the Mesopotamian clay tablets, the Chinese writings on bamboo and silk as well as the Egyptian writings on papyrus. For *search and retrieval*, methods for systematic archiving of complete documents in a library were developed by monks and by the clerks of emperors and kings in several cultures. However, the technology of *editing* an existing document by local addition and correction of text elements has a much younger history. Traditional copying and improvement of text was a painstakingly slow process, sometimes involving many man years for one single document of importance. The invention of the pencil and eraser in 1858 was one of the signs of things to come. The advent of the typing machine by Sholes in 1860 allowed for faster copying and a simultaneous on-the-fly editing of text. The computer, finally, allowed for a very convenient processing of text in digital form. However, even today, methods for generating a new document are still more advanced and mature than are the methods for processing an existing document.

** Appeared as Chapter 1 in: Digital Document Processing (2007). B. Chaudhuri (Ed.). Springer, pp. 1-28, ISBN 978-1-84628-501-1 (Advances in Pattern Recognition Series)

This observation may sound unlikely to the fervent user of a particular common word-processing system, since creation and correction of documents seems to pose little problems. However, such a user has forgotten that his or her favorite word-processor software will only deal with a finite number of digital text formats. The transformation of the image of an existing paper document - without loss of content or layout - into a digital format which can be textually processed is mostly difficult and often impossible. Our user may try to circumvent the problem by using some available software package for optical-character recognition (OCR). Current OCR software packages will do a reasonable job in aiding the user to convert the image into a document format which can be handled by a regular word-processing system, provided that there are optimal conditions with respect to:

- image quality;
- separability of the text from its background image;
- presence of standard character-font types;
- absence of connected-cursive handwritten script and
- simplicity of page layout.

Indeed, in those cases where strict constraints on content, character shape and layout do exist, current methods will even do quite a decent job in faithfully converting the character images to their corresponding strings of digital character codes in ASCII or Unicode. Examples of such applications are postal address reading or digit recognition on bank checks.

On the other hand, if the user wants to digitally process the handwritten diary of a grandparent or a newspaper snippet from the eighteenth century, the chances of success are still dim. Librarians and humanities researchers worldwide will still prefer to manually type ancient texts into their computer while copying from paper rather than entrusting their material to current text-recognition algorithms. Not only is the word processing of arbitrary-origin text images a considerable problem. Even if the goal can be reduced to a mere search and retrieval of relevant text from a large digital archive of heterogeneous text images there are many stumbling blocks. Furthermore, surprisingly, not only the ancient texts are posing problems.

Even the processing of modern, digitally created text in various formats such as web pages with their mixed encoded and image-based textual content will require "reverse engineering" before such a digital document can be loaded into the word processor of the recipient. Indeed, classification of text within an image is so difficult that the presence of human users of a web site is often gauged by presenting them with a text fragment in a distorted rendering which is easy on the human reading system but an insurmountable stumbling block for current OCR systems. This weakness of the artificial reading system thus can be put to good use. The principle is known as "CAPTCHA: Completely Automated Public Turing Tests to Tell Computers and Humans Apart [5]. During recent years, yet another exciting challenge has become apparent in pattern-recognition research. The reading of text from natural scenes as recorded by a camera poses many problems, unless we are dealing with a heavily constrained application such as, e.g., the automatic recognition of letters and digits in snapshots of automobile license plates. Whereas license-plate recognition has become a 'mere' technical problem, the camera-based reading of text in man-made environments, e.g., within support systems for blind persons [8], is only starting to show preliminary results.

Non-technical users will often have difficulties in understanding the problems in digital-document processing (DDP) and in optical character recognition. The human reading process evolves almost effortlessly in the experienced reader. Therefore, it is difficult to explain to users that machine reading of documents involves a wide range of complicated algorithms, which in one way or another must emulate the processing of text by the human brain. Where the human eye samples the text opportunistically from syllable to syllable and from word to word, the algorithms in DDP will scan a complete page image and will need to segment its layout at a number of hierarchical levels. Where the human reader eclectically takes into account several layers of contextual information on language and text topic, systems in DDP are still limited in their analytic and associative capabilities to handle arbitrary textual input. Rather than aiming to develop a *Universal Reading Machine*, most engineering efforts today are aimed at dealing with a structured subset of all possible text-processing problems. In some cases, the machine already outperforms humans. A state-of-the art check reading system has been trained on more variants of handwritten digit shapes and styles than most humans will ever encounter during a lifetime. For the first few minutes of reading digits, human and machine will have comparable digit-classification performances. However, after half an hour, the machine will have won, not being distracted by problems of perceptual-cognitive concentration and/or tedium. For other tasks, where the human will prove to be the better reader, the machine still provides an attractive alternative due to the speed of processing and the massive amounts of data that can be processed, albeit in a crude manner. Let us take a look at the differences between human and machine reading. In Table 1, a comparison is made between aspects of the human and artificial reading system. We will focus on each of the numbered items in this table.

### (a) Text sensing

Humans are able to read texts easily as flat 2D patterns in orthonormal projection or as filled 3D objects from a wide range of poses. In machine reading, usually flat-bed scanners are used. There is an increased interest in the challenging problem of camera-based text reading from photographs or video recorded in a man-made environment with text patterns embedded within a natural scene. The input consists of an image $I(x, y)$ where the intensity $I$ may be scalar, as in gray-scale images, or vectorial, as in RGB color images. In some systems and application settings for automatic handwriting recognition such as, e.g., an enhanced mobile phone, recordings of pen-tip movement and writing force (pressure) can be recorded, yielding a signal $(x_t, y_t)$ or $(x_t, y_t, p_t)$.

### (b) Sensor scope

Although it is in conflict with our subjective experience, human vision does not function at all like the camera. Cameras and scanning devices are designed to record images with an even resolution over the whole visual field. The eyes, on the contrary, view a limited part of the visual field and are sampling the surrounding world in a series of saccades, i.e., eye jumps, and fixations. The illusion of a permanent 'Ganzfeld' is not in the eye, but is constructed by the occipital part of the brain. The human system has the advantage of not having

***Table 1.*** *A comparison between human and machine reading. Letters in the first column refer to corresponding paragraphs in the text.*

| Aspect | Human | Machine |
|---|---|---|
| (a) **text sensing** | any visual modality 2D,3D | -flat-bed scanner<br>-digitizer tablet<br>-camera |
| (b) **sensor scope** | opportunistic, limited-field panning sensor | full-page scan |
| (c) **sensor grid** | "log polar" | Cartesian |
| (d) **pre-processing power** | powerful, flexible and adaptive in hypothesizing 'layers' of foreground and background | designed for known ink and background colors and textures, difficulties with touching or overlapping shapes |
| (e) **affine invariance** | high | usually some position, size, slant and perspective sensitivity |
| (f) **ink thickness invariance** | high | often some ink-thickness sensitivity |
| (g) **shape features** | on-demand, various | prefixed, limited |
| (h) **processing type** | analytic and associative power | fixed, probability and/or shape-based models, brute-force computation |
| (i) **computing architecture** | open-ended | fixed processing pipeline |
| (j) **computing strategy** | satisficing | fixed thresholds and goals |
| (k) **knowledge base** | broad | narrow |
| (l) **cognitive reliability** | stochastic | deterministic |
| (m) **response in case of difficult input** | graceful degradation | counter-intuitive output |
| (n) **classification accuracy** | | |
| - machine print | very high | very high |
| - isolated hand print | very high | high |
| - cursive handwriting | | |
| - in context | high | low to medium |
| - out of context | medium to high | very low |
| - 3D multicolor text | very high | low |
| (o) **energy, concentration** | limited | indefatigable |
| (p) **speed** | limited | high |
| (q) **volume processing** | limited | massive |

to compute image transformations over the whole image. There is a price to be payed, naturally. The movements of the sensor must be guided by computed estimates of what is the most informative jump of the eye towards the next particular spot in the text. Additionally, the system must retain, more centrally in the cortex, all information in the image which is pertinent to reading but which is not in central, foveated view. Whereas the engineered system is forced to sense the complete field, the biological system uses 'opportunistic sampling'. This means that in machine reading, there exists a problem of layout analysis and segmentation which is completely different from the problems which are encountered and solved by the human reader. Figure 1 shows a recording of human eye movement during reading.

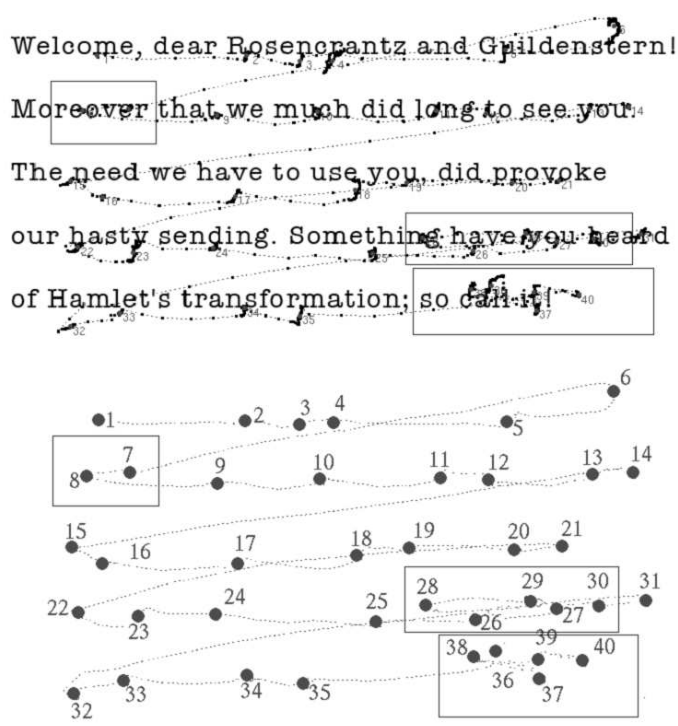


**Fig. 1.** Eye movements in the human reading of a text fragment from Shakespeare's Hamlet (Dutch subject). The reader is referred to the text for further details. Data courtesy H. van Rijn and L. van Maanen.

The top half of Figure 1 shows a text fragment from Hamlet, with dotted lines depicting the saccadic movements. In the bottom half of Figure 1, the eye movements are isolated from the text and the order of the fixation points is represented by their number. The regular reading behavior is characterized by a focus, which is located mostly just below the text line. Words and syllables are scanned, usually from left to right. At the end of the line there is an 'eyes return' (cf. 'carriage return') movement, to a horizontal position corresponding to about the fourth letter, in this font. The exceptions to the rule are informative. The word *'Moreover'* is fixated first in its rightmost half, probably due to a stochastic undershoot of fixation 7, which is compensated by a leftward saccade and fixation 8. Where there are syntactical surprises, the regular pattern is clearly disturbed. The Shakespearean phrasing surprises the Dutch reader, such that the scanning behavior is characterized by backtracking: fixations 25,26,27, leftwards to 28 etc. (in *'Something have you heard'*); and fixations 35,36,37 then leftwards to 38 (in *'so call it'*). Such recordings illustrate the purposive nature of reading. Contrary to the intuition in traditional views on human information processing, syntax and semantics are not 'late' end stages of a long and unidirectional processing pipeline, they directly influence the low-level sensing behavior. It can also be observed that the foreign name *'Rosencrantz'* elicits three fixations (2,3, and 4). In Dutch spelling, the nearest neighbour would be spelled 'Rozenkrans'. In order to decipher the exact spelling, iterative fixations seem to be required here to solve the reading puzzle at the level of orthography.

### (c) Sensor grid

Not only is human reading characterized by active control of selective attention, as opposed to the indeterminate recording of a large rectangular portion, i.e., a whole page, from the visual field by the machine. The distribution of receptive cells in the retina is determined by a dense region in the retina, the fovea, with a high resolution. As one moves away from the center, the resolution decreases, while at the same time temporal acuity, i.e., the sensitivity to changes in luminance over time, will increase. The organization of the receptors is dense, but not in any way regular, let alone Cartesian. Some have proposed that the most appropriate interpretation of the density of rods and cones in the retina is a log polar representation [28]. In any case, the human eye saves the brain from an undue amount of optic computation. The cost which is payed here is that an internal controlling model is needed to guide the eye muscles to aim the fovea at the angle in the visual field where the most informative next piece of information can be found.

### (d) Preprocessing

A second characteristic which saves the human visual cortex from undue computation is that preliminary forms of image filtering and edge enhancement already take place within the eye itself. What is more, the biological system in primates makes a distinction between "Where" and "What" channels for processing optical information. These two pathways start within the eye, i.e., a slow, parvocellular system representing the "What" information for pattern classification in area "V4"[1] of the visual cortex and a fast magnocellular system representing information to solve "Where" questions from motor control in area "MT" of the cortex [33]. Within area "V1", with its early projections coming from the optic nerve, additional optical preprocessing takes place on the basis of a distribution of center-surround kernels which are optimal for adaptive foreground/background separation and orientation detection by 'grating' neurons [18].

Although many improvements have been made in scanner and camera technology over the last decade, a lot can be improved. The reader is challenged to try to make a scan from a personal, old, coffee-stained, folded, crackled and lineated page with conference notes and attain a perfect foreground/background separation of the handwritten text. Obviously, local thresholding methods for binarization exist. Methods which are exclusively oriented on local luminance will fail in case complex texture and colors with low luminance-differences are involved. Humans appear to be able to maintain

[1] Brodmann (1868-1918) started a systematic analysis of the histology (tissue typing) of the brain, suggesting names for brain areas which usually coincide with a particular function. Modern brain atlases [26] are in use within functional magnetic-resonance imaging (fMRI) research, but reference to Brodmann is still common.

active hypotheses on layers of visual information yielding, subjectively, a solid percept of 'hovering layers' each representing paper texture, lineation, stains and, of course, the ink that makes up the shape of a character. In current OCR systems on the contrary, a problem as 'simple' as the touching characters will pose a problem. Usually heuristics are applied, which may work in some conditions but will fail miserably elsewhere. While some researchers are interested in biologically motivated methods for textural background removal [18] using Gabor functions, others exploit a combination of mathematical morphology and genetic algorithms [9] for trained foreground/background separation. Figures 2 and 3 give examples of, respectively, difficult and manageable historic material as regards foreground-background separation.

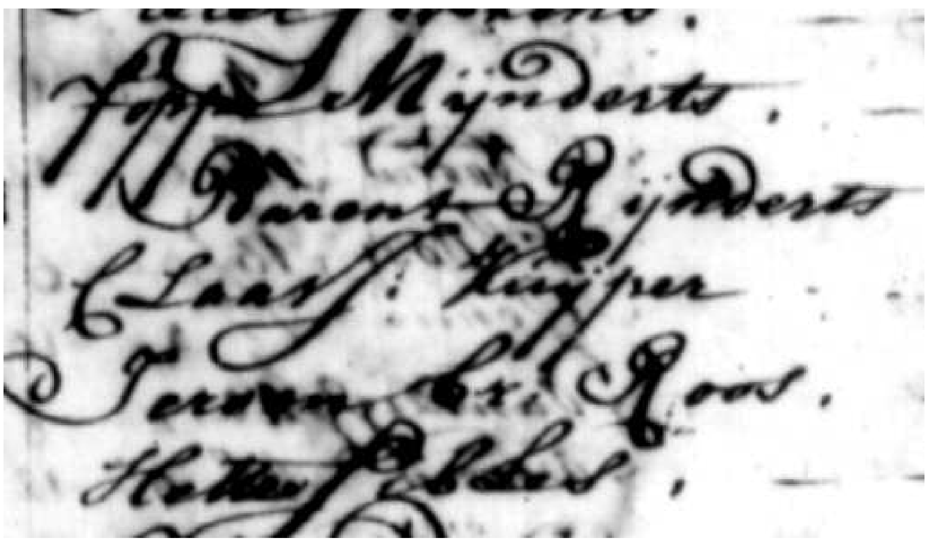

**Fig. 2.** Infeasible foreground-background separation in an eighteenth century Dutch shipping list. Even a native Dutch historian will have problems reading this material.

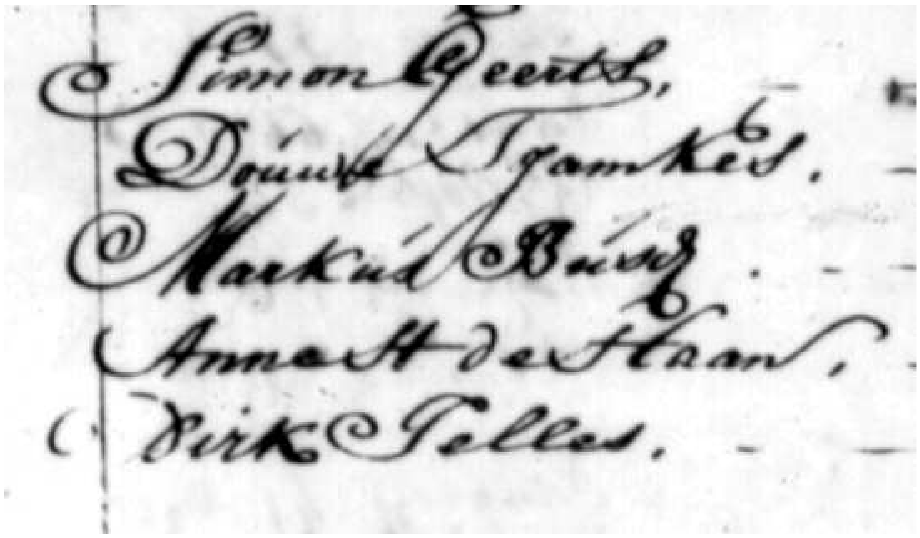

**Fig. 3.** Although not easy, foreground-background separation seems feasible on this eighteenth century Dutch sample. Both figures 2 and 3 are digitized from microfiche of a collection of harbour-tax registers: Paalgeldregisters, selected years 1744-1748, Gemeente Archief Amsterdam, courtesy of dr. George Welling. Automatic transcription of text and amounts in these financial documents is still impossible.

### (e) Invariance to affine transforms

Human reading is not completely insensitive [32] to affine transforms of character shape (position, scale, orientation, shear). This indicates that shape normalization in the human vision system might not be as simple as a matrix-vector multiplication: There exists a confounding between geometric processing and the statistics of the common text pose. As an example, reading speed will remain constant for a wide range of orientations, but reading text which is upside down will demand increased cognitive efforts. Similarly, there is a range of an acceptable shear transform of text, beyond which it becomes illegible. Position invariance is solved because the eye-movement control will lead to a relative coding of the image information. Notably, it is the output signal of the muscle-control signals which represents the position of a foveated syllable within the text. This demonstrates that perceptual facilities cannot be decoupled from motor facilities [22]. Scale invariance in human vision is very good for characters larger

than the lower limit of visual acuity and small enough to be foveated and focused as a complete pattern.
In systems for digital document processing we usually have to be pragmatic in order to realize a tolerance for affine transforms. Algorithm design in the field of pattern recognition is on the one hand characterized by the quest for a comprehensive mathematical formulation of an invariant geometric space for relevant shape features which would provide the necessary robustness against natural variation. This appears to be rather difficult. Therefore, there also exists, on the other hand, a quest for automatic geometric normalization tricks. Examples are the methods to estimate baseline pose, character size and slant, followed by a normalization stage after which non-invariant features may be used. Reasonable results can be obtained in this area, but also here, the flexibility of the machine is limited in comparison to biological vision with its head start of millions of years of evolution in a harsh environment. Here, as in the other levels of processing there is a 'chicken and egg problem': we need to have an idea about the content of, e.g., a camera-based snapshot of text to be able to unslant it while we need to unslant the pattern to be able to classify it reliably (Figure 4).


**Fig. 4.** A challenge to camera-based reading systems: perspective deformation, low contrast and glare make this a difficult sample. The addition of color would only partly alleviate the preprocessing problems (the text is: "HEMA").

### (f) Invariance to ink-trace thickness

In traditional OCR systems, it makes a difference whether one recognizes 12pt Times Roman in normal as compared to bold rendering. To the human reader, boldness in font rendering is an irrelevant superficial property of text. Some multi-font approaches, however, actually require that several rendering styles of a font family and size are represented within the recognition engine. Alternatively, omnifont character-recognition methods may attempt to minimize the influence of ink thickness by the use of thinning and skeletonization techniques. Although the stroke-width variation is reduced, new problems emerge due to thinning. It is especially the case in handwriting that the skeletonization to an ink-trace thickness of only one pixel may complicate recognition rather than facilitating it due to the addition of structural features in the residual trace where the pattern was simple, before. A simple crossing in the digit *8* may be replaced by a forked pattern of unpredictable orientation after thinning. Chapter 11 will contain an example of problematic skeletonization in Arabic script. Historical documents in particular will pose problems. Each time the ink pen or quill is dipped into the ink bottle, the trace of the subsequently written characters will be wide and dark. As the ink flows into the absorbing paper, the trace usually gets thinner and less satiated. For metal pens which make use of capillary ink flow, the trace may actually consist of two dark lines which are formed by the metal legs which are pressed apart too widely, the middle part of the trace being less satiated (Figure 5). Ink-trace thickness invariance remains an important problem.

**Fig. 5.** Realistic problems of ink width and grey-level variation. There are black boundaries in the horizontal line, where the metal pen scratched the paper, leading to increased ink absorption, Kabinet van de Koningin, Nationaal Archief, Den Haag, 1903).

### (g) Shape features

A plethora of feature representations has been developed in our research field to describe the shape of characters or words. The design of, e.g., a character-feature vector is still an art rather than a technology. Successful approaches are usually copied: the evidence for success is often derived from empirical evaluation rather than being obtained through principled design. No amount of mediocre features or classification algorithms is a substitute to an excellent feature. A good feature, i.e., a source of information with generalizing power as regards all instances of a target class together with a discriminative power as regards instances of irrelevant classes, will already yield reasonable performances on even simple nearest-neighbour search. There are some general criteria as regards features.
A good shape feature:

- is informative as regards class separation;
- is robust to noise;
- is invariant to affine transforms and ink width;
- is homogeneously scaled with other features in the feature vector;
- can be computed efficiently.

Character-based features are useful but they assume that characters can be segmented out of the input pattern stream. Word-shape features may avoid character segmentation but they must be powerful enough to give each lexical word its own unique and separable volume in the word-feature space. Unfortunately, little is known about the features which humans use in reading. The informative value of ascenders and descenders is so high [23], that it would be strange if human readers would not use it in reading. In reality, however, it is difficult to detect human usage of word-shape features if the individual characters are clearly separable from the background. There is some support for the notion that human perception is drawn to singularities [25] in the script pattern [23]. A clear disadvantage of many approaches to pattern recognition is the fact that all features must be computed for all classes in a fixed-dimensionality feature vector classification method. Some techniques, such as decision trees, allow for the ad hoc and on demand computation of feature values, per class. However, tree-based approaches suffer from the 'premature commitment' problem: An erroneous decision in the beginning leads to a certain misclassification in the end. Furthermore, while class-specific features may be desirable, the manual design of features per class quickly becomes prohibitive in free-style handwriting or international scripts with many shape classes. In this respect, human perception may be exquisitely equipped with mechanisms for opportunistic feature extraction, i.e., exploiting the presence of features as they are required to obtain a solid decision.

Currently a hot topic in neuro-scientific analysis of the reading process [30] is the question whether and how word-shape analysis plays a separate role next to character-level shape analysis in human reading [17, 23]. Recently, a brain region has been detected, dubbed "Visual Word Form Area" or VWFA. It is located in the left brain hemisphere, where language resides, notably in the left fusiform gyrus. The VWFA area is active while literate subjects are reading [6]. Also, this brain region shows a stronger activation in response to real letters and real words as compared to letter strings or pseudo-fonts of comparable visual complexity. Underscoring the separation between Where and What routes of processing, the activation in the VWFA region is invariant to the spatial location. Furthermore, the specific case or font type which is used to present words has no noticeable effect [6]. It has been suggested that the VWFA region contains orthographic representations of written words [4]. Dehaene et al. [7] recently proposed an explanation for the neural coding scheme of words which stresses a gradual transition between low-level features and high-level view-independent features. Rather than committing itself to "symbolic" codes, the neural activity simultaneously represents, from low to high-level features: local contrast; oriented bars; local contours; case-specific character shapes; case-unspecific letter detectors; local character bigrams upto small words and recurring ('reusable') substrings or morphemes. Such a model also would explain more easily than the strict left-to-right Markov models why a sequence such as "fsat cempoutr porgarm" can be decyphered as "fast computer program" by human readers mastering the English language [21, 15]. It is argued [7] that a minimalistic, symbolic code for visual words is unlikely. Rather, a diverse and redundant repertoire of reading-related neurons are expected to be involved, which make use of and modify the information delivered by the pre-existing primate visual system. After all, the phenomenon of reading only exists five to six thousand years. It is unlikely that structural, genetic changes have led to reading-specific modules in the brain, as is the case in speech, in such a short period.

### (h) Processing type

Human and machine reading are evidently different at the level of implementation, i.e., neural wetware vs. silicon hardware. Within the current context, the functional differences in processing are more relevant. Engineered systems provide a a level of reading performance which can be attributed both to (a) massive training with labeled data and (b) the availability of exact, 'brute' computations according to heavily structured algorithms. The process of human reading, however, is characterized by a limited amount of training on stylized characters provided by a single teacher and a schoolbook. Punctuated supervision, where character image and its spoken representation are coupled explicitly by the teacher, only lasts a relatively brief period at primary school. Generalization over script styles is realized in an independent manner by the reading child itself. In case of illegible script, human readers activate a number of cognitive facilities. If superficial association fails, the reader may solve the underlying reading puzzle by analytically combining elements from (a) the linguistic context, (b) the shape context [19, 31] as well as (c) the pragmatic context. At the very moment of being handed a handwritten shopping list, the human reader is precued to a very limited subset of the total lexicon, even before having attempted to read the first character. Similarly, the envelope with the logo of the tax department (IRS in the United States) will precue the reader as regards the linguistic context to be expected. Once reading has started, linguistic structure determines the interpretation of the patterns in a top-down manner. At the same time, the shape of characters in flanking words to the current gaze provides cues as regards the peculiarity of a current font or script style. It is the flexible manner in which the human reader actively hunts for the necessary information which contrasts distinctly with the stereotyped and rigid processing in machine reading.

### (i) Computing architecture

This brings us to the computing architecture. Although not enough is known about the details of human cognition, its architecture is very different from the architecture of a check-reading system or a word recognizer on a hand-held pen-based computer. To a large extent, our artificial systems are still built according to the factory metaphor: a pipeline where raw data comes in at the left, being refined in successive stages and dropping a perfect product in the form of a symbolic character string at its output on the right. Feedback processes have been tried, e.g., for the purpose of optimization of a geometric normalization stage. However, the variable response time which is the result of feedback is often unacceptable in industrial or interactive applications. Also, the uncertainty which is involved in incremental learning or live adaptation is considered much too large to yield an efficient and stable system behavior. Indeed, most reading systems implicitly rely on the fact that a human reader is still somewhere in the loop. Modern interactive handwriting-recognition systems - as are in use in the Tablet-PC [2] software by Microsoft - allow for interactive corrections to written words by adding or deleting elements in the input pattern. In such an approach, the design and implementation of the word classifier is heavily determined by considerations of computing architecture.

### (j) Computing strategy

There is an additional distinction between human and machine computing which needs to be taken into account. This distinction is both related to the intention of the computing process itself and to the way in which purposive computation proceeds, strategically. In machine reading, the designer of the algorithm strives for the perfect response, which is either a correct answer or a *Reject* response in case of unknown input material, given constant thresholds on probability or distance values. Human cognition, on the other hand, is more liberal and sloppy in its nature. The usual operating strategy in many aspects of human cognition is called "*satisficing*" [24]. The human brain, much as the electronic computer, has a finite computing

[2] Pen Computing Magazine, July 2001, `http://www.pencomputing.com/`

power. Perfect rationality requires an undue amount of computation such that a best possible response often cannot be given in a predictable and limited amount of computing time. It is therefore assumed that the human cognizer has adopted the strategy to be content with “bounded rationality”, yielding reasonably correct answers with an acceptable time delay. In order to survive, the trick is to apply heuristics in a predetermined order until a sufficient level of certainty is reached [11]. Such a sloppy strategy may indeed seem irrational if the reading puzzle can be solved in a closed and perfect manner, given the evidence which is present in the patterns. However, if there are many problems of uncertainty due to low pattern quality, computing the perfect answer becomes virtually impossible. Here, a satisficing reader may very well yield an average performance which is higher than that of a well-trained but bare-bones hidden-Markov model (HMM). The latter system has no intrinsic cognitive functionality to take a step aside and conclude, e.g., that it makes no sense to compute the posterior probabilities for all characters due to the bad image quality in the middle zone of the given word pattern. In this matter, natural cognition and practical systems might be closer than one would expect. Where performance counts, a large portion of an *engineered* reading system is devoted exactly to such smart use of heuristics and multiple strategies. It may come as a disappointing surprise to academicians, but as a rough estimate, less than 5% of the source code of reading systems is concerned with (statistical) shape classification, per se, even if the user-interface code is not counted. Of course it is important to strive for the best possible character classifier. However, without a proper embedding in a powerful goal-directed computing strategy, the isolated HMM classifier is as practical as a combustion engine on a shaky test bench.

### (k) Knowledge base

From the scientific viewpoint, keeping in mind Occam’s razor, it is very unfortunate, but the massive amount of parameters representing human knowledge is hard to beat by an artificial reading system. The availability of explicit knowledge allows an address-reading system to take into account diverse information such as word shape or the list of allowable ZIP codes. Unfortunately, the slightest cultural change will require human labour in the form of newly entered knowledge. In license-plate reading, once the character-classification problems have stabilized, a company will spend a considerable amount of time in manual knowledge engineering if their systems are sold in a new, hitherto unknown country with its own license-plate formats. One of the reasons why the *Universal Reading Machine* cannot be built, as yet, is that we cannot simply glue together knowledge from different areas and expect to have a more powerful system. With the addition of each new category of text (numbers, license plates, addresses, traffic-sign text, handwritten material etc.) current artificial systems will be clogged rather than helped by the growing amount of knowledge. In such a *Universal Reading Machine*, there will be a new problem: How to know which subset of all the available knowledge pertains to the problem at hand?

### (l) Cognitive reliability

Beside all its strengths, human cognition also has flaws. Apart from its intrinsic "satisficing" sloppiness, there is the ever-present problem of noise. To persons with a mathematical mind and a fondness for logic it may be disconcerting but the basic computing element of the human brain, the neuron, can be viewed as a generator of a stochastic point process in time [29]. By virtue of the population statistics of the firing behavior of groups of neurons, a behavior results which is partly deterministic, however, with an intrinsic amount of residual noise which is ever present. This can be observed from the trembling fingers after two cups of coffee or the numerous typing, writing and speaking errors which are made during the day by an average human being. The laser pointer which is used to point at items on the screen during a presentation at a conference relentlessly amplifies such motor noise to a wandering trajectory of the light spot, even in the relaxed speaker. This intrinsic system noise leads to problems as diverse as handwriting variability in writers and typing errors in data-entry operators. In this area, fortunately, we will prefer silicon hardware, to yield the same response on the same scanned word, deterministically, regardless of time.

### (m) Response in case of difficult input

For real-life practical systems, a proper response in case of difficult input is highly important. Users and journalists often like to joke about misrecognitions in reading systems as well as in automatic speech recognition. However, inappropriate system behavior is a serious problem and still constitutes a major stumbling block to widespread acceptation of document processing systems in the market. Human reading is characterized by graceful degradation [20]. If there is an error, the Levenshtein or edit distance between the response and the input word usually will be low. In artificial reading systems, again, the state of affairs is less fortunate. On the one hand, the use of lexicon-based response improvement may enhance the reading performance. On the other hand, an error in such a system will concern a full wrong word. Full-word errors are hard to accept by human users and difficult to correct if the original document is not present for inspection. Depending on the application and the preferences of the user, a *Reject* response may be most appropriate. However, it depends on the details of the user interface in interactive applications whether a system failure will be accepted by the human users.

### (n) Classification accuracy

Human reading of machine printed text is near perfect, apart from fatigue problems to be mentioned in the next paragraph. Therefore it is not useful to use recognition error rates. The differences between the regular font types are usually negligible. Reading speed decreases when font size decreases to 10pt and smaller. Legge et al. [13] have compiled a series of chapters on the *Psychophysics of reading*, addressing the legibility of fonts. The conditions in reading of handwritten material are different. Here, human reading has limitations, too (Table 2), yielding imperfect recognition of isolated words or words from a small and restricted context window.

**Table 2.** Human recognition rates of handwritten words on paper (H.L. Teulings & L. Schomaker, unpublished). Experiments: A: single-word recognition; B-D: three-word sequences, middle-word recognition.

| Exp. | Context | Style | Writers | Target Words | Readers | Words | Recog-nized |
|---|---|---|---|---|---|---|---|
| A | frequently-used Dutch words | handprint | 1 | 30 | 12 | N=360 | 98% |
| | frequently-used Dutch words | neat cursive | 1 | 30 | 12 | N=360 | 88% |
| B | sentence fragments, same writer | cursive | 3 | 15 | 12 | N=180 | 85% |
| C | sentence fragments, same writer | cursive | 4 | 13 | 20 | N=260 | 85% |
| | unrelated words, same writer | cursive | 4 | 13 | 20 | N=260 | 77% |
| D | unrelated words, same writer | fast cursive | 12 | 12 | 15 | N=180 | 72% |
| | unrelated words, different writers | fast cursive | 12 | 12 | 15 | N=180 | 54% |

As can be observed in Table 2, only in handprint does human recognition reach a high recognition rate of 98%, whereas the recognition of cursive words varies from 54-88%. The combination of sloppy writing and absence of linguistic and shape context leads to a poor 54%. These and other [2] findings suggest that it is unrealistic to assume that the asymptote for cursive recognition is 100%. This poses an interesting problem concerning the reliability of the ground truth provided for handwritten words in free style. Usually, multiple truthers will be called in to increase the reliability of text labels to the raw handwritten-image data. As regards human reading performances on free-style text in natural scenes, there are currently no reference data. Finally, as regards measuring classification performance, these human-reading results should make us skeptical when considering overly optimistic results published in academic research. Fortunately, system benchmarking has improved from an earlier stage in research where a PhD student worked for several years on a single training/test set combination. In order to report reliable classification performance values, it is common to use k-fold cross-validation of systems [12], today. It would even be better if thesis advisors would keep back unseen data, for an objective measurement at project conclusion.

### (o) Energy and mental concentration

The reading process in human and machine require energy both for (a) document handling, scanning and (b) computing. Both document transport and the physical movement of the sensors require energy and imply wear and tear. Where human and machine differ is in the recognition performance over time, which is constant in the machine. In humans, recognition performance degrades over time due to fatigue. For this reason, data typists and check-reading operators are only allowed to work for limited time spans (20-30 min.) if the consequences of an error are grave, as is the case in financial applications.

### (p) Processing speed

Human reading speed from screen varies from 160-200 words per minute (wpm) on average [1]. Fast readers may attain speeds above 200 wpm. Such a speed corresponds to 160 bits per second for a language with an average word length of eight characters. It should be noted that the reading speed from paper is much faster, i.e., 250 to 300 wpm. The cause of this phenomenon is unknown. The contrast and resolution of printed documents on paper still outperforms current electronic display technologies (CRT or LCD). Typing (keying) speed in transcription is 27-40 wpm, on average. The speed of optical character recognition (OCR) can exceed 1500 words per minute. However, such a rating does not take into account the manual handling, i.e., the scanning process itself, manual annotations and manual image adjustment and finally, the manual correction of recognition errors which may be required. For difficult text images, human typing may often be attractive, especially if the volume in terms of number of pages is small.

### (q) Volume processing

Indeed, the strongest point for digital document processing and OCR is the ability to process massive amounts of data. Today, several companies refrain from promising perfect text recognition. Since text retrieval is even more important than document editing in many office applications, a product which aims at an acceptable performance on keyword-based retrieval will be more attractive than a product promising veridical transcription which cannot be achieved in practice. If proper text-matching algorithms are used, such a keyword-based search and retrieval system may provide a valuable functionality in allowing access to an otherwise unwieldy archive of digital-text images. As stated earlier, the applications of check reading and address reading may provide quite good recognition rates, in addition to volume processing. In these latter applications, human routine labour can be saved.

### Summary on human versus machine reading

As can be concluded from Table 1, both human and machine reading have powerful characteristics and it cannot be denied that the machine has attractive performances in terms of speed and volume, especially in the case of constrained-shape and constrained-content text images. Still, the robustness of the human reading system poses a challenge to system developers. Recent results in human-reading research and brain imaging provide an increasingly detailed picture of the architecture and processes in the human reading system. These findings may provide inspiration for system designers.

## Book overview

This book will describe in detail many of the aforementioned topics in digital document processing, and more, by experts from within the difference subfields of our research area.

### Chapter 2, *Document Structure and Layout Analysis* by Anoop M. Namboodiri and Anil K. Jain

This chapter gives an overview of layout-analysis methods, mostly concentrating on scanned, off-line, image analysis. Furthermore it describes the case of layout analysis of on-line handwritten text obtained sampling pen-tip coordinates in time. Both traditional methods such as the analysis of ink-density profiles as well as more advanced and recent approaches based on connected-components are handled. Bottom-up and top-down approaches in document-structure analysis are described, also introducing important concepts such as physical versus logical document structure. In order to evaluate hard-coded or trained layout-analysis systems, it is important to use reliable and valid methods for performance evaluation. It is noted that we are still far away from having a truly generic document-layout analysis. Also here, the development of the universal reading system remains an exciting scientific challenge.

**Chapter 3 *OCR Technologies for Machine Printed and Hand Printed Japanese Text*, by Fumitaka Kimura**

This chapter describes the complete processing pipeline for optical character recognition in Japanese text, assume machine-printed or hand-printed characters. Japanese texts pose interesting problems due to the fact that different scripts (Kanji, Katakana, Hiragana, English and Arabic numerals) are contained in one document. Additionally, the orientation of text lines may be vertical or horizontal. The Asian characters are 'monospaced', while the English text contains character boxes of another size, potentially also in a font with a proportional horizontal spacing. The preprocessing stages of deskewing, slant normalization and non-linear size normalization are described. For the feature-extraction phase, directional features are presented. Finally, the paper compares a number of classification methods and addresses the problem of feature scaling and dimensionality reduction.

**Chapter 4 *Multi-Font Printed Tibetan OCR*, by Xiaoqing Ding and Hua Wang**

The authors introduce a complete system for the recognition of multi-font Tibetan script. By deskewing pages, detecting the base line, which in this script is about at the top of the characters and by a subsequent normalization of the upper and lower parts of the character, statistical classification can be realized on features derived from the segmented character boxes. Feature selection is guided by testing the entropy-based mutual information measure and by using linear-discriminant analysis in order to obtain a compressed feature set. The actual classification method is a modified quadratic discriminant function.

### Chapter 5 ***On OCR of a Printed Indian Script*, by Bidyut Baran Chaudhuri**

The diversity of scripts on the global scale is large. Multi-script, international digital encoding of text is a fairly recent development, after a long stage in information-technology which was dominated by the use of the ASCII code which is only really suitable for a limited number of Western scripts. Not surprisingly, OCR of a multitude of scripts is just starting to develop since the acceptation of Unicode has become widespread. There are a billion people in India, writing in a dozen of scripts. In this chapter, an OCR method for the Bangla script is described. The chapter explains the structure of this alphabetic-syllabic script. The similarities between Bangla and the Devanagari script allow for a use of similar OCR approaches. Based on character-shape knowledge, a distinction can be made between upper, middle and lower zones. The presence of elongated horizontal structures aids in the preprocessing stage. A method for solving touching characters in Bangla is introduced. Words can be represented as a graph where the nodes constitute the basic recognized parts and the edges represent the geometric organization of structural elements. The large number of shape classes for the latter makes it difficult to exploit statistical methods (ANN,HMM,SVM) beyond nearest-neighbour matching. More research in this area is deemed necessary.

### Chapter 6 ***Off-line Roman Cursive Handwriting Recognition*, by Horst Bunke and Tamás Varga**

This paper describes digital document processing with Roman cursive handwriting recognition as its main purpose. It is noted that, while isolated character recognition is a mature field, the recognition of words and word sequences are still a topic of research. After preprocessing and normalization stages, such as the removal of the predominant slant in the handwritten patterns (which is jointly determined by pen grip and preferred movement direction [14], LS), a number of classification (reading) strategies are possible: Isolated character recognition; Cursive word recognition; (Over)-segmentation-based approaches bridging the gap between grapheme sequences and words; Hidden-Markov Model based recognition; and, finally cursive word sequence recognition, possibly using statistical language models. As emerging topics, the chapter identifies (1) an ongoing trend towards standardized databases and performance evaluation, (2) the possibilities of solving the data starvation problem of machine learning approaches by using original data with synthetic mixtures based on proper distortion models, (3) the usefulness of using multiple classifiers in handwriting recognition.

### Chapter 7 *A Bayesian Network Approach for Online Handwriting Recognition*, by Sung-Jung Cho and Jin Hyung Kim

This chapter presents a novel method in modeling on-line handwriting. It is noted that traditional hidden-Markov modeling misses long-range dependencies between strokes in handwriting. Durational modeling is far from ideal in such models. By using an explicit modeling of inter-stroke relationship using Bayesian nets, assuming conditional Gaussian distributions, the authors are able to show very high recognition rates on three databases as compared to two traditional approaches. The model also allows for a generation of clear character prototypes which are considerably more natural than the HMM-generated patterns.

### Chapter 8 *New Advances and New Challenges in On-line Handwriting Recognition & Electronic Ink Management*, by Eric Anquetil and Guy Lorette

On-line handwriting recognition is concerned with the 'live' transform of pen-tip movements of handwriting into character codes, as is used on pen-based digital appliances and pen-based notebook computers. This application consequently poses a number of constraints on a handwriting-recognition system which are noted in this chapter. Whereas on-line handwriting recognition usually goes along with higher recognition rates than is the case in off-line handwriting recognition and techniques may even address writer dependence, the on-line field lagged with respect to the aspects of document-layout analysis which is highly advanced in off-line systems. However, the increased availability of hardware and the emergence of standards for 'electronic ink' have provided an impetus for the development of methods for the interpretation of on-line generated document structure. The chapter addresses lazy versus eager interpretation of document structure. Additionally problems of small devices and their consequences for the human-machine interface are addressed.

### Chapter 9 ***Robustness Design of Industrial Strength Recognition Systems***, by Hiromichi Fujisawa

Fujisawa provides an overview on performance-influencing factors (PIF) in postal-address analysis systems. The general theme of this chapter is "know thy enemy". By an explicit modeling of all hostile factors which influence recognition performer, robust systems can be designed for operation in the real world. A detailed enumeration of noise factors, defects, variations, imperfections and distortions is given. An example of a particular problem class would be the 'touching characters'. An argument is made for a number of robustness enhancing techniques. Although the chapter is not explicitly oriented towards agent-based computing, the design philosophy is based on a clear encapsulation of document related expertise in specialized modules which are only activated when needed. An important and useful warning is issued concerning the reliance on posterior probabilities for generating rejects on irrelevant or unknown input. Indeed, a good recognizer should ideally first recognize if the input pattern belongs to the classes on which it has been trained. Many systems are flawed in this respect. In real systems in the real world such 'meta intelligence' is needed. Theoretical work in this area [27] may be used for defining a single-class classifier which encompasses all targeted classes on which it is trained.

**Chapter 11 *Arabic Cheque Processing System: issues and future trends*, by Mohamed Cheriet M., Al-Ohali Y., Ayat N.E., and Suen C. Y.**

There is a clear trend towards handling more and more international scripts. This chapter deals with the interesting case of reading text on Arabic checks. The availability of an Arabic check-processing system will have an impact on this application in at least twenty countries. Interestingly, numerals may be written as Indian numerals, not Arabic numerals in some regions. Using a data set of realistic samples from a bank, the study describes processing steps, feature choice and classification technology. The goal is to use discrete-HMM models for sub-words. The ISODATA algorithm is used for searching shape clusters, and vector quantization is used to adopt the discrete model. The chapter illustrates how an analysis of the errors provides insight in the strengths and weaknesses of the approach. For the special case of the Indian numerals, the suitability of the support-vector machine is successfully demonstrated. Finally, the chapter discusses the requirements for future scalable Arabic check processing systems.

**Chapter 12 *Bank check data mining: integrated check recognition technologies*, by Nikolai Gorski**

A comprehensive view on an actual check-reading system is provided in this chapter. Check reading started with attempts to read legal and courtesy amounts on checks. However, the continued use of handwritten and signed checks and the need for automation necessitated the development of recognition of other fields, as well. The visual content of a check contains of machine printed and handwritten items: payee name, payer's address, date, check number, legal amount, bank name, address and logo, a memo line, a courtesy amount and the field containing the payer's signature. Each of these fields requires a specialized approach. Especially in case of handwritten fields the amount of information is not sufficient to allow for perfect classification. Domain knowledge is needed to segment and recognize the visual content. As a more elaborate example, the handling of name aliases is described. Payee or payer's name often occur in a number of variations ("aliases"). By training a neural network on real and generated name aliases, the resulting alias detector can be used to improve system performance. The chapters also gives an insight in the process of internationalization and the concomitant improvement of the described check-reader as a product.

### Chapter 13 ***OCR of Printed Mathematical Expressions***, by Utpal Garain and Bidyut Baran Chaudhuri

With the success of optical character recognition on general texts, it also becomes clear where there are limitations. In scientific documents, the presence of mathematical expressions will pose a problem to current OCR systems. Such expressions contain a multitude of additional characters and symbols. Additionally, their organization in the plane is vastly different from the regular line structure of normal text. The chapter gives a detailed overview on the detection of page zones containing mathematical expressions, the recognition of symbols, the segmentation of touching symbols and the interpretation of expression structure. Performance evaluation poses problems in this domain, since annotated reference data is hard to obtain and the measurement of correctness in terms of element classification and structure detection is not trivial.

### Chapter 14 ***The State of the Art of Document Image Degradation Modeling***, by Henry S. Baird

In the chapter by Baird, the disconcerting problem of the fragility of many systems for digital document processing is brought to our attention. Small defects in image quality of have detrimental effects on text-recognition performance. Fortunately, however, new methods have evolved that allow for a controlled generation of degraded text images in order to improve recognizer robustness during training or in order to test the sensitivity to image degradation in a given OCR system. The chapter gives an overview of the state of this art in literature. Two basic approaches are introduced: (1) physics-based modeling of defects and (2) statistical modeling of text-image defects. Current degradation models are based on bi-level document images. Still, the availability of degradation models allows for an effective enlargement of a given training set. Indeed, two important concluding remarks of this chapter are, (1) "real data is corrupting: it is so expensive that we reuse it repeatedly [...]", and (2) training on a mixture of real and synthetic data may be, today, the safest [method]".

### Chapter 15 *Advances in Graphics Recognition*, by Josep Lladós

This chapter provides an overview of the subdomains of Digital Document Processing which focus mostly on non-textual two-dimensional graphical patterns. Here, the goal is to regularize and abstract images of schematic drawings by means of pixel-to-vector transforms and shape classification. Application examples concern: electrical and logic diagrams, geographic maps, engineering drawings, architectural drawings, musical scores, logo recognition, table and chart recognition. A distinction can be made between off-line analysis of scanned images of graphics and on-line processing of drawing and sketching movement in pen-based systems. Whereas the problem of vectorization of bitmapped images has been solved to a large extent, it is this success itself which has revealed more clearly the residual and fundamental problem of extracting a meaningful structural description from two-dimensional graphic patterns. The field of graphics recognition is particularly evolved in the area of integrating low-level processing modules within a knowledge-based framework for the interpretation of shape. The availability of benchmarks for performance evaluation in graphics recognition has provided a major impetus to the maturity of the field.

### Chapter 16 *An Introduction to Super-Resolution Text*, by Céline Mancas-Thillou and Majid Mirmehdi

The introduction of camera-based OCR has reintroduced an interest in image-processing techniques. While digital document processing has advanced from bitonal 200 dpi scans to 300 dpi and higher-resolution grey-scale or color scanning from paper, camera recordings contain jagged low-resolution text images. The regular structure of characters allows for advanced reconstruction schemes known as super-resolution processing. In still images, local bilinear interpolation can be used. In video recordings, super-resolution can also be obtained using frame-to-frame information. The additional information can be used in order to determine a warping transform in conjunction with a deblurring and denoising step. The amount of parameters makes this problem ill-posed, still. A number of methods are described. The authors introduce an advanced variant, using filtering, Taylor series and bilinear interpolation in a robust framework. Super-resolution techniques are shown to improve OCR performance from 72 up to 91 %.

### Chapter 17 *Metadata Extraction from Bibliographic Documents for Digital Library*, by Abdel Belaïd and D. Besagni

The authors of this chapter describe knowledge-based methods for the analysis of text representing bibliographic information. By using linguistic and structural constraints, it becomes possible to automatically generate metadata for bibliometric and retrieval purposes in Digital Library applications. The modeling methods may use, for instance, linguistic background information as made explicit through part-of speech (POS) tagging. Examples are given of experimental results.

### Chapter 18 *Document information retrieval*, by S. Klink, K. Kise, A. Dengel, M. Junker, and S. Agne

This chapter provides an introduction to information retrieval. The field of information retrieval is concerned with the search for documents in large electronic collections (i.e., ASCII or Unicode text) on the basis of keywords. The basic vector-space method (VSM) is explained, as introduced by Salton. Subsequently, the chapter continues to address more advanced topics. Relevance feedback concerns the guided search and improved retrieval on the basis of user feedback on the relevance of found hit lists. A specialty of the authors is *Passage Retrieval*, where the goal is to find relevant passages rather than whole documents. Collaborative Information Retrieval (CIR) is introduced: A later user with similar information needs to earlier searchers can profit from automatically acquired knowledge in several ways. A CIR system keeps tracks of queries and the returned documents. This information can be used to adapt a given query to an expanded or more precise format. Methods for textual ASCII/Unicode retrieval of electronic documents can be generalized to retrieval of documents which are archived as sequences of page images by means of an OCR front-end indexer.

**Chapter 19 *Biometric and Forensic aspects of Digital Document Processing*, by Sargur N. Srihari, Chen Huang, Harish Srinivasan, and Vivek Shah**

The advances in digital document processing not only have an impact on OCR, per se. Using current methods for preprocessing and feature extraction, new applications are possible. In forensic applications, one would like to find the identity of the unknown writer of a given sample of handwriting. Another possible application concerns the verification of a claimed identity for a given sample of handwriting or handwritten signature. In traditional forensic handwriting analysis, human experts perform manual measurements and judgments. Using current techniques, a more objective semi-automatic approach becomes feasible. The chapter introduces a number of interactive processing methods, handwriting features and matching schemes for writer and signature verification, using Bayesian modeling. Explicit statistical modeling allows for the use of probabilities and likelihood ratios, which is essential for the acceptance of this technology by forensic practitioners.

**Chapter 20 *Web Document Analysis*, by Apostolos Antonacopoulos and Jianying Hu**

The chapter by Antonacopoulos and Hu provides an eye opener for those who think that life is easy if a document is already in an encoded electronic format beyond the flat image. Even if a document is decoded at the character level its content must be analysed for repurposing or retrieval. An example is the induction of the structure of tables. Additionally, the image material in web pages which contains text poses new problems which are different both from the problems encountered in scanned document images and problems in camera-based text recognition. Figures 2 and 3 in this chapter provide example of the large difficulties of analysing text of machine-based origin in images on web pages.

### Chapter 21 *Semantic Structure Analysis of Web Documents*, by Rupesh R. Mehta, Harish Karnick and Pabitra Mitra

Even in cases where a document exists in digitally encoded and enriched form, a reading system must be able to structure the content. Web documents are usually encoded in HTML, representing text with rendering attributes in a particular layout, as well as referring to non-textual page elements such as images. The human reader is well able to direct attention to the textual core content of a web page. Information Retrieval systems which aim at finding relevant documents on the basis of key words or example texts will benefit from a thorough analysis of layout for semantically homogeneous text passages. In this chapter, an analysis of visual elements on a page is proposed in order to determine topical text elements. Examples are visual elements of background color, lines and images used as separators, font changes, and so on. Bayesian modeling can be used to estimate the predictive value of visual page attributes in the determination of semantic categories for a web page.

## Conclusion

This book will focus on technological problems of digital-document processing from different angles and at different levels of the document processing pipeline. The motivation for the development of new algorithms in this area is fed by the strong conviction that DDP will not only solve a number of existing problems in the ways we manipulate text, it will also allow us to develop new methods of working with documents, extracting detailed and often hidden pieces of information. New algorithms will be available to determine writer identity, historical dating of a text and even estimation of the author's age in handwritten documents. New data-mining methods will be able to uncover hidden relations between documents in huge collections, not only in terms of textual content but also in terms of layout and typographic styles. Improved modeling of character shapes will lead to a more general applicability of optical character recognition on artistic or three-dimensional shapes. Even if it is unlikely that we will be able to construct a *Universal Reading Machine* anytime soon within the next decade, it is certain that research in digital document processing will help to pave the road towards understanding perceptual intelligence. At the same time the book will show that an important part of the work devoted to this exciting topic has an actual impact on the ways in which documents are processed in the world of real practical applications.

# References


1. Bailey, R.W. (1996). Human Performance Engineering: Designing High Quality Professional User Interfaces for Computer Products, Applications and Systems. Upper Saddle River (NJ): Prentice Hall.
2. Barrière, C. and Plamondon, R. (1998). Human identification of letters in mixed-script handwriting: An upper bound on recognition rates, *IEEE Trans. on System, Man and Cybernetics - (B), 28(1)*, pp. 78-82.
3. Brodmann, K. (1909). Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Leipzig: Barth.
4. Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T. B., & Mesulam, M.M. (2002). Functional anatomy of intra- and cross-modal lexical tasks. *Neuroimage, 16(1)*, pp. 7-22.
5. Coates, A.L., Baird, H.S., Fateman, R.J. (2001). *Pessimal Print: A Reverse Turing Test*, Proc. of the 6th International Conference on Document Analysis and Recognition, Seattle, WA, USA, September 10-13, pp. 1154-1158, Los Alamitos: IEEE Computer Society, ISBN 0-7695-1263-1.
6. Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain, 125(5)*, pp. 1054-1069.
7. Dehaene, S., Cohen, L., Sigman, M. & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in cognitive sciences, 9(7)*, pp. 335-341.
8. Ezaki, N., Bulacu, M. & Schomaker, L. (2004). *Text detection from natural scene images: Towards a system for visually impaired persons.* Proc. of ICPR 2004, Cambridge, UK, IEEE Computer Society, pp. 683-686. ISBN: 0-7695-2128-2.
9. Franke, K., Köppen, M. (1999). *Towards an universal approach to background removal in images of bank checks*, In: S.-W. Lee (Ed.) Advances in Handwriting Recognition, Singapore: World Scientific, pp. 91-100.
10. Georges Jean (1997). *Writing: The story of alphabets and scripts*, London: Thames and Hudson Ltd.
11. Gigerenzer, G., Todd, P.M. et al. (2000). *Simple Heuristics That Make Us Smart.* (Evolution and Cognition Series), Oxford Univ. Press.
12. Goutte, C. (1997), Note on free lunches and cross-validation, *Neural Computation, 9*, pp. 1211-1215.
13. Mansfield J.S., Legge G.E. & Bane M.C. (1996). Psychophysics of reading. XV. Font effects in normal and low vision. *Investigative Ophthalmology & Visual Science, 37*, 1492-1501.
14. Maarse, F.J., & Thomassen, A.J.W.M. (1983). Produced and perceived writing slant: Difference between up and down strokes. *Acta Psychologica,* 54, pp. 131-147.
15. McCusker, L.X., Gough, P.B., Bias, R.G. (1981). Word recognition inside out and outside in. *Journal of Experimental Psychology: Human Perception and Performance, 7(3)*, pp. 538-551.
16. Nagy, G., Nartker, T.A. & Rice, S.V. (2000). Optical Character Recognition: An Illustrated Guide to the Frontier, invited paper in Proceedings of SPIE: Document Recognition and Retrieval VII, Volume 3967, San Jose, California, 2000.
17. Pammer K, Hansen PC, Kringelbach ML, Holliday IE, Barnes G, Hillebrand A, Singh KD & Cornelissen PL. (2004). Visual word recognition: the first half second *Neuroimage 22(4)*, pp. 1819-1825.

18. Petkov, N. and Kruizinga, P. (1999). Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells, *Biological Cybernetics, 76(2)*, 1997, 83-96.
19. Plamondon, R., Lopresti, D.P., Schomaker, L.R.B. & Srihari, R. (1999). *On-line handwriting recognition.* In: J.G. Webster (Ed.). Wiley Encyclopedia of Electrical & Electronics Engineering, pp. 123-146, New York: Wiley.
20. Plaut, D.C. & Shallice, T. (1994). *Word Reading in Damaged Connectionist Networks: Computational and Neuropsychological Implications.* In R. Mammone (Ed.) Artificial neural networks for speech and vision (pp. 294 323). London: Chapman & Hall.
21. Rawlinson, G. (1999). “Reibadailty” [Letter to the Editor], *New Scientist, 162(2188)*, p. 55.
22. Schomaker, L.R.B. (2004). Anticipation in cybernetic systems: A case against antirepresentationalism. IEEE SMC Delft, The Netherlands, 2004.
23. Schomaker, L. & Segers, E. (1999). Finding features used in the human reading of cursive handwriting. *International Journal on Document Analysis and Recognition, 2*, pp. 13-18.
24. Simon, H. (1957). *Models of Man*, New York: Wiley.
25. Simon, J.C. & Baret, O. (1990). Handwriting Recognition as an application of Regularities and Singularities in Line Pictures. *Proceedings of the International Workshop on Frontiers in Handwriting Recognition (IWFHR).* (pp. 23-37). Montreal: CENPARMI Concordia.
26. Talairach, J. & Tournoux, P. (1988). Co-planar Stereotaxic Atlas of the Human Brain. New York: Thieme Medical.
27. Tax, D.M.J. & Duin, R.P.W. (2002). Uniform object generation for optimizing one-class classifiers, *Journal of Machine Learning Research, 2*, pp. 155-173.
28. Tistarelli, M. & Sandini, G. (1993). On the Advantages of Polar and Log-Polar Mapping for Direct Estimation of Time-To-Impact from Optical Flow *IEEE Transactions on Pattern Analysis and Machine Intelligence archive, 15(4)*, pp. 401-410.
29. Truccolo, W., Eden, U.T, Fellows, M.R., Donoghue, J.P. & Brown, E.N. (2005). A point-process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects, *Journal of Neurophysiology, 93*, pp. 1074-1089.
30. Turkeltaub, P.E., Weisberg, J., Flowers, D.L. Basu, D & Eden, G.F. (2004). *The Neurobiological Basis of Reading: A Special Case of Skill Acquisition*, In: Developmental Language Disorders: from Phenotypes to Etiologies. Rice and Catts (Eds.), Lawrence Erlbaum.
31. Veeramachaneni, A. & Nagy, G. (2005). Style Context with Second-Order Statistics, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1)*, pp. 14-22.
32. Whitney, C. (2001). *An explanation of the length effect for rotated words.* In: E. Altmann, A. Cleermans, C. Schunn, & W. Gray (Eds.), Proceedings of the Fourth International Conference on Cognitive Modeling. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 217-221.
33. Wilson, F.A.W., Scalaidhe S.P.O. & Goldmanrakic P.S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science 260(5116)*, pp. 1955-1958.