Information Extraction from Document Images using Attention Based Layout Segmentation

Gerd Maderlechner, Angela Schreyer[1] and Peter Suda

Siemens AG, Corporate Technology
Otto-Hahn-Ring 6, D-81730 Munich, Germany
Tel. +49-89-63653389, Fax.+49-89-63640153
email: gerd.maderlechner@mchp.siemens.de
[1)] University of Technology Munich,
Institute for Human-Machine-Communication,
Arcisstr. 21, D-80290 Munich, Germany

Introduction

The attention of a human reader and the reading speed strongly depends on the layout of a document The term layout is used for the geometrical arrangement of document components (i.e. text, graphics and figures) on the page as well as for the typographic features of the text (i.e. font type, style, size, alignment and line spacing). Although the human visual and cognitive perception of documents is not exactly known, there are interesting results from psychophysical experiments [1, 2, 3] which can be used in document layout segmentation.
Under the assumption, that the authors and publishers make use of the layout to emphasize the relevant content, a fast extraction of this information should be possible by reverse determination of the emphasis from the layout.
This paper describes an approach to attention based layout segmentation using general principles of the human visual perception to achieve this goal. The segmentation is generic and not limited to specific document classes and models. The resulting regions of interest may be used for further interpretation. The overall speed of browsing and searching large volumes of scanned documents can be increased considerably.

Stream model

According to the communication theoretical model, a sender (author) forms a one-dimensional stream of information (message string M with emphasis and structural tags), and the writing or printing process generates a two-dimensional document. The emphasis and structure is represented by the layout features. The communication channel corresponds to all processes between the ideal document image and the scanned image, like printer and scanner distortions, paper and ink quality and manipulations. The receiver (reader) converts the 2-dim image back to a 1-dim message M' including the emphasis and structural tags (hopefully M = M', which requires e.g. that the reader understands the language, meaning, layout etc. from common sense). In document image analysis the message and emphasis has to be reconstructed from the 2-dimensional image. In this work the main effort is to recognize the emphasized parts from the layout features.
Recently Doerman et. al. [4] introduced the term *function* of a document to describe the efficiency with which a document transfers its information to the reader. A similar messaging model between author and reader is proposed independent on the class of document.

Attention measures

An attention value (*attractiveness*) is given to every word or graphics part contained in a document. Each of these objects displays layout features with a significance value which is normalized to [0,1]. The measure of attention is calculated for each object by weighted summation of the significance values.
The weights were derived from a survey. 46 persons were asked about their perception of document layout, in order to establish a set of relevant layout features and their ranking with respect to attractiveness. Table 1 shows the main questions and the results of the survey, where weight ≈ 1/ranking of respective layout feature.

The determination of the attention measure is performed in three steps:
1. Segmentation into text and graphics blocks according to texton theory and texture model (see below)
2. Determination of the following layout features:
- for text blocks and words: surrounding white space, font size, indentation, bullets, and font styles bold italics and all-capitals,
- for graphic blocks: surrounding white space, black intensity, and area.

(In the survey the persons suggested further layout features like color, rulers, frames, and columns which could be considered in an extended version of the system.)

3. Calculation of attractiveness by normalization and summation for words and graphic blocks.

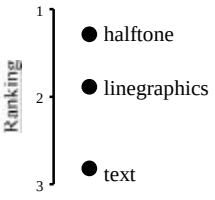
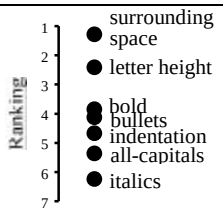
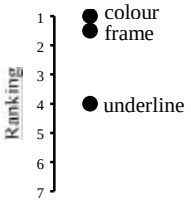
Question	Give ranking of blocks	Give ranking of layout features	Suggest missing layout features and respective ranking
Mean deviation of answer	0.39	0.96	-
Results			

Table 1: Results of the survey on layout features (1 = ranking of the highest attention)

Segmentation and layout features

We start with binary images scanned at 300dpi in run-length representation, and perform the texture analysis and segmentation steps on results of connected component analysis, in contrast to [6]. The segmentation is based on the assumption that text consists of texture regions with *textures* at different resolution levels. In [3] text is also treated as texture in four characteristic spatial scales corresponding to the text block, text lines, words and letter levels. The letter level is dominated by small vertical strokes. By psychophysical experiments using synthetic textures made from regularly arranged letters or other simple strokes (see Figure 1), Julesz could discriminate a global and fast pre-attentive visual perception of textures from a slower focal attention which uses local search [2]. In his texton theory he explained this phenomenon by the visual stimulus caused by texture elements (*textons*), which are defined as elongated blobs (strokes) with orientation, width, length and density. We adopted this concept and developed a hierarchical bottom up segmentation starting from strokes (as parts of connected components), letters, words, lines to blocks. Details are given in [5] which also describes a texture-based font style classifier.

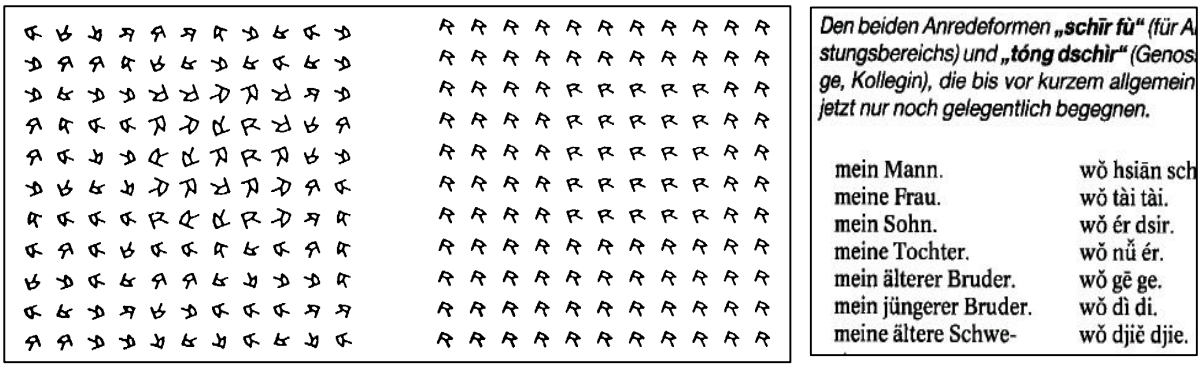


Figure 1: Examples of texture pattern used by Julesz to study the pre-attentive vision, and to develop the texton theory [2]. For comparison an example from a real document showing similar pre-attentive features for the text blocks and the bold words.

In addition to the hierarchical extension of the texton theory we applied the Gestalt principles proximity and continuation to the grouping process of the basic document objects.

As a side effect of the implementation of psychophysical principles we get an understanding of proven layout rules (e.g. the alignment of text lines or blocks).

An alternative approach to attention based segmentation is described in [6], which mainly relies on physiological principles of human vision, like eye movement fixations, analysis of gazing points and space-variant recep-

tive fields (log-polar mapping). It is applicable to arbitrary images using features of similarity, salient angles, etc., whereas our approach is more specific to document images and document layout structures.

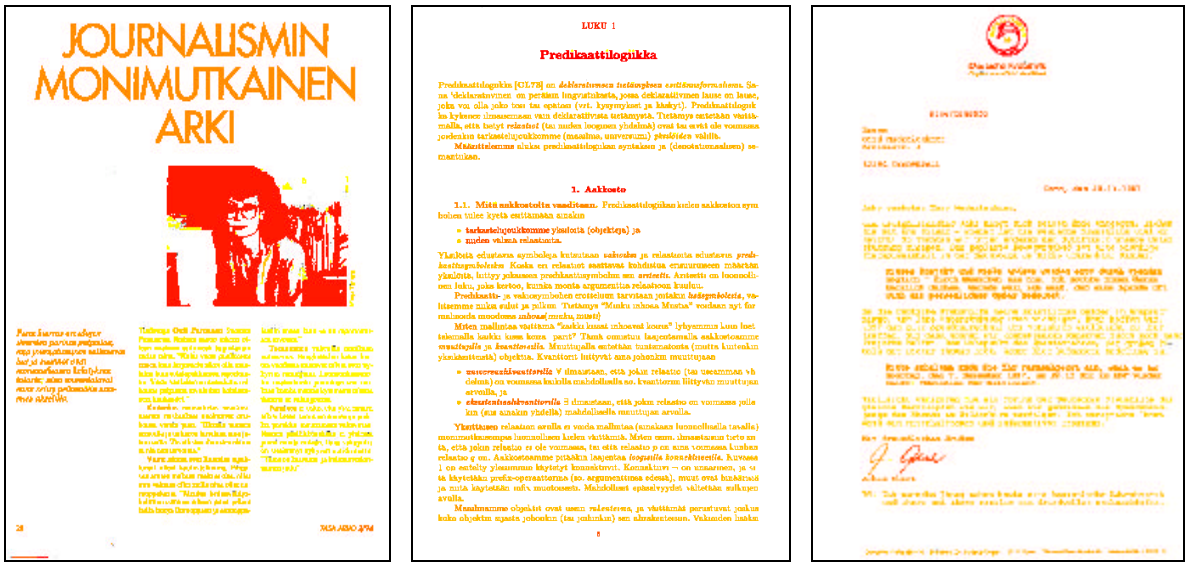


Figure 2: Results of the attention based segmentation. The measure of attractiveness is indicated with increasing gray scale values (from light to dark). The results are plausible for journals, reports and business letters.

Results and discussion

Figure 2 shows three examples for the classification of attractiveness on scanned document. The Finnish journals were chosen because the test persons (not Finnish speaking) should not be influenced by the meaning of the text, but the visual appearance is like other Latin languages. Note, that the values of attractiveness are normalized to the highest value on the respective document.

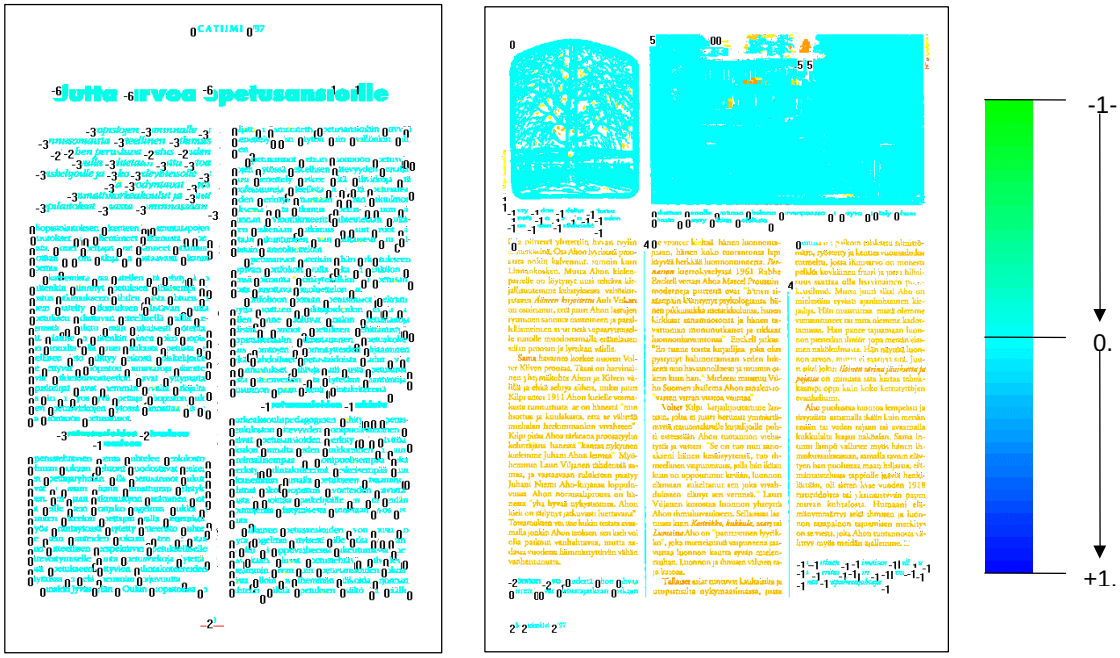
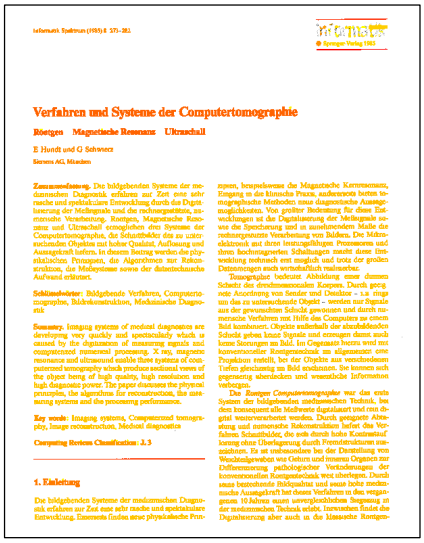


Figure3: Numerical comparison between the computed attractiveness A and the human ranking H measured as the difference A-H. Darker values correspond to higher A than human ranking. Figures and graphics are slightly more emphasized and font styles are less emphasized by the algorithm.

A second survey with interactive display of these results has indicated very reasonable behavior of the measures of attention (Fig. 3).
The textual segments (blocks and words) with high attractiveness have been processed by OCR and arranged in a list of keywords and sentences, sorted in decreasing attractiveness values (Fig. 4). The text between two



Verfahren und Systeme der Computertomographie

Röntgen Magnetische Resonanz Ultraschall

E Hundt und G Schwierz

Siemens AG, München

SCHLÜSSELSÄTZE:
****
Verfahren und Systeme der Computertomographie
Röntgen - Magnetische Resonanz - Ultraschall
****
1. Einleitung
****
Informaik-Spektrum (1985)8: 273282
****
E. Hundt und G. Schwierz
Siemens AG, München
****
©
****
Computing Reviews Classification: J. 3
****
Key words: Imaging Systems, Computerized tomogra phy, Image reconstruction, Medical diagnostics
****
****
Zusammenfassung.
****
Summary.
****
Schlüsseiwörter: Bildgebende Verfahren, Computerto mographie, Bildrekonstruktion, Medizinische Diagno stik
****
Die Röntgen-Computerfomographie war das erste System der bildgebenden medizinischen Technik, bei dem konsequent alle Meßwerte digitalisiert und rein di gital weiterverarbeitet werden.
****
Tomographie bedeutet Abbildung einer dünnen Schicht des dreidimensionalen Körpers.

Figure 4: Example of an extracted summary in ASCII text form after character recognition in the text blocks with the highest measures of attractiveness $A > A_{thr}$ ($A_{thr} = 0.5$ with $0 < A < 1.0$ and $A_{median} = 0.36$).

delimiters (***) corresponds to a text block or a keyword followed by the sentence in which it is included.
The performance of the whole extraction process, including connected component analysis, segmentation, layout feature extraction, font classification, and calculation of attractiveness is less than 1 sec per page on a 200MHz Pentium PC. The accuracy of the attractiveness measures is within the variance of the human evaluation.

Conclusion and future work

The layout of a document was successfully used to extract relevant text and non-text information from a document image according to a communication theoretical stream model. Texture analysis using a hierarchy of textons was applied to layout segmentation. Measures of attractiveness are based on psychophysical experiments and were adapted and proven in several surveys.
The most attractive document parts can be extracted as rectangular regions of interest according to their ranking, thus enabling a more efficient interactive browsing through large document archives. Further automatic processing using OCR, document type information and application specific knowledge is in progress. First results have been shown. The attention based layout segmentation and information extraction can be used in browsing, searching and abstracting tools.

References

[1] J. R. Anderson: *Cognitive Psychology and its Implications*, W. H. Freeman and Company, San Francisco, 1947

[2] B. Julesz, J. R. Bergen: *Textons: The Fundamental Elements in Pre-attentive Vision and Perception of Textures*, The Bell System Technical Journal, Vol. 62, No. 6, 1983, pp. 1619 - 1645

[3] R. Watt: *The Visual Analysis of Pages of Text* in R. Sassoon: Computers and Typography, Intellect Books, Oxford, 1993, pp. 179 – 201

[4] D. Doermann, E. Revlin , A. Rosenfeld: *The Function of Documents*, Image and Vision Computing, 16, 1998, pp. 799 - 814

[5] A. Schreyer, P. Suda, G. Maderlechner: *Font Style Detection Using Textons*, Proc. DAS, Nov. 1988 in Nagano, pp. 99 - 108

[6] V. Eglin, H. Emptoz: Logarithmic Spiral Grid and Gaze Control for the Development of Strategies of Visual Segmentation on a Document, Proc. ICDAR, Aug. 1997 in Ulm, Germany, 1997, pp. 689 - 692