

Name Block Location in Facsimile Images Using Spatial/Visual Cues

Laurence Likforman-Sulem

Ecole Nationale Supérieure des Télécommunications
Signal and Image Processing Laboratory & CNRS LTCI, 46, rue Barrault, 75013 Paris, France
E-mail : likforman@tsi.enst.fr

ABSTRACT

We discuss here the name block location process of a messaging system which deals with facsimile images including handwritten items. This process is the first stage of an information extraction task which aims at extracting condensed information from such documents. The approach is based on the use of spatial/visual cues for selecting regions in the image where sender and recipient information can be found. These cues are based on the human visual grouping process. The start of the process is the binary image and its electronic version produced by an OCR system. A document processing stage first extracts layout components at a pseudo word level and detects handwritten components. The electronic version enables the selection of layout components including predefined printed keywords. These layout components are then grouped under spatial constraints to initialize a logical description. The logical objects searched for are the main field headers relative to the sender and the recipient and the sender name. Experimental results on the performance of the approach are presented.

1. Introduction

Messaging Systems already allow voice, e-mail and facsimile message consulting on a same support and in a unified way. Document, speech and language processing techniques may help to convert messaging systems into intelligent assistants. For instance extracting key information such as sender names or message subjects would enable the recipient to browse quickly through messages and sort them by priority level. Besides messaging systems, there is the general need for accessing the content of document images for effective retrieval and electronic transmission [7].

We are concerned here with a messaging system called *Majordomo* which will serve as a telephone attendant [4].

Dealing with facsimile images or documents attached to e-mails, the following information can be searched : name, address, fax/phone numbers of the sender, name of the recipient, object of the message and date. This enables the recipient to browse quickly through messages which may be especially convenient for remote consultation. We aim here at extracting the location of the name of the sender which may appear in the message box as a text string if printed or as an image if handwritten.

Extracting specific items from a document image is an information extraction task which requires document understanding. Such a process involves extracting the layout structure of the document searching for layout objects (lines, blocks). Then layout objects are labeled into significant components (logical objects) which constructs the logical structure of the document.

Understanding the structure of business and financial documents such as invoices, forms, business letters, facsimile, needs such layout and logical labeling processes. Traditional methods for extracting logical objects rely on numerical location assumptions. For instance in [3] invoice fields are extracted according to predefined models and to a description of the relationship between a field and its content. In [9] [5] [2] [12], systems are trained from sample documents to derive models and corresponding object locations. A matching scheme enables the closest model to be found and to extract logical objects.

In this paper, we do not make strong assumptions about the position of fields as there is a great variance in field positions. Logical objects that are searched for are : main headers (*sender_header*, *recipient_header*), the header corresponding to the name (*name_header*) and the name of the sender (*sender_name*) which is the final result.

We exploit spatial/visual cues for the extraction of the sender name block location from the fax image and its OCR'd electronic version. These spatial properties exploit the human visual grouping process. Such spatial cues rely on the alignment of

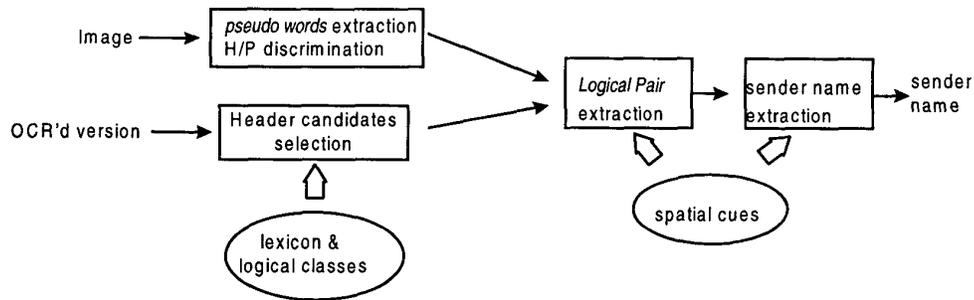


Figure 1: Architecture

fields, other cues rely on the proximity of a field with its content for the retrieval of field contents.

We consider here images of fax front covers : field headers are printed but their content may be printed or handwritten. Fax images are of low quality which involves character recognition errors. Layouts are highly variable as they are not normalized. There is a variety for field designation (synonyms) which involves the use of a lexicon. Keywords themselves are homonymous to other words : many candidates will be generated which involves a process for selecting the right ones. This selection of the right candidate for the sender field is based on spatial properties.

2. Architecture

Before logical labeling, a low level analysis is necessary to extract layout components and discriminate between handwritten and printed ones. This classification enables the extraction of an image for the sender name if handwritten and the text string if printed. Components at the word level, called Pseudo Words (PWs), are searched using the Run Length Smoothing Algorithm. PWs may be words, parts of text lines, or even lines. This level of segmentation was chosen in order not to merge printed and handwritten items possibly lying on the same line into one component. A neural network is in charge of classifying PWs into 2 classes (printed or handwritten). More details on this previous work are described in [11].

Parallel to document analysis, the electronic version provided by an OCR system is a structured text file which contains the text strings of the printed text and the position of the words (figure 2). Because of the low quality of fax images, the electronic version contains recognition errors. Pseudo words including characteristic words related to a sender and a recipient are candidates for being potential field headers. These characteristic words belong to logical classes *sender_class*, *recipient_class*. For instance *sender_class* may include keywords such as 'from', 'sender', 'de', 'expediteur' belonging to a French/English lexicon. At this early stage of the process, recognition is used. Many candidates are generated and the selection process uses only spatial/visual knowledge. The selection of the right pair (sender-recipient) of candidates is performed using general spatial cues between field headers. One cue for instance is : «Sender and recipient headers are aligned - horizontally or vertically ». The right pair of layout blocks is labeled as one logical object called a logical pair. The position of the logical pair is used to delimit the corresponding zones in the image for the sender and the recipient respectively.

Once the sender and recipient zones are delimited, the name of the sender is searched. Another spatial cue consists in labeling as field content the nearest neighbor component of the field header. And then all aligned components with this nearest neighbor will create a logical object composed of a header and its content. Sometimes the nearest neighbor has to be searched near header *name_header* instead of header *sender_header* but the spatial cues are the same (cf. section 3.2).

```

[e;12][s;12;198;0;0;40;t;1;0;1][c;1]DEC...20...2000[h;415;42;1]09:12[h;555;181;6;8;1]U[h;753;22;7]OF
[h;813;22;8]T[h;853;22;9]MPTHEMPTICS[h;1092;441;2;19;1]41[h;1590;22;3]978[h;1670;22;4]4107[h;1769;62;5]P.01[y;2012;106;40;3;H]
[s;12;198;582;10;355;c;2;14][c;2]UNIVERSITY[h;1145;22;11]OF[h;1243;26;12]TORONTO[y;2012;450;355;1;H]
[s;12;198;482;13;428;c;2;11]DEPARTMENT[h;1088;25;14]OF[h;1188;25;15]MATHEMATICS[y;2012;360;428;3;H]
[s;12;198;790;16;613;c;3;28][c;3]Tel:[h;1053;20;17](416)[h;1160;21;18]978-3323[y;2012;665;613;1;H]
[s;12;198;786;19;687;c;3;28]Fax:[h;1058;20;20](416)[h;1165;20;21]978-4107[y;2012;658;687;3;H]
[s;12;198;577;22;871;c;2;14][c;2]FAX[h;886;22;23]TRANSMISSION[h;1368;22;24]FORM[y;2012;453;871;0;H]
  
```

Figure 2 : Fragment of an electronic version (OCR'd version)

3- Sender Name Location

3.1 Extraction of recipient/sender main headers

Facsimile images share similar conventions. However there is a large variance in layouts. The location of main headers (recipient, sender) is significantly different for such documents. To retrieve the content of fields, several approaches may be proposed. In [10] a bottom up approach consists in selecting headers on image criteria, then applying OCR to select the right ones. A top down approach may also be used as in [1] searching first for proper name strings in the electronic version using linguistic criteria, then selecting the recipient name on image criteria. Our approach is also top down as we search for keywords in the electronic version, then we select headers on image criteria. But these criteria do not rely on locations but rather on alignments.

To avoid numerical criteria and implementing a large set of heuristic rules, we rather exploit the symmetry that occurs in most fax front covers between the recipient and sender information. This symmetry implies that sender and recipient fields are both introduced by main headers that define regions where to find sender and recipient information respectively. A general structure for fax front covers can be described as : field headers are aligned and field content is in the neighborhood of its header.

The strategy exploiting these cues consists in grouping pseudo words under spatial constraints to form the logical pair, i.e. the right pair of main headers. Only pseudo words which contain a keyword belonging to the lexicon and to one of the logical classes *sender_class* and *recipient_class* are candidates for this grouping process. The spatial constraints are inspired from perceptive grouping : alignment (principle of direction continuity), proximity and similarity. This scheme has been used in document analysis to group layout components into objects of higher level [8]. [6] also uses both visual and keyword detection for the logical labeling of scientific papers.

The visual constraints used are : main headers are aligned horizontally or vertically. Alignments are checked from the position of the bounding boxes of layout components.

This grouping enables logical pair candidates to be formed. A score is computed for each candidate pair to select the best hypotheses. The score is computed as follows : the more the position of the header is on the left of the layout component, the higher the score and the score lowers when the position of the header reaches the

extreme upper or the lower part of the image. High scored hypotheses are kept for further analysis.

Each hypothesis enables main headers to be extracted corresponding to the sender and the recipient respectively. These main headers form the logical pair and enable the delimitation of two regions in the image where potential sender and recipient information can be found. Two layout configurations are considered : one or two columns (figure 3)

This approach enables the position of the logical pair in the image to be found without assuming an a priori location of the main headers. In the case the logical pair is not found (Ocr errors, headers exist but are not aligned, headers do not exist) an alternative strategy consists in generating the sender region from one header only, belonging to class *sender_class*.

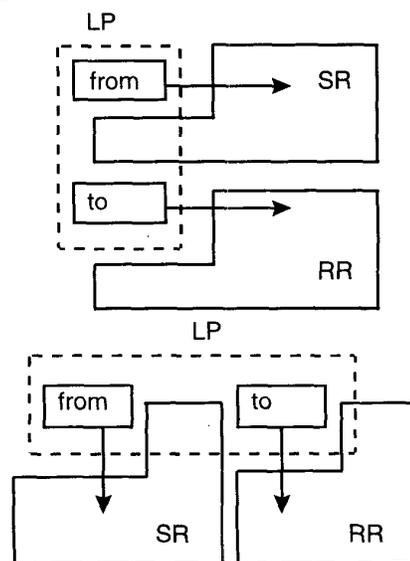


Figure 3 : Logical pair (LP) delimiting sender and recipient regions (SR, RR)

3.2 Sender name extraction

The name of the sender can be found in the sender region around object *sender_header*. The sender name, can be accessed directly through the main header or indirectly through a header of class *name_class* (figure 4). A header of class *name_class* is first searched in the sender region. If it does not exist, the main header of class *sender_class* and belonging to the logical pair, is considered. The sender name can be found in header neighborhood. This neighborhood is a region which lies to

the right and under the header itself. Location of header content has also a great variance so that proximity is the visual cue used to find layout components corresponding to header content.

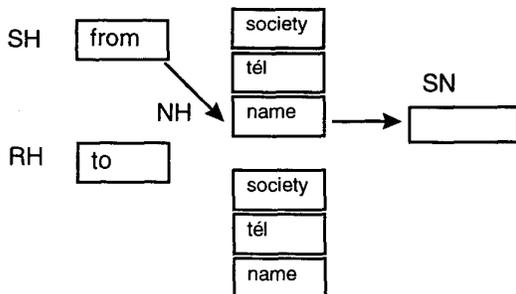


Figure 4 : example of a visual hierarchy

To retrieve header content, the pseudo word, the nearest neighbor of the header component is searched. As the logical object *sender_name* can be included in one or more layout components, the nearest neighbor is considered as an anchor point which is grouped to other layout components to build an alignment.

This process enables the retrieval of contents not necessarily aligned with the header.

4. Experiments

First experiments have been performed on a sample of 130 fax images. This sample contains fax front covers with preprinted fields. For these sample documents, 76 % of the existing logical pairs were accurately extracted and 80% of the *sender_name* objects were accurately located. Missed logical pairs are due to document skew and to missed headers because of Ocr segmentation/recognition errors. In a few cases, headers are not aligned and the alternative strategy is triggered which improves the extraction of sender names by 4% compared to the extraction of logical pairs.

In this experiment, we considered all highly scored hypotheses for logical pairs : typically one, sometimes two hypotheses are generated for one image. 80 % of correct logical pairs are top choice results.

The result is the sender name location in the image. If this name is printed, the text string can be provided. If the name is handwritten, the image of the name can be provided for instance through a structured html file that can be visualized even on a little screen (figure 7).

Another experiment was performed including also letter-like type faxes. On 90 % of these documents, no

logical pair was found which means that the method detected the lack of symmetry for sender and recipient information. This could be used for classifying facsimile images into letter-like and form-like documents.



Figure 5 : original fax document

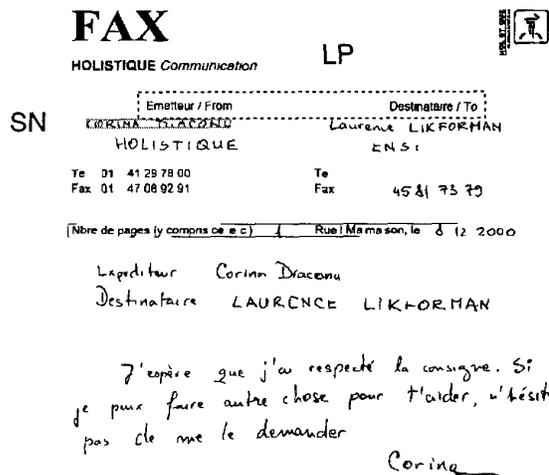


Figure 6 : logical pair (LP) and sender name (SN) extracted

```

<html>
<br> You have received a new fax from

<br> Sender name : <IMG SRC="recu006.gif">
</html>

```

You have received a new fax from

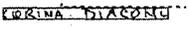
Sender name : 

Figure 7: html file and visualisation

5. Conclusion

We have described an analysis procedure which aims at extracting the location of the sender name from a fax image and its OCR'd version. These images are fax front covers which may include handwritten items. As layout variance is high, we avoid numerical criteria by exploiting visual/spatial cues such as symmetry, proximity, and alignment. This enables us to associate main headers into logical pairs and to retrieve their content. Language processing is reduced to the extraction of predefined keywords. It could be extended to language identification in order to load the appropriate lexicon. The processing of letter-like faxes will need deeper language processing such as proper name detection.

Finally, the use of HCR (handwritten character recognition) and name data bases will result in the recognition of handwritten names.

Acknowledgments : we thank the French Ministry of Economy (MEFI) for supporting this research, with grant no 01 2 93 0268

6. References

- [1] H. Alam et al., *FaxAssist : an automatic routing of unconstrained fax to email location*, IS&T/SPIE Conference on Document recognition and retrieval, January 2000, San Jose, California, pp. 148-156.
- [2] S. Baumann et al., *Message Extraction from printed documents*, 4th ICDAR, 1997, Ulm, pp. 1055-1059.
- [3] F. Cesarini, M. Gori, S. Marinai, G. Soda, *INFORMys : a flexible invoice-like form reader system*, IEEE PAMI, Vol 20, no 7, july 1998, pp. 730-745.
- [4] G. Chollet, J. Cernocky, J. Hennerberg, P. Petrovska, F. Yvon, *Towards fully automatic speech processing techniques for interactive voice servers*, in Speech

Processing, Recognition and Artificial Neural Nets, E. Caianiello ed., Springer Verlag, 1999, pp. 297-326.

[5] A. Dengel, *Anastasil : a system for low level and high level geometric analysis of printed documents*, in Structured Document Analysis, (Baird, Bunke, Yamamoto eds), Springer-Verlag, 1992, pp. 70-98.

[6] C. Faure, *preattentive Reading and selective Attention for Document Image Analysis*, proceedings of ICDAR 99, Bangalore, october 1999 , pp. 577-580.

[7] G. Guiffrida, E. Shek, J. Yang, *Knowledge-Based Metadata Extraction from Postscript Files*, Digital Libraries 2000, San Antonio, pp. 77-84.

[8] M. Holstege, Y. Inn, L. Tokuda,, *Visual Parsing : An aid to text understanding*, proceedings of RIAO'91, Barcelone, August 1991, pp. 175-193.

[9] T. Kochi, T. Saito, *User-defined template for identifying document type and extracting information from documents*, proceedings of ICDAR 99, Bangalore, october 1999 , pp. 127-130.

[10] J. Lii, S. Srihari, *Location of name and address on fax cover pages*, 3th ICDAR, 1995, Montreal, pp. 756-759.

[11] L. Likforman-Sulem, B. Cuenca, *facsimile processing for a messaging server*, proceedings of the 10th Int. Work. On Database and Expert Systems Applications -DEXA' 99, Florence, september 1999 , pp. 539-543.

[12] H. Walischewski, *Automatic knowledge acquisition for spatial document interpretation*, 4th ICDAR, 1997, Ulm, pp. 243-247.