Project-Team Imadoc

Interprétation et Reconnaissance d'Images et de Documents

Rennes

THEME COG

Activity Report

2004

Contents

1 Team

Team Leader

Jean Camillerapp [Professor Insa]

Administrative Assistant

Myriam David [TR Inria]

Staff member (Inria)

Jacques André [Research Director]

Bertrand Coüasnon [Research scientist (secondment)]

Staff member (Insa)

Éric Anquetil [Associate Professor]

Ivan Leplumey [Associate Professor]

Yann Ricquebourg [Associate Professor]

Nicolas Ragot [Lecturer from 01 Oct 2004]

Staff member (Université of Rennes 1)

Guy Lorette [Professor]

Charles Quéguiner [Associate Professor]

Hélène Richy [Associate Professor]

Nicolas Ragot [Lecturer to 30 Sep 2004]

Ph.D. Student

Sabine Carbonnel [Menrt research grant]

Laetitia Rousseau [Insa research grant]

François Bouteruche [Insa research grant from 01 Oct 2004]

Isaac Martinat [Insa research grant from 01 Oct 2004]

Harold Mouchère [CNRS/Région research grant from 01 Oct 2004]

Solen Quiniou [Menrt research grant from 01 Oct 2004]

Technical staff

Grégory Maitrallain [Contract with the *Conseil Général des Yvelines*]

2 Overall Objectives

The research topics of the project concern *the written communication* and *the engineering of documents* under various aspects: *analyzis, recognition, interpretation of digitized documents, man-document interaction*. This research relates to the writing and the documents under all their forms (manuscript, printed paper form, graph, images, composite documents, etc.) as well as the linked activities.

Facing the multiplicity and the diversity of the applications, the usual solution consists in bringing answers dedicated to each type of application. The project privileges, on the contrary, the development of generic approaches. With this intention, our research is carried out according to five interdependent topics, two of more methodological nature and three of applicative nature.

- **Image processing, early vision**
  Low level processing, also called early vision processing, has an unquestionable influence on the total performances of the data processing sequences.The digitalization of document images must carry out a compromise between the total size of the images and the apparent size of the characters. In this compromise, the width of the feature can be reduced to two or three pixels. The document image processing thus requires a very good precision: in the assignment of a pixel either to the background or the layout and in the extraction of the visual indices characteristics.

- **Modeling and management of knowledge**
  In order to develop automatic systems of *interpretation* of signals and of document images, the project studies methods of dynamic and multi-contextual recognition based on visual indices. Three complementary approaches for the modeling of contextual knowledge are considered for this purpose: one based on rules for symbolic knowledge systems, another one based on fuzzy inference systems for vague knowledge, and the third one based on neural networks for numerical data.

  The modeling of knowledge is carried out, using either models of *a priori* knowledge or machine learning from experimental data.

  In order to lead to the final interpretation, contextual knowledge is dynamically managed, according to the respective types of modeling, either by modified syntactic analysis, or by a soft computing method (fusion of fuzzy data or perceptive cycles).

  These two approaches have the advantage of offering several possible levels of reading, to locate where the errors remain and to lead to transparent systems which way of working can be analyzed and which performances can be optimized.

- **Handwritten writing**
  We study recognition systems for *handwritten* documents, either during their phase of creation (recognition of *on-line* writing), or afterwards (recognition of *off-line* writing).

  The operation of these systems is based on *a priori* general knowledge about handwritten writing, about the styles of writing and about the visual indices resulting from the layout.

This enables us to overcome the complex problem of the segmentation of the words in separate letters and facilitates the recognition of the words of a large size vocabulary (*analytical approach*). The use of such systems can be of interest to a very large audience (*multilingual and omni-script writers recognition systems*). Nevertheless, it is always possible to specialize a given system to build from it a dedicated application.

- **Numerical document and semantic Web**
  The recent technological developments (development of the Internet, high flow communications, large storage capacities, effective image compression algorithms, policy of digitalization on a large scale of the patrimonial or industrial files) make it possible to place enormous corpora of digitized documents at the public disposal.

  However, an intelligent access to the images (*semantic Web*) supposes the existence of annotations describing with precise enough details the contents of these images.

  The studies which we currently undertake relate to the development of models and techniques aiming at facilitating the automatic annotation of digitized images.

  Our research relates to the transformation of images of existing paper documents (reverse conversion of old documents) into structured numerical documents. The generic approach relates not only on the description of the forms (geometry), but also to the definition of models of annotation and ontologies on the contents. This approach makes it possible to control the various stages of the transformation of the documents:
  - localization: location of the zones in the images,
  - reverse conversion: extraction of information and structuring,
  - annotation: analyze and interpretation of the contents,
  - indexing: search for information (tools to recompose images, filtering, etc).

- **Pen-based interaction**
  As most electronic devices are now provided with a pen interface, a number of applications are expanded where the pen can be used as a convenient and natural form of input. Moreover handwriting is a very familiar modality of input. Thus, our work focuses on the adaptation of recognition algorithms for an interactive environment.

3 Scientific Foundations

3.1 Early vision

Keywords: early vision, binarization, skeletonization, Kalman filtering.

Adaptative binarization

Documents are generally digitized at a minimal sampling rate, which implies that important elements for the recognition can take only a few pixels. For example, the non-detection of a very small white area can mask the presence, however very informative, of the loop in the letter *e*.

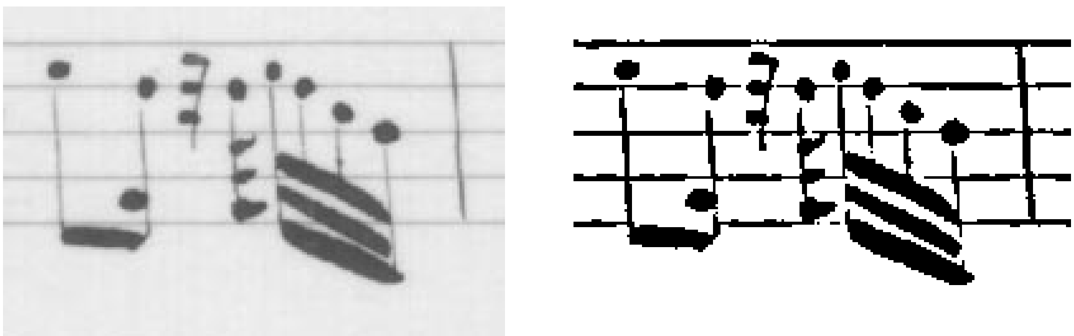
(a) Threshold of 210 : allowing the detection of thin lines

(b) Threshold of 195 : the vertical elements remain

(c) Threshold of 110 : the bars of the notes are separated

Figure 1: Binarization of the gray levels image of the figure 2 with a global threshold.

A digitalization in gray levels allows to implement adaptative binarization algorithms, often based on cooperation between edge detection and region growing. They give a local interpretation of gray values which separate well the various objects in the document (cf figures 1 et 2).



(a) Initial image in gray level

(b) Binarized image

Figure 2: Binarization with an adaptative threshold

Measurements of objects characteristics are sensitive to space sampling noise. Hence increasing space resolution improves the relevance of parameters related to visual indices. With a local interpretation of gray level, it is possible to locate the edge between the objects and the background with a *subpixel* precision and thus increase the relevance of these parameters.

Gray level image skeletonization

Skeletonization algorithms have been developed to thin images of lines. But in documents images, it is necessary to distinguish in the layout (figure 3) the area really skeletizable (also called regular area) from the singular areas, principally localised in crossings in which the concept of median axis does not have any practical interest [11].

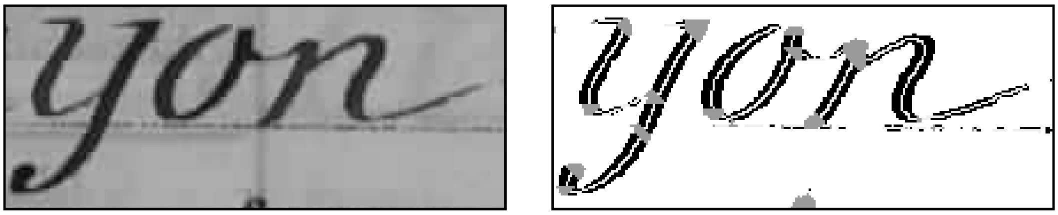


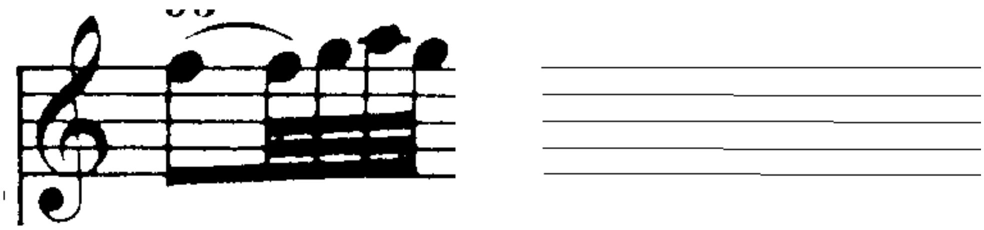
Figure 3: Processing of an image from an old document. The skeleton appears in white in the medium of the line and singular areas are in dark gray.

In order to limit the binarization noise, one can obtain a sub-pixel localization of points of the median axis while working when gray level images.

Segmentation with Kalman filters

Linear structures have a great importance in documents, but they are likely to interfere between them or with the drawing. However, the simplicity of their structure makes it possible to plan simultaneously segmentation and recognition.

Kalman filtering is a technique of parameters identification from an ordered series of measures. In the case of lines, the model is reduced to the thickness of the line, its slope and the equivalent of the ordinate at zero point. Measure results from the position and the thickness of the black run-length[1] orthogonal to the drawing.



(a) Original image (b) Extracted stave lines

Figure 4: Detection of stave lines with Kalman filters

[1] Set of consecutive black pixels according to one of the four directions: horizontal, vertical or diagonals

The Kalman filtering also calculates the covariance matrix of its estimate of the parameters of the model. This matrix makes it possible to evaluate on one hand the probability to assign a measure to a segment and on the other hand the probability to continue an assumption disturbed by the presence of another object [6].

The real difficulty of this approach does not lie in the equations of Kalman, but in the association of the extracted run-length from the image with the predicted run-length when linear structures cross or overlap (figure 4).

Thus a run-length in the image can result from overlap of several drawings. At this level the information brought by the covariance matrix is of great importance.

In this approach we can also use a measure of a run-length with a sub-pixel precision.

3.2 Dmos, a Generic Method for Structured Document Recognition

Keywords: syntactical recognition, grammar, definite clause grammar, segmentation, structured document recognition, musical scores, mathematical formulae, table structure, form.

Literature shows that structured document recognition systems are long and difficult to develop. Moreover, we usually find systems specifically developed for one kind of document - for example mathematical formulae - which cannot be easily adapted to another kind of document, like table forms. A new kind of document often requires the complete development of a new recognition system, which is a real loss of energy. Defining a generator of recognition systems for structured documents is a real challenge to avoid this costly rewriting work.

Moreover, improving recognition quality is vital for an industrial processing of documents. This can be done, for example, by solving segmentation problems usually found in documents.

In fact, we can consider that those problems are linked to the management of *a priori* knowledge. Indeed:

- a generic system must allow to change only what changes from a document to another: the *a priori* knowledge on each document;

- improving segmentation on documents needs to introduce as much as possible *a priori* knowledge in the system.

Therefore, we proposed the Dmos method (Description and MOdification of Segmentation), a generic recognition method for structured documents made of:

- the new grammatical formalism EPF (Enhanced Position Formalism), which can be seen as a description language for structured documents. EPF makes possible at the same time a graphical, a syntactical or even a semantical description of a document;

- the associated parser which is able to change the parsed structure during the parsing. This allows the system to try other segmentations with the help of context to improve recognition.

EPF Language

We can find in the literature various bi-dimensional extensions of mono-dimensional grammars defined for object and document recognition. However, they offer a too weak expressiveness (trees and *web grammars*), or the syntax is too complex for dealing with large and difficult *a priori* knowledge (plex grammars, graph grammars).

We developed the EPF formalism to be able to describe a bi-dimensional structure, in a graphical as well as in a syntactical way. Instead of having characters as terminals in mono-dimensional grammars, EPF uses two kinds of terminals: line-segments and pixel arrays (components, connected or not, which represent a symbol). In EPF several operators are added. Here are some examples:

Position Operator (encapsulated by `AT`):

```
A && AT(pos) && B
```

means `A`, and at the position `pos` in relation to `A`, we find `B`.

Where `&&` is the concatenation in the grammar, `A` and `B` represent a terminal or a non-terminal.

Factorization Operator (`##`, in association with the position operators):

```
A && (AT(pos1) && B ##
      AT(pos2) && C)
```

means `A && AT(pos1) && B` and `A && AT(pos2) && C`

With this syntax it is possible, for example, to describe a simplified beamed note (a beamed eighth note, with only one beam - a segment - that links the notes):

```
beamedNote ::= beam &&
         (AT(leftTip) && noteGr ##
          notesInBetween ##
          AT(rightTip) && noteGr)
```

Where `::=` is the constructor of a grammar rule. The writer of the grammar can define, as much as necessary, position operators as well as he can for non-terminals.

Save Operators (`--->` and `<---`): To be able to represent the same instance of a terminal or a non-terminal `A`, we propose to save an instance of it by using the operator `--->`. This backup of `A` gives then the possibility to refer (using the operator `<---`) to `A` as much as necessary. It is then possible to describe a rectangle:

```
rectangle ::=
   (segV ---> segLeftSide) &&
   AT(touchUp) && segH &&
```

```
AT(touchRight) && segV &&
AT(touchDown) && segH &&
AT(touchLeft) &&
(segV <--- segLeftSide).
```

Associated Parser

The EPF language allows to define a description of a document. From this description, we produce by compilation, a parser with the specific properties needed for parsing bidimensional documents. Compared to classical monodimensional parser, the main properties of the bidimensional parser we develop are:

- changing the parsed structure during parsing for contextual segmentation. The parsed structure can then be changed to be like it should be if there was no segmentation error;
- detecting the next element to parse. Indeed, in classical parsers, the next element is simply the head of the parsed string. In two dimensions, the next element can be everywhere in the image, hence everywhere in the parsed structure. The parser uses the position operators to select the next element to parse;
- dealing with noise. We can consider that dealing with noise corresponds to finding the next element to parse, even when there is a lot of noise in the parsed structure.

In conclusion, with this new EPF language we have developed a generic recognition method for structured documents. We implemented this method to obtain a generator of structured document recognition systems. This generator can automatically produce new recognition systems. It is just necessary to describe the structure of the document with an EPF grammar which is then simply compiled. In this way we have developed various recognition systems: on musical scores, on mathematical formulae, on recursive table structures...

3.3 Pattern recognition by fuzzy inference systems

Keywords: Explicit modeling of the knowledge, fuzzy inference system, learning, classification.

Faced with the variability and the complexity of patterns to be modeled, many recognition systems are often based on "opaque" algorithms which cannot be easily interpreted after the learning process. This "opaque" aspect often hinders an explicit description of the process. Therefore, the modification and the interpretation of error causes of complex system is very difficult. Handwriting recognition is a typical example of this kind of problem. Different methodologies are used to try to solve this difficult problem like approaches based on neural networks or methodologies based on hidden Markov models that lead typically to "opaque" systems. Up to now, it is commonly believed that fuzzy set theory is an interesting approach to design "transparent" modeling (i.e. readily understandable) capable of dealing with the input imprecision. More precisely, we use fuzzy rules to design robust, compact and transparent classifiers. A fuzzy inference system is composed of N_r fuzzy rules to model the C classes.

These rules can be expressed explicitly by:

$\boldsymbol{R_i}$: **IF** x_1 is M_{i1} **and** x_2 is M_{i2} **and** ... **and** x_n is M_{in}
THEN the input shape belongs
to the class 1 with the degree $\{b_{i1}\}$ **and** ... **and**
to the class k with the degree $\{b_{ik}\}$ **and** ... **and**
to the class C with the degree $\{b_{iC}\}$.

- the inputs $x = (x_1, ..., x_n)$ are defined in a n dimensional space;
- the premise part of the rule is a conjunction of (x_j is M_{ij}) that formalizes the membership degree $\mu_{M_{ij}}(x_j)$ of the input x_j, to the fuzzy set M_{ij} modeling a property of a class;
- the outputs Y_k are fuzzy sets B_{ik} reduced here to singletons $\{b_{ik}\}$;
- the consequent part of the rule (Y_k *is* b_{ik}) formalizes the membership degree of the input shape, according to the rule R_i and the class k.

Fuzzy inference is then defined as below :

$$\mu_{B'_k}(y) = \perp_{i=1}^{N_r} \; I(\beta_i, \mu_{B_{ik}}(y)) \; with \; \beta_i = \mathbf{T}_{j=1}^{n} \; \mu_{M_{ij}}(x_j),$$

where

- $\mathbf{T}$ is a T-norm corresponding to the conjunction in the premise of the rule;
- β_i is the activation degree of the rule R_i;
- I is the fuzzy implication[BM95];
- $\perp$ represents the aggregation operator.

The output of the fuzzy inference system provides the adequacy of the input shape to each class.

The automatic generation of membership functions is still an open problem. We use here the possibilistic clustering developed by Krishnapuram [Kri94] to automatically generate fuzzy rules. Krishnapuram has shown that possibilistic clustering can be successfully used to solve the determination of membership functions which can be interpreted as degrees of "typicality", contrarily to classical clustering where membership functions are relative to each other and interpreted rather as degrees of "sharing". Faced with the problem of complex modeling and to keep an explicit description of each model, we structure hierarchically fuzzy rules relatively to the robustness and to the pertinence of the modeled primitive. The internal structure of

[BM95] B. Bouchon-Meunier, *La logique floue et ses applications*, Addison-Wesley, 1995.
[Kri94] R. Krishnapuram, "Generation of membership functions via possibilistic clustering", *in: IEEE World congress on computational intelligence*, p. 902–908, 1994.

the models is automatically generated during the learning phase and is consequently adapted to current class which is going to be modeled. Based on the previous observations, we set out a methodology to generate "transparent" pattern recognition systems based on qualitative modeling by hierarchical fuzzy rules automatically deduced from possibilistic clustering.

This methodology has been applied to the difficult problem of on-line handwritten character recognition. For each character class, robust and stable properties have been automatically extracted and modeled in a qualitative manner. They are organized in a hierarchical structure according to their robustness and guided by *a priori* knowledge on cursive handwriting. This led to the RESIF software specifically elaborated for on-line cursive recognition (cf. section 5.1). Thanks to its compact fuzzy modeling, RESIF is able to absorb the variability of handwriting shapes and can cope with limited computing and memory resources. The RESIF software has successfully been integrated on a Smart Phone device with limited resources (cf. section 7.1). Recently, a new classification approach, named Mélidis, has been elaborated. This new system can be considered as an extension of RESIF. The main goals are to obtain a more generic recognition system able to deal with different kinds of classification problems without needed *a priori* knowledge and without loss of transparency so that optimization can be done afterwards by experts (cf. section 6.7).

4 Application Domains

4.1 Paper document analysis

Applications of our research in the context of paper document analysis are very wide. Indeed, the generic approach we have developed (the Dmos method) allows a quite fast adaptation to new kinds of documents. Documents we have already worked on are:

- Orchestra scores with polyphonic staves;
- Mathematical formulae;
- Table structures, forms with recognition of the hierarchical organization;
- Archives documents: old forms more or less structured.

The Dmos method could also be applied to various other kinds of documents: chemical formulae, title blocks, newspaper pages, invoices, delivery notes...

Moreover, a grammatical description of a kind of document (made with EPF) can be either general, in order to recognize a class of document (title blocks for example), or specific, to compensate a lack of information (damaged title blocks).

Of course, it is possible to merge grammatical descriptions of documents. For example, we can simply build a system able to recognize table structure containing mathematical formulae.

With the help of handwriting recognition those recognition systems can be used in many ways:

- Retroconversion of paper documents to avoid a manual input to get a usable electronic version. It can for example be used for a new edition, or to produce a Braille document;

- Kind of document identification, to make for example an automatic management of faxes;
- Automatic production of indexes and annotations for an automatic access by content to documents (section 6.1);
- Detection of specific areas in a document to improve its access.

4.2 Handwriting recognition systems

Recently, there was a new increasing of the applications of handwriting recognition in both domains of automatic processing of paper documents (*off-line recognition*) and in the new modality of man-machine interaction based on the use of a pen and a touch-screen (*on-line recognition*).

To date, in *off-line recognition*, industrial needs are huge and for this reason the design of robust and performant recognition systems is highly needed. The application domain range is very large. It concerns the problem of the automatic processing of every kind of paper documents, e.g. order lists, social security forms or fax. In this domain, we mainly brought our efforts to guaranty a high degree of robustness and confidence in the results in the case of automatic processing of handwriting. This was done to avoid any risk of error.

In *on-line recognition*, a huge market has arised due to the recent economic development of the following devices:

- pocket computers (Palm, PocketPC),
- tabletPCs (computers without keyboard and with a sensitive LCD),
- mobile phones of the new generation (¡¡ smartphones ¿¿) which integrate the interaction modalities of the pocket computers (LCD coupled with a sensitive screen) and their principal functions (agenda, text editor, Internet access, etc.).

To date, the ResifCar software was embedded in such devices (cf. subsections 5.1 and 7.1).

4.3 Pen-based interaction

More electronic devices with pen interface are now available for entering and manipulating information. A number of applications are expanded where the pen can be used as a very convenient and natural form of input. Handwriting is an input modality that is very familiar for most users since every one learns to write at school. Pen based interfaces capture information as the user composed it, including text, graphs or commands.

The Imadoc project focuses on the interpretation of pen input (electronic ink) as captured by the pen based interface system. The pen input may be available on any kind of device where pen interface is available: mobile-device's screens, tactile screens, electronic paper, or tablet.

Topics of interest of these experimentations are the adaptation of the recognition algorithms for an interactive environment, the validation of the generic approach, and the analyze of the benefits of multimodal interface. So, several applications are developped and experimented using gestual commands, or intuitive editing commands:

- writing a musical score,
- drawing a graph,
- annotating documents (photo, archive) for indexing (digital library),
- editing web pages.

Most applications will provide the users with a multimodal interface that enables the choice among various modes for commands.

5 Software

5.1 RESIF : Handwriting Recognition by Hierarchical Fuzzy Inference Systems

Participant: Éric Anquetil.

Keywords: Handwriting Recognition, smartphone, fuzzy logic.

RESIF technology is today composed of three main software to analyze, model and recognize handwritten characters and words:

- RESIFCar is specialized to recognize handwritten characters: latin alphabet, digits and special symbols.
- RESIFMot is the software for unconstrained cursive handwritten words recognition.
- RESIFApp is the automatic learning process that generates from a handwritten character database the hierarchical fuzzy models used by the recognition systems: RESIFCar and RESIFMot.

RESIFCar and RESIFApp are already in their fourth version. In 2001, through an industrial collaboration with PurpleLabs company [2], RESIFCar has been successfully integrated into mobile devices (smartphones) which are characterized by their limited computing and memory resources. Theses mobile phones are nowadays commercialized in Europe (cf. section 7.1).

Work is in progress to extend the capability of the recognition system RESIFMot which is today in a beta version. The aim is to optimize lexical post-processing to deal with large vocabulary and to reduce time computing and memory resources.

[2] PurpleLabs : cf. http://www.purplelabs.com

5.2 DocRead : an Automatic Generator of Recognition Systems on Structured Documents

Participant: Bertrand Coüasnon.

Keywords: Recognition, structured document, musical scores, mathematical formulae, table structures, forms, archives.

DocRead is an automatic generator of recognition systems on structured documents. It has been developed thanks to the Dmos method. It is made of a compiler of the EPF language (with which it is possible to describe a document), a parser associated to this language, an early vision module (binarization and line segments detection) and a classifier able to reject.

This generator allows us a fast adaptation to a new type of document. Indeed, it is only necessary to define a new grammar in EPF, which describes the new type of document. Then, if necessary, a new learning of the classifier is done to make it able to recognize new symbols. The new recognition system adapted to a new structured document is produced by compilation.

With this generator, we already have been able to produce recognition systems of structured documents:

- ScoRead: a prototype for musical scores recognition;
- MathRead: a prototype for mathematical formulae recognition;
- TennisRead: a prototype for tennis court detection in videos;
- TabRead: a prototype for table structures recognition;
- DecRead: a prototype for naturalization decrees of the 19th century recognition. These are fully handwritten forms;
- FormuRead: a software for military forms of the 19th century recognition despite their deterioration. This software has been successfully tested on 88,954 pages of the *Archives de la Mayenne* and *Archives des Yvelines* (section 6.3);

5.3 Library to extract visual indices

Participant: Jean Camillerapp.

Keywords: Early vision, adaptative binarization, vectorisation, Kalman filtering, squeletonization, multi-resolution..

It is the library on which the DocRead software described in the section 5.2 is based.

It is composed of various methods of adaptive binarization. Those use either local determinations of threshold or edge detection and region growing cooperation. Some of these methods can provide results with a sub-pixel resolution.

Using Kalman filters, as described in section 3.1, this library carries out detection of rectilinear segments in complex images.

It also contains a skeletonization method from gray level images.

The selected resolution for the digitalization of the documents is not always adapted at best to the recognition of all the structures. This is why the library also provides possibilities of multi-resolution processing.

5.4 FDT-L : Fuzzy Decision Trees Library

Participants: Nicolas Ragot, Éric Anquetil.

Keywords: Fuzzy decision tree, fuzzy logic, fuzzy clustering.

The FDT library is a classification software for extracting and exploiting fuzzy decision trees. Fuzzy sets are automatically extracted by fuzzy clustering algorithms in one dimension or more. Depending on the clustering algorithm used, they can either describe the classification problem based on a discrimination point of view or on an intrinsic point of view. The resulting classifiers can be more accurate than other traditional classifiers such as radial basis function network or multi-layer perceptron with a higher legibility. This first version is also able to construct a forest of decision trees to be more accurate on multi-classes problems.

FDT-L is currently used by the Mélidis system (cf. section 6.7). Work is also in progress to extend the capability of the library by taking in account more clustering algorithms and to provide a visual interface.

5.5 A platform for image document annotation

Participants: Bertrand Coüasnon, Grégory Maitrallain, Ivan Leplumey.

Keywords: Annotation, access by content, document retrieval, handwritten documents, archives.

We propose a platform to improve the access by content on Archives documents with handwritten text. To make this access, it is necessary to associate annotations to the images of documents. Annotations for Archives documents can be geometric or textual. With those annotations, it is then possible to make an automatic selection of images. The platform we propose to manage annotations has the interest of producing annotations in two complementary ways: automatically with document recognition and collectively with the help of the readers during their readings.

We propose to build a platform for Archives document retrieval which could deal with textual and geometric annotations at the same level. Moreover this platform is able to manage relations between textual and geometric annotations to specify that a textual information is at this specific location in the image of document. In a different approach, a textual annotation can be linked to different locations in different pages of document *e.g.* in different images.

We choose to develop the platform with a classical architecture: a web server (Apache) with a servlet container (TomCat); the Java servlet access to a relational database (PostGreSQL) to store annotations and send them to the client: a java applet running in a web browser (figure 5).

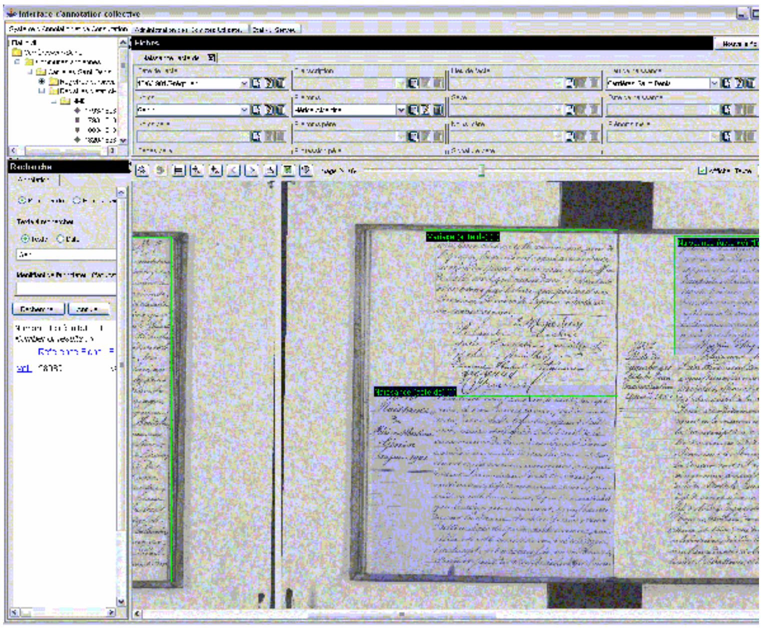


Figure 5: A unique interface to consult, annotate, retrieve Archives documents

Figure 6: A pen-based interface on the platform for annotations

With the platform, on a web browser, a user can leaf through images of Archives documents. When a page is displayed, all the associated annotations are presented on the interface: geometric annotations are drawn on the image, the textual annotations are presented in tabs for the forms (marriage certificate ...) and in field boxes for the fields in the form (name, date ...)(figure 5). The reader can consult annotations, add or modify annotation (if he has the right to), but is limited by the allowed annotation structure given by the DTD configuration file, according to the kind of document. The system can also store various interpretations if readers do not agree.

Structured search or full text search is possible on all the annotations whatever the way they have been produced: automatically or manually.

Since January 2004, the platform is running in the building of the Archives départementales des Yvelines. 28,000 military forms are available for the public, through the graphical user interface installed in reading rooms. This quantity will reach 150,000 during the year 2005, and will be followed by 1,200,000 double pages of register of births, marriages and deaths, which are the most requested documents.

The platform is applied presently on Archives documents, but of course it could be used on every kind of images.

A pen-based interface has been defined on this platform. Using specific gestures (section 4.3) and on-line handwriting recognition (RESIFCar, section 5.1), it has been possible to design a new way to interact with digital documents and paper documents (figure 6).

5.6 Handwriting Ink Grabbing Application

Participants: Grégory Maitrallain, Eric Anquetil, Sabine Carbonnel.

Keywords: Data Base Aquisition, Handwriting, InkML, Pen-based Interface.

To improve its different character and word recognition techniques, our project needs much handwriting data, from numerous writers. To gather these data, an ink managing library and handwriting grabbing application have been developped in Java.

The ink managing library permits to handle and store extended ink data in two file formats, a raw data file format (Imadoc), and a XML based file format (ImadocML). This ImadocML format includes such data as timestamped strokes, data about the writer (sex, age ...), and can deal with references to other traces, in order to handle segmentation data.

The handwriting ink grabbing application uses the ImadocML file format to design pen-input forms for differents types of inputs (letters, words, sentences, dates ...) (fig 7). The writer fills in the forms, and the data are then saved in ImadocML format.

A transcoding module has also been developped in order to handle raw Imadoc file format and external formats such as Unipen or InkML.

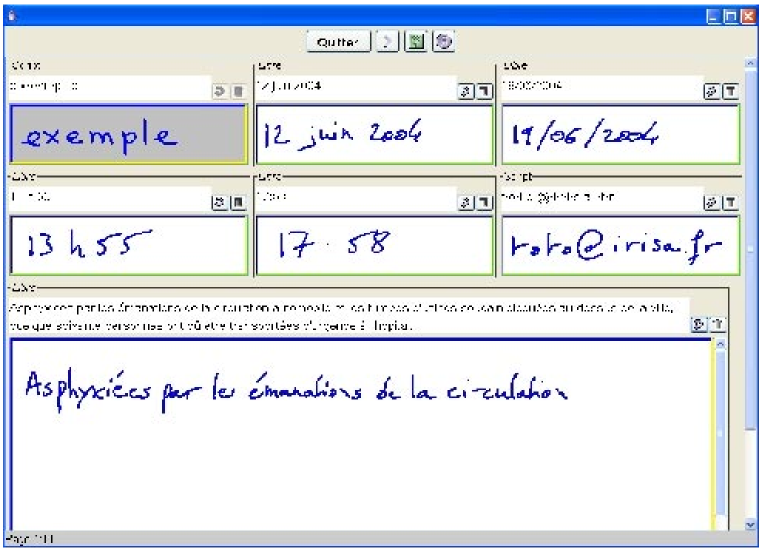


Figure 7: Handwriting Grabbing Application

6 New Results

6.1 Improvements in indexing old printed forms with handwritten last name

Participants: Jean Camillerapp, Bertrand Coüasnon.

Digitizing old documents is important to protect originals and to display them on Internet. However, it does not offer an improvement on document retrieval by content. To do so, it is necessary to apply document image analysis and handwriting recognition. Handwriting recognition in old documents is, most of time, extremely difficult or impossible to do. Nevertheless on some kind of old documents it is possible to think about automatic indexing of handwriting.

In structured documents as printed forms, indexation of some cells would be a very useful tool to select documents. Military forms of incorporation of the 19th century, of which there are millions of pages, are an example of indexable documents by their handwritten contents. Indeed these forms have a *name* cell in which the last name is not too badly handwritten.

The DMOS method (section 3.2) makes it possible to locate in the image the various cells included in the form and in particular the *name* cell (figure 12 in section 6.3). After that only this specific cell is processed.

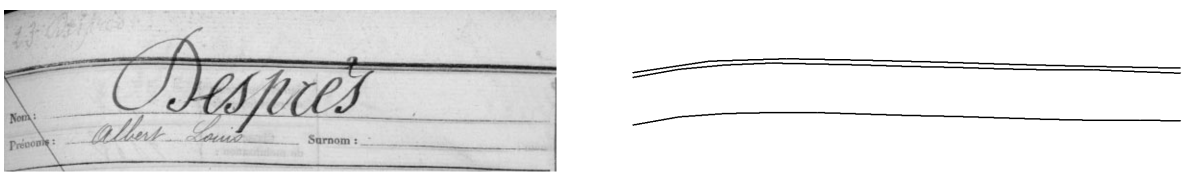
In 2003, a first application was built to index this kind of documents [19]. Last names are written by using an alphabet quite standardized: the slanting round-hand. In this handwriting style, the letters are systematically decomposed using a very limited vocabulary of shapes: the *graphemes*. To index images of military forms we chose to represent them with a segmentation in *graphemes* of the last name.

To retrieve images which correspond to a given patronymic, a user gives his request with the string of characters which spells the name. Then each character of the string is systematically translated into *grapheme*. The request is then compared to indexes of all images using an edit distance. It is then possible to order the list of images according to this distance to select the closest images to the request.

In 2004, various improvements described below were added.

Curvature

First series of documents were digitized with flatbed scanners. So the documents were flat and the rulings in images were straight lines. In new series, documents are now digitized with a camera and because of book binding, the rulings become strongly curved (figure 8).



(a) The name cell (b) Detected rulings and base line

Figure 8: Example of curvature

Our system was adapted to this new digitizing conditions. To do this, the choice to detect line-segments with Kalman filtering (cf. 3.1) was a strong advantage compares with classical projective methods.

Representation of an image

In the first version *graphemes* were classified in only seven classes. In the second one, the number of classes was increased to about forty. For each *grapheme* in an image, instead of keeping all answers of the classifier for all classes, we keep the only answers with a good enough score. An answer is kept if the score is greater than an absolute threshold and is also greater than a fraction of the score of the best answer. With these conditions an average of 2.5 answers per *grapheme* is kept. This limitation reduces drastically disk space and also speeds up comparison between request and images, while having the same quality in indexation.

Comparison between a request and images index

The comparison between one user request and all images of one register (about 500 images) is based on an edit (or Levenshtein) distance. This comparison uses a different weight for every operation such as *deletion*, *insertion* or *substitution*. In the current year, these weights have been optimized and the majority of them have been automatically deduced from the confusion matrix produced by the classifier.

Actually to find images in one register with a given patronymic, the good response is in the 5 top positions in 80% of requests.

(a) A crossed out name

(b) A struck off name

Figure 9: Example of damaged images

Detection of damaged images

Some images can be badly damaged (figure 9) : name can be crossed out when the person is dead during his military period, or the name is struck off when it was misspelled. Such images are very difficult to index correctly, so it is important to detect and mark them in the database. Thus it is possible to present separately to readers those unreadable images (about 5%).

This research was carried out with a grant of the *Conseil Général des Yvelines* (section 7.6). A version of this application is available for public at the *Archives départementales des Yvelines.*

6.2 Recovery of drawing order

Participants: Laëtitia Rousseau, Jean Camillerapp, Éric Anquetil.

IMADOC team has already designed an on-line handwriting recognition system called RESIF (cf. section 5.1). Our project is to unify the modelling of on-line and off-line handwriting. The first step of our approach consists in generating an equivalent on-line signal from a letter or a word image. To do this, the off-line drawing order has to be recovered by using writing knowledge.

Representation

To generate an on-line signal from an image, the first step is the handwritten drawing extraction from image. After using an image extractor developed by IMADOC project-team (cf. section 3.1), handwriting drawing is formalized by a graph where nodes correspond to singular areas (intersections, end points...) and edges correspond to regular areas (threadlike drawing).

Kato and Yasuhara's reconstruction method

The proposed approach is inspired from the Kato and Yasuhara's reconstruction method [KY99]. This method is a generic approach that was initially designed for drawing order recovery in handwriting images in general. It consists in finding the good path in a graph knowing the start and the end points. It uses topological criteria in graphs. This method is limited by several constraints that have to be respected:

- there exists only one start point and one end point;
- the start and the end points have to be specified;
- there exists no junction where more than two lines intersect;
- an edge can be crossed only once or twice.

This method is interesting because it doesn't cause a combinatorial explosion and it is generic. We have evaluated this method on letters and we propose an extension to resolve remaining errors and to relax constraints.

Evaluation

The Kato and Yasuhara's algorithm was tested on letters from IRONOFF base [VGLKB99]. This base includes both on-line and off-line signals. This particularity allows comparison between on-line signal and on-line signal deducted from image. The start and end points are deduced from the original on-line signal.

We tested this algorithm on 5000 letters images traced in one hand, and we obtained (cf. [26]) 94% of good paths (we call good a path identical to the on-line signal one). These results were rather good, so we analysed the error cases. Some letters don't respect constraints of the algorithm and several cases were not expected.

Extension

An extension (cf. [26]) was designed to resolve the remaining errors and have better reconstruction results. Some constraints were relaxed and new rules were introduced. The relaxed constraints are:

- there exists no junction where more than four lines intersect instead of two lines;
- an edge can be traversed only once, twice or three times instead of once or twice.

[KY99] Y. Kato, M. Yasuhara, "Recovery of Drawing Orders from Scanned Images of Multi-Strokes Handwriting", *in: International Conference on Document Analysis and Recognition 5,*, p. 261–264, Bangalore, 1999.

[VGLKB99] C. Viard-Gaudin, P. M. Lallican, S. Knerr, P. Binter, "The IRESTE On/Off (IRONOFF) Dual Handwriting Database", *in: Proceedings of the Fifth International Conference on Document Analysis and Recognition (ICDAR'99)*, p. 455–458, 1999.

We have introduced new rules to treat nodes with degrees greater than four, tangences of strokes, edges traversed three times, double-traced line not expected. Moreover, we propose several paths with one start point and one end point when the context isn't sufficient to take a decision.

With the same test protocol, the good paths rate was increased to 96.8%. The number of images in error was reduced of about 47%.

Conclusion

The future work will focus on the detection of the start and the end points, and on the handwriting made in several strokes, letters and words.

6.3 Table Structure Recognition

Participant: Bertrand Coüasnon.

We have worked this year on one of the numerous interests of designing a generic recognition system i.e. the possibility of producing either general or specific systems. We proposed the DMOS method which is made of a new grammatical language (EPF) and an associated parser able to deal with noise (section 3.2). From an EPF description of a kind of document structure, a new recognition system is produced by compilation. This method has been successfully used to produce recognition systems on musical scores, mathematical formulae, archives documents like naturalization decrees of the 19th century[23] and even tennis courts in videos.

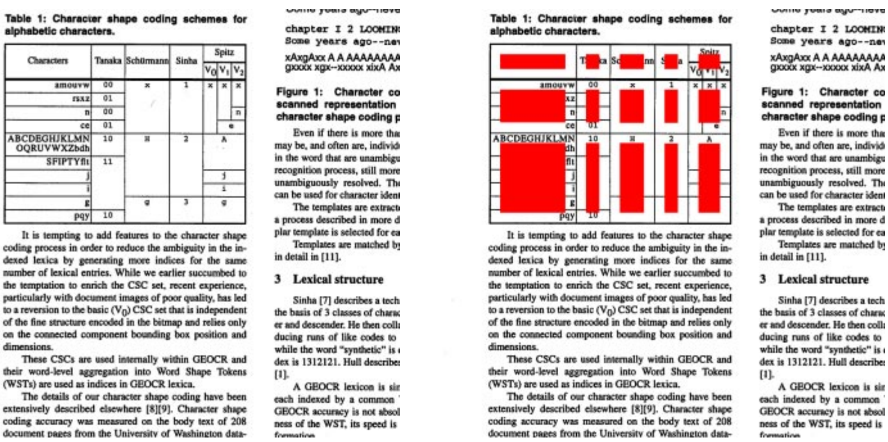


Figure 10: General description: example of a recursive table recognition in the middle of a page, first level of hierarchy

With this generic method, we have also been able to produce a general recognition system of table structures. It can recognize the hierarchical organization of a table made with rulings, whatever the number/size of column/rows and the deep of the hierarchy contents in it, as soon as the document has a not too bad quality (no missing rulings for example). This system has been produced from a description in EPF which is general enough to cover very different table organizations. This description can start by:

```
tableStructure ::=
    firstLineOfTable &&
    middleOfTable &&
    lastLineOfTable.
```

The produced recognition system is then able to recognize very different table structure hierarchies, assuming that they they are built on rulings (figure 10).

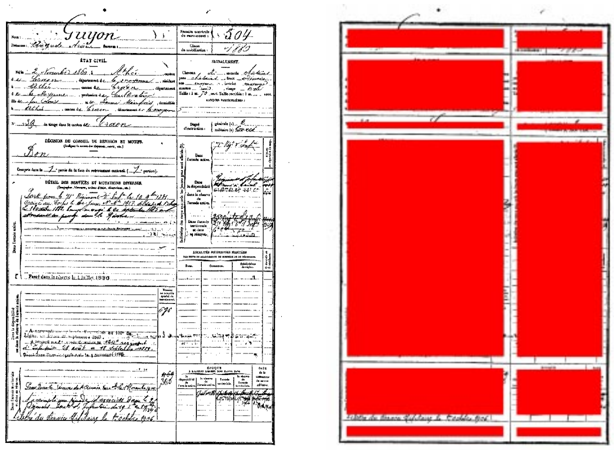


Figure 11: General description: example of recursive table recognition in an old document with a poor quality (first level of hierarchy)

The same general recognition system can also work on old documents. Figure 11 gives the first level on a military form of the 19th century found in French Archives. Even if the rulings are broken, not rectilinear, of poor quality, with intersection with handwritten text, with some skew ...

But when some information concerning the table is missing, due to a real bad quality or some hidden part of the structure, this system produced from a general description could not recognize the structure any more. This is because it is necessary to compensate this lack of information by a more specific description of the table.

With the help of the DMOS method and its genericity, we can easily define a specific description of the table structure of quite damaged military forms of the 19th century. The description starts by:

```
militaryForm ::=
    (DECLARE(centralVseg) (
            centralRuling && (globalFrame ##
                              allCells))).
```

From this quite precise description of those military form, we produced by compilation a specific recognition system which detects the cells positions (figure 12). This system has been successfully validated on 88,745 images, showing that this DMOS generic method can be used at an industrial level.

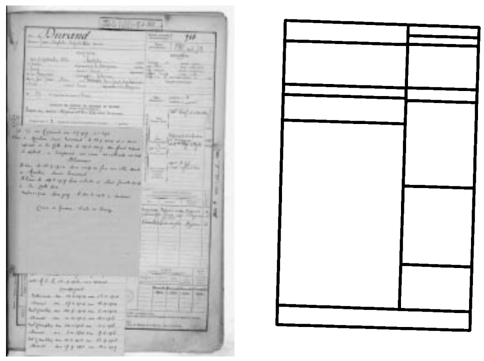


Figure 12: Specific description for military forms of the 19th century. Left: example of document with added sheets of paper, stamps ... Right: recognized structure even when it is partially hidden

We can consider that there is a continuous change between general descriptions and very specific descriptions of documents. A more or less precise description can then be done according to the necessity of compensating a lack of information, usually due to a very bad document quality. This strong adaptability is only possible with a generic recognition method like DMOS, because it offers a way to separate the *a priori* knowledge from the program.

6.4 ILib: a feature extraction library

Participants: Yann Ricquebourg, Bertrand Coüasnon, Grégory Maitrallain.

Concerning the studies of efficient classification and recognition methods, the Imadoc team lacks its own library of low-level extraction of information that should feed the developed classifiers. Therefore, we decided to federate our works and join our algorithms extracting and computing features from images in a common framework. This library has to be complete enough, with a stable API and easily reusable.

This platform represents a missing link in the components of the team to build entire processing chains.

To achieve this goal, we have been working on the definition and the progressive implementation of a portable C++ API. First concerning “classical” features often cited as required by common recognition systems (as surface, gravity center, curvature...), the library tends now to integrate more original or promising characteristic functions as Zernike moments, (concerning plain pixel images, see Figure 13), or elliptic Fourier descriptors (concerning contour curves, see Figure 14). Both these functions have interesting hierarchical information properties, thus leading to a progressive separation of major factors and details, associated with a high power of information concentration since only a few descriptors lead to a quite total representation of the characterized prototypes. We have also implemented auxiliary data structures that are required by some feature extractors (Freeman chains, run-length representation...) as well as utility functions (adaptative binarisation, connexted region labelling...). Finally we have specified a generic representation of our feature extraction functions, so as to be able to operate classical "zoning" computations for the classifiers using any of the underlying available feature extractors.

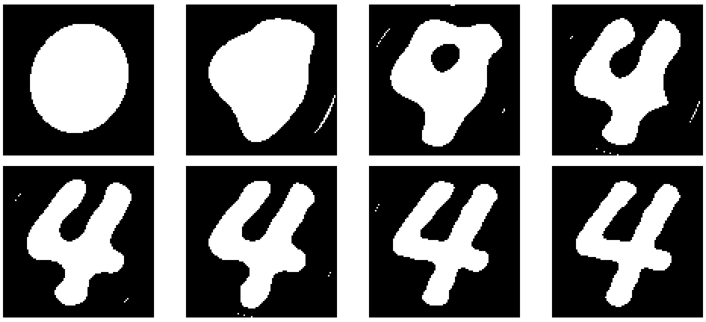


Figure 13: Progressive synthesis of the image of a digit 4, using its Zernike moments of order up to 2, 6, 8, 10, 12, 14, 18 and 20

Besides, we are developing a graphical user interface (see Figure 15), for the test and validation of those feature extraction functions. The GUI intends to be as generic as possible since it has been based on the common API we defined in the framework of the library.

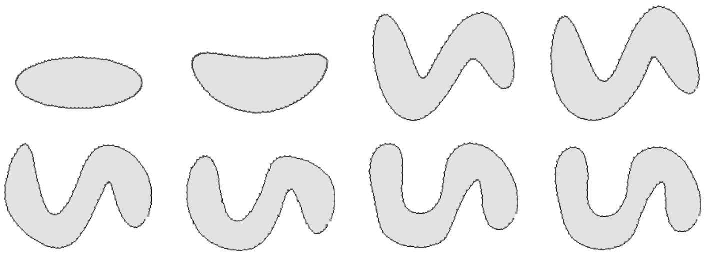


Figure 14: Progressive synthesis of a contour curve, using its elliptic Fourier coefficients of order up to 1, 2, 3, 4, 5, 6, 7 and 8

6.5 Pen-based user interface design

Participants: François Bouteruche, Éric Anquetil, Guy Lorette.

The RESIFCar system (cf. section 5.1) target small-size mobile devices such as smartphone or PDA. These devices present two main hardware constraints that we must keep in mind. The first one is their low resources (CPU speed, memory). The second one is the

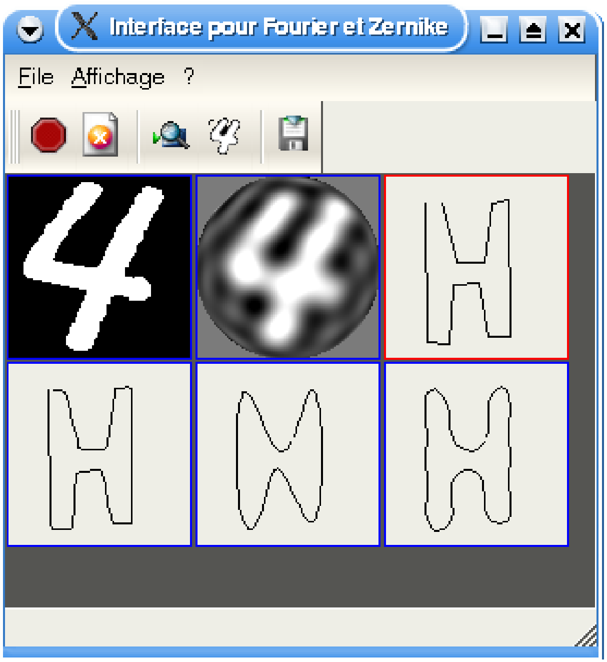


Figure 15: The test and validation graphical interface showing a Zernike moments application on a digit 4 and a elliptic Fourier coefficients computation on a letter H contour curve

small size of their touch screen. The first constraint has guided us to an isolated handwritten character recognition system. Indeed, response times of a handwritten word recognition system wouldn't be acceptable for users. The second constraint acts on the size of the input area. Considering this, we have decided to study and design interfaces which are adapted to this kind of systems. The objective of this research is to propose an intuitive and efficient interaction to the user.

Our approach takes into account previous works in ergonomic psychology on the effect of the information presentation and on text - entry tasks. Although they are not specific to handwritten character entry tasks on mobile devices, several of their conclusions can be used in our case:

- the positive effect of the spatial contiguity between several information sources;
- the influence on user's performances and satisfaction of the number of attention areas and the distance between them.

As a result of this analysis, a set of design principles for pen-based handwriting input interfaces has been identified [18]. Based on these principles, we have designed an interface which we have called ¡¡digital ink micro editor¿¿ [17, 18]. To go further on the validation of the proposed interface ergonomic quality, a collaboration with the Research Center in Psychology, Cognition and Communication of the University of Rennes 2 has started (cf. section 7.3). Experiments are being led to test two hypotheses which are implicit in our design principles:

- the spatial contiguity respect between the visual feedback area and the input area increases user performances (in terms of handwritten words per minute);
- the conservation of the previously written character in the input area ease the user's working memory from a reminding effort what increase user's performances and satisfaction.

6.6 Study of adaptation mechanisms for fuzzy classifiers optimization

Participants: Harold Mouchère, Éric Anquetil, Nicolas Ragot, Guy Lorette.

The Imadoc team has already designed powerful fuzzy systems for on-ligne handwriting recognition like RESIF (see section 5.1) and MELIDIS [25]. RESIF is already used in cell phones sold in Europe. We propose [24] to improve the recognition rate by designing a technique for on-line adaptation to the writer. Due to the on-line context, the adaptation has to be automatic, progressive and embedded.

In this study we use a Fuzzy Inference System (FIS), like in the existing system, but simplified, in order to focus the adaptation strategies and to transpose these adaptation methods to the original systems. This FIS is composed of N rules, each rule defined by:

$$\mathbf{R}_i : \textbf{IF} \;\; \vec{X} \;\; \text{is } \vec{P_i} \; \textbf{THEN} \; S_1^i = a_1^i \; \textbf{and} \; \ldots \;\; \textbf{and} \; S_C^i = a_C^i$$

The premise is defined by $\vec{X}$ which is the vector of characteristics of the form to recognize and by the fuzzy prototype $\vec{P_i}$. This prototype represents the intrinsic properties of the class described by this rule and its activation is written β_i. The conclusion describes each class by associating them a score S_C^i. The decision of the FIS is taken by a combination of scores of each rule by the inference called "sum - product" (sum for the aggregation and product for the implication, cf. section 3.3) which compute a score S_c for each class.

This year, the adaptation has concerned both the conclusions of the rules (by gradient descent) and the fuzzy sets of the premises (by moving their representative prototypes). We study different existing techniques as the LVQ[3] and the AVQ[4]. The common drawback of these methods is that they do not use all information available, like the score of the rule for each class. So we have developed a new one, ADAPT, which takes into account all the information. The movement of the center $\vec{\mu_i}$ of the prototype is given by:

$$\Delta\vec{\mu_i} = \lambda * \beta_i * \left(\sum_{c=1}^{C} (b_c - s_c) * s_c^i \right) * (\vec{X} - \vec{\mu_i}),$$

[3] Laaksonen98
[4] Dickerson1993

where b_c is the target score for the class c and λ is the adaptation parameter which represents the speed of the adaptation. This moving method combined with the gradient descent method is called *ADAPT + DG*.

The system was tested on 8 different writers. We compare the ADAPT + DG method with existing method LVQ and AVQ combined with DG. The LVQ+DG method (as the AVQ+DG one) rises the recognition rate from 87.95% to 90.72% but the ADAPT+DG rises it from 87.95% to 96.06%. The benefits of the method in term of error rate reduction is 67% in average (i.e. 2 on 3 errors are avoided) and up to 80% for the best writer.

6.7 Mélidis improvements

Participants: Nicolas Ragot, Éric Anquetil, Guy Lorette.

The Mélidis system, previously developed, is a classifier which can be considered as an extension of RESIFcar [3]. The main goals for the elaboration of this system were to obtain a generic recognition system able to deal with different kinds of classification problems without need of *a priori* knowledge and without a loss of transparency so that *a posteriori* optimization can be done by experts. The system should also be as compact as possible for integration on devices with limited resources such as PDAs or Smart Phones.

The resulting classifier, in its improved version, is fully described in a paper that was submitted to the IEEE Transactions on Pattern Analysis and Machine Intelligence and that is actually in its second round of review. The system is based on an original hybrid architecture composed of two complementary modeling levels driven through a focusing mechanism. The first level models each class to be recognized according to its intrinsic characteristics. Using this kind of modeling, the system is able to find automatically stable subclasses that compose the initial classes. The intrinsic models (description of the subclasses by fuzzy prototypes) are generated by an adapted fuzzy clustering algorithm like the Possibilistic C-Mean (PCM) [Kri94] on each class separately.

Based on this modeling, a focusing mechanism determines which symbols of a given class are confusing with others (this not necessarily entire classes that are a source of confusion).

The aim of the second level of the hybrid system is then to operate a progressive discrimination of the classes based on the result of the focusing mechanism. This is done by specific Fuzzy Decision Trees (FDT) dedicated to the discrimination task.

For recognition of unknown shapes, both levels are formalized homogeneously by fuzzy inference systems that are merged for final decision.

Previous experiments have shown the ability of the Mélidis system to adapt itself to different classification problems and to provide high recognition rates with few parameters. To explain these results, new experiments were carried out to evaluate more precisely the interest of each part of the system and especially of the focusing mechanism [25]. It can be seen that the second level is always more or equally accurate than the first level and in all cases, the final decision provides always higher recognition rates than both levels. Moreover, as expected, the

[Kri94] R. Krishnapuram, “Generation of membership functions via possibilistic clustering”, *in: IEEE World congress on computational intelligence*, p. 902–908, 1994.

high performances of the second level are not only a consequence of the classification technic used by itself (FDT) but also of the way it is used through the focusing mechanism. Indeed, this mechanism provides at least 5% and up to 40% of relative error reduction, which represents an average of 24% on four different classification problems. These experiments apart, we are also currently working on the integration of adaptation mechanisms that are described in section 6.6.

6.8 Lexical Post-processing for Handwritten Word Recognition

Participants: Sabine Carbonnel, Éric Anquetil, Guy Lorette.

We study an optimized lexical post-processing designed for handwritten word recognition. The aim of this work is to correct recognition and segmentation errors using lexical information from a lexicon. The lexical post-processing is based on two phases: in the first phase a lexicon organization is made to reduce the search space into sub-lexicons during the recognition process. The second phase develops a specific edit distance to identify the handwritten word by using a selection of the sub-lexicons. We present here a new approach to automatically learn an edit distance specifically adapted to the properties of the on-line handwritten word recognition. Experimental results show the impact of this learning process of the new edit distance.

A classic edit distance does not allow to solve specific problems of handwriting recognition (segmentation errors). Some adaptations have been carried out to take this problem into account [CA03]. Even if good results are obtained, an empiric and manual estimation of the different edit operations and their respective costs is needed. We present here an automatic method to define edit operations and costs according to the recognition system properties: it gives the possible to follow the evolution and improvement of the recognition module.

Moreover these operations depend on the recognizer properties, that is why improving the recognizer requires to make a new empirical adaptation of the edit distance. In order to automatically build all the edit tables and their associated costs, we have designed an automatic learning process (see figure 16).

The main idea is inspired by the concept of "boosting". The learning process is based on three iterative steps: firstly, learn a new edit distance using the precedent learned distance (the specific case of the first iteration is detailled in the following), secondly compute the errors for this distance on the learning base, thirdly modify the weights of the learning base elements according to errors. At each iteration a new distance is computed according to the precedent distance. The incremental process allows to modify the impact of the counted edit operations according to the error rate: the importance of an edit operation is decreased if the current edit distance can not correct the hypothesis, but is increased otherwise. That means the learning process focuses on the errors that the current distance does not correct. After N iterations of this improvement algorithm we obtain N edit distances and their associated error rates. For the post-processing, a new edit distance ($dist_{autoInc}$) is computed by a combination

[CA03] S. Carbonnel, E. Anquetil, "Lexical Post-Processing Optimization for Handwritten Word Recognition", *in: International Conference on Document Analysis and Recognition (ICDAR'03)*, p. 477–481, Edinburgh, August 2003.

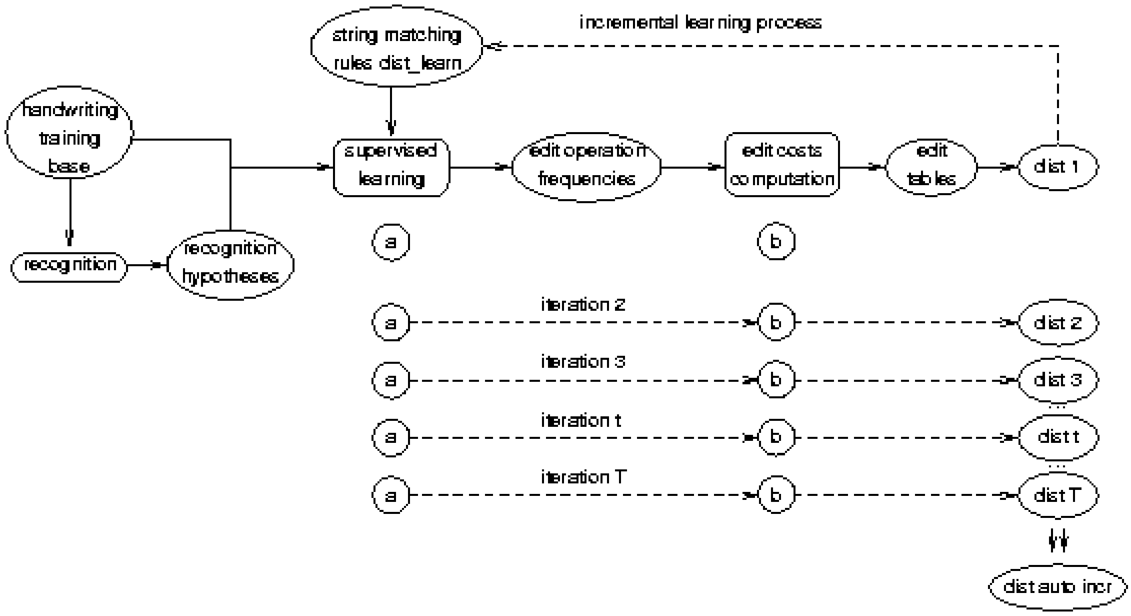


Figure 16: Principle of the edit distance learning.

of these N distances in order to give them an importance inverse to their error rate.

The first step uses the recognition system in a supervised mode to determine the set of edit operations for $dist_t$ with the matching rule of $dist_{t-1}$; and to compute the edit cost according to their frequencies in the training base. This step permits to learn the parameters of the new edit distance: $dist_t$.

The second step consists in the computation of the word error rate for the training base and the new learned distance $dist_t$.

The third step corresponds to up-date the weights of the words in the training base according to the recognition result: the weight of a word w increases if the recognition result for w is not correct, and decreases if the recognition result for w is correct. The aim is to give more importance to the words not recognized for the next distance learning ($dist_{t+1}$).

For the first iteration of the learning process, the learning is initialized by a first estimation using basic matching rules to learn $dist_1$. The basic matching rules are defined according to the word shapes. The learning distance $dist_{learn}$ is composed of six character edit operations: substitution, insertion, deletion, fusion, division and pair substitution. An initial cost is allowed to each operation according to the character shapes. This basic distance $dist_{learn}$ is based on fundamental word structures and allows a matching between the learning pair: a word label and an hypothesis.

The automatically learned distance $dist_{auto}$ permits to obtain slightly better results (error reduction rate: 5.7%) than with the modified distance $dist_{modif}$ empirically computed. These results are very interesting because they introduce the possibility to follow the recognition module evolutions by an automatic adaptation of the edit distance. The first impact of the edit distance learning improvement $dist_{autoInc}$ is relatively low on the considered benchmark (error rate reduction: 6.9% after 15 iterations) but this method allows to reach the edit distance limits: the study of the remaining error cases shows us that they can not be solved by an edit distance. Our perspectives are to introduce statistical informations (letters n-grams) directly in the segmentation graph of the recognition process. A first exploration for the combination of statistical informations and lexicon is described in 6.9.

6.9 Use of linguistic knowledge with statistical language models

Participants: Solen Quiniou, Éric Anquetil, Sabine Carbonnel, Guy Lorette.

Our works deals with the use of statistical linguistic knowledge in an online handwriting recognition system (see section 5.1) to improve the recognition rate. These statistical informations are used for handwritten words recognition in combination with lexical knowledge (see section 6.8 for the use of lexical knowledge) and also for handwritten sentences recognition. The representation of such knowledge is made by the use of statistical language models.

A first study was lead on the use of statistical language models of characters (namely n-gram models of character) in handwritten isolated words recognition in combination with the use of a lexicon during a lexical post-processing phase [CA03]. In that case, the language model is used to reorganize words hypothesis (given by RESIF - see section 5.1) in order to favour hypothesis consisting of most probable sequences of characters. The top hypothesis are then used during the lexical post-processing phase in order to generate real words, thanks to the use of the lexicon.

Tests were done on about 3 000 words, written by 3 different writers (each of the 1 000 words was written by the 3 writers). Thanks to a trigram model of character (build on a dictionary consisting of about 25 000 english words), the average gain (for the 3 writers) on word error rate was about 9 %.

The second study concerned handwritten words in the context of sentences. Here, the language model of words is used to find the best sentence corresponding to the handwritten sentence. Such a sentence contains both the most probable words (given by the word recognition system) and the most probable sequences of words. In that case, several n-gram models of words were designed: a bigram model of words, a bigram model of statistical classes (classes were automatically made using Brown algorithm [BPdSL92]) and a bigram model using some morpho-syntactic classes (classes for numbers, classes for proper nouns and other words each forming a class).

The Susanne Corpus [Sam00] was used for both construction of language models (about 5 500 sentences) and tests (about 1 400 sentences). For the tests, artificial sentences were build where words belonging to our base (*i.e.* the 1 000 words used for tests on handwritten isolated words recognition) were replaced by the corresponding handwritten word of our base

[CA03] S. Carbonnel, E. Anquetil, "Lexical Post-Processing Optimization for Handwritten Word Recognition", *in: International Conference on Document Analysis and Recognition (ICDAR'03)*, p. 477–481, Edinburgh, August 2003.

[BPdSL92] P. F. Brown, V. J. D. Pietra, P. V. de Souza, J. C. Lai, "Class-Based N-Gram Models of natural Language", *Computational Linguistics 18*, 4, 1992, p. 467–479.

[Sam00] G. Sampson, "The Susanne Corpus", 2000, available at "http://www.grsampson.net".

and others words (not in our base) were considered as recognized. Because the 1 000 words were each written by 3 writers, tests were done on each of them.
Thus, using the bigram model of words lead to an average reduction on word error rate was about 52 % (only considering the words belonging to our base). The use of the bigram model of statistical classes gave little less reduction of that rate (an average of 50.5 % when using 400 classes) but this model is more compact than the bigram model (30 000 parameters for the bigram model of 400 statistical classes versus 86 000 for the bigram model of words). Finally the use of a bigram model using morpho-syntactic classes for numbers and proper nouns achieves similar results than those with the bigram model of words but its size was also reduced (79 000 parameters).

These results were presented during the "Journée des Jeunes Chercheurs" organized by the S.A. 5.2 of the GDR I3 which took place in La Rochelle in June 2004.
These works are being extended to lead to the recognition of handwritten texts with using linguistic knowledge notably thanks to statistical language models.

7 Contracts and Grants with Industry

7.1 Integration into a Smart Phone Device of RESIFCar : an On-line Handwritten Character Recognition System.

Participant: Eric Anquetil.

- Company : PurpleLabs
- Software licence : INSA 3014

The collaboration with PurpleLabs company has started in 2001. PurpleLabs offers fully integrated solutions for handset designs (http://www.purplelabs.com). The ultimate goal of this collaboration is to implement an accurate handwriting recognizer into mobile devices with limited computing and memory resources. RESIFCar software (cf. section 5.1), developed by IMADOC Project, has been optimized to achieve this challenge. The hierarchical fuzzy modeling of RESIFCar offered a compact and robust knowledge representation. Moreover the decision process that is based on adapted fuzzy inference systems has allowed a reduction of the computing requirements without involving a decreasing in the performance.

Today several mobile phones design by PurpleLabs integrate RESIFCar software so that user can communicate directly with his smart phone by writing on the tactile screen of the mobile device using his personal handwriting.

7.2 Evaluation of the RESIF technology : Integration into Smartphone Device and Extension of the On-line Handwritten Word Recognition System to Large Word Vocabulary.

Participants: Eric Anquetil, Sabine Carbonel.

- Company: France Telecom R&D
- Contract: INSA 3012, INSA 3015

A collaboration with France Telecom R&D company started in 2003. The aim of this research collaboration is to evaluate RESIF technology according to two main axes:

- The first axis consists in the study of the integration possibilities of the on-line handwriting recognition systems RESIFCar and RESIFMot (cf. section 5.1) into smartphone devices;
- The second axis focuses on the task of lexicon post-processing (cf. section 6.8) to deal with large word vocabulary in the context of limited resources. The aim of this work is to explore the combination of different approaches to reduce time computing and memory resources involved in lexicon post-processing.

7.3 Automatic Context Modelling for On-line Pen-Based Interface Design.

Participants: Eric Anquetil, François Bouteruche.

- Company: France Telecom R&D
- Contract: INSA 4016

This new collaboration with France Telecom R&D company started in October 2004. We work on the interpretation of pen input (electronic ink) as captured by pen based interfaces (cf. section 4.3). The interpretation of handwritten shapes requires the knowledge of their specific context : the recognition process is driven by the context. The aim of this research is to study the possibilities of an automatic context modelling and its impact on the interpretation process. The first considered approaches are based on fuzzy logic modelling and especially fuzzy decision trees.

7.4 Evaluation of Pen-based Interface Ergonomic Quality.

Participants: Eric Anquetil, François Bouteruche.

- Company: Research Center in Psychology, Cognition and Communication of the University of Rennes 2
- Contract: INSA research collaboration

We have worked in collaboration with the Research Center Psychology, Cognition and Communication of the University of Rennes 2 since January 2004. The aim of this research collaboration is to validate experimentally the quality of the pen-based interfaces for handwriting input on mobile devices. This year, the study focuses on the impact of two components on user's performances and satisfaction(cf. section 6.5): the spatial contiguity respect between the visual feedback area and the input area and the conservation of the previously written character in the input area.

7.5 Musical score Pen-based Editor.

Participants: Eric Anquetil, Sébastien Macé, Francois Bouteruche.

- Company: Research department in Arts, Humanities and Communication of the University of Rennes 2
- Contract: INSA research collaboration

The collaboration with the research department in Arts, Humanities and Communication of the University of Rennes 2 started in January 2004. The aim of this research collaboration is to design an efficient pen-based editor for writing musical score. This collaboration with experts on music has conducted this years to the development of a first operational prototype of a musical score pen-based editor.

7.6 Access to Handwritten Archives Documents

Participants: Bertrand Coüasnon, Eric Anquetil, Jean Camillerapp, Ivan Leplumey, Grégory Maitrallain, Isaac Martinat.

- Partner : *Conseil Général des Yvelines, Archives des Yvelines*
- Contract : INSA 3009, INRIA 1 02 C 0602 00 31406 01 1

This research project started 2002 and is planed until October 2007. It is about making handwritten archives documents accessible to public. In collaboration with the *Archives des Yvelines*, Imadoc is working on three axes:

- document structure recognition and handwriting recognition to automatically produce annotations. These annotations offer the possibility to access by handwritten content to archives documents;
- a web platform to manage images, automatic annotations and collective annotations made by readers;
- a pen-based interface build on this platform to offer a new way of interacting with digital documents and paper documents.

The platform for image document annotations (section 5.5) with an automatic access by handwritten last names on military forms (section 6.1) is available in the reading room of the *Archives départementales des Yvelines*. Readers, after retrieving the right document by a request on handwritten names are able to add collective annotation on it. For the moment 28,000 pages are open to public. During 2005, 150,000 pages will be available on Internet.

The pen-based interface of the platform is also running on two reading desks of the future. These prototypes are made of a graphical tablet built in a wooden desktop, and a digital pen to interact with the interface.

This system has been officially presented during the inauguration ceremony of the new building of the *Archives départementales des Yvelines* on January, 17.

Other new results on this project are presented in sections 6.1 and 6.2.

8 Other Grants and Activities

8.1 National initiatives

8.1.1 ACI Madonne (Ministry Grant)

Imadoc is involved in the ACI (*Action Concertée Incitative*) Madonne (*MAsse de DOnnées issues de la Numérisation du patrimoiNE*). This project is about large data set produced by digitizing cultural heritage. Partners of this ACI are: L3I (La Rochelle), PSI (Rouen), LI (Tours), Loria (Nancy), LIRIS (Lyon), Irisa (Rennes).

8.1.2 RTP

Participation of Imadoc to the *Réseaux Thématiques Pluridisciplinaires* :

- RTP-DOC (STIC) :
  - AS “Numérisation et valorisation des collections” : Jacques André, Bertrand Coüasnon
  - AS “ Modèle(s) de publication sur le web” : participation of Jacques André and managing of the *AS* by Irisa.

8.2 International initiatives

Guy Lorette is in charge of international relations at IFSIC, and member of the international relations group of the Rennes 1 University. He was also the French deleguate at *Governing Board* of the IAPR *(International Association of Pattern Recognition)* and a member of the *Constitutions and Bylaws Committee.*

9 Dissemination

9.1 Leadership within scientific community

9.1.1 Editorial board

Members of the team are in editorial board of:

- *Document numérique*,
- *Les Cahiers GUTenberg*,
- Encyclopaedia *Techniques de l'ingénieur – Informatique* ;
- *Typography series* from Addison-Wesley.

9.1.2 Chairman

Members of the team are chairman of:

- CHIR 2004 (Colloque Histoire de l'Informatique)
- IWFHR'9 (International Workshop on Frontiers in Handwriting Recognition), 2004.

9.1.3 Programme comittee

Members of the team are in Programme comittee of:

- 2nd International Conference on Typography, Thessaloniki (Greece), june 2004,
- WEDELMUSIC 2004 (International Conference on Web Delivering of Music),
- CHIR 2004 (Colloque Histoire de l'Informatique),
- CIFED 2004 (Conférence Internationale Francophone sur l'Ecrit et le Document),
- ICPR 2004 (International Conference on Pattern Recognition),
- EuroTEX'2005 (TEX European Conference), March 2005,
- GREC 2005 (International Workshop on Graphics Recognition),
- IGS 2005 (International Graphonomics Society)
- IWFHR'9 (International Workshop on Frontiers in Handwriting Recognition), 2004,

9.1.4 Organizing comittee

Members of the team are in organizing comittee of:

- IWFHR'10 (International Workshop on Frontiers in Handwriting Recognition), 2006.

9.1.5 Member of scientific society

Jacques André is a member of the board of directors of the society GUTenberg.

Guy Lorette is :

- member of the scientific council of PSI (Université-INSA de Rouen) and of E3I computer laboratory (E3I-Université de Tours);
- vice president in load of international relations of GRCE;
- delegated at IAPR by AFRIF : ¡¡ Association Française de Reconnaissance et d'Interprétation de Formes ¿¿, de son CA ;
- member of the ASTI society.

Eric Anquetil, Bertrand Coüasnon, and Guy Lorette take part in the working group ¡¡ 5.2-Écrit ¿¿, topics 5: Communication, of GDR-PRC I^3 (Information, Interaction, Intelligence).

Jean Camillerapp, Guy Lorette, Eric Anquetil, and Bertrand Coüasnon take part in the activities of the society GRCE : ¡¡ Groupe de Recherche en Communication Écrite ¿¿.

9.1.6 Forum list

Jacques André has the responsability of the list *typographie* (`sympa@irisa.fr`), stored at URL : `https://www.irisa.fr/wws/info/typographie`.

9.2 University education

The team is mainly made up of teachers and those are very implied in activities of teaching. But a majority of lectures are not rattached to this research topic, so they are not quoted here.

- Guy Lorette gives lectures at DEA *d'informatique* and DESS Mitic at l'Ifsic, University of Rennes 1.

- Hélène Richy is in charge of the DESS Mitic (*Méthodes Informatiques et Technologies de l'Information et de la Communication*) at Ifsic, University of Rennes 1 and gives lectures in this cursus.

9.3 Participation to conferences, seminars, invitations

- Jacques André gave invited lectures at the International Conference on Typography (Thessaloniki, Greece, June 2004) and at the International Conference on Unicode and Amazighe Writing System (Rabat, Marocco, september).

- Bertrand Coüasnon presented research results at the official opening ceremony of the new building of the *Archives départementales des Yvelines* on January, 17

- Bertrand Coüasnon did a tutorial and a presentation at *Numérisation et Patrimoine* during SDN'04 (*Semaine du Document Numérique*), June 2004

- Nicolas Ragot and Éric Anquetil were invited to present their research works at the IAPD seminar at the LIP6, Paris (march 2004)

- Éric Anquetil and Guy Lorette have piloted the IRISATECH meeting on pen-based computers and interfaces (end of october 2003)

- Solen Quiniou presented works on the use of linguistic knowledge with statistical language models during the Journée des Jeunes Chercheurs organized by the S.A. 5.2 of the GDR I3 which took place in La Rochelle on June 2004

10 Bibliography

Major publications by the team in recent years

[1] J. André, M.-A. Chabin (editors), *Les documents anciens*, Numéro spécial de *Document numérique*, vol. 3, n°1-2, Hermès, 1999.

[2] J. André, H. Hudrisier (editors), *Unicode, écriture du monde ?*, Numéro spécial de *Document numérique*, vol. 6, n°3-4, Lavoisier, 2002.

[3] E. Anquetil, G. Lorette, "Automatic generation of hierarchical fuzzy classification systems based on explicit fuzzy rules deduced from possibilistic clustering: application to on-line handwritten character recognition", *in : Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, (IPMU'96)*, p. 259–264, Grenade, Spain, July 1996.

[4] B. Coüasnon, J.-P. Dalbéra, H. Emptoz (editors), *Numérisation et patrimoine, Numéro spécial de Document Numérique*, *7*, 3-4, Hermes-Lavoisier, 2003.

[5] B. Coüasnon, "DMOS: A Generic Document Recognition Method, Application to an Automatic Generator of Musical Scores, Mathematical Formulae and Table Structures Recognition Systems", *in : ICDAR, International Conference on Document Analysis and Recognition*, p. 215–220, Seattle, USA, September 2001.

[6] V. P. d'Andecy, J. Camillerapp, I. Leplumey, "Analyse de partitions musicales", *Traitement du Signal 12(6)*, 1996, p. 653–661.

[7] G. Lorette, E. Anquetil, "Théorie des catastrophes, géométrie différentielle et segmentation de l'écriture cursive", *in : Actes du 4ème Colloque National sur l'Écrit et le Document, (CNED'96)*, p. 1–6, Nantes, France, July 1996.

[8] G. Lorette, J.-P. Crettez, "Reconnaissance de l'écriture manuscrite", *in : Traité informatique*, Techniques de l'Ingénieur, 1998, ch. H 1358, p. 1–15.

[9] G. Lorette, "Handwriting recognition or reading? What is the situation at the dawn of the 3^{rd} millenium?", *International Journal on Document Analysis and Recognition 2*, 1, 1999, p. 2–12.

[10] L. Pasquer, Éric Anquetil, G. Lorette, "Coherent Knowledge Source Integration through Perceptual Cycle Framework for Handwriting Recognition", *in : Advances in Handwriting Recognition*, S.-W. Lee (editor), World Scientific, 1999, ch. 5 in PART2 : On-Line Handwriting Recognition, p. 59–68.

[11] J. Pettier, J. Camillerapp, "Script representation by a generalized skeleton", *in : Proceedings of the 2nd International Conference on Document Analysis and Recognition, (ICDAR'93)*, p. 850–853, Tsukuba, Japan, October 1993.

Books and Monographs

[12] *Actes du septième colloque d'histoire de l'informatique – Cesson, novembre 2004*, INRIA-Rennes, 2004.

Articles in referred journals and book chapters

[13] J. André, *Histoire d'Œ, histoire d'@ – rumeurs, typographie et enseignement*, INRP, octobre 2004, p. 19–34.

[14] B. Coüasnon, *Dealing with noise in DMOS, a Generic Method for Structured Document Recognition: an Example on a Complete Grammar*, *LNCS*, *3088*, Springer-Verlag, 2004, p. 38–49.

Publications in Conferences and Workshops

[15] J. André, "IBM 1620 – un mini avant la lettre – et les universités françaises", *in: Actes du septième colloque d'histoire de l'informatique – Cesson, novembre 2004*, p. 171–172, Rennes, France, novembre 2004.

[16] J. André, "Préhistoire de l'informatique à l'Université de Rennes (1954-1970)", *in: Actes du septième colloque d'histoire de l'informatique – Cesson, novembre 2004*, p. 228–236, Rennes, France, novembre 2004.

[17] E. Anquetil, F. Bouteruche, "Conception d'un micro éditeur d'encre électronique et embarquement d'un système de reconnaissance d'écriture manuscrite sur téléphone mobile", *in: Conférence Francophone: Mobilité & Ubiquité'04*, p. 151–157, Nice, France, june 2004.

[18] F. Bouteruche, E. Anquetil, "Interfaces stylo pour la saisie d'écriture manuscrite sur systèmes mobiles de petite taille", *in: 2ndes Rencontres Jeunes Chercheurs en Interaction Homme-Machine*, p. 39–42, Lacanau, France, octobre 2004.

[19] J. Camillerapp, L. Pasquer, B. Coüasnon, "Indexation automatique de formulaires anciens par reconnaissance du patronyme manuscrit", *in: RFIA, Reconnaissance des Formes et Intelligence Artificielle*, p. 1493–1502, Toulouse, France, janvier 2004.

[20] S. Carbonnel, E. Anquetil, "Apprentissage automatique d'une distance d'édition dédiée à la reconnaissance d'écriture manuscrite", *in: Colloque International Francophone sur l'Écrit et le Document (CIFED'04)*, *3*, p. 105–110, La Rochelle, France, juin 2004.

[21] S. Carbonnel, E. Anquetil, "Lexicon Organization and String Edit Distance Learning for Lexical Post-Processing in Handwriting Recognition", *in: 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR 9)*, p. 462–467, Tokyo, Japon, october 2004.

[22] S. Carbonnel, E. Anquetil, "Modélisation et intégration de connaissances lexicales pour le post-traitement de l'écriture manuscrite en-ligne", *in: 14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA'04)*, *3*, p. 1313–1322, Toulouse, France, janvier 2004.

[23] B. Coüasnon, J. Camillerapp, I. Leplumey, "Making Handwritten Archives Documents accessible to Public with a Generic System of Document Image Analysis", *in: International Workshop on Document Image Analysis for Libraries (DIAL'04)*, p. 270–277, Palo Alto, USA, January 2004.

[24] H. Mouchère, E. Anquetil, N. Ragot, "Étude des mécanismes d'adaptation pour l'optimisation de classifieurs flous dans le cadre de la reconnaissance d'écriture manuscrite", *in: 12es rencontres francophones sur la Logique Floue et ses Applications (LFA'04)*, p. 93–100, Nantes, France, novembre 2004.

[25] N. Ragot, E. Anquetil, "MELIDIS: Pattern recognition by intrinsic/discriminant dual modeling based on a hierarchical organization of fuzzy inference systems", *in: 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, (IPMU'04)*, *3*, p. 2069–2076, Perugia, Italy, juillet 2004.

[26] L. Rousseau, E. Anquetil, J. Camillerapp, "Reconstitution du parcours du tracé manuscrit hors-ligne de caractères isolés", *in: 8ème Colloque Internationnal Francophone sur l'écrit et le Document, (CIFED'04)*, p. 123–127, La Rochelle, France, juin 2004.

Miscellaneous

[27] B. Coüasnon, J. Camillerapp, I. Leplumey, "Making Handwritten Archives Documents accessible to the Public with a Document Image Analysis System", *in: DigiCULT.info Newsletter*, *8*, August 2004.