# Multiresolution cooperation makes easier document structure recognition

Aurélie Lemaitre, Jean Camillerapp, Bertrand B. Couasnon

**International Journal of Document Analysis and Recognition (IJDAR) manuscript No.**
(will be inserted by the editor)

**Aurélie Lemaitre · Jean Camillerapp · Bertrand Coüasnon**

# Multiresolution Cooperation Makes Easier Document Structure Recognition

Received: date / Accepted: date

**Abstract** This paper shows the interest of imitating the perceptive vision to improve the recognition of the structure of ancient, noisy and low structured documents. The perceptive vision, that is used by human eye, consists in focusing attention on interesting elements after having detecting their presence in a global vision process. We propose a generic method in order to apply this concept to various problems and kinds of documents. Thus, we introduce the concept of cooperation between multiresolution visions into a generic method. The originality of this work is that the cooperation between resolutions is totally led by the knowledge dedicated to each kind of document. In this paper, we present this method on three kinds of documents: handwritten low structured mail documents, naturalization decree register that are archive noisy documents from the 19th century and Bangla script that requires a precise vision. This work is validated on 86,291 documents.

**Keywords** Structure recognition · Multiresolution · Perceptive vision · Grammar

## 1 Introduction

The idea of imitating human vision has been long used for image processing [13]. In the field of document structure recognition, it seems particularly interesting to imitate the perceptive vision. This mechanism, used by the human vision, consists in combining a global vision of a document with focuses on points that attracts the eye interest.

We work on structure recognition of low structured documents and of damaged archive documents. Thus, we are faced with a large variety of structure and with the difficulties linked with archive documents: irregular printing or writing, noise, ink which bleed through paper.

The use of a multiresolution approach seems well adapted for this kind of document. Thus, having a first global vision before focusing on zones of interest is helpful when few information is available concerning the structure. Moreover, in a global vision, the influence of noise is reduced. Then, we can concentrate on the study of the organization of elements, without being disturbed by noise.

We propose to implement the mechanism of cooperation between multiresolution visions in the context of a generic method: DMOS. This method [9] is based on a two-dimension grammatical formalism: EPF, Enhanced Position Formalism. This language is used to describe the knowledge dedicated to a kind of document by expressing the logical, syntactic and even semantic organization of the document. Then, the associated parser is automatically produced by a compilation step. The main interest of this method is the separation of the knowledge from the recognition system. Working in the context of DMOS method offers the possibility to implement the mechanism of multiresolution vision in a generic way, and to apply it to many kinds of documents. Moreover, thanks to this method, the cooperation between resolutions is totally guided by the knowledge and specific to each problem.

In this paper, we will show how the multiresolution cooperation can improve the recognition of documents that present specific problems. We first present related work on the use of perceptive vision in image processing and the applications of multiresolution to improve the quality of document structure recognition. Then, we propose three kinds of cases where the structure recognition can be improved by the perceptive vision. Each one of these problems is illustrated by a specific kind of document. In order to solve these problems in a generic way, we introduce our work in the general DMOS method, presented in section 4. Then, we present the results obtained at a large scale (more than 86,000 images) on

Aurélie Lemaitre
IRISA/INSA
Campus de Beaulieu
35043 Rennes Cedex, France
Tel.: +33 2 99 84 75 39
Fax: +33 2 99 84 71 71
E-mail: alemaitr@irisa.fr

three kinds of documents, that shows how the use of a multiresolution cooperation improves document recognition.

## 2 Related work

### 2.1 Interest of perceptive vision for image processing

One of the first use of perceptive vision found in the literature of image processing is proposed by Bajcsy and Rosenthal [2]. Their idea is based on the observation that people do not look at a whole scene with the same intensity but focus visually on the objects where their attention is attracted. In order to use this mechanism, they set up a system based on a visual hierarchy and a conceptual hierarchy. The visual hierarchy is made of different sized images. The conceptual hierarchy expresses the idea that the knowledge varies according to the level of analysis. In this work, the link between the visual and the conceptual hierarchy is realized by a control structure that manages interactions between them. However, this work has only been applied for dedicated images.

The idea of imitating the human vision has been studied by Burt who proposes in [4] the Pyramid Vision Machine. This system is based on a fine-to-coarse algorithm to generate a Gaussian pyramid of images. Then this images are used in a coarse-to-fine strategy that rapidly locates objects within a scene. This mechanism is controlled by a high level mechanism that guide data gathering even as visual interpretation is being interpreted. However their method is dedicated to object recognition for smart surveillance and mechanisms of alerting.

In 1987, Dyer synthesizes in [12] the key points of a multiscale analysis that are the persistence of a property over a range of scales, the possibility to make a coarse-to-fine search for a given feature, the detection of global content using operations at coarse scales and the hierarchical organization linked with perception. He presents low level mechanism for analyzing multiscale representations in image processing. The main application of this work is a model-based object recognition.

Silberberg [20] gives a framework for the multiresolution vision of objects. He stores data in a Multiresolution Symbolic Pixel Array that contains pixels properties, object properties and object spatial relationships for each resolution. He proposes an interesting way to interpret the image according to the multiresolution data. For each resolution, hypotheses are made about the presence of an element. Then, an interpretation is realized from the smallest resolution where this element has been detected. Successive focusing of attention leads to confirm or not the initial hypothesis. The philosophy of this work is very interesting but it has been applied only on two images: a submarine image and an airplane image. Thus, the implementation seems very dedicated to those two images.

Jolion and Rosenfeld [13] give an overview of frameworks for multiresolution computer vision, mainly based on pyramidal structures, used for image processing. However, these works are mainly dedicated to early vision treatments, and does not include any high level knowledge.

### 2.2 Usage of multiresolution in document structure recognition

In the field of document structure recognition, several works have been based on multiresolution. Bloomberg proposes in [3] a multiscale approach to detect shapes and textures. He builds the different resolutions using a morphological tool. The method is applied for font style identification but remains very close to low level image processing.

Another method based on low level analysis is proposed in [19]. The resolutions are built using a Dynamic Local Connectivity Map (DLCM), with a varying connectivity threshold. This method has been applied for the segmentation of column blocks and graphics in journals, magazines and books. As this method remains close to image processing, it seems difficult to include special knowledge required by more complex documents.

Déforges and Barba present in [11] the idea that at a certain resolution, text lines appear as regular strokes. Then they propose to build a pyramid of images, and to find in the pyramid the convenient image for text line detection. This approach has been applied for text line detection in postal documents. However, once the convenient resolution is found, there is no cooperation with the other levels. The process of perceptive vision is not totally exploited.

Cantoni *et al.* [5] build a pyramidal representation of images, made of four features maps: the average, the variance, the threshold and the median. Each feature map is built at four resolutions. Then, they apply three fixed processes: background analysis, graphics analysis, text analysis, based on the contents of the features maps. However, results found at low resolution are not directly used to analyze high resolution. Elements found at various resolutions are only compared to validate the segmentation. There is not really a cooperation between resolutions.

The work proposed in [7] is dedicated to newspaper pages. This work is a high level analysis that aims at extracting the logical decomposition of the documents into paragraphs, headings... This analysis is based on two steps: first extract blocks using a global vision, then detail each block at low resolution. The main limit of this work is the specialization for one kind of document.

A more generic method is proposed in [6]. The authors present a Bayesian approach for the segmentation of magazine pages into blocks, and their classification. The method is based on the modeling of the transition

probabilities between adjacent levels in the multiscale structure. This model has to be trained, which requires a quite homogeneous base of document. It seems difficult to include a specific knowledge for special kinds of documents.

### 2.3 Position of our approach

In the case of document structure analysis, we want to use a cooperation between resolutions, guided by high level knowledge about the structure, and to introduce this mechanism into a generic method in order to apply it to various kinds of documents.

We propose to combine in our method several interesting points that can be found separately in the literature. Thus, our method is based on the idea of visual and conceptual hierarchies proposed in [2]. We use as data a pyramid of images, built like [4] by a fine-to-coarse algorithm. We apply the idea from [7] that a high level analysis makes it possible the introduction of a larger knowledge.

Moreover, we realize an implementation that is independent of the studied kind of document, in order to apply our work to various problems. At last, we propose a global approach that makes it possible to indifferently treat handwritten or printed documents with various structures.

## 3 Intuitive need of multiresolution vision

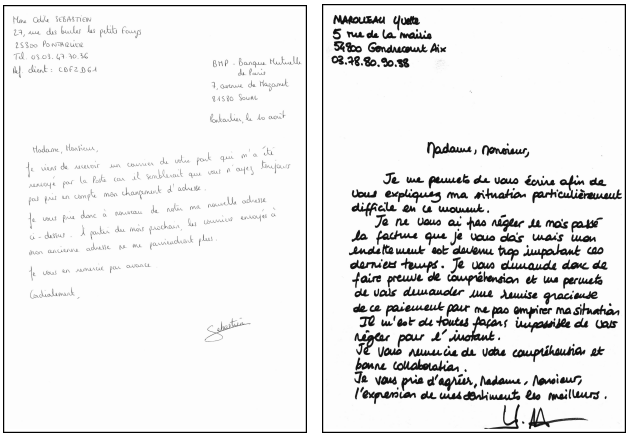
We turn our attention to different problems: low structured documents, noisy archive documents, and more generally the study of elements that are not directly perceptible in the original resolution. We present how the use of multiresolution vision could improve document recognition for all these problems.

### 3.1 Low structured documents

We call low structured documents the ones which structure is not materialized by any line or regular block. However, the human eye is able to detect instantly the whole organization of such a document, having a global view. Thus, we propose to use the notion of perceptive vision to show how the cooperation between a global view and detail analysis can simplify the structure detection.

Handwritten mail documents are an example of low structured documents. In those documents, the goal is to extract text blocks as sender details, addressee details, date and place and so on. However, those blocks are not always present and their disposition can vary (figure 1) even if each tongue gives rules of structuring for the writer. The perceptive vision gives a global vision of the text lines and of the blocks.

Then, the task consists in classifying these blocks. The main difficulty is that the logical segmentation depends on the classification. Using a generic system makes it possible to introduce knowledge that is dedicated to each language.



(a) (b)

**Fig. 1** Examples of handwritten mail documents

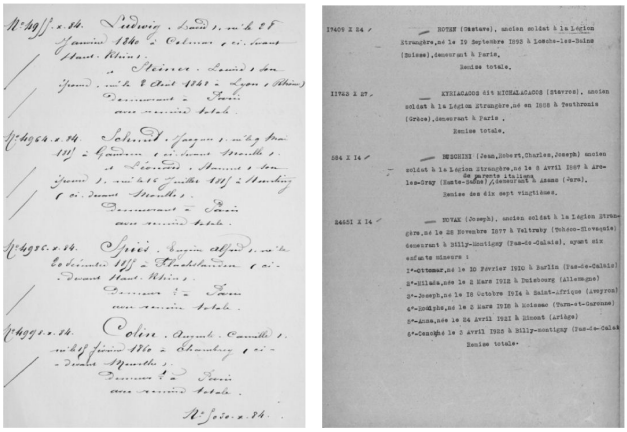
### 3.2 Noisy printed and handwritten documents

We consider the case of noisy printed and handwritten documents. We call noise the elements that come from the bad quality of the image and more generally all the information that is outside of the recognition system vocabulary. For example, archive documents gather several problems for structure recognition: bleed through ink, overlapping text lines, spotty documents. The local analysis based on connected components is difficult because connected components are biased by overlapping lines and noise. However, having a global vision of the document makes it possible to decrease the influence of noise in the document.

For example, we study naturalization decree registers from national French archives, that are printed or handwritten archive documents. Examples are presented in figure 2. The structure of the document is not materialized by any line, but it is stable for all the collection of documents. The goal of the analysis is to detect the decomposition of the page into successive acts.

The main difficulty with this kind of document is to deal with both printed and handwritten text, and to treat noise and damaged paper. We think that the use of perceptive vision makes it possible to consider equally printed and handwritten text in a global vision. Then, we can combine a global view that simplifies the detection of

the structure (separation of the margin and text body) and an accurate view that makes it possible to detail characters contained the words.



(a) Handwritten page from 1884

(b) Typewritten page from 1928

**Fig. 2** Example of naturalization decree pages from the national French archives

### 3.3 Search of absent headline

Another application of the perceptive vision if for the cases when the searched element is not perceptible at a high level, but only at low resolution. In that case, the presence of the element can be found at low resolution, and the high resolution vision makes it possible, in a second step, to find the exact position of the element.

We present as an example the case of Bangla documents (figure 3). Bangla is a language of eastern India. Bangla script recognition is complex, due to the large number of characters combinations, but can be improved by the detection of the headline (figure 4). The headline is a fictive line located in the upper part of the word.

This headline is not always perceptible at the original resolution but can be perceived in a global vision. Then, the use of perceptive vision makes it possible to extract the global slope of each text line, and a focusing makes it possible to determine the exact vertical position of each headline.

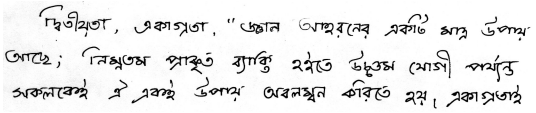


**Fig. 3** Example of Bangla document

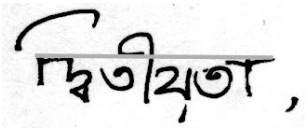


**Fig. 4** The headline of the first word of figure 3

## 4 General DMOS method

In order to implement this mechanism of multiresolution cooperation, and to easily describe the context of analysis for each kind of document, we introduce our work in the generic DMOS method.

### 4.1 Principle of the method

DMOS method [8](Description and MOdification of Segmentation) is a generic method for structured document recognition. It is made of a new grammatical language (Enhanced Position Formalism - EPF) and an associated parser able to deal with noise.

Once a description of a kind of document has been realized in EPF language, the associated parser is automatically produced by a compilation stage. This parser is bi-dimensional and presents the following properties:

- changing the parsed structure during parsing for contextual segmentation;
- detecting the next element to parse, anywhere in the image, using two dimension position operators;
- dealing with noise.

Moreover, this parser tries successively the different possible combinations until one succeeds.

We distinguish two levels of analysis in this method. The digital level is made of features that are extracted directly from the image, and that represent the terminals of the grammar. Those features are the connected components and the line segments. Our line segment extractor, presented in [18], is based on Kalman filtering. Its main properties are the ability to deal with dotted lines, curvature and skew, line segments running into each other.

The symbolic level contains the knowledge associated to each kind of document. It is made of a description of the organization of elements based on digital features, realized with the grammatical formalism EPF. EPF makes it possible a structural, syntactic and semantic two dimension description of any studied kind of document. In this generic method, the knowledge is grouped in EPF description and separated from the recognition system.

This method has been applied for several kinds of documents: music scores, mathematical formulae, table structures [8] or archive documents [15], which shows its genericity. It also has been validated at a very large scale, on more than 500,000 pages of documents.

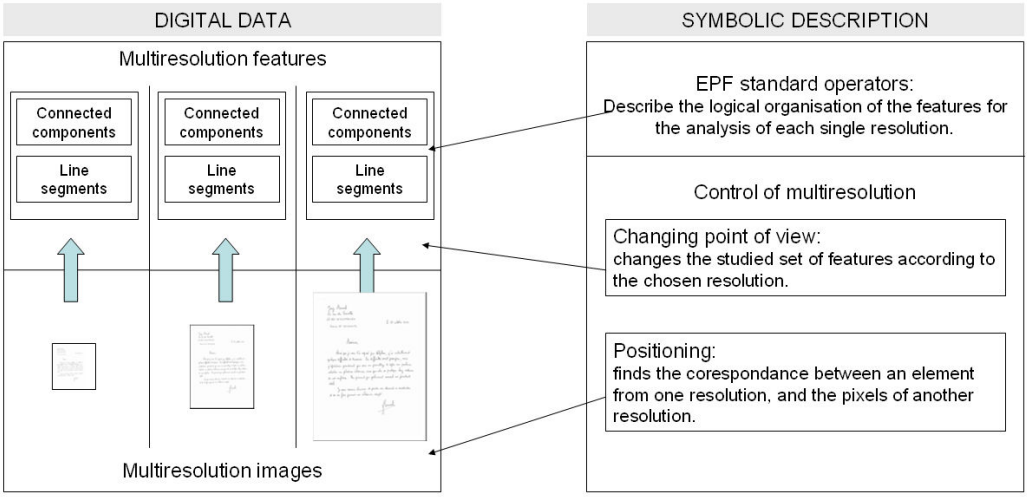



**Fig. 5** Overview of the mechanism of multiresolution cooperation in DMOS method

### 4.2 EPF language

EPF language makes it possible to describe the relation between features from the digital level. We give an example of description for the kind of document presented on figure 1(b).

```
mailPage::=
   AT(topLeftPage) &&
      senderDetails &&
   AT(middlePage) &&
      opening &&
   AT(underOpening)&&
      textBody &&
   AT(underBody) &&
      signature.
```

It means that a mail page is made of four elements: sender details, opening, text body and signature, that are disposed in the page at special relative positions. The positions are defined using the formalism below: let `A` and `B` be terminals or non-terminals of the grammar, and `&&` the concatenation operator.

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

means that we have `A`, and at the position `pos` in relation to `A`, we find `B`. The user can define as many position operators (like `pos`) as necessary. For example `topLeftPage` is a search zone. In this zone, we apply the rule `senderDetails` that looks for sender details, and so on to find the described elements in the page.

Each element, like `senderDetails` or `opening`, is detailed in a specific rule that independently describes it. The lower level of description is based on terminals, that are the digital features. The *terminal operators* are used to detect a connected component `Cmp` or a segment `Seg`. Their syntax is:

```
TERM_CMP PreCond PostCond Cmp
TERM_SEG PreCond PostCond Seg
```

They can accept pre or post conditions on the searched terminal, `Cmp` or `Seg`.

Here is a toy example. However, EPF language is very rich and offers complex mechanisms, for example to deal with noise in documents. More details on the language are presented in [9].

## 5 Multiresolution cooperation

DMOS method is very generic and can be applied to many kinds of document: it is well adapted to treat the various cases we presented in section 3. Consequently, we propose to introduce the mechanism of multiresolution cooperation in DMOS method. The figure 5 presents an overview of our work.

We improved the existing digital level to produce a multiresolution digital level, made of the digital features that are extracted from images at various resolutions. It means that the analysis is based on a set of levels of resolutions. Each level of resolution is made of an image and the associated connected components and line segments, extracted with a method based on Kalman filtering [18]. This features are the multiresolution terminals of the grammar.

The cooperation is controlled according to two axes: changing the point of view and positioning elements. Indeed, during the analysis, we have to change the point of view from one resolution to another. This is possible thanks to the *Changing point of view* operators. We must also set up a correspondence between elements that are seen in a resolution, and the associated pixels from another resolution. This is possible thanks to the *Positioning* tool.

We detail each element in the following sections.

### 5.1 Multiresolution digital features

In our improved version of DMOS method, the user can use as many resolutions as required for the description of a precise kind of document. The different resolutions are built successively using a low-pass filtering from the initial image. This initial image is also called resolution `1`. We call the resolution `-n` an image of which dimension have been divided by `n`. We obtain a pyramid containing several resolutions (figure 6). Thanks to the use of the low-pass filtering, this pyramid imitates the various resolutions that are perceived by the human eye.

Then, for each chosen resolution, we can apply our feature extractors for connected components or line segments [18]. Results of extractions are presented in figures 7 and 8. The features are presented on images of the same size. Indeed, even if they are extracted from various resolution levels, the features are all described in the same level of coordinates in order to simplify the cooperation between resolutions. Concerning the connected components, in this example, the extracted features slightly vary from one resolution to another. We can see that there is really few noise detected at low resolution as a connected component. The line segments are very rare in the upper resolutions, whereas they clearly correspond to text lines at low resolutions.

This shows that, depending on the kind of document, the user will have to choose what features at what resolutions are significant for the description of the structure of the document. However, we experimentally notice that the perceived features are stable, inside a set of documents, for a given resolution. Thus, the choice of the appropriate resolution is quite clear when watching the extracted features in a few documents of the collection. For example, if we want to work with line segments that represent text lines in the kind of document presented on figure 8, the resolution -16 seems obvisously the best.

We also experimentally observe that a step of at least four between resolution levels is convenient in order to observe significant differences between the images, while keeping a strong enough link between data perceived at successive resolutions.

The cooperation must be set up between at least one low resolution level (`-16` for example) and one high resolution level (initial image), in order to obtain a mechanism of perceptive vision based on both global features and local precise details. For example, we present in table 1 the digital features we used for the documents presented on section 3. An intermediate level is sometimes useful, depending on the kind of document.

In the case of the mail documents and of naturalization decree registers, we decide to use only the connected components extracted at high resolution (figure 7(b)) which correspond to characters, and line segments extracted at low resolution (figure 8(d)) which correspond to text lines. In the case of Bangla script, we also use the line segments extracted at high resolutions, that correspond to straight parts of characters.

| Resolution level | 1 | -16 |
|---|---|---|
| Mail documents | CC | Seg |
| Naturalization decree | CC | Seg |
| Bangla script | CC, Seg | Seg |

**Table 1** Resolution levels and digital features ( CC: connected components, Seg: line segments) used for each kind of document

Once the choice of the useful resolutions is done for a kind of document, a cooperation must be set up between those resolutions to implement the mechanism of perceptive vision. This cooperation is totally led by the symbolic level.

### 5.2 Control of multiresolution

The cooperation is based on digital features extracted from the different resolutions. The symbolic level of DMOS method, made of EPF formalism, must set up a cooperation between those elements. Our objective is to keep the possibility to change the point of view of the analysis, that is to say the resolution, whenever we want in the description of the document. Thus, the symbolic level totally supervises the mechanism of perceptive vision. Moreover, we propose a tool to set up a correspondence between elements that are seen in a resolution, and the associated pixels from another resolution. This points are detailed in the following sections.

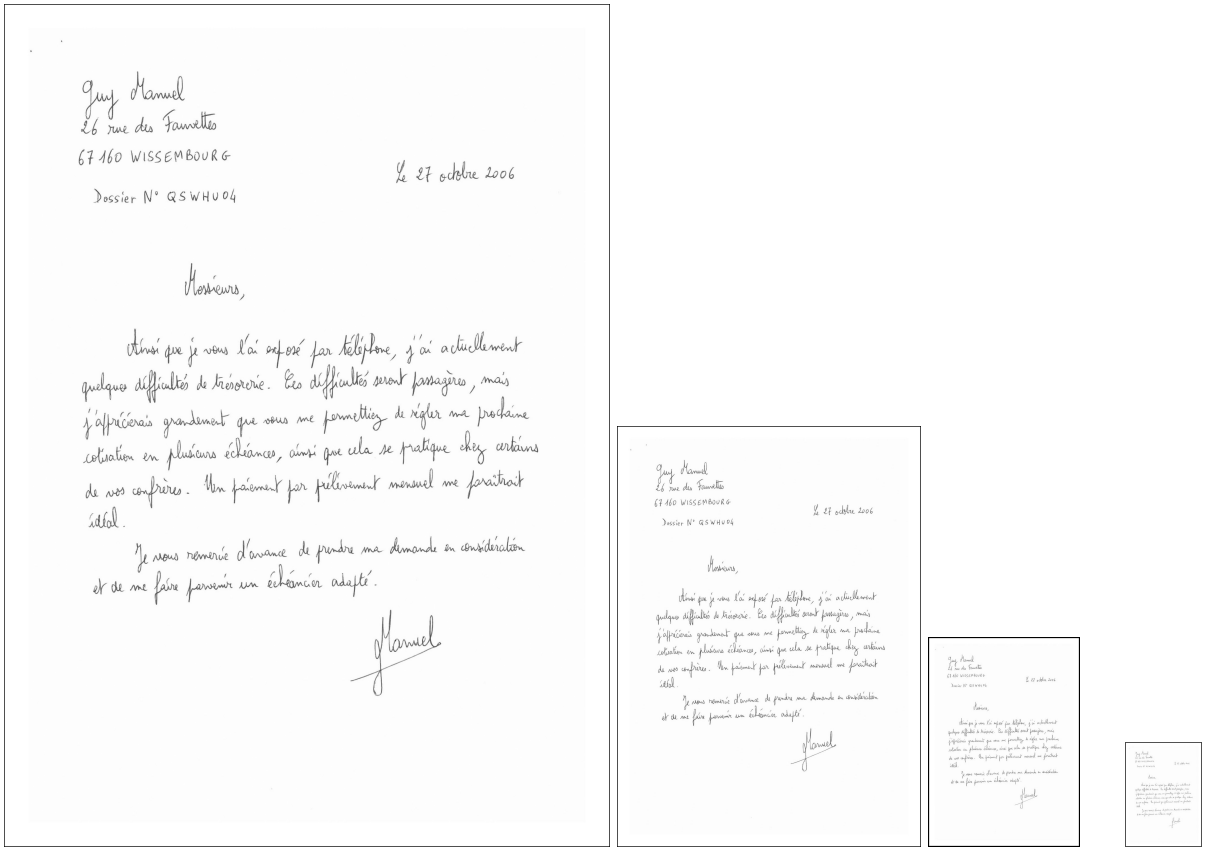
#### *5.2.1 Changing point of view*

The multiresolution vision requires to change the point of view during the analysis. Thus, we propose to begin the analysis at the lowest resolution and give the possibility to focus on an interesting zone for detailing it. We create a new operator in EPF formalism:

```
A && FOCUSING ON(resol) FOR(B)
```

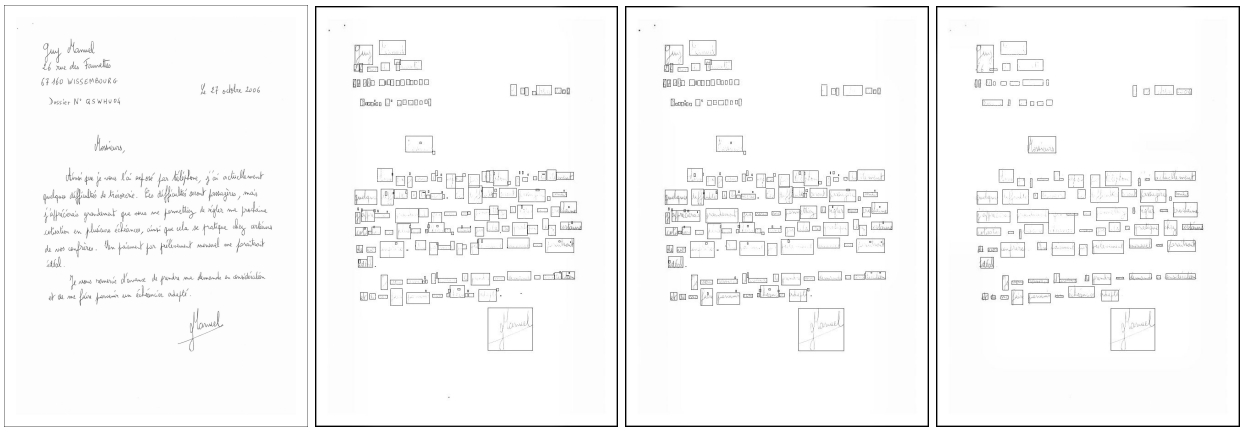
It means that at the current resolution, we find `A`, which is an element we want to detail. Then, we focus on resolution `resol`, in order to detect `B`, relatively to `A`. `A` and `B` can be terminals or non-terminals. `A` is based on features from the image at the current resolution. `B` is based on features from the image at resolution `resol`. The description of `B` can also be recursively based on other focuses.

For example, this operator is used for a first application: the vision of text line as line segment. Indeed, a text line can be seen at low resolution like a line segment, and at high resolution like a succession of connected components. We presented preliminary work on this topic in [14] and apply this concept for mail documents. Thus, we



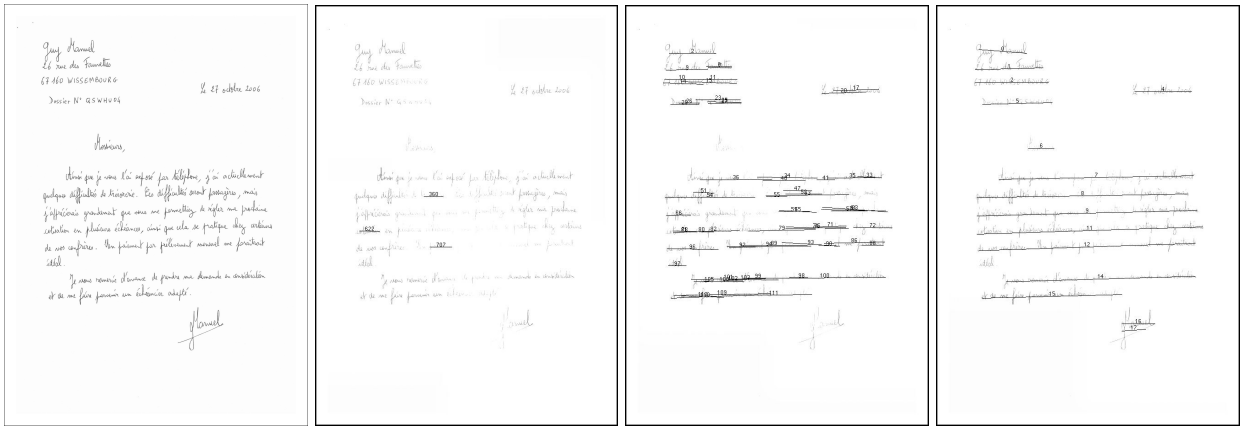
(a) Resolution 1 (b) Resolution -2 (c) Resolution -4 (d) Resolution -8

**Fig. 6** Multiresolution images built with low-pass filtering, from the initial image to the image divided by 8.



(a) Initial image (b) Resolution 1 (c) Resolution -4 (d) Resolution -16

**Fig. 7** Connected components extracted at various resolutions



(a) Initial image (b) Resolution 1 (3 line segments) (c) Resolution -4 (61 line segments) (d) Resolution -16 (15 line segments)

**Fig. 8** Line segments extracted with a method based on Kalman filtering, applied at various resolutions

extract each text line in two steps: first find a line segment at low resolution, and then focus attention at high resolution to detail connected components. This concept is translated in EPF language by the following rule. The analysis begins at low resolution.

```
textLine::=
  lineSegment &&
  AT(lineSegmentZone) &&
  FOCUSING ON(upperResolution)
    FOR(groupOfComponents).
```

This rule is applied on figure 9. The interest of using perceptive vision for text line detection is to determine a global position of the text line, without any problem of noise, and to use this position to locally assign the components of the text line. Morevover, as our line segment extractor is able to deal with skew, curvature, and line segments running into each other, our decription can face with curved, skewed, and overlapping text lines. Thanks to this line definition, we produce a description of mail documents (figure 1) as an arrangement of blocs made of text lines. Results are presented in section 6.1

The change of point of view can be employed several times in one description to produce a recursive navigation between resolution levels. The result is then obtained from the combination of features coming from different resolutions. This is employed for naturalization decree registers (figure 2). The global mechanism consists first in separating the page into a margin and a body, and then in detecting acts. Each act is made, in the margin, of a registration number, and in the text body, of a paragraph that begins with the surname of the concerned person. The margin extraction is realized at low resolution:

1. Detect text lines as line segments.

(a) Detection of *lineSegment* number 12 at low resolution

(b) Focusing at initial resolution on this zone and detection of *groupOfComponents*

**Fig. 9** Application of rule *textLine*

2. Deduce the position of the margin (figure 10(a)).

The recognition of each act requires a cooperation between resolutions:

1. Focus at original resolution in the margin, find a registration number (figure 10(b)) as a succession of aligned connected components.
2. Go back to low resolution, find the text line as line segment in the body, in front of the previous number.
3. Focus at original resolution, over the previous text line, detail connected components and deduce the position of the searched surname (figure 10(c)).

The successive navigation between resolutions represents an use of the perceptive vision. This can be easily implemented using EPF formalism and particularly the `FOCUSING` operator, with the following rules. The analysis of the page consists in finding a margin and then in extracting acts:

```
pageOfDecree::=
    margin && AT(topPage) && setOfActs.
```

Finding the margin requires to recognize the text lines as line segments. The analysis begins at low resolution.

```
margin::=
    AT(topPage) && setOfTextLine &&
    computeAverageMargin.
```

The set of text lines is extracted recursively with:

```
setOfTextLine::=
    TERM_SEG noCond noCond FoundSeg &&
    AT(underSeg FoundSed) &&
    setOfTextLine.
```

The set of acts `setOfActs` is found recursively and each act is described as below:

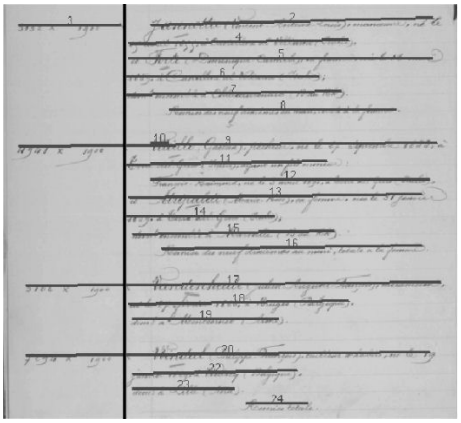
```
act::=
    AT(marginZone) &&
    FOCUSING ON(originalResolution)
        FOR(numberDetail Nb) &&
    AT(inFrontOfNumber Nb) &&
    TERM_SEG noCond noCond NameLine &&
    AT(nameLineZone NameLine) &&
    FOCUSING ON(originalResolution)
        FOR(nameDetail).
```

It is important to see that the analysis is realized at low resolution except for the elements that are contained in the operator `FOCUSING`. Thus, the progress of analysis depends on what is found successively at the convenient resolution.

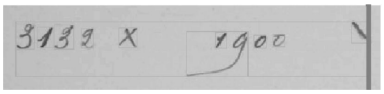
Thanks to the genericity of this method, the mechanism of focusing can be used for the description of many kinds of documents.

### *5.2.2 Positioning*

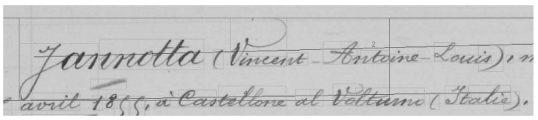
When a document is described at different resolution levels, it is sometimes difficult to find a correspondence between elements that come from various resolutions. This is the case for a text line seen as a line segment at low resolution: at high resolution, it corresponds to several



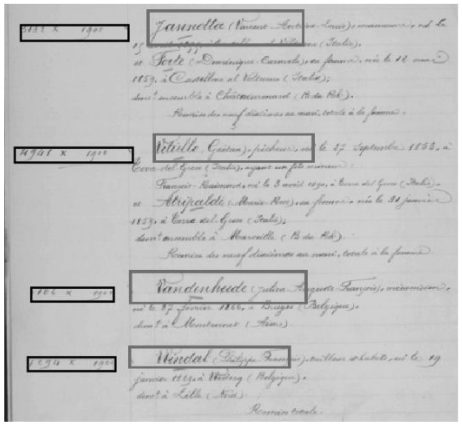
(a) Text lines at low resolution and the computed margin



(b) Focusing on margin for registration number



(c) Focusing on text line for name



(d) Final result: numbers and names

**Fig. 10** Analysis mechanism for naturalization decree registers

connected components, but the exact correspondence is not clear.

In the case of Bangla script, presented on figure 3, we want to find locally the exact position of the headline, whereas the headline is only perceived globally. Thus, we assume that a global vision of the text line, as a line segment, gives the global slope of each headline. In a second step, we have to find the exact position of the line segment, on the upper part of pixels. We propose to use the black pixels that are present in the image as an indication of the exact position of the headline. Thus, we set up a digital tool that makes it possible to locally position a line segment detected at low resolution on pixels that are present at high resolution. This is an adaptation of projection methods based on the position and the local slope of the line segment found at low resolution. However, the particularity is that we can follow the slope and the curvature of the line that has been detected at low resolution.

We introduce in EPF language the following tool:

```
POSITIONING_LINE
   LineSegment ResolutionName Position
```

This mean that, assuming we detected a `LineSegment` at low resolution, we want to translate it into the resolution `ResolutionName`. The parameter `Position` makes it possible to choose if we want a positioning at the top or at the bottom of words.

For example, the word on figure 11(a) is seen as a text line at low resolution. Then the symbolic level can order a digital positioning on existing black pixels, for example on the top of the word (figures 11(b) and 11(c)).

This tool is applied for Bangla documents, according to the following steps:

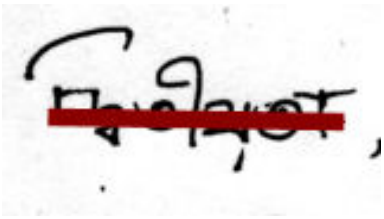
1. At low resolution, extract text lines as line segment (figure 12(a)).
2. At high resolution, group the connected components of each line into words (figure 12(b)).
3. For each word, call digital positioning to accurately place the headline from low resolution on the upper pixels of the high resolution image (figure 12(c)).

This mechanism is expressed using EPF formalism. Thus, for each word, we obtained a rule (simplified here):

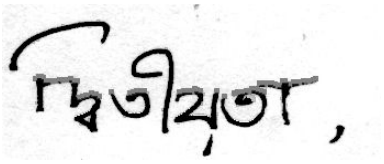
```
word::=
   lineSegment S &&
   AT(lineSegmentZone S) &&
   FOCUSING ON(UpperResolution)
      FOR(detailedHeadline S).
```

`S` is the line segment obtained by `lineSegment`. `S` is used as a parameter of `lineSegmentZone` and of `detailedHeadline`. The detailed headline is obtained by:

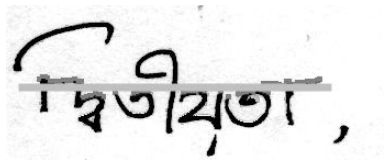
```
detailedHeadline S::=
   POSITIONING_LINE S UpperResolution Top.
```



(a) Line segment extracted at low resolution



(b) Digital positioning using upper black pixels (represented in grey)



(c) Result for a positioning on the top of the word

**Fig. 11** Digital positioning of a line segment on existing black pixels

We introduced the possibility to accurately position the elements using the digital level when more precision is required. The main interest is that the global vision makes it possible to reduce the amount of pixels involved at high resolution.

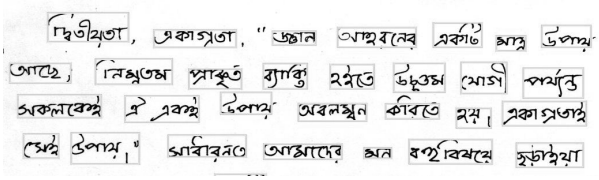
## 6 Application to different kinds of documents

We want to show how the use of multiresolution cooperation can improve document recognition. We introduced this mechanism into a generic method. Consequently, it can be applied for many kinds of documents. We validate our approach on three kinds of problems: the analysis of low structured documents (section 3.1) is validated with mail documents, noisy printed and hanwdritten documents (section 3.2) are studied with naturalization decree registers and the search of absent headline (section 3.3) is validated in Bangla handwritten text.

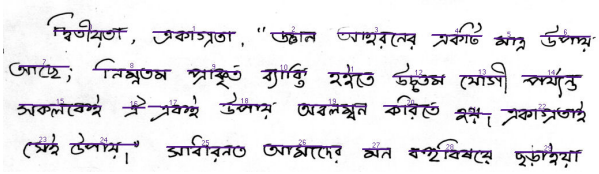
| Class | Involved pixels | Learning base 300 images | | Test base 850 images | | Whole base 1150 images | |
|---|---|---|---|---|---|---|---|
| | | Recall | Precision | Recall | Precision | Recall | Precision |
| Text body | 61.5% | 98.0% | 96.8% | 97.0% | 97.0% | 97.2% | 96.9% |
| Sender details | 15.1% | 93.4% | 93.0% | 91.5% | 91.7% | 92.0% | 92.1% |
| Addressee details | 9.0% | 87.3% | 85.5% | 83.0% | 83.3% | 84.1% | 83.9% |
| Signature | 4.2% | 84.4% | 90.2% | 90.6% | 90.9% | 88.9% | 90.7% |
| Subject | 4.0% | 68.4% | 70.5% | 65.7% | 72.7% | 66.4% | 72.1% |
| Date, Place | 3.2% | 53.1% | 79.1% | 54.3% | 78.2% | 54.0% | 78.4% |
| Opening | 2.9% | 85.8% | 80.4% | 79.5% | 74.5% | 81.1% | 76.0% |
| Global | - | 92.6% | 92.7% | 91.4% | 92.6% | **91.7%** | **92.6%** |

**Table 2** Results on 1,150 handwritten mail documents

(a) Text lines at low resolution



(b) Focusing to group connected components into words



(c) Global results

**Fig. 12** Headline detection in Bangla script

### 6.1 Mail document structure recognition

We studied mail documents in the context of the national French project RIMES [1] [16]. RIMES stands for "Reconnaissance et Indexation de données Manuscrites et de fac similES", which means "handwritten data and fax recognition and indexing". This project is financed by the French Ministries of Research and Defense. Its goal is to build a large handwritten document database (more than 5000 mail documents), to define criteria and metrics for the evaluation campaign using those data, and to drive this campaign. Nine French research teams took part in this project as participants on different tasks linked with document recognition and retrieving.

We took part on the task of mail structure recognition. For the first campaign, the base supplied by RIMES is made of 1,150 French mail images for learning systems. Images have a size around 2500*3500 pixels. They have been manually annotated. Indeed, each zone to label is represented by the coordinates of a rectangle and its associated class. Annotations are stored as a list of boxes in an XML file for each image, called *validation*. The objective is to produce as a result a list of labeled box describing the document, in an XML file called *hypothesis*. Results are computed by comparison of *validation* and *hypothesis* files.

The metric is based on a binary image. The recognition rate correspond to the number of black pixels that have been correctly classified. In this case, *Recall* and *Precision* are defined below:

$$Recall = \frac{NbCorrectPixelForTheClass}{NbExpectedPixelForTheClass}$$

$$Precision = \frac{NbCorrectPixelForTheClass}{NbFoundPixelForTheClass}$$

We use two resolutions: initial images have a size around 2500*3500 pixels, low resolution images have a size around 150*320 pixels.

We detail results obtained with our method in table 2. Results are presented for three bases. Indeed, we used 300 images of 1150 as a kind of learning base, as we partly watched them to produce the rules of our system in EPF language. The whole base is made of 1,150 images.

We obtain a global recognition rate (recall) of 91.7% for the base of 1150 images, which correspond of the percentage of pixels correctly labeled, out of the 254,207,489 black pixels we have to label. The global precision is 92.6%. The second column gives the importance, in number of pixels, of the classes. "Text body" represents the main part, 61.5% of pixels. We can see that recognition rate is globally proportional to the importance of the class. "Subject" recognition is difficult because it can be easily confused with several other classes: "Date, Place", "Sender details" or "Opening". Improving these three classes would automatically improve "Subject" recogni-

tion rate. It is important to see that the results on the test base are close to the results on the learning base.

The official results of RIMES contest have not been published yet, but our method with multiresolution obtained the best results for the first test in 2007.

### 6.2 Naturalization decree register treatment

The analysis of naturalization decree registers aims at extracting register numbers and surnames for each act contained in the document. This work has been realized in collaboration with the French national archives. We applied our method on 15,699 registers, dated between 1883 and 1930, which represents 85,088 pages. Initial images were at a resolution of 240 dpi (image size around 2000*3000 pixels) and stored in JPEG. The so called *low resolution* was 15 dpi (about 120*190 pixels). We detected on this base 433,230 acts {number, surname} (5 per page on the average).

| | Monoresolution | Multiresolution |
|---|---|---|
| Number of pages | 347 | 347 |
| Number of acts | 3,186 | 3,186 |
| Recognition rate | 92.69% | **98.31%** |
| Average time/act | 6.4 s | 1.2 s |

**Table 3** Recognition rate on a representative validation base: better results with multiresolution

The validation base is built with taking one image out of 250, in the chronological order. Thus, it is representative of the ratio handwritten/printed and of the different problems of the whole base.

In a previous work [10], a recognition system was made for naturalization decree registers *without* multiresolution. For example, the margin research was based on the detection of globally aligned connected components, which was sometimes vague and very sensitive to noise. Moreover, this method was adapted to handwritten documents. When we applied it to the *representative* base, we only obtained a rate of 92.69% recognition (table 3). Indeed, this method could not be generic enough to deal with both handwritten and printed documents.

Using multiresolution, we really improved the recognition, mainly thanks to the generic aspect of our description that is not restrictive to handwritten documents. Thus, we obtain a recognition rate of 98.31% for the *representative* base, instead of 92.69% with the previous version. This is possible due to the ability to equally extract handwritten or printed text lines as line segments at low resolution, with a better vision than at initial resolution. Moreover, working at low resolution decreases noise and a global structure can be easily extracted. Then, this extracted structure is a strong support for high resolution analysis. At last, even if it was not the first goal, we reduced the execution time by 85%. Indeed, the use of multiresolution decreases the number of hypothesis that have to be studied by the parser.

This method has been evaluated on a larger base of 2666 pages, and obtains a recognition rate of 98.63% for the analysis of 19,788 acts with the multiresolution. This result correspond to those presented in table 3. Due to the quality of the results, this method has been applied on more than 85,000 pages for a collaboration with French national archives.

### 6.3 Headline localization in Bangla handwritten text

As presented in section 3, the work on Bangla documents aims at extracting headlines in words, as a preprocessing tool for handwriting recognition.

We applied our mechanism on 53 document images of handwritten Bangla text pages, from 26 different writers, who had to write the same text. Initial images were at a resolution of 300 dpi, with an initial size around 2000*3200 pixels for half documents and 2000*1900 pixels for other part of documents. The so called -8 resolution corresponds to 38 dpi images (image size around 250*400 or 250*240 pixels) and resolution -16 corresponds to 19 dpi images (image size around 125*200 or 125*120 pixels).

We proposed two method for headline extraction: the first method, detailed in [17], is an application of the cooperation between resolution, without positioning. The second method, presented in section 5.2.2, is based on the digital positioning.

We manually estimate results on 26 pages, one per writer. Results are presented in table 4.

| | No positioning | With positioning |
|---|---|---|
| Words | 2114 | 2114 |
| False headline | 198 | 107 |
| Recognition rate | 90.6% | **94.9%** |
| Average time/word | 1.2 s | 0.1 s |

**Table 4** Positioning improves recognition rate and running time

The main interest of the multiresolution cooperation is the ability to have first a global view before text recognition. This is the key point for skewed images and varied slope inside a single document, which is common for handwritten data. Indeed, the slope is computed independently for each line. Thanks to the positioning tool, we obtain a correct extraction of the headline in 94.9% cases for 2114 detected words. We can see that the digital positioning decreases the error rate of 46% and the running time of 78%. The extracted headline can be used as a preprocessing for handwritting recognition. That is why it is necessary to have a precise positionning at high resolution (300 dpi).

## 7 Conclusion

This works shows that using a cooperation between multiresolution visions for document structure recognition makes it possible to combine a large vision in order to detect a global structure and to reduce the importance of noise, and a close vision that gives the precise details of elements. Thus, the precise detection of the structure will make it possible to apply locally dedicated systems like handwriting recognition. This precision is particularly important when handwritten text crosses structural elements (cell borders for example).

We introduced the multiresolution in the generic DMOS method. The originality of our approach is that the cooperation between resolutions is entirely led by the grammatical description of symbolic level. Moreover, this cooperation can be bidirectional: it may combine a low-to-high level and a high-to-low level of analysis. Thanks to this generic context of DMOS method, we applied our work to three kinds of documents with various problems and objectives.

We show that the multiresolution cooperation is very useful for low structured documents, and noisy or damaged documents. Indeed, in the case of naturalization decree register, the use of a multiresolution approach improves the recognition rate from 92.7% to 98.3%, and divides the execution time nearly by 6. In this kind of document, the multiresolution cooperation increases the reliability to deal with noise. Moreover, using multiresolution vision makes it possible to detect at low resolution elements that are not perceptible at high resolution. The digital positioning has shown its interest where a strong precision was asked: we studied the case of headline position in Bangla script. We estimate that a correct headline is found in 94.9% cases in our base. Our method has been evaluated on a large scale with more than 86.000 documents.

As a conclusion, we introduced a multiresolution approach in a generic method and applied it on a large scale with various documents. Thus, we shown that the use of multiresolution cooperation gives more reliability in noisy documents, decreases running time and improves the recognition rate in various problems of structure recognition.

## References


1. Augustin, E., Carre, M., Grosicki, E., Brodin, J.M., Geoffrois, E., Preteux, F.: Rimes evaluation campaign for handwritten mail processing. In: Proceedings 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR06), pp. p.231–235. La Baule, France (2006)
2. Bajcsy, R., Rosenthal, D.A.: Visual and Conceptual Focus of Attention, pp. 133–149. Academic Press (1980)
3. Bloomberg, D.: Multiresolution morphological approach to document image analysis. In: ICDAR 1991, pp. 963–971 (1991)
4. Burt, P.J.: Smart sensing with a pyramid vision machine. Proceedings of the IEEE **76**, 1006–1015 (1988)
5. Cantoni, V., Cinque, L., Lombardi, L., Manzini, G.: Page segmentation using a pyramidal architecture. In: Workshop on Computer Architectures for Machine Perception, p. Session 6 (1997)
6. Cheng, H., Bouman, C.: Multiscale bayesian segmentation using a trainable context model. IEEE Transactions on Image Processing **10**(4), 511–525 (2001). URL citeseer.ist.psu.edu/cheng01multiscale.html
7. Cinque, L., Forino, L., Levialdi, S., Lombardi, L., Tanimoto, S.L.: Understanding the page logical structure. In: 1oth International Conference on Image Analysis and Processing (ICIAP 1999), pp. 1003–1008 (1999)
8. Coüasnon, B.: DMOS: A generic document recognition method to application to an automatic generator of musical scores, mathematical formulae and table structures recognition systems. In: Proceedings of International Conference on Document Analysis and Recognition (ICDAR'01), pp. 215–220 (2001)
9. Coüasnon, B.: DMOS, a generic document recognition method: Application to table structure analysis in a general and in a specific way. International Journal on Document Analysis and Recognition, IJDAR **8(2)**, 111–122 (2006)
10. Coüasnon, B., Camillerapp, J., Leplumey, I.: Making handwritten archives documents accessible to public with a generic system of document image analysis. In: International Conference on Document Image Analysis for Libraries (DIAL), pp. 270–277 (2004)
11. Déforges, O., Barba, D.: A fast multiresolution text-line and non text-line structures extraction. In: International Conference on Image Processing (ICIP), pp. 134–138 (1994)
12. Dyer, C.R.: Multiscale image understanding, pp. 171–213. Academic Press Professional, Inc., San Diego, CA, USA (1987)
13. Jolion, J.M., Rosenfeld, A.: A Pyramid Framework for Early Vision: Multiresolutional Computer Vision. Kluwer Academic Publishers, Norwell, MA, USA (1994)
14. Lemaitre, A., Camillerapp, J.: Text line extraction in handwritten document with kalman filter applied on low resolution image. In: Document Image Analysis for Libraries (DIAL'06), pp. 38–45 (2006). URL http://dx.doi.org/10.1109/DIAL.2006.41
15. Lemaitre, A., Camillerapp, J., Coüasnon, B.: Contribution of multiresolution description for archive document structure recognition. In: Proceedings of International Conference on Document Analysis and Recognition (ICDAR'07), pp. 247–251 (2007)
16. Lemaitre, A., Camillerapp, J., Coüasnon, B.: A generic method for structure recognition of handwritten mail documents. In: Document Recognition and Retrieval (DRR XV) (2008)
17. Lemaitre, A., Chaudhuri, B.B., Coüasnon, B.: Perceptive vision for headline localisation in bangla handwritten text recognition. In: Proceedings of International Conference on Document Analysis and Recognition (ICDAR'07), pp. 614–618 (2007)
18. Leplumey, I., Camillerapp, J., Queguiner, C.: Kalman filter contributions towards document segmentation. In: Proceedings of International Conference on Document Analysis and Recognition (ICDAR'95), pp. 765–769 (1995)
19. Shi, Z., Govindaraju, V.: Multi-scale techniques for document page segmentation. In: ICDAR '05: Proceedings of the Eighth International Conference on Document Analysis and Recognition, pp. 1020–1024. IEEE Computer Society, Washington, DC, USA (2005). DOI http://dx.doi.org/10.1109/ICDAR.2005.165
20. Silberberg, T.M.: Multiresolution aerial image interpretation. In: Image Understanding Workshop, pp. 505–511 (1988)