

Complex Handwritten Page Segmentation Using Contextual Models

Stéphane Nicolas Thierry Paquet Laurent Heutte
Laboratoire PSI – Université de Rouen
UFR des Sciences et Techniques
Avenue de l'Université, Technopôle du Madrillet
76801 Saint Etienne du Rouvray cedex
{Stephane.Nicolas, Thierry.Paquet, Laurent.Heutte}@univ-rouen.fr

Abstract

In this paper we address the problem of segmenting complex handwritten pages such as novelist drafts or authorial manuscripts. We propose to use stochastic and contextual models in order to cope with local spatial variability, and to take into account some prior knowledge about the global structure of the document image. The models we propose to use are Markov Random Field models. After a formal description of the theoretical framework of Markov Random Fields and the principles of image segmentation using such models, we describe the implementation of our model and the proposed segmentation method. Then we discuss the results obtained with this approach on the drafts of the French novelist Gustave Flaubert, for different segmentation tasks. In conclusion, an extension of this work towards the use of discriminative models is discussed.

1. Introduction

These last years improvements have been made in the field of handwriting recognition, especially in the context of industrial applications such as check reading, postal address recognition or form processing,... These applications have been mainly focused on word or phrase recognition [1]. However with the advance of digital technologies, numerous institutions are moving towards the use of digital document images rather than traditional paper copies of the original documents. This situation raises new needs for indexing and accessing to these numerical sources [2].

In this context of shared access to our cultural and historical heritage, the Bovary Project, a digitization program of the manuscripts of the famous novel "Madame Bovary" of Gustave Flaubert, aims at providing a numerical web edition¹ of the genesis of this novel, to browse the original manuscripts of Flaubert associated with diplomatic textual transcriptions respecting as much as possible the layout of the original manuscripts. Such a numerical edition will be of great interest for researchers in literary.

However the production of the textual transcriptions of the 4127 manuscripts that constitute the Bovary directory is a challenging task. Considering the state of the art of document image analysis techniques, as well as the extreme variability of Flaubert's drafts, full automation of the process cannot be envisaged yet.

For this reason a network of volunteers has been recruited. However, it is assumed that their work could be greatly facilitated thanks to the use of automatic document analysis techniques. This is why we investigate the use of such methods with the aim of applying them to archived handwritten documents. First goal is to identify the regions of interest such as marginal annotations or deleted paragraphs, by extracting the layout of the manuscripts and allowing further manual or automatic indexing using layout information, transcription production or text/image coupling for genetic edition.

Many methods dedicated to machine printed document segmentation have been proposed [3], but these methods cannot be directly applied to handwritten documents because of the spatial variability of handwriting. The few existing methods

¹ <http://www.univ-rouen.fr/psi/BOVARY>

dedicated to handwritten documents focus on a particular type of documents or a particular segmentation task (word or line extraction only). Furthermore these methods are based on a local analysis, and sometimes fail to find the correct solution. It is the reason why we propose to use a general formalism that could be adapted to different types of documents, and which takes into account some contextual information. Hidden Markov Random Field formalism has been retained for this purpose. Markov Random Field models have been widely used during the last twenty years for different tasks of image analysis such as denoising, restoration, binarization and segmentation. Little work has been done however on document image analysis, and as far as we know, never on handwritten documents.

We propose to use Markov Random Field for the task of complex handwritten document image segmentation, such as authorial drafts or historical documents, and we present here an application consisting in the segmentation of Flaubert's manuscripts into their elementary parts, namely: text lines, erasures, punctuation marks, inter-linear annotations, marginal annotations (just to mention the most important of them) or to detect area of interest such as text body, header, margins, footer, ... In the Markov Random Field framework, segmentation is addressed as an image labeling problem. This problem is solved using optimization techniques. In section 2 we describe the theoretical framework of Markov Random Fields, then in section 3 we present our implementation for authorial manuscripts segmentation and we discuss the results obtained in section 4. Then we propose in section 5 a discussion about the evolution of our segmentation system towards the use of Conditional Random Fields to cope with some limitations of Markov Random Fields, and we present our first preliminary works with this type of models. We conclude in section 6.

2. Theoretical framework

Each document image is considered to be produced by implicit layout rules used by the author. While these rules cannot be formally justified, it is however experimentally verified by literacy experts that Flaubert's manuscripts exhibit some typical layout rules characterized by an important text body occupying two thirds of the page and containing a lot of erasures; and a marginal area with some text annotations, as can be seen on figure 1.

As there exist some local interactions between these layout rules, a Markov Random Field (MRF) seems to be adapted to model the layout of a manuscript. Furthermore we deal with handwritten documents characterized by some local spatial variability in the layout, therefore a stochastic model appears to be well suited to cope with the spatial variability of layout elements.

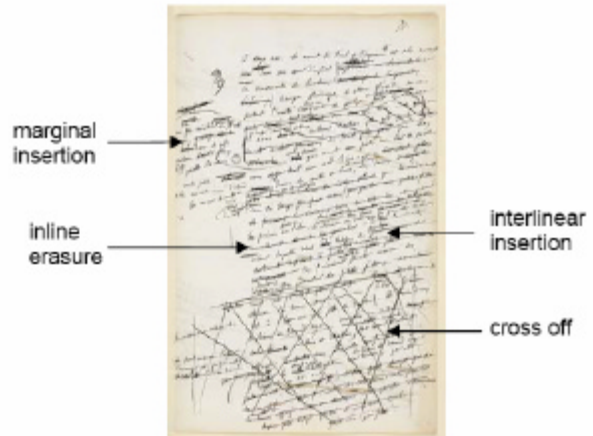


Figure 1. One example of Flaubert's manuscript layout.

According to MRF formalism [4], the image is associated with a rectangular grid G of size $n \times m$. Each site s , or location in the image, is associated to a cell on the grid defined by its coordinates over G and is denoted $g(i, j)$, $1 \leq i \leq n$ $1 \leq j \leq m$. The site set is denoted $S = \{s\}$.

Following the stochastic framework of Hidden Markov Random Fields, the image gives access to a set of observations on each site of the grid G denoted by $O = \{o(i, j), 1 \leq i \leq n, 1 \leq j \leq m\}$. Furthermore, considering that each state X_s of the Markov Field X is associated to a label l that takes its value in a discrete and finite set of q label $L = \{l_i\}$, $0 < i < q$, and corresponding to a particular layout rule or class pattern, the problem of layout extraction in the image can be formulated as finding the most probable state configuration among all the possible labeling E of the field X that can be associated to the image, i.e. finding:

$$\hat{X} = \arg \max_{X \in E} (P(X, O)) = \arg \max_{X \in E} (P(O|X)P(X))$$

which results in the following formula when applying Markovian hypothesis and independence assumption of observations:

$$\hat{X} = \arg \max_{X \in E} (P(X, O)) = \arg \max_{X \in E} (P(O|X)P(X)) \quad (1)$$

where $N_G(s)$ is the neighborhood of the site s .

While in this expression the term $\prod_s P(o_s | x_s)$ can

be computed using Gaussian mixtures to model the conditional probability densities of the observations, the calculation of the second term (i.e.

$\prod_s P(x_s | x_{s'}, s' \in N_G(s))$), which represents the

contextual knowledge introduced by the model or prior model, appears to be intractable due to its non causal expression i.e. interdependence between neighboring states. To overcome this difficulty, one generally uses simulation methods such as Gibbs sampling or Metropolis algorithm [5]. Another possibility is to restrict the expression to a causal neighboring system. In any case however, finding the optimal segmentation solution requires a huge exploration of the configuration set E . This consideration is especially important because handwritten document images are particularly large. Image decoding using Markov Random Field models is an optimization problem. It consists in finding the realization \hat{x} of the label field X which maximize the posterior probability $P(X/O)$ of the observations set O and the label field X , or similarly an energy function. This is related as MRF-MAP framework.

Indeed according to the Hammersley-Clifford theorem, a MRF is equivalent to a Gibbs distribution [4], so that the second term of equation 1, formally the prior model $P(X)$, can be rewritten as follows:

$$P(X) = \frac{1}{Z} \exp \left(- \sum_{c \in C} V_c(X) \right)$$

where C is the set of all cliques over the image, defined according to the chosen neighboring system $N_G = \{N_G(s), s \in S\}$. V_c is a potential function associated to the clique c and Z is a normalization constant called partitioning function in the context of MRF framework. This allows to introduce the joint energy $U(X, O)$ of a configuration of the field, by calculating the negative logarithm of the joint probability:

$$U(X, O) = \sum_s -\log(P(o_s | x_s)) + \sum_{c \in C} V_c(X)$$

Thus in the MRF-MAP framework, decoding or image labeling involves minimizing the joint energy function:

$$\hat{x} = \arg \min_x U(X, O)$$

It is a non trivial combinatorial problem, because the energy function may be non convex and exhibits many local minima. Different optimization techniques can be used to find the optimal configuration of the label field by minimizing the energy function. We distinguish methods based on relaxation and methods based on dynamic programming. Furthermore some are deterministic and other stochastic. Another important criterion is their ability to find the global optimum of the energy function. The following table gives a classification of the main approaches of the literature, namely Simulated Annealing (SA), Genetic Algorithms (GA), Ant Colony System (ACS), Iterated Conditional Modes (ICM), Highest Confidence First (HCF) and Region Merging method, according to these criteria.

	relaxation methods		dynamic programming
	deterministic	stochastic	
optimal		SA[5], GA[12], ACS [13]	
suboptimal	ICM[7], [8]	HCF	region merging method [9]

Tab 1. classification of decoding methods

We describe in the following subsections the techniques most used in practice.

2.1. Simulated Annealing

This optimization method has been proposed by Kirkpatrick [6] in 1983 and introduced in computer vision by Geman and Geman [5] for image restoration using MRF-MAP framework. It is a stochastic relaxation algorithm based on Metropolis sampling method, which in theory allows to find global minima of the energy function. This algorithm uses a so called "temperature" parameter which controls random label flipping, even if these label changes do not decrease the global energy. This process allows a random exploration of the search space and prevents convergence to local minima. The higher the temperature parameter the higher the probability of label changes. On the contrary if the temperature is low only label changes which decrease the energy are

authorized. During the relaxation process the temperature parameter is gradually decreased starting from a high value, according to a predefined cooling function. In order to provide converge to global optimum one has to set the initial value of the temperature parameter high enough, and has to use an adapted temperature decreasing function, that is slow enough. In theory a logarithmic cooling function is recommended. For each temperature value, several relaxation iterations are done on the entire image site set. Sites are visited randomly or according to a predefined strategy. The number of iterations has to be high enough too. In the simulated annealing method only the temperature and the number of iteration must be predefined. The optimality of the final solution and the computational cost depend closely on the setting of these parameters. The main problem is that there is no theoretical rule for determining the correct values. In practice they are determined empirically. Another main drawback of this algorithm is its prohibitive computational cost. In fact, this algorithm explores the search space « blindly » and therefore requires many updates of the label configuration before convergence. The main advantage of this algorithm is that it doesn't require any particular initialisation of the label field, the exploration of the search space can start from any configuration.

```

Choose an initial temperature  $T=T_0$ 
choose any initial configuration  $x(0)$  of
the label field  $X$ 
repeat
   $i=0$ 
  repeat
    Choose a site  $s$  (according to any
    visit strategy or randomly) and
    randomly change its
    label  $x$  into  $z$  .
    Compute  $\Delta U = U(X_s = x) - U(X_s = z)$ 
    if  $\Delta U > 0$  replace  $x$  by  $z$ 
    else replace  $x$  by  $z$  only if
     $p < \exp(\Delta U/T)$   $p \in [0,1]$ 
    (uniform distribution)
     $i++$ 
  until  $i = Niter$ 
 $T = f(T) = \alpha \cdot T$  until  $0 < \alpha < 1$ 
until  $T < e$  (freezing)

```

Algorithm 1. simulated annealing algorithm

2.2. Iterated Conditional Modes (ICM)

The Iterated Conditional Mode (ICM) Algorithm has been proposed by Besag [7] in 1986. It is an iterative and deterministic relaxation algorithm based on

gradient descent strategy which converges quickly to a local minimum of the energy function. The ICM algorithm can be considered as a special case of the simulated annealing with a null temperature, that is with no energy increase allowed. For each image site, the label which gives the largest local energy decrease is chosen. The principle of the algorithm is the following. Starting from an initial configuration of the label field, all the image sites are visited according to a predefined strategy, and their label are updated by the one that gives the largest local energy decrease, thus causing a decrease of the global energy of the label field. As the modification of a site label may modify the local energy of the neighbouring sites, the process is repeated until convergence to a local minimum of the global energy of the label field or until a predefined stop condition is reached. This stop condition can be for example the number of modified labels during one iteration or the number of performed iterations. This method is very fast but the final quality of the segmentation depends strongly on the initial configuration of the label field, since only local minima of the energy function can be reached. This algorithm is recommended with a performant initialization process which is able to produce initial configurations near the global minima of the energy function.

```

Label field initialization  $x(0)$ 
Computation of the new label field
configuration  $x(n+1)$  from previous
configuration  $x(n)$ 
  1. sites  $s$  are visited according to a
  predefined visit strategy
   $x_s(n+1) = \arg \min_{l \in L} U_s(X_s(n) = l, Y_s = y_s)$ 
  with  $L = \{l_i\}$   $0 < i < q$ 
   $n=n+1$ 
  2. Back to step 1. Until stop criterion
  is reached.

```

Algorithm 2. ICM algorithm

2.3. Highest Confidence First (HCF) algorithm

The Highest Confidence First (HCF) algorithm introduced by Chou and Brown [8], is a variant of the ICM algorithm, and is therefore a deterministic relaxation algorithm too.

Considering the fact that the visit order of the sites influences the convergence of the relaxation process. Instead of examining all the sites systematically without any prior knowledge to lead the process, the main idea of this algorithm is to use a particular strategy.

The site labels are updated successively according to a stability criterion, starting from least stable sites. With this strategy, only unstable sites are considered. The stability measure is calculated as the difference between the current local energy of a site, and the possible lowest local energy for this site.

The least stable site is always processed first, because it is the most likely to change its label and a label change may have some incidence on the labeling of neighbouring sites. A heap is used to control the visit order. In this heap sites are ordered according to their stability measure. The site at the top of the heap is the least stable. The algorithm stops when all sites are stable.

```

All sites are marked as "uncommitted"

Compute the stability G of all sites and
create a heap P according to stability
measure, where the least stable site is
placed at the top.

Take the first s at the top of the heap
P.

if Gs >= 0, where Gs is the stability
measure of site s end
else
    if s is an uncommitted site,
    commit it and associate label l which
    gives the lowest local energy
    else change its label l to the
    label lmin which gives the largest local
    energy decrease
    end if

Update stability measure of site s and of
its neighbours

Put the site s in heap again.
Adjust heap P according to stability
measures.
end if

```

Algorithm 3. HCF optimization method

2.4. 2D Dynamic Programming

In 2003 Geoffrois proposed to adapt the Dynamic Programming principle which is one-dimensional in nature, to two-dimensional or n-dimensional spaces in

general [9]. Before that the principle of the Dynamic Programming had already been applied by Derin and al. [10] for MRF-based image segmentation in computer vision. This principle of 2D Dynamic Programming has been applied by Geoffrois in the context of MRF-MAP framework to energy minimization, for handwritten digit recognition using MRF [11]. This algorithm allows to solve this problem efficiently, by exploiting the grid structure of random fields. The main idea is to divide the image recursively into subregions. The n best configurations of a region are determined among the $n \times n$ configurations obtained by merging the two subregions it contains. This merging process is repeated iteratively starting from unitary subregions corresponding exactly to one site, and for which initialization is simple (Maximum Likelihood of the data term of the energy function), until a region covering all the image is obtained. This method uses the "divide and conquer" principle of dynamic programming.

Assume that two neighboring rectangular regions O_1 and O_2 are associated to their respective state configuration $X_1(i, j)$ and $X_2(i, j)$. Then, the joint probability of region $O = O_1 \cup O_2$ and its associated state configuration $X(u, v)$ defined by:

$$X(u, v) = \begin{cases} X_1(i, j) & \text{if } (u, v) = (i, j) \\ X_2(k, l) & \text{if } (u, v) = (k, l) \end{cases}$$

can be derived as follows:

$$P(X, O) = P(X_1, O_1)P(X_2, O_2)I(X_1, X_2)$$

where the expression

$$I(X_1, X_2) = \prod_{g \in G_1, h \in G_2} P(x_g | x_h, h \in N_{G_1}(g))$$

denotes the interactions between the two state configurations. If we take into account the contextual local dependance between sites in Markov Random Fields, it appears that there are only interactions between sites belonging to the boundaries of the regions during the merging process. In consequence it is not necessary to determine all the possible configurations of entire regions, but simply all the configurations of region boundaries, and for each of them the best configuration of the inside. This allows to reduce the number of configurations to memorize during the merging process. However in practice the number of configurations to memorize is high, especially if the image is large. This is the reason why in practice a pruning strategy is applied to reduce the number of configurations to store. Only the n best

configurations with minimal energy are stored. This parameter n is the pruning threshold. This is the only parameter of this method, however it is very hard to determine. In fact, if this threshold is too low, the optimality of the final solution is not guaranteed, and on the contrary if it is too high the number of intermediate configurations to store and the combinatorial complexity become intractable. Due to this pruning, this method is suboptimal and the quality of the result depends on the merging order. We may consider different merging orders, as for example line horizontal merging, column vertical merging, or alternative merging (fig. 2). Many other merging strategies can be used.

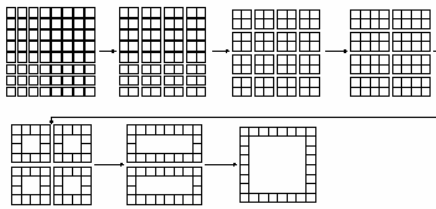


Figure 2. alternative merging strategy

Other optimization techniques have been proposed in the literature for image labeling using MRF-MAP framework, among which Mean Field Annealing[12], Genetic Algorithms[13], or Ant Colony System [14]. We do not describe these methods in this paper.

3. Application of MRF labeling to handwritten document segmentation

When using MRF-MAP labeling framework to segment images, one has simply to make some choices concerning the modelling of the probability density function of observation emission, the clique potential function and the optimization method used to minimize the energy function. In this work we are interested in segmenting handwritten documents, such as drafts or authorial manuscripts, into their elementary parts using a prior MRF model. We describe here our implementation choices to solve this task.

- Probability densities

The probability densities are modeled by gaussian mixtures. The parameters of the mixtures are learned on manually labelled images, using the EM algorithm. The number of gaussians is determined automatically using the Rissanen criterion. We use Bouman's

CLUSTER software² to learn the number of gaussian components and mixture parameters.

- Clique potential functions

We consider the second order cliques associated to a 4-connected neighboring:

$$C = C_1 \cup C_2 \cup C_3$$

where

$$C_1 = \{(i, j), 1 \leq i \leq n, 1 \leq j \leq m\}$$

$$C_2 = \{(i, j), (i+1, j), 1 \leq i \leq n, 1 \leq j \leq m\}$$

$$C_3 = \{(i, j), (i, j+1), 1 \leq i \leq n, 1 \leq j \leq m\}$$

The interaction terms are defined as mutual information terms taking into account only the horizontal and vertical directions (4-connectivity):

$$I_H = \frac{P(w_k | w_l)}{P(w_k)P(w_l)} \quad I_V = \frac{P\left(\frac{w_k}{w_l}\right)}{P(w_k)P(w_l)}$$

where

$$P(w_k | w_l) = P(w_{(i,j)} = w_k, w_{(i+1,j)} = w_l) \text{ and}$$

$$P\left(\frac{w_k}{w_l}\right) = P(w_{(i,j)} = w_k, w_{(i,j+1)} = w_l)$$

As for the gaussian mixture parameters, these probabilities are learned on some labeled examples, by counting the frequency of each possible transition. If a rule transition does not appear in the learning examples, its probability is not set to zero but to a very low value, making it not impossible but very unlikely.

Finally, the clique potential functions are defined as follows:

$$V_c(w) = \begin{cases} -\log(P(w_k)) & \text{if } c \in C_1 \\ -\log(I_H(w_k, w_l)) & \text{if } c \in C_2 \\ -\log(I_V(w_k, w_l)) & \text{if } c \in C_3 \end{cases}$$

In a similar way, according to these definitions, the use of 2-order cliques with 8-connected neighboring is very simple. One has only to take into account diagonal interactions too. The different n-order clique forms associated to 4-connected and 8-connected neighboring are illustrated on figure 3.

² <http://dynamo.ecn.purdue.edu/~bouman/software/cluster>







neighbor. system \ clique order	clique order			
	1	2	3	4
4-connected				
8-connected				

Figure 3. Neighboring systems of order 1 and 2 with their corresponding cliques.

- Observations

Observations are features that are extracted on each site s at the position $g(i, j)$ on the grid G applied on the image. As we work on binary images, we have chosen to extract for each site s a bi-scale feature vector based on pixel density measurement. This vector contains 18 features. The first 9 are the density of black pixels in cell $g(i, j)$ and its 8-connected neighbors at the first scale level. Based on the same principle, the remaining 9 features are the density of black pixels extracted at the second coarser scale level (see figure 4). Each cell at this scale corresponds to a 3×3 window at the previous scale. Note that the size of the cells $g(i, j)$ on the grid G must be adapted to the size of the smallest objects or layout elements we want to extract in the image. The choice of this size is necessarily the result of a compromise between the segmentation quality and the computational efforts. The smaller the cells are, the more labeling is fine, but more there will be sites, so more complicated will be the energy minimization process. On our images, depending on the considered segmentation task, we are using different cell sizes.

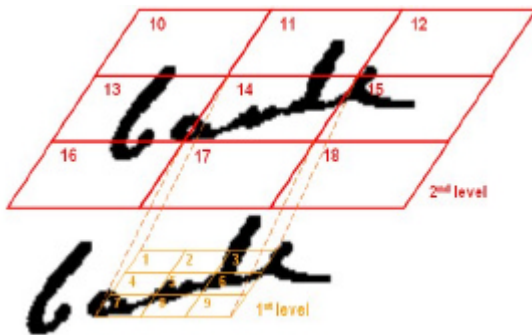


Figure 4. Multiresolution pixel density feature extraction

- Decoding strategy

To proceed to the decoding of the image by means of minimization of the energy function, we have implemented several of the methods described in the literature, mainly ICM, HCF, and 2D dynamic programming. We have tested and compared these methods. The results are provided in the next section. We have implemented simulated annealing too, but we have tested this algorithm only on very small images fragments, not on entire manuscript images, because of the very high computational complexity of this algorithm, so no results are provided.

4. Results

The analysis of the results of a document image segmentation algorithm is a difficult and not always a well defined task, since there exist very few protocols and image databases for performance evaluation [15]. The few existing ones are only designed for machine printed documents for which the proposed methodologies and metrics used to compare the algorithms are dedicated to well defined classes of methods or documents (newspaper, mail, form, postal address). To the best of our knowledge, there do not exist such methodologies and metrics in the field of handwritten documents or historical documents.

As our approach is able to produce labelings at different analysis level using different grid sizes, we present here the results obtained on two different segmentation tasks working at two different scales. The first task consists on labeling large areas of interest in manuscript images, such as text body, margins or text blocks, working at a coarse resolution. For this task we provide quantitative results in term of labeling rates and processing times, for several decoding methods. The results obtained are also illustrated visually and discussed. With the second task, which consists on text line labeling, we show the ability of this approach to perform at finer level, in order to extract and separate small entities such as words or word fragments and erasures. For several reasons we explain, we provide for this task qualitative results only obtained on few images of full page of handwriting or parts of pages from the Bovary database.

4.1. Zone Labeling

In order to evaluate precisely the performance of our approach and compare the decoding methods

according to labeling rate and processing time, we have first considered a segmentation task where a simple coarser labeling is possible. In this case, it is easy and fast to label a database of Flaubert manuscript images manually for model learning and groundtruthing. The task we consider consists in labeling the main regions of the manuscripts such as text body, margins, header, footer, page number, and marginal annotations (see Figure 5.a). The model contains 6 labels. The database contains 69 manuscript images at 300 dpi. The average dimensions of the images are 2400×3700. All the images of the database have been binarized and manually labeled according to the defined 6 labels.

The database has been divided into 3 parts: one for the learning of the model parameters (parameters of the gaussian mixtures, clique potential functions), an other for model setting, and the last one for testing. We have a regular grid where the dimensions of each cell are 50×50 pixels. We compare the results obtained with a Mixture Model using Maximum Likelihood criterion and the results obtained with ICM, HCF and 2D Dynamic Programming (2D DP) decoding with the groundtruth labelings manually produced (Figure 5) For each decoding method we evaluate the global labeling rate (GLR) by counting the number of well-labeled sites and the normalized labeling rate (NLR) by counting the average number of well-labeled sites for each label class.

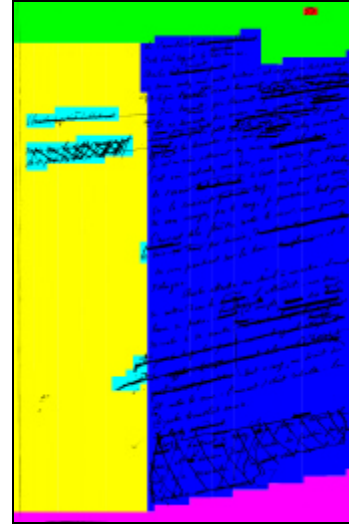
$$GLR = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

$$NLR = \frac{\sum_{i=0}^{q-1} \left(\frac{TruePositive}{TruePositive + FalseNegative} \right)_{l_i}}{q}$$

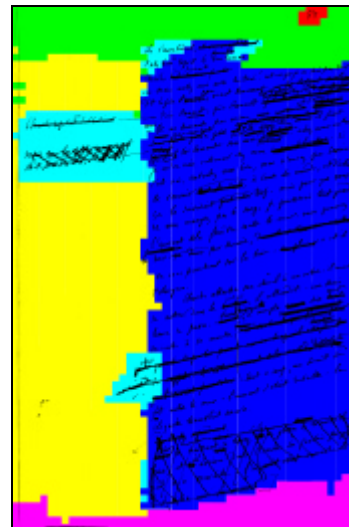
For each decoding methods we also give the average processing time in seconds for one page decoding. This time is only related to the decoding process, probability distribution estimation is not taken into account. Results are provided in table 2.

	Mixture s	ICM	HCF	2D DP
GLR (%)	88,0	86,6	90,3	84,6
NLR (%)	83,7	87,5	88,2	87,4
time (s)	-	0,21	0,29	0,61

Tab 2. labeling rates obtained with different decoding method



(a)



(b)

Figure 5. Zone labeling at a coarser scale: (a) groundtruth (b) result with Markovian labeling using the following color/label convention: red = page number, green = header, blue = text body, pink = footer, cyan = text block, yellow = margin.

These results show that the use of a MRF model allows to increase the normalized labeling rate and that the HCF algorithm outperforms the other decoding methods. Furthermore HCF algorithm is faster than 2D dynamic programming method. The difference between GLR and NLR are due to non homogeneous class repartition in dataset (see Figure. 6).

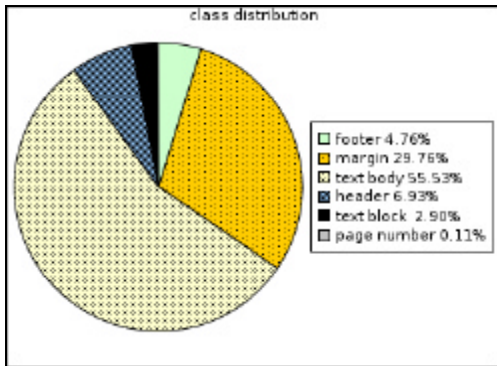
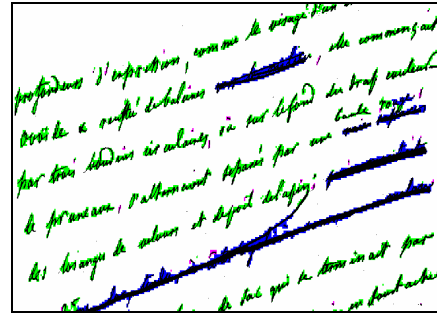


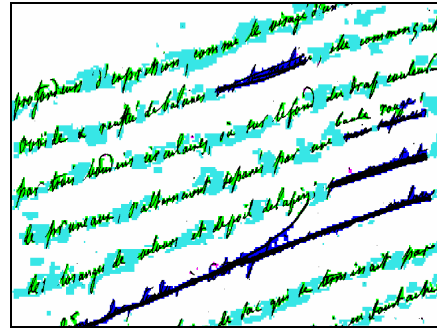
Figure 6. Class distribution in dataset

4.2. Text Line Labeling

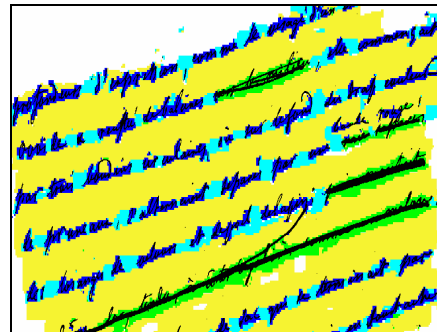
Let us recall that Flaubert's manuscripts contain a lot of deletions and crossed out words or lines (see figure 1). Therefore, in this second experiment, we have tried to evaluate the capabilities of our method to work at finer analysis level, on a specific task which consists in separating words (or parts of words) and deletions, and to extract text lines using a prior model which integrates several states. For this purpose we have first defined a model made up of 4 states: "pseudo-word", "deletion", "diacritic" and "background". We work at the pixel level using a regular grid of 1×1 cells and we use the 2D Dynamic Programming method of Geoffrois for decoding. For this task we provide qualitative results only because it is very hard to manually label images at a pixel level for groundtruthing. Figure 7.a. presents the results obtained with this model on a page fragment. Figure 8.a. shows a zoom on a deletion area where word and deletion strokes are completely connected. One can see on this result that the deletion lines are well separated from the strokes below. This result highlights the superiority of this method on the approaches working at the connected component level. Indeed, the fact of working at the pixel level allows us to segment different objects which are connected together. Figure 8.b. shows similar results on a fragment containing a word and an erasure connected by a descending loop. Both components are well separated.



(a)

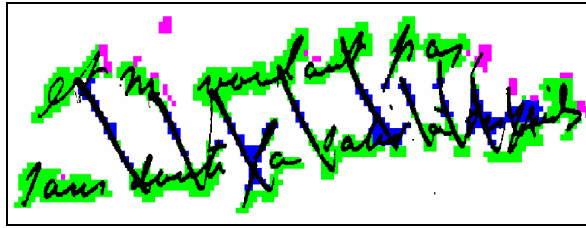


(b)

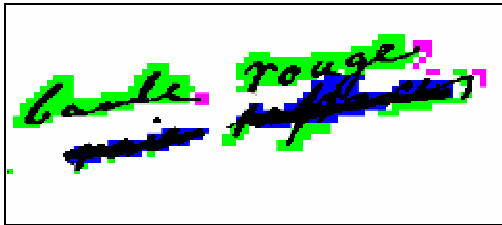


(c)

Figure 7. Segmentation results obtained on a page fragment: (a) using a 4-state model; (b) using a 5-state model; (c) using a 6-state model, with the following color/label convention: white = background, green = textual component, blue = erasure, pink = diacritic, cyan = interwords spacing, yellow = interline.



(a)



(b)

Figure 8. Segmentation results obtained on some complex page fragments using the 4-state model.

This model allows to extract word fragments and erasures, but does not model text lines, so we have refined it by introducing an additional "inter pseudo-word space" state. The addition of this state makes it possible thereafter to extract the text lines because one can define a text line as a sequence of "pseudo-words" separated by "inter-word space". Thus from the results returned by the method, it is possible to extract text lines or other objects of higher level (such as text blocks for example), by applying label merging rules. Globally the results are promising, the inter-word spaces are well segmented (see figure 7.b).

Finally in the same way, we have defined a third model with 6 states by adding an "interlines" state to the previous model, in order to model also the interlinear spacings. The knowledge of interlines allows to better segment text lines, and to detect text blocks. The result obtained with this model on the same page fragment is shown on figure 7.c and the result obtained on a full page is shown on figure 9.

For these three models the results are globally satisfactory. However if we look locally at the results, we can see that some pixels are misclassified. One has to keep in mind that the 2D dynamic programming algorithm with pruning procedure is a sub-optimal decoding algorithm. It means that the final segmentation obtained is not the optimal one. Some configurations of the label field can be locally less probable and thus be pruned during the merging procedure, whereas they could be globally the optimal

ones. If the size of the image is large and if there are a lot of states in the model, the number of possible configurations of the label field is very large. In this case, it is not possible to store all the possible intermediate solutions, so the pruning threshold should not be too high. On the other hand, if this threshold is too low, the final configuration retained may be one of the least probable ones (because involving not probable transitions during the region merging). The choice of the merging strategy is important for the final segmentation result, but we think that the choice of features extracted on the observations is important too.

However these experiments on two different labeling tasks show the ability of the approach to work at different level of analysis and to extract area of interest in complex documents such as authorial manuscripts. The system could benefit of user interactions and multiscale approach to improve the labeling results.

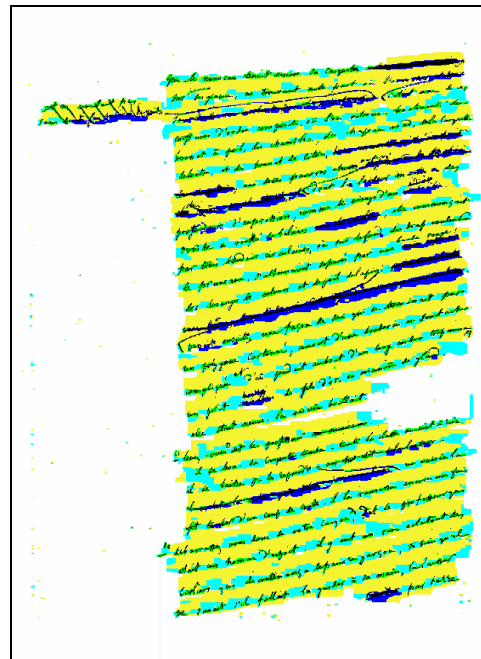


Figure 9. Segmentation results obtained on a complete Flaubert manuscript page using a 6-state model with the following color/label convention: white = background, green = textual component, blue = erasure, pink = diacritic, cyan = interwords spacing, yellow = interline.

5. Future works

As pointed out by He in [16], Markov Random Fields have two main drawbacks. First, they make hypothesis about the independence of observations, for inference tractability reasons. These hypothesis are too strong for image labeling. For these reasons only local relationships between neighboring nodes are incorporated into the model. The second one is their generative nature. Markov Random Field attempt to model the joint distribution of the observed image and the corresponding label field, so a great effort is spent to model the observation distribution. However during the decoding the problem is to estimate the conditional distribution of the label field according to the observed image, there is no need to try to model the joint distribution, which may be very complex, but only the conditional distribution of the label field given observations. In consequence, less training data are needed. It is the reason why discriminative models, such as Conditional Random Fields, have been proposed recently to directly model this conditional distribution. Conditional Random Field have been introduced first by Laferty and Mc Callum [17], for part-of-speech tagging, that is to segment one-dimensional sequences, and have been adapted to image segmentation. In [15], He and al. propose a MLP-based CRF implementation for image labeling, which aim to take into account and to learn features that operate at different scales of the image. Pointing the fact that labeling process needs contextual information because of the dependance of the labels across the image, the authors propose a multiscale conditional random field model considering three analysis levels: a local analysis, a regional analysis and a global analysis.

Up to now Conditional Random Field have not yet been applied to document image segmentation. In future work, and starting from our MRF model, we propose to transform it to a discriminative conditional model.

6. Conclusion

In this paper we have proposed to use Markov Random Field models to segment complex handwritten manuscripts into their elementary parts, such as text body, margins, header, footer, page numbers, deletions, ... by means of image labeling using different optimization techniques such ICM, HCF and dynamic programming. We have tested the approach on a dataset of manuscripts of french writer

Gustave Flaubert. The proposed approach provides interesting results especially with HCF algorithm. The main advantages are the ability of Markov Random Fields to deal with local variability, to model prior knowledge and the learning possibilities which allow an easier adaptation to different type of documents. However due to their generative nature, Markov models suffer from several limitations. For this reason we plan in future works to provide our system an evolution towards Conditional Random Fields which are discriminative models.

7. References

- [1] H. Bunke, "Recognition of Cursive Roman Handwriting, Past, Present and Future", *Proceeding of the seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, pp. 448-459, Edinburgh, Scotland, 2003.
- [2] H. Baird, "Digital Libraries and Document Image Analysis", *Proceeding of the seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, pp. 2-14, Edinburgh, Scotland, 2003.
- [3] Nagy, G, "Twenty years of document image analysis in pamiTwenty Years", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, n°1, pp. 38-62, 2000.
- [4] Chellappa, R. et Jain, A., editors (1993). *Markov Random Fields - Theory and application*, Academic Press.
- [5] S. Geman, D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 6, pp. 721-741, 1984.
- [6] S. Kirkpatrick, C. D. Gellatt and M. P. Vecchi, "Optimization by Simulated Annealing", *Science* (1983), no. 220, pp. 671-680, 1983.
- [7] J. E. Besag, "On the statistical analysis of dirty pictures", *Journal of the Royal Statistical Society B*, vol. 48, no. 3, pp. 259-302, 1986.
- [8] P. Chou, C. Brown, "The theory and practice of Bayesian image labeling", *International Journal of Computer Vision*, 4, pp.185-210, 1990.
- [9] E. Geoffrois, "Multi-dimensional Dynamic Programming for statistical image segmentation and recognition", *International Conference on Image and Signal Processing*, pp. 397-403, Agadir, Morocco, 2003.
- [10] H. Derin, H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields",

IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 9, n°1, pp. 39-55, 1987.

[11] E. Geoffrois, S. Chevalier, F. Prêteux, "Programmation dynamique 2D pour la reconnaissance de caractères manuscrits par champs de Markov", *proceedings of RFIA, Reconnaissance de Formes et Intelligence Artificielle*, pp. 1143-1152, Toulouse, France, Janvier 2004.

[12] S.Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer, Tokyo, 1995.

[13] E.Y. Kim, S.H. Park, H.J. Kim, "A genetic algorithm-based segmentation of Markov Random Field modeled images", *IEEE Signal processing letters* 11(7): 301-303, 2000.

[14] S. Ouadfel, M. Batouche, "MRF-based image segmentation using Ant Colony System", in *Electronic Letters on Computer Vision and Image Analysis (LCVIA)*, vol. 2, n°1, pp. 12-24, August 2003.

[15] "Performance Evaluation: Theory, Practice, and Impact", T. Kanungo, H. S. Baird, R.M. Haralick, Guest Editors, special issue of *International Journal on Document Analysis and Recognition*, vol. 4, n°3, march 2002.

[16] X. He, R.S. Zemel, M. A. Carreira-Perpinan, "Multiscale Conditional Random Fields for Image Labeling", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, pp. 695-702, Washington DC, USA, 2004.

[17] J. Lafferty, A. McCallum, F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data", *18th International Conference on Machine Learning*, pp 282-289, Williamstown, USA, 2001.