Document Ink bleed-through removal with two hidden Markov random fields and a single observation field

Christian Wolf

Technical Report RR-LIRIS-2006-019

Laboratoire d'informatique en images et systèmes d'information - UMR 5205

INSA-Lyon, Bât. J.Verne; 20, Av. Albert Einstein
69621 Villeurbanne cedex, France
Tel.: 0033 4 72 43 63 08 Fax.: 0033 4 72 43 71 17
Email: christian.wolf@insa-lyon.fr
Web: http://liris.cnrs.fr/christian.wolf

July 2, 2007

Abstract

We present a new method for blind document bleed through removal based on separate Markov Random Field (MRF) regularization for the recto and for the verso side, where separate priors are derived from the full graph. The segmentation algorithm is based on Bayesian Maximum a Posteriori (MAP) estimation. The advantages of this separate approach are the adaptation of the prior to the contents creation process (e.g. superimposing two hand written pages), and the improvement of the estimation of the verso pixels through an estimation of the verso pixels covered by recto pixels. Optimization is carried out with the simulated annealing algorithm. The labels of the initial recto and verso clusters are recognized without using any color or gray value information. The proposed method is evaluated on synthetic images as well as scanned document images. The results on real scanned data have been evaluated using statistical evaluation on an empirical test performed by 16 people.

Keywords
Markov Random Fields, Bayesian estimation, Document Image Restoration, Bleed-Through Removal

1 Introduction

General image restoration methods which do not deal with document image analysis have mostly been designed to cope with sensor noise, quantization noise and optical degradations as blur, defocussing etc. (see [22] for a survey). Document images, however, are often additionally subject to further and stronger degradations:

1. non stationary noise due to illumination changes.
2. curvature of the document.
3. ink and coffee stains and holes in the document.
4. ink bleed through : the appearance of the verso side text or graphics on the scanned image of the recto side. This is an important problem when very old historical documents are processed.
5. low print contrast.
6. errors in the alignment of multiple printing or imaging stages.

In this paper we concentrate on the problem of ink bleed through removal, i.e. the separation of a single scanned document image into a recto side and a verso side. We assume that a scan of the verso side is *not* available (blind separation). In this case, the task is equivalent to a segmentation problem: classify each pixel as either *recto*, *verso*, *back ground*, or eventually *recto-and-verso* (simultaneously). This means that the vast collection of widely known segmentation techniques can be applied directly. On the other hand, document images are a specific type of images with their own properties and their own specific problems. It is desirable to design algorithms which exploit their specific properties in order to improve the segmentation performance.

At first thought it might be a good idea to interpret the task as a blind source separation problem similar to the "cocktail party" problems successfully dealt with by the (audio) signal processing community. Independent components analysis (ICA) is one of the techniques which are most widely used and so it is no surprise that it also has been applied to document bleed through removal [25]. However, the issue which makes this formulation questionable is that ICA assumes a linear model:

$$\boldsymbol{d}_s = \boldsymbol{A}\boldsymbol{f}_s$$

where $\boldsymbol{d}_s$ is the observation vector, $\boldsymbol{f}_s$ is the source vector and $\boldsymbol{A}$ is the mixing matrix. In the case of documents, each vector corresponds to a pixel at site s. The source vector is mostly

chosen to be three dimensional, the dimensions corresponding to the recto signal, the verso signal and an additional signal adding the background color [25]. In this case, the column vectors of the mixing matrix become the color vectors for, respectively, recto pixels, verso pixels and background pixels, as can be seen easily by setting $\boldsymbol{f}_s = [\ 1\ 0\ 0\ ]^T, [\ 0\ 1\ 0\ ]^T$ and $[\ 0\ 0\ 1\ ]^T$ and $\boldsymbol{d}_s$ to the respective color vector and solving for $\boldsymbol{A}$.

We can easily verify that the linear hypothesis cannot be justified for ink bleed through by calculating the color of an observed pixel created by a source pixel which contains overlapping recto and verso pixels ($\boldsymbol{f}_s = [\ 1\ 1\ 0\ ]^T$) : according to the model the observed color vector is the sum of the color vectors for the recto and the verso pixel, which cannot be true in reality.

Sharma presents a non-blind restoration algorithm, i.e. a method which requires a scan of the recto as well as the verso side of the document [23]. The two images are aligned using image registration techniques. A reflectance model taking into account a bleed-through spread function is created, approximated and corrected with an adaptive linear filter.

Another non-blind method is proposed by Dubois and Pathak [10]. The emphasis is set to the image registration part, the restoration itself is performed using a thresholding-like heuristic.

Tan et al. propose a non-blind method where the alignment is done manually [24]. Foreground strokes are detected using both images and enhanced using a wavelet decomposition and restoration. The same authors also present a blind method, which is based on the hypothesis that the handwriting is (very) slanted, and therefore that the strokes of the recto and the verso side are oriented differently [27]. A directional wavelet transform is employed to identify the origin of each stroke.

Nishida and Suzuki describe a method based on the assumption that high frequency components of the verso side are cut off in the bleeding process [21]. Their restoration process uses a multi-scale analysis and edge magnitude thresholding. Leydier et al. propose a serialized (i.e. adaptive) version of the k-means algorithm[19]. Drira et al. propose an iterative algorithm which recursively applies the k-means algorithm to the image reduced with principal components analysis [9]. The recursive calls produce a tree of image layers corresponding to different color clusters in the image. The leaf containing the verso layer is chosen by histogram analysis.

The method presented by Don [7] is justified by a noise spot model with Gaussian spatial distributions and Gaussian gray value distributions. Under this assumption, thresholds near the class means produce spurious connected components. A histogram of connected component counts is created and thresholded using standard techniques.

MRF regularization has already been used for this kind of problem. For instance, Tonazzini et al. present a document recognition system which restores selected difficult parts of the document image with MRF regularization [26]. As prior model they chose an Ising model with non-zero clique types of 1, 2, 3, and 9 pixels. The observation model contains a convolution term with an unknown parameter. The optimization procedure alternates between the segmentation procedure and the estimation of the parameter of the observation model. The authors do not specify how they estimate the parameters of the prior model.

Donaldson and Myers apply MRF regularization with two priors to the problem of estimating a super-resolution image from several low-resolution observations [8]. However, as opposed to our solution, there is only one field : the first prior measures smoothness, whereas the second prior measures a bi-modality criterion of the histogram.

In this approach we ignore degradations no. 1 and 3 and propose an approach based on a stationary model. Non homogeneous models will be developed in further publications. This restriction allows us to choose the well known MRF-MAP Framework (Bayesian maximum a posteriori estimation with a MRF prior) and to formulate the problem in terms of two different models:

- the *a priori* knowledge on the segmented document is included in the prior model. In our case, the prior model consists of two MRFs, one for each side of the document.

- the knowledge on the document degradation process is included in the observation model.

In a previous paper, we described a MRF model for document image segmentation [28]. The goal, however, was to learn the spatial properties of text in document images in order to improve the binarization performance. The clique potentials of large 4×4 cliques were determined by strict supervised learning from training images. In this paper, however, the emphasis is set to regularization. Therefore, a parametric prior model has been chosen.

The contribution of this paper is threefold:

1. Creation of a double MRF model with a single observation field and the corresponding inference algorithm.

2. Design of an algorithm for the initial recognition of recto and verso labels without using any color or gray value information.

3. Design of a hierarchical algorithm for the calculation of the background gray value replacing the verso pixel.

This paper is organized as follows. Section 2 proposes a dependency graph for the joint probability density of the full set of variables (the hidden recto and verso variables as well as the observed variables) and derives the prior probability. Section 3 proposes the observation model. Section 4 describes the posterior probability and its optimization (the Gibbs sampler of the model). Section 5 outlines the estimation procedure for the prior parameters and the parameters of the observation model. Section 6 illustrates the initialization of the different fields as well as the recognition of the lables of the initial clusters and section 7 describes the restoration process. Section 8 presents the experiments we performed on synthetic and real scanned document images in order to evaluate the performance of the proposed method. Section 9 finally concludes.

2 The prior model

MRFs capture the spatial distribution of the pixels of an image by assigning a probability (or an energy) to a given configuration, i.e. a given segmentation result. This is normally used to regularize the segmentation process, i.e. to favor certain configurations which are considered more probable. One of the most widely used assumptions is the smoothness criterion - homogeneous areas are considered more probable then frequent label changes.

This assumption is normally justified[1] considering that very often high frequencies in the image content correspond to noise and assuming that the MRF model has been adapted to the prior knowledge on image content. However, this changes when the observed image is the result of the superposition of two or more "source" images, which is the situation dealt with in this work. In this case, *a priori* knowledge may be available for each of the source images, but not for the mixture of these images. Applying a simple regularization on the combined image may over-smooth areas which should actually contain high frequency edges due to the superposition process.

We therefore propose to create a prior model with two different label fields : one for the recto side (F^1) and one for the verso side (F^2). Instead of a segmentation problem with a configuration space of 3 or eventually 4 labels for each site (*recto*, *verso*, *back ground*, and eventually *recto-and-verso*), we get a segmentation problem where each pixel corresponds to two different hidden labels, one for each field, and where each label is chosen from a space of two labels: *text* and *back ground*. The advantages of this formulation are two-fold:

- the separation into two different label fields creates a situation where the priors regularize fields which directly correspond to the natural process "creating" the contents (e.g. hand writing letters), as opposed to the single field case, where the prior tries to regularize a field which is the result of overlapping two content fields.

- Correctly estimating verso pixels which are shadowed by recto pixels, which is only possible with two separate fields, is not just desirable in the case where the verso field is needed. More so, a correct estimation of the covered verso pixels, through the spatial interactions encoded in the MRF, helps to correctly estimate verso pixels which are *not* covered by a recto pixel, thus increasing the performance of the algorithm.

Note, that the same result could be achieved with a single hidden label field and by adapting the prior model such that its regularization handles different label interactions differently. In general this produces rather complicated energy functions equivalent to rather simple interactions in the respective fields.

In the following and as usual, uppercase letters denote random variables or fields of random variables and lower case letters denote realizations of values of random variables or of fields of random values. In particular, $P(F = f)$ will be abbreviated as $P(f)$ when it is convenient.

Markov Random Fields have a long history, we refer the reader to the seminal work and very often cited paper by Geman and Geman [12] and to the book written by Li for a large yet profound overview of the theory [20]. MRFs are non causal models on undirected graphs which treat images as stochastic processes. A field F of random variables F_{s_1}, F_{s_2}, ... F_{s_N} is a MRF if and only if

$$P(F{=}f) > 0 \quad \forall f \in \Omega \text{ and}$$

$$P(F_s{=}f_s|F_r{=}f_r, r \neq s) \;=\; P(F_s{=}f_s|F_r{=}f_r, r \in N_s)$$

where f is a configuration of the random field, Ω is the space of all possible configurations and N_s is the neighborhood of the site s. In other words, the conditional probability for a pixel of the image depends only on the pixels of a pre-defined neighborhood around this pixel.

On a graph, each neighborhood defines a set of cliques, where a clique is fully connected sub graph. According to the Hammersley-Cifford theorem [13] [2], the joint probability density functions of MRFs are equivalent to Gibbs distributions defined on the maxima cliques, i.e. are of the form

$$P(f) = \frac{1}{Z} \exp\{-U(f)/T\} \tag{1}$$

where $Z = \sum_f e^{-U(f)/T}$ is a normalization constant, T is a temperature factor which can be assumed to be 1 for simplicity, $U(f) = \sum_{c \in \mathcal{C}} V_c(f)$ is a user defined energy function, $\mathcal{C}$ is the set of all possible cliques of the field and $V_c(f)$ is the energy potential for the realization f defined on the single clique c.

Given the nature of the problem, the three different label fields (two hidden and one observed) should be considered in a holistic way in order to precisely describe the interactions between the two fields and to define a joint probability distribution on the full set of labels. In the rest of this paper we therefore consider a full graph $\mathcal{G} = \{V, E\}$ with a set of nodes V and a set of edges E. V is partitioned into three subsets: the two fields of hidden variables F^1 and F^2 and the field of observed variables D. The three fields are indexed by the same indices corresponding to the pixels of the image, i.e. F^1_s, F^2_s and D_s denote, respectively, the hidden recto label, the hidden verso label and the observation for the same pixel s. The set of edges E defines the neighborhood on the graph, i.e. there is an edge between to nodes r and s if and only if $r \in N_s$ and $s \in N_r$.

The model described in this work is generative, i.e. it tries to explain the process of creating the observed variables from the hidden ones. Normally, when one creates contents for a page consisting of a recto and a verso page, one considers the recto content to be "independent" of the verso content since the two different pages do not necessarily influence each other — they may even be created by different authors. Statistically speaking, however, the two fields are not independent. More so, they are not even conditionally independent given the observed field, since, given the observed field, knowledge

[1] Label changes on the borders of regions can be ignored, dealt with by a separate line processes[12] or directly in the main process [5].

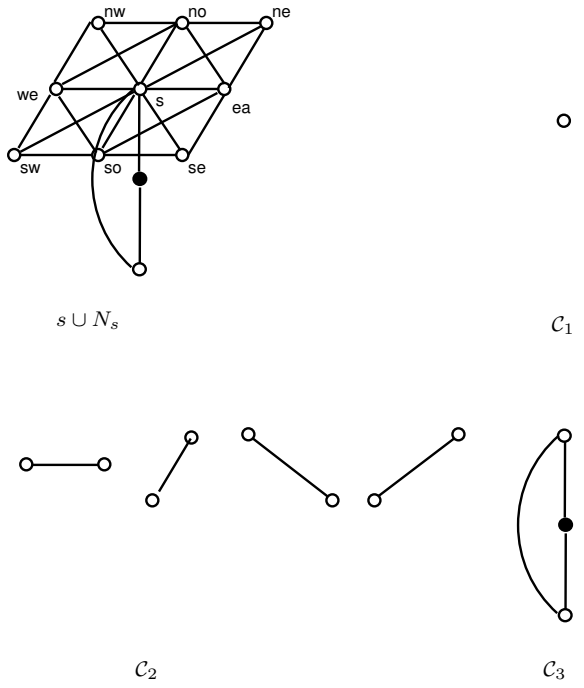


Figure 1: A site $s \in F^1$ and its neighborhood N_s and the non zero clique types: "intra-field" single site ($c \in \mathcal{C}_1$), pair site ($c \in \mathcal{C}_2$) and "inter-field" three-node cliques ($c \in \mathcal{C}_3$).

of the recto field influences inference of the verso field and vice versa.

Considering the relationships between the observed variables and the hidden variables, i.e. the degradation processes, we assume a first-order MRF, which means that the following two conditions hold (a common assumption in the MAP-MRF framework):

1. The random variables D_s are independent conditional to the hidden label fields F^1 and F^2.
2. $P(D_s|F^1, F^2) = P(D_s|F^1_s, F^2_s)$

As a consequence, the dependency graph (see figure 2) contains the following cliques types: first order and second order "intra-field"[2] cliques in the subgraph F^1, first order and second order "intra-field" cliques in the subgraph F^2 (we will assume the 3-node clique potentials to be zero) and finally the "inter-field" cliques between F^1, F^2 and D. For reasons which will become clear in section 3, we will set the potentials for the pairwise inter-field cliques to zero, i.e. the second order cliques with one node $\in F^1$ and one node $\in D$ as well as the second order cliques with one node in $\in F^2$ and one node $\in D$. The only contributing cliques are therefore three-node cliques with one node of each respective field (F^1, F^2 and D); see figure 1.

The joint probability distribution of the whole graph can therefore be given as follows:

[2]The reader may have noticed that we frequently denote the subsets of sites F^1, F^2 and D as "fields" and will excuse the slight ambiguity with the "full" Markov random field which consists of all three fields.

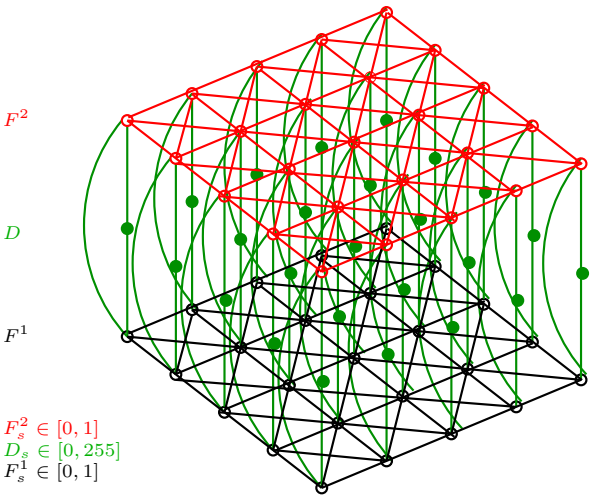


Figure 2: The dependency graph of the model consisting of the two label fields F^1 and F^2 ("empty" nodes) and the single observation field D (shaded nodes).

$$
\begin{aligned}
& P(f^1, f^2, d) \\
= \; & \frac{1}{Z} \quad \exp\left\{-\left(U(f^1) + U(f^2) + U(f^1, f^2, d)\right)/T\right\} \\
= \; & \frac{1}{Z} \quad \exp\left\{-\left(U(f^1) + U(f^2)\right)/T\right\} \cdot \\
& \qquad \exp\left\{-\left(U(f^1, f^2, d)\right)/T\right\} \\
= \; & \frac{1}{P(d)} \quad P(f^1, f^2) P(d|f^1, f^2)
\end{aligned}
\tag{2}
$$

The last equality indicates the Bayesian interpretation of the problem: the first factor corresponds to the prior knowledge and the second factor corresponds to the data likelihood determined by the observation/degradation model. Infering the set of hidden labels from the observed labels corresponds to a maximizaton of the posterior probability (see section 4).

Following (2) we can see that the prior probability is actually the product of the two probabilities of the two fields F^1 and F^2:

$$P(f^1, f^2) = P(f^1)P(f^2)$$

This can also directly be seen in the dependency graph: the two hidden label fields F^1 and F^2 do not share any common nodes nor edges.

In image processing applications, a widely used prior model for MRFs on rectangular grids is the logistic model [20], which we slightly adapted:

$$U(f_s, f_{N_s}) = \sum_{\{s\} \in \mathcal{C}_1} \alpha f_s + \sum_{\{s,s'\} \in \mathcal{C}_2} \beta_{s,s'} \gamma(f_s, f_{s'}) \tag{3}$$

where $U(f_s, f_{N_s})$ is the *local evidence* for s, i.e. the potential calculated on the subset of cliques which contain s, $\mathcal{C}_1$ is the

set of single site cliques, $\mathcal{C}_2$ is the set of pair site cliques (see figure 1) and γ is defined as follows:

$$\gamma(L_1, L_2) = \begin{cases} +1 & \text{if } L_1 = L_2 \\ -1 & \text{else} \end{cases}$$

The labels f_s may take values from $\mathcal{L} = \{0, 1\}$. We chose a stationary and anisotropic model, therefore the single site parameter α depends on the label f_s of the corresponding site s whereas the pair site parameters $\beta_{s,s'}$ depend on the direction of the clique (horizontal, vertical, right-diagonal, left-diagonal) and its labeling.

The whole prior energy defined on both hidden fields is given as the product (sum) of two (adapted) logistic models:

$$\begin{aligned} & U(f_s^1, f_{N_s}^1, f_s^2, f_{N_s}^2) \\ = & \sum_{\{s\}\in\mathcal{C}_1} \alpha^1 f_s^1 + \sum_{\{s,s'\}\in\mathcal{C}_2} \beta_{s,s'}^1 \gamma(f_s^1, f_{s'}^1) \\ + & \sum_{\{s\}\in\mathcal{C}_1} \alpha^2 f_s^2 + \sum_{\{s,s'\}\in\mathcal{C}_2} \beta_{s,s'}^2 \gamma(f_s^2, f_{s'}^2) \end{aligned}$$

Note that only the intra-field cliques from the sets $\mathcal{C}_1$ and $\mathcal{C}_2$ are used in the prior model, the clique potentials from the set $\mathcal{C}_3$ are part of the observation model and will be defined in the next section.

This choice results in a prior parameter vector θ_p which consists of 10 parameters (5 for the recto field and 5 for the verso field):

$$\theta_p = [\alpha^1, \beta_1^1, \ldots, \beta_4^1\, \alpha^2, \beta_1^2, \ldots, \beta_4^2\,]^T$$

3 The observation model

The document degradation model can be seen as a two step process: first the two sides (recto and verso) are subject to separate degradation processes ϕ_1 and ϕ_2, possibly with different parameters, and then they are combined in a second stage:

$$D = \phi_c(\phi_1(F^1), \phi_2(F^2))$$

where D is the observation field and F^1 and F^2 are the two hidden label fields.

As already mentioned in section 2, we assume a first-order MRF for the first stage degradation processes ($\phi_i,\ i = 1..2$). In particular, we assume additive Gaussian noise with different parameters for each class (text and background). The Gaussian assumption may seem to be an over simplification of the complex process involved in the degradation of historic documents which very often have been stored for centuries in not optimal conditions. However, the choice is motivated by several reasons : the simplicity of the Gaussian function makes the mathematical formulation of the model easy and very often the oversimplifications of the observation model are compensated by the regularizing effect of the prior.

Degradation models designed for document images do exist and are widely used in the document image community. Unfortunately most of them have been developed for the evaluation of document analysis algorithms and therefore have been designed as binary operations, e.g. a series of morphological operations [1] [30]. In [14], Kanungo et al. propose a degradation model which takes into a account the page bending process as well as the perspective distortion and the illumination change which results from it. These formulations make it impossible to use them in a statistical estimation framework.

For the second stage degradation (ϕ_c), we assume ink which is 100% opaque, i.e. that in the observation field a recto text pixel totally covers the corresponding verso pixel, whereas a recto background pixel does not. Combining the two, we can write the likelihood as follows:

$$\begin{aligned} P(d|f^1, f^2) &= \prod_s G(d_s;\ \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) \\ &= \prod_s \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}_s|^{1/2}} \\ &\quad \exp\left\{-\frac{1}{2}(d_s - \boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_s^{-1}(d_s - \boldsymbol{\mu}_s)\right\} \end{aligned} \tag{4}$$

where $\boldsymbol{\mu}_s$ is the mean for class f_s and $\boldsymbol{\Sigma}_s$ is the covariance matrix for class f_s given as follows:

$$\boldsymbol{\mu}_s = \begin{cases} \boldsymbol{\mu}_r & \text{if } f_s^1 = \textit{text} \\ \boldsymbol{\mu}_v & \text{if } f_s^1 = \textit{background} \text{ and } f_s^2 = \textit{text} \\ \boldsymbol{\mu}_{bg} & \text{else} \end{cases}$$

$$\boldsymbol{\Sigma}_s = \begin{cases} \boldsymbol{\Sigma}_r & \text{if } f_s^1 = \textit{text} \\ \boldsymbol{\Sigma}_v & \text{if } f_s^1 = \textit{background} \text{ and } f_s^2 = \textit{text} \\ \boldsymbol{\Sigma}_{bg} & \text{else} \end{cases}$$

where $\boldsymbol{\mu}_r, \boldsymbol{\mu}_v$ and $\boldsymbol{\mu}_{bg}$ are, respectively, the means for the recto class, the verso class and the background class, and the covariances are denoted equivalently.

4 The posterior probability and its maximization

Applying the Bayes rule to (2) we get the posterior probability of the two label fields:

$$\begin{aligned} P(f^1, f^2|d) &= \frac{1}{Z} P(f^1, f^2) P(d|f^1, f^2) \\ &= \frac{1}{Z} P(f^1) P(f^2) P(d|f^1, f^2) \end{aligned} \tag{5}$$

Combining (3), (4) and (5) we can see that the posterior probability is a MRF on the same neighborhood as the prior MRF and with the following energy potential function:

$$\begin{aligned} & U^P(f_s^1, f_{N_s}^1, f_s^2, f_{N_s}^2) \\ = & \sum_{\{s\}\in\mathcal{C}_1} \alpha^1 f_s^1 + \sum_{\{s,s'\}\in\mathcal{C}_2} \beta_{s,s'}^1 \gamma(f_s^1, f_{s'}^1) \\ + & \sum_{\{s\}\in\mathcal{C}_1} \alpha^2 f_s^2 + \sum_{\{s,s'\}\in\mathcal{C}_2} \beta_{s,s'}^2 \gamma(f_s^2, f_{s'}^2) \\ + & \frac{1}{2}(d_s - \boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_s^{-1}(d_s - \boldsymbol{\mu}_s) \end{aligned} \tag{6}$$

To estimate the binary image, equation (5) must be maximized. Unfortunately, the function is not convex and standard gradient descent methods will most likely return a non global solution. Simulated Annealing has been proven to return the global optimum under certain conditions [12].

Simulated Annealing received its name from physical processes, which decrease temperatures to allow particles (e.g. atoms in an alloy) to relax into a low energy configuration. Similarly, for the optimization of a non-convex function, the simulated annealing process lowers a — virtual — temperature factor. During the annealing process, pixels of the estimated binary fields are flipped in order to bring the estimations closer to the model. However, a certain amount of randomness is included in the optimization process, which allows the system to flip to more unfavorable estimates at certain times. This amount of randomness depends on the temperature factor: it is set relatively high at the beginning to allow the system to "jump" out of local minima, and is gradually lowered together with the temperature factor.

More precisely, during the annealing process, for each pixel the energy potential is calculated before and after chosing a new state. The decision whether to keep the new state or not is based on the value

$$q = e^{-\Delta/T} \tag{7}$$

where Δ is the difference of the posterior energy potentials (6) before and after the change. If $q > 1$ then the change is favorable and accepted. If $q < 1$ then it is accepted with probability q, which depends on the global temperature factor T. For the cooling schedule we used the suggestions in [11] (page 356), where the temperature T is set to

$$T^{(k)} = T^{(1)} \cdot c^{k-1}$$

where c is a constant controlling the speed of the cooling process and k denotes the current iteration. The start temperature must be sufficiently high to switch to energetically very unfavorable states with a certain probability.

In our case, a concurrent estimation of two fields is necessary, therefore there is not one but several possible state changes for each pixel:

- change pixel f_s^1
- change pixel f_s^2
- change pixels f_s^1 and f_s^2

At each iteration, a random state change is chosen and its energy change is evaluated by (7). A summary of the annealing algorithm is given in figure 3.

Figure 3: Simulated annealing and Gibbs sampler for two label fields.

Input: f^1, f^2 (initialized label fields), $T^{(1)}$ (start temperature), C (cooling speed), k_{max} (number of iterations)
Output: f^1, f^2 (estimated label fields)

for $k \leftarrow 1$ **to** k_{max} **do**
  $T \leftarrow T^{(1)} \cdot C^{k-1}$
  for $s \leftarrow 1$ **to** m **do**
    $\mathcal{E}_b \leftarrow U^P(f_s^1, f_{N_s}^1, f_s^2, f_{N_s}^2)$
    $\mathcal{E}_1 \leftarrow U^P(f_s'^1, f_{N_s}^1, f_s^2, f_{N_s}^2)$
    $\mathcal{E}_2 \leftarrow U^P(f_s^1, f_{N_s}^1, f_s'^2, f_{N_s}^2)$
    $\mathcal{E}_{12} \leftarrow U^P(f_s'^1, f_{N_s}^1, f_s'^2, f_{N_s}^2)$
    $\mathcal{E}_a \leftarrow$ random choice among $\{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_{12}\}$
    $q \leftarrow e^{-(\mathcal{E}_a - \mathcal{E}_b)/T}$
    if $q > 1$ **then**
      flip pixel s according to the choice of $\mathcal{E}_a$
    else
      flip pixel s according to the choice of $\mathcal{E}_a$ with probability q
    end
  end
end

5 Parameter estimation

Since realizations of the label fields F^1 and F^2 are not available, the parameters of the prior model and the observation model must be estimated from the observed data or from intermediate estimations of the label fields. In this work we chose to estimate the parameters in a supervised manner on the median filtered label fields. Alternatives would be, for instance, iterated conditional estimation [4] or the mean field theory [29].

5.1 The MRF parameters

For the supervised estimation of the MRF parameters we use least squares estimation, which was first proposed by Derin et al. [6]. For a single MRF the estimation procedure may be described as follows.

The potential function for a single site s may be given as

$$U(f_s, f_{N_s}, \theta_p) = \theta_p^T N(f_s, f_{N_s}) \tag{8}$$

where N_s are the intra-field neighbors of s: $N_s = \{f_{we}, f_{ea}, f_{no}, f_{so}, f_{nw}, f_{ne}, f_{sw}, f_{se}\}$ (see figure 1), θ_p is the prior parameter vector and $N(f_s, f_{N_s})$ can be derived from (3) as follows:

$$N(f_s, f_{N_s}) = [\ \begin{array}{l} \delta_{f_s,1}, \\ \gamma(f_s, f_{we}) + \gamma(f_s, f_{ea}) \\ \gamma(f_s, f_{no}) + \gamma(f_s, f_{so}) \\ \gamma(f_s, f_{ne}) + \gamma(f_s, f_{sw}) \\ \gamma(f_s, f_{nw}) + \gamma(f_s, f_{se}) \end{array} \ ]^T$$

where $\delta_{i,j}$ is the Kronecker delta given as

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

From (8) and the basic definition of conditional probabilities on MRFs:

$$P(f_s|N_s) = \frac{e^{-U(f_s, f_{N_s}, \theta_p)}}{\sum_{f_s \in \mathcal{L}} e^{-U(f_s, f_{N_s}, \theta_p)}}$$

the following relationship can be derived [6]:

$$\theta_p^T[N(f'_s, f_{N_s}) - N(f_s, f_{N_s})] = \ln\left(\frac{P(f_s, f_{N_s})}{P(f'_s, f_{N_s})}\right) \quad (9)$$

where f'_s and is a label different of f_s. The RHS of (9) can be estimated using histogram techniques [6], counting the number of occurrences of the clique labellings in the label field. Considering all possible combinations of f_s, f'_s and f_{N_s}, (9) represents an over determined system of linear equations which can be rewritten in matrix form as follows:

$$\boldsymbol{N}\boldsymbol{\theta}_p = \boldsymbol{p} \quad (10)$$

where $\boldsymbol{N}$ is a $M \times 6$ matrix, M being the number of data points, i.e. the number of combinations of L_1, L_2 and f_{N_s}. The rows of $\boldsymbol{N}$ contain the transposed vectors $[N(L_1, f_{N_s}) - N(L_2, f_{N_s})]^T$. The rows of the vector $\boldsymbol{p}$ contain the corresponding values from the RHS of (9). The system (10) can be solved using standard least squares techniques, as for instance the pseudo inverse.

For practical purposes, note that labeling pairs with one or both of the probabilities $P(f_s, f_{N_s})$ and $P(f'_s, f_{N_s})$ equal zero cannot be used. Furthermore, Derin et al. suggest to discard equations with low labeling counts in order to make the estimation more robust.

5.2 Stability issues

In the case of document images we noticed an additional problem: the resulting clique sample is not representative enough. In particular the first order statistics seem to be skewed which severely affects the estimation of the single clique parameter α. We therefore decided to estimate this parameter directly from the histogram. Assuming second order cliques with zero energy and from (3) we can write the energy for a single site f_s:

$$U(f_s) = \alpha f_s$$

From (1) we can estimate the potential corresponding to a given probability for site s having label 1:

$$\alpha \;=\; -\ln P(f_s) \;=\; -\ln\left[\frac{\text{n.o. set pixels}}{\text{n.o. pixels}}\right]$$

The other parameters are estimated with Derin et al.'s procedure. The already estimated parameters for the single clique cliques are injected into the system. Instead of solving (10), the following system is solved:

$$\boldsymbol{N}_u\boldsymbol{\theta}_u = \boldsymbol{p} - \boldsymbol{N}_k\boldsymbol{\theta}_k$$

where $\boldsymbol{N}_u$, $\boldsymbol{\theta}_u$, $\boldsymbol{N}_k$ and $\boldsymbol{\theta}_k$ are, respectively, the known (already estimated) and unknown parts of $\boldsymbol{N}$, $\boldsymbol{\theta}_p$:

$$\boldsymbol{\theta}_p \;=\; \begin{bmatrix} \boldsymbol{\theta}_k \\ \boldsymbol{\theta}_u \end{bmatrix} \qquad \boldsymbol{N}_k \;=\; [\, \boldsymbol{N}_k \; \boldsymbol{N}_u \,]$$

5.3 The double MRF case

Adapting the estimation procedure for a double MRF is straight forward. We estimate the parameters on the recto field only, since this field is more stable — all its labels are directly related to the observation field. The parameters of the verso field are directly calculated from the parameters of the recto field based on the assumption that, statistically speaking, the verso field is a flipped version of the recto field. In this case the first order statistics stay the same, while some second order statistics are affected:

$$\begin{array}{rcl} \alpha^2 & = & \alpha^1 \\ \beta_1^2 & = & \beta_1^1 \\ \beta_2^2 & = & \beta_2^1 \\ \beta_3^2 & = & \beta_4^1 \\ \beta_4^2 & = & \beta_3^1 \end{array}$$

where the $\beta_i, i = 1..4$ are, respectively, the pairwise clique parameters for horizontal, vertical, right-diagonal and left-diagonal cliques. Basically the two diagonal clique parameters are exchanged. This operation mostly concerns documents containing cursive handwriting and skewed lines.

5.4 The parameters of the observation model

The parameters of the observation model are estimated using the classical maximum likelihood estimators (the empirical means and covariances):

$$\begin{array}{rcll} \hat{\boldsymbol{\mu}}_i & = & \frac{1}{N}\sum_{s\in S_i} d_s & \\ \hat{\boldsymbol{\Sigma}}_i & = & \frac{1}{N}\sum_{s\in S_i} (d_s - \boldsymbol{\mu}_i)(d_s - \boldsymbol{\mu}_i)^T & i \in \{r, v, bg\} \end{array}$$

where S_i is the set of sites which has label i. Note that for the label of a site the information of both label fields is used.

6 Initialisation of the label fields

The iterative algorithm described in the previous sections needs to be initialized. More specifically, an initial estimation of the two label fields f^1 and f^2 is needed. A natural choice is to apply a segmentation technique without regularization, e.g. a k-means segmentation, in order to classify the pixels into three clusters. Then we determine for each cluster whether it is background, recto or verso. For most images that could be done using gray level information only, background being the lightest cluster and recto being the darkest one. However, this fails for some images, e.g. the ones where the text on the verso side is printed in very much darker color than the one on the recto side. We therefore developed a cluster labeling method which does not use the gray level of the pixels. Instead, it is based on the following two assumptions, in our opinion much more justified than assumptions on the color or gray values of the clusters:

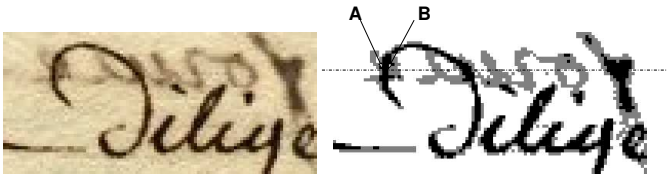


Figure 4: (a) an input image; (b) the result of the k-means clustering. As described by assumption 2, the word in the upper part of the image, which belongs to the verso side of the document, has been cut into several connected components by a letter written on the recto side.

Assumption 1 *Most space on the document page is occupied by background.*

Assumption 2 *The ink is 100% opaque and therefore a recto text pixel completely covers a verso pixel.*

The first assumption is used to determine the background cluster as the one having most pixels, which is rather straightforward and very efficient. The second assumption is used to determine which one of the two remaining cluster labels — henceforth denoted label a and label b — is the recto label. The basic idea is the following: since recto pixels cover verso pixels, connected components in the (unobservable) verso label field are often cut into several connected components in the observation field when they interact with connected components from the (unobservable) recto label field. Since we do not have the unobservable label fields — which would make the task trivial — we use histogram statistics on the initial segmentation to exploit this fact.

First we perform a connected components analysis on the k-means clustered observation image. Then we search for all the places in the image where the two labels interact, i.e. where there are two neighboring pixels (p_a, p_b), one having label a and the other having label b. We then get the corresponding connected component for each the two neighboring pixels. If label a corresponds to the recto label, then the corresponding connected component tends to be a whole letter or even a whole word, whereas the connected component corresponding to label b — the verso label — is very often only a part of a letter or a word, which has been cut into several components by the recto component (see figure 4). We can exploit this fact by numerating all connected components and collecting them in two different sets, S_a for the components of label a and S_b for the components of label b. The id of a connected component is inserted in a set for each transition which involves the connected component and the label of the corresponding set. Note, that these are sets in the mathematical meaning, i.e. they do not contain multiple elements.

As an example, consider the two transitions indicated by the two points A and B in figure 4. For these two transitions, 2 connected components are inserted into the set corresponding to the gray-ish label, while only one connected component is inserted into the set corresponding to the darker label.

In order to make the algorithm more robust against spurious noise, we perform the operations on a median filtered version on the image and we only take into account label transitions where the size of each involved connected component exceeds a certain threshold T. Setting T to around 20 pixels assures that only characters and parts of characters are considered. After a full traversal of the image, the recto label can be determined as the label having the minimum number of component ids. A summary of the algorithm in a scanline order version is found in figure 5.

Figure 5: Scanline algorithm for the recognition of the recto and verso cluster labels.

Input: z (the k-means clustered observation field), a, b (the two non-background cluster labels), T (a threshold)
Output: l_r (the label of the recto cluster)

$S_a, S_b \leftarrow \{\}$
Connected component analysis(z)
foreach *scanline in* z **do**
  foreach *pixel in scanline* **do**
    if *label change from* a *to* b *or from* b *to* a **then**
      $C_a \leftarrow$ component with label a
      $C_b \leftarrow$ component with label b
      if $|C_a| > T$ *and* $|C_b| > T$ **then**
        $S_a \leftarrow S_a \cup C_a$
        $S_b \leftarrow S_b \cup C_b$
      end
    end
  end
end
$l_r \leftarrow \arg\min_{i=a,b} |S_i|$

7 Restoration

The principle of the restoration algorithm is simple: replace the color or gray value of the pixels classified as verso by the color or gray value of the background. Directly using the mean of the background class will produce visible artifacts due to the noise in the image. A better solution is to use the mean of the neighboring pixels classified as background.

Searching these pixels might be laborious in cases where we need to fill larger areas of verso pixels. We therefore propose a hierarchical pyramid structure for the calculation of the replacement values. The pyramid is characterized by a 2×2 reduction function and a receptive field of 3×3 children for each parent site (see figure 6). Each site s of the hierarchical structure contains 2 values: O_s is an estimation of the background color at this site and M_s is the count of observed background pixels used for the calculation of M_s. The base level of the pyramid is initialized as follows:

$$M_s = \begin{cases} 1 & f_s^1 = \textit{background} \text{ and } f_s^2 = \textit{background} \\ 0 & \text{else} \end{cases}$$

$$O_s = d_s$$

The parent levels are calculated from their child levels:

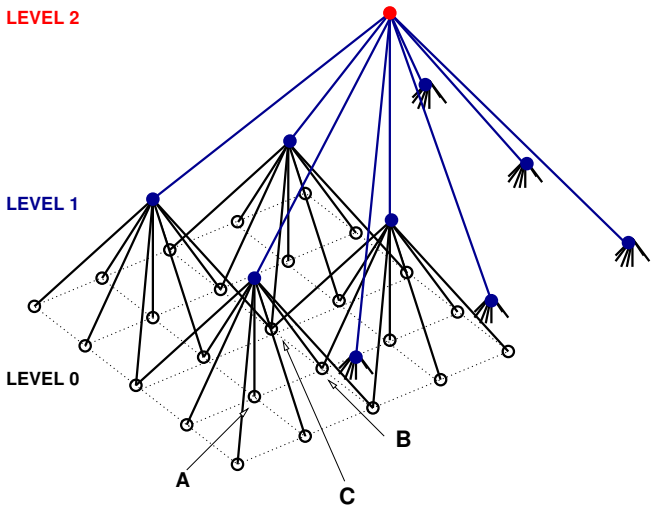


Figure 6: The pyramid structure for the calculation of the replacement color or gray value.

$$M_s = \sum_{s' \in s_-} M_{s'}$$

$$O_s = (\sum_{s' \in s_-} M_{s'})^{-1} \sum_{s' \in s_-} M_{s'} O_{s'}$$

where s_- is the set of children of site s. In other words, M_s is the weighted mean of the available observation sites where each value O_s is weighted by the number M_s of source observation pixels which influenced it. Consequently, at a given site, O_s is an approximation of the mean of the observed background pixels included in it's support at the base of the pyramid, since the calculation of M_s is an approximation of the true number of pixels.

The aim of the restoration process is to set the verso pixel to a color or gray value which is calculated using a minimum of T values, where setting the parameter T to 4-5 pixels is sufficient for a robust estimation. Intuitively speaking, after building the pyramid, the segmented image (i.e. the base level of the pyramid) is traversed and the replacement value for each verso pixel can be taken from the parent site or going up even further in the pyramid if the parent site does not contain enough background pixels.

More precisely, depending on its position on the grid, each site s may have 1, 2 or 4 parents, as can be seen at the example points A, B and C, respectively, in figure 6. In the following, we denote the set of parents of a site s by s^-. Instead of simply climbing the pyramid and chosing a value, we will need to combine the values of the parents, e.g. calculating the mean. If the combined number of used observation pixels does not fulfill the requirement of being greater than T, then we continue to the climb the pyramid. The algorithm is given in figure 7.

Figure 7: Calculation of the gray value or color replacing a verso pixel.

Input: M, O (the restoration pyramid), s (the site of the pixel to replace)
Output: V (the gray value or color of the replacement)

$F \leftarrow \{s^-\}$ {F is a queue}
while $\sum_{s' \in F} M_{s'} < T$ **do**
  $s' \leftarrow$ first element in F
  $F \leftarrow F \setminus s'$
  $F \leftarrow F \cup s'^-$ {add at the end}
end
$V = (\sum_{s' \in F} M_{s'})^{-1} \sum_{s' \in F} M_{s'} O_{s'}$

8 Experimental results

Evaluating document restoration algorithms is a non trivial task since ground truth is very hard to come by. Short of manually classifying each pixel in a scanned image, the only way to get reliable ground truth data is to test the algorithm on synthetic data. These tests, on the other hand, may not be realistic enough to capture all the subtleties of a real environment. To evaluate our algorithm we therefore decided to perform tests on synthetic data with ground truth as well as real data. The results on the latter have been evaluated applying a statistical test on empirical evaluation results.

In all cases we performed the experiments on gray scale images only. If the images were available in color, we transformed them to gray scale first.

8.1 Experiments on synthetic images

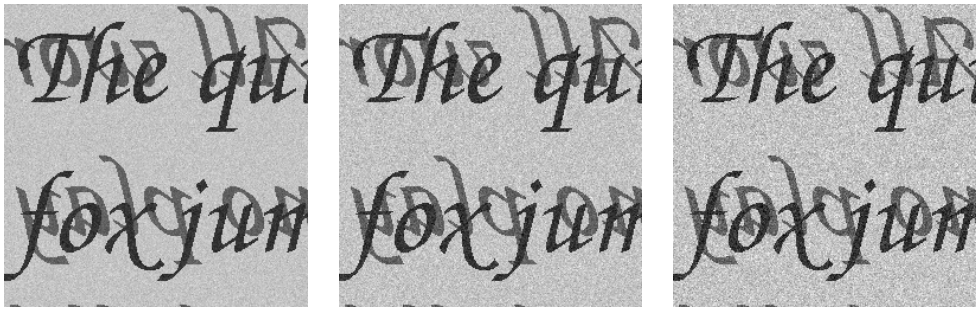
We created synthetic images according to the degradation model described in section 3. Two perfect images have been superimposed and Gaussian noise with different variances has been added (see figure 8a).

Table 1 shows the results of our algorithm as well as set of other algorithms : a simple k-means algorithm, a MRF segmentation with a single label field, a double MRF segmentation with normal least squares estimation and the double MRF segmentation where the parameter α has been estimated directly from the histogram (see section 5.2). The methods are tested against 2 different types of ground truth :

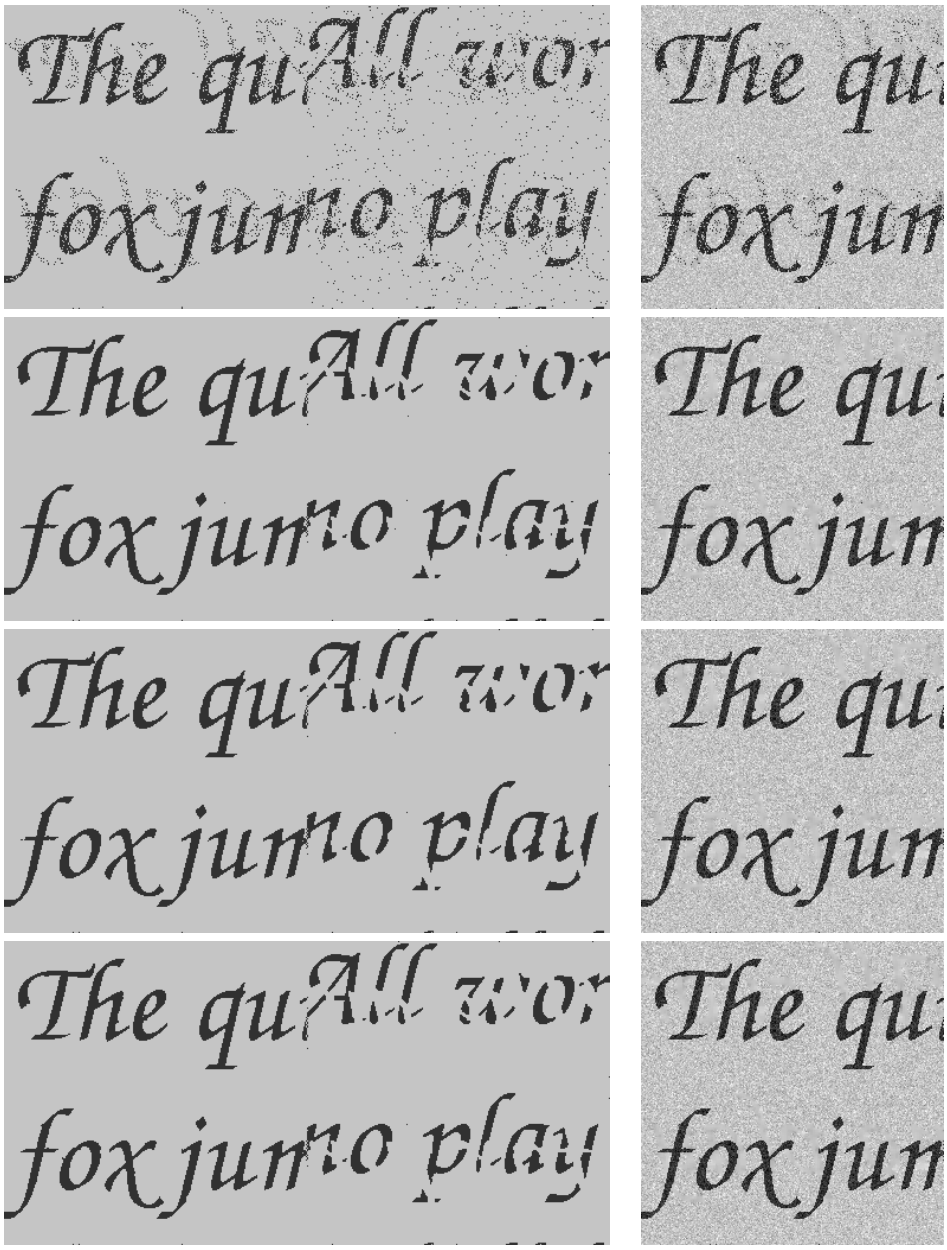
4 classes GT	the four classes are: *recto, verso, background* and *recto-and-verso*
3 classes GT	the three classes are: *recto, verso* and *background.*

Only the double MRF methods can be evaluated against the 4 class ground truth as the other methods are not capable of estimating the *recto-and-verso* class. On the other hand, the double MRF methods can't be evaluated directly against the 3 class ground truth. However, an estimation is possible if the segmentation output is transformed into a 3 class output by replacing all *recto-and-verso* pixels with *recto* pixels (c.f. the third column of table 1). Note that this evaluation is the most relevant one, since it measures the performance of the restoration process, during which *recto-and-verso* pixels are treated just as recto pixels.

In a direct (that is, "unfair") comparison the the performance of the 4 class double MRF method is lower than the ones of the two three class methods. This is expected, since the task of estimating 4 classes is much more difficult,



(a)



(b)

Figure 8: (a) the synthetic input images used in the experiments, created with Gaussian noise and respective variances of $\sigma = 10, 15, 20$; (b) segmentation (left) and restoration results (right) on the synthetic image with $\sigma = 20$. From top to bottom: k-means, single MRF, double MRF, double MRF with partial Derin et al.

Nr. of classes	gt:4	gt:3	gt:3 seg→3
K-Means (k=3)	n.a.	0.25	0.25
Single MRF	n.a.	0.03	0.03
Double MRF	2.00	n.a.	**0.01**
Double MRF-ParD	2.13	n.a.	**0.01**

(a)

Nr. of classes	gt:4	gt:3	gt:3 seg→3
K-Means (k=3)	n.a.	1.40	1.40
Single MRF	n.a.	0.23	0.23
Double MRF	2.04	n.a.	0.10
Double MRF-ParD	2.25	n.a.	**0.08**

(b)

Nr. of classes	gt:4	gt:3	gt:3 seg→3
K-Means (k=3)	n.a.	3.56	3.56
Single MRF	n.a.	0.73	0.73
Double MRF	2.60	n.a.	0.46
Double MRF-ParD	2.46	n.a.	**0.31**

(c)

Table 1: Classification error (in %) against ground truth with different numbers of classes and on synthetic images with different amount of noise: (a) $\sigma = 10$; (b) $\sigma = 15$; (c) $\sigma = 20$.

especially since the *recto-and-verso* class is unobserved and can only be estimated through the spatial interaction, thus through the Markov prior. However, looking at the "fair" comparison shown in the third column of table 1, where all methods are evaluated against the same 3 class ground truth, we see that the double MRF method outperforms the other methods.

This positive result confirms the objectives of the double MRF prior described in section 2, namely the increase of the regularization performance due to two different facts: the adaptation of the prior to the contents creation process, and the improvement of the estimation of the verso pixels through an estimation of the verso pixels covered by recto pixels, which is only possible with two different label fields.

8.2 Experiments on scanned document images

The method has also been tested on real document images as shown in figures 9 and 10. As can be seen, the MRF regularization is capable of removing many artifacts present in the k-means segmented image. The double MRF method further decreases the number of artifacts. There is no ground truth for these kind of images, so we presented the images in figures 9 and 10 to 16 different people (of course after randomly shuffling the result images) and let them rank the 3 result images by perceived quality. The results of these $N = 64$ tests (16 people evaluated 4 images each) is shown in table 2a. Our method has been ranked first 33 times against 18 times and 13 times for the K-Means and the single MRF, respectively.

It might be surprising that the K-Means algorithm has been ranked first more often than the single MRF algorithm

Method	Ranked 1	Ranked 2	Ranked 3
K-Means	18	10	36
Single-MRF	13	39	12
Double-MRF	33	15	16
Total	64	64	64

(a)

Method	Ranked 1	Ranked 2	Ranked 3
K-Means	21	43	0
Double-MRF	43	21	0
Total	64	64	0

(b)

Table 2: Results of the emperical tests on real data: (a) the complete results; (b) the results ignoring the method *Single-MRF*.

since its output is visibly more noisy than the images produced by the regularized methods. However, apparently the good ranking is due to the fact, that the K-Means algorithm tends to keep more recto pixels than the single MRF one.
To test the statistical significance of this result, more particularly of the result "The method *Double MRF* is ranked first 33 times, therefore it outperforms the other methods", let us assume the following hypothesis:

H_0 **(null hypothesis)**	The method *Double MRF* is as efficient as the other two methods.
H_A **(alternative)**]	The method *Double MRF* is either more or less efficient as one or both other two methods.

We can conclude from the data that the method is not less efficient, it suffices therefore to reject the null hypothesis. Our test statistics will be U = the number of times the method *Double MRF* is ranked first. Assuming H_0, the probability for a given method to be ranked first for a given image is $\pi = \frac{1}{3}$, U is therefore distributed Binomial, more precisely $U \sim B(N, \pi)$. The probability of the actual value of $U = 33$ is given as:

$$P(U = 33) = \left(\begin{array}{c} N \\ 33 \end{array} \right) \pi^p (1-\pi)^{N-3} = 0.00111$$

Given a standard significance level of $\alpha = 0.05$, the null hypothesis is therefore rejected.

This only proves that the method is better than *one* or both of the alternatives, not that the method is better than the second ranked one. This can be examined by looking at the results after ignoring the method ranked as third, shown in figure 2b, and creating a new null hypothesis:

H_0 **(null hypothesis)**	The method *Double MRF* is as efficient as the method *K-means*.
H_A **(alternative)**]	The method *Double MRF* is either more or less efficient as the method *K-means*.

Again, our test statistics U will be the number of times the method *Double MRF* has been ranked first, the actual value

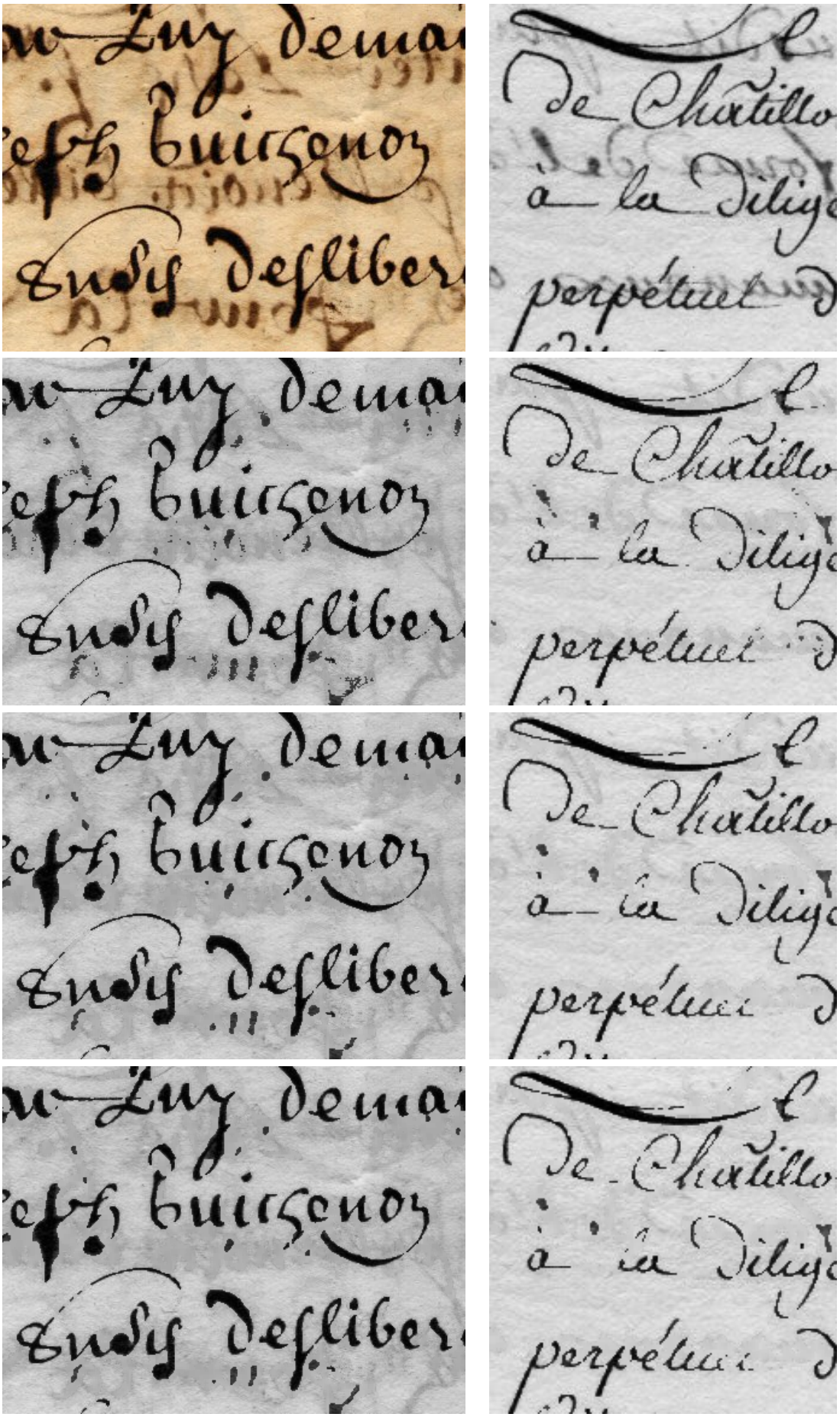


Figure 9: Restoration results on real data. From top to bottom: input image, k-means, single MRF, double MRF.

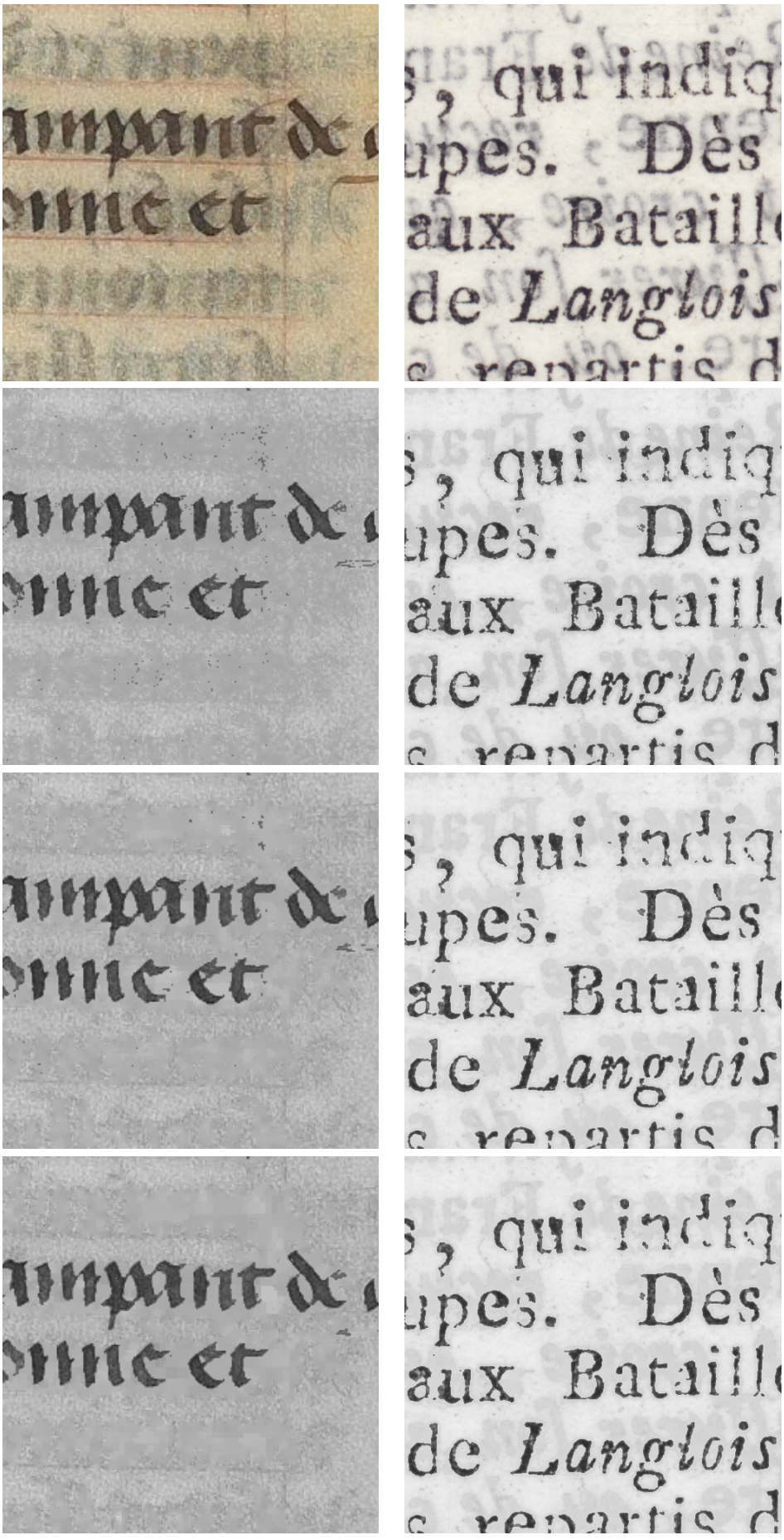


Figure 10: Restoration results on real data. From top to bottom: input image, k-means, single MRF, double MRF.

being $U = 43$. Again U is distributed Binomial, this time with parameters $N = 64$ and $p = 0.5$. The probability of the actual value is $P(U = 43) = 0.00222$, which means that the null hypothesis is again rejected. The *Double MRF* is thus more efficient than the *K-Means*.

9 Conclusion and Outlook

In this paper we presented a method to separate the verso side from the recto side of a single scan of document images. The novelty of the method is the separation of the MRF prior into two different label fields, each of which regularizes one of the two sides of the document. This separation allows to estimate the verso pixels of the document which are covered by the recto pixels, which, again through the MRF prior, improves the estimation of the verso pixels not covered by recto pixels, thus increasing the performance of the regularization.

The performance of the method has been evaluated using synthetic images with known ground truth as well as scanned document images. The latter experiments have been evaluated using empirical tests performed by 16 different people. Statistical tests have been carried out to check the significance of the results.

Involved in several digitization projects around the world, our team is currently looking into the following perspectives of this work:

- Creation of a homogeneous (into document analysis terms: adaptive) observation model, which increases the performance on larger images. This model needs to take into account several text colors, as well as the page bending process and other different kinds of degradation (see section 1).
- Creation of a hierarchical Markov model which is able to take into account larger neighborhood structures. Hierarchical models do exist and are widely used (e.g. [3][17][15]), the description of their various shortcomings is beyond the scope of this work.
- Creation of a discriminative model, as for instance a conditional random field [18][16] adapted to the nature of the problem.

References

[1] H.S. Baird. Document image defect models and their uses. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 62–67, 1993.

[2] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36(2):192–236, 1974.

[3] C.A. Bouman and M. Shapiro. A Multiscale Random Field Model for Bayesian Image Segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, 3 1994.

[4] B. Braathen and W. Pieczynski. Global and Local Methods of Unsupervised Bayesian Segmentaiton of Images. *Machine Graphics and Vision*, 2(1):39–52, 1993.

[5] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computated imaging. *IEEE Transactions on Imagea Processing*, 6(2):298–311, 1997.

[6] H. Derin and H. Elliott. Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):39–55, 1987.

[7] H.-S. Don. A noise attribute thresholding method for document image binarization. *International Journal on Document Image Analysis and Recognition*, 4(2):131–138, 2000.

[8] K. Donaldson and G.K. Myers. Bayesian super-resolution of text in video with a text-specific bimodal prior. *International Journal on Document Analysis and Recognition*, 7(2-3):159–167, 2005.

[9] F. Drira, F. LeBourgeois, and H. Emptoz. Restoring ink bleed-through degraded document images using a recursive unsupervised classification technique. In *Proceedings of the 7th Workshop on Document Analysis Systems*, pages 38–49, 2006.

[10] E. Dubois and A. Pathak. Reduction of bleed-through in scanned manuscript documents. In *Proceedings of the Image Processing, Image Quality, Image Capture Systems Conference*, pages 177–180, 2001.

[11] R. Duda, P. Hart, and D. Stork. *Pattern Classification, 2nd Edition*. Wiley, New York, NY, November 2000.

[12] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 11 1984.

[13] J.M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. unpublished manuscript, 1968.

[14] T. Kanungo and R.M. Haralick an I. Philips. Global and local document degradation models. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 730–734, 1993.

[15] Z. Kato, M. Berthod, and J. Zerubia. A hierarchical markov random field model and multitemperature annealing for parallel image classification. *Graphical Models and Image Processing*, 58(1):18–37, 1996.

[16] S. Kumar and M. Hebert. Discriminative random fields. *International Journal of Computer Vision*, 68(2):179–201, 2006.

[17] J.-M. Laferte, P. Perez, and F. Heitz. Discrete markov image modelling and inference on the quad tree. *IEEE Transactions on Image Processing*, 9(3):390–404, 2000.

[18] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling data. In *International Confernence on Machine Learning*, 2001.

[19] Y. Leydier, F. LeBourgeois, and H. Emptoz. Serialized Unsupervised Classifier for Adaptative Color Image Segmentation: Application to Digitized Ancient Manuscripts. In *Internationl Conference on Pattern Recogniton*, pages 494–497, 2004.

[20] S.Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer Verlag, 2001.

[21] H. Nishida and T. Suzuki. Correcting show-through effects on document images by multiscale analysis. In *Proceedings of the International Conference on Pattern Recognition*, volume 3, pages 65–68, 2002.

[22] M.I. Sezan and A.M. Tekalp. Survey of recent developments in digital image restoration. *Optical Engineering*, 29(5):393–404, 1990.

[23] G. Sharma. Show-through cancellation in scans of duplex printed documents. *IEEE Transactions on Image Processing*, 10(5):736–754, 2001.

[24] C.L. Tan, R. Cao, and P. Shen. Restoration of archival documents using a wavelet technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1399–1404, 2002.

[25] A. Tonazzini and L. Bedini. Independent component analysis for document restoration. *International Journal on Document Analysis and Recognition*, 7(1):17–27, 2004.

[26] A. Tonazzini, S. Vezzosi, and L. Bedini. Analysis and recognition of highly degraded printed characters. *International Journal on Document Analysis and Recognition*, 6(4):236–247, 2003.

[27] Q. Wang, T. Xia, C.L. Tan, and L. Li. Directional wavelet approach to remove document image interference. In *Proceedings of the 7th International Conference on Document Analysis and Recognition*, pages 736–740, 2003.

[28] C. Wolf and D. Doermann. Binarization of Low Quality Text using a Markov Random Field Model. In *Proceedings of the International Conference on Pattern Recognition*, volume 3, pages 160–163, 2002.

[29] J. Zhang. The mean field theory in em procedures for markov random fields. *IEEE transactions on image processing*, 40(10):2570–2583, 1992.

[30] Q. Zheng and T. Kanungo. Morphological degradation models and their use in document image restoration. In *Proceedings of the International Conference on Image Processing*, volume 1, pages 193–196, 2001.