Better restore the recto side of a document with an estimation of the verso side: Markov model and inference with graph cuts

Christian Wolf

Laboratoire d'InfoRmatique en Image et Systèmes d’information
LIRIS UMR 5205 CNRS / Université de Lyon
INSA de Lyon, bâtiment J. Verne
20, Avenue Albert Einstein - 69622 Villeurbanne cedex
http://liris.cnrs.fr
Plan

1. Introduction
2. The prior model
3. The observation model
4. Iterative graph cuts - the posterior probability and its maximization
5. Pre- and postprocessing
6. Results
Plan

1 Introduction

2 The prior model

3 The observation model

4 Iterative graph cuts - the posterior probability and its maximization

5 Pre- and postprocessing

6 Results
The problem

**Blind** ink bleed-through removal (scans of the verso side are **not** available).

The verso contents must be replaced by background.
Problem: **image segmentation.** The classification decision for each pixel is based on
- the color/gray value of each pixel
- the **local structure** of the image.

The framework: Bayesian estimation with two hidden Markov random fields (MRF)
Bayesian estimation

PRIOR KNOWLEDGE

LIKELIHOOD OF THE DATA GIVEN THE HIDDEN VARIABLES

ESTIMATION
Plan

1. Introduction
2. The prior model
3. The observation model
4. Iterative graph cuts - the posterior probability and its maximization
5. Pre- and postprocessing
6. Results
A classical hidden Markov random field

\[
P(f) = \frac{1}{Z} \exp \left\{ - U(f) / T \right\}
\]
**Why two hidden label fields?**

- The prior should regularize fields which directly correspond to the natural process “creating” the contents.
- 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.
The double hidden Markov random field

\[ F^1 \setminus D \setminus F^2 \]

- Observed variable
- Hidden variable
- Clique of the Prior model
- Clique of the observ. model

Christian Wolf - http://liris.cnrs.fr/christian.wolf
The joint probability: dependence of recto and verso

\[ P(f^1, f^2, d) \]

\[ = \frac{1}{Z} \exp \left\{ - \left( U(f^1) + U(f^2) + U(f^1, f^2, d) \right) / T \right\} \]

\[ = \frac{1}{Z_1} \exp \left\{ - U(f^1, f^2) / T \right\} \cdot \frac{1}{Z_2} \exp \left\{ - U(f^1, f^2, d) / T \right\} \]

\[ = P(f^1, f^2)P(d|f^1, f^2) \]

\[ = P(f^1)P(f^2)P(d|f^1, f^2) \]

- w/o observations, \( f^1 \) and \( f^2 \) are independent
- w/ observations, they are dependent
The prior of a single field: the Potts model

\[ U(f) = \sum_{\{s\} \in C_1} \alpha f_s + \sum_{\{s, s'\} \in C_2} \beta_{s, s'} \delta_{f_s, f_{s'}} \]

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

- Discontinuity preserving
Plan

1. Introduction
2. The prior model
3. The observation model
4. Iterative graph cuts - the posterior probability and its maximization
5. Pre- and postprocessing
6. Results
The observation model (Likelihood)

The observation model characterizes the degradation process which is applied to the theoretical “perfect” image resulting in the observed image.
We suppose that there are two independent degradation processes for each side, followed by a combination of the two “virtual” observations:

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

The combination process \( \phi_c \) assumes 100% opaque ink: a recto pixel \( f^1_s = 1 \) completely covers the verso pixel \( f^2_s \), whatever its label.
A Gaussian observation model

\[
P(d|f^1, f^2) = \prod_s \mathcal{N}(d_s; \mu_s, \Sigma_s)
\]

\[
\mu_s = \begin{cases} 
\mu_r & \text{if } f^1_s = \text{text} \\
\mu_v & \text{if } f^1_s = \text{background} \text{ and } f^2_s = \text{text} \\
\mu_{bg} & \text{else}
\end{cases}
\]

\[
\Sigma_s = \begin{cases} 
\Sigma_r & \text{if } f^1_s = \text{text} \\
\Sigma_v & \text{if } f^1_s = \text{background} \text{ and } f^2_s = \text{text} \\
\Sigma_{bg} & \text{else}
\end{cases}
\]
The posterior probability and its maximization

The posterior probability and its maximization

\[ P(f^1, f^2 | d) = \frac{1}{P(d)} \cdot P(f^1, f^2) P(d | f^1, f^2) \]

\[ \propto P(f^1, f^2) P(d | f^1, f^2) \]

\[ U(f^1, f^2, d) \]

\[ = \sum_{\{s\} \in C_1} \alpha^1 f^1_s + \sum_{\{s, s'\} \in C_2} \beta^1_{s, s'} \delta f^1_s f^1_{s'} \]

\[ + \sum_{\{s\} \in C_1} \alpha^2 f^2_s + \sum_{\{s, s'\} \in C_2} \beta^2_{s, s'} \delta f^2_s f^2_{s'} \]

\[ + \sum_{\{s\} \in C_1} \frac{1}{2} (d_s - \mu_s)^T \Sigma_s^{-1} (d_s - \mu_s) \]
The posterior probability and its maximization

Rewrite of the energy in terms of unary functions $U_1$ and two types of binary functions $U_2$ and $U'_2$

$$U(f^1, f^2, d)$$

$$= \sum_{\{s\} \in C_1} \left[ \alpha_1 U_1(f^1_s) + \alpha_2 U_1(f^2_s) \right]$$

$$+ \sum_{\{s, s'\} \in C_2} \left[ \beta_1 U_2(f^1_s, f^1_{s'}) + \beta_2 U_2(f^2_s, f^2_{s'}) \right]$$

$$+ \sum_{\{s\} \in C_1} U'_2(f^1_s, f^2_s; d_s)$$
[Kolmogorov et al., PAMI 2004] : a function of binary variables composed of unary terms and binary terms is graph-representable, i.e. it can be minimized with algorithms based on the calculation of the maximum flow in a graph, if and only if each binary term $E(x, y)$ is regular, i.e. it satisfies the following equation:

$$E(0, 0) + E(1, 1) \leq E(0, 1) + E(1, 0)$$

Depending on the observation $d_s$ of site $s$ this may or may not be the case for the observation model.
Minimization with minimum cut/maximum flow

Iterative optimization algorithm:

Fix/estimate

- Initialize labels, e.g. with k-means
- repeat
  - Fix $f^1$, estimate optimal $f^2$
  - Fix $f^2$, estimate optimal $f^1$
- until happy/dead/running out of patience

- Sub problem can be solved exactly with minimum cut/maximum flow
- Equivalent to the $\alpha$-expansion move algorithm for a single MRF with adapted interaction potentials.
Joint estimation of a subset of the graph $G$

- Determine **regular** sites:
  $$H_s = 1 \iff U'_2(f^1_s, f^2_s, d_s)$$ is regular

- Initialize lables, e.g. with k-means

- repeat
  - Fix $f^1_s$ for all $s$ and $f^2_s$ for $H_s = 0$, estimate optimal $f^2$ for all $s$ and $f^2_s$ for $H_s = 1$
  - Fix $f^2_s$ for all $s$ and $f^1_s$ for $H_s = 0$, estimate optimal $f^1$ for all $s$ and $f^1_s$ for $H_s = 1$

- until happy/dead/running out of patience

$\implies$ larger standard move!
Minimization with minimum cut/maximum flow

Iterative algorithm: one field is fixed (e.g. $f^1$, the other field is estimated (e.g. $f^2$).

Regular sites ($H_s \leftarrow 1$) are jointly estimated.
The cut graph

*Cut graph $\alpha$-exp move*

*Cut graph proposed algorithm*
Plan

1. Introduction
2. The prior model
3. The observation model
4. Iterative graph cuts - the posterior probability and its maximization
5. Pre- and postprocessing
6. Results
Estimation of the model parameters

Estimation of all parameters from the median filtered initial segmentation (k-means).

- Estimation of the likelihood parameters: empirical mean and variance (covariance matrix).
- Estimation of the MRF parameters: least squares (Derin et al.)

Alternatives: Iterated Conditional Estimation, Expectation-Maximization, the meanfield theory etc.
Estimation of the MRF parameters

Least squares method [Derin et al., 1987]. Input data: pairs of different labels $f_s$ and $f_s'$ with the same neighborhood $N_s$

$$\theta_T^p [N(f'_s,f_{N_s}) - N(f_s,f_{N_s})] = \ln \left( \frac{P(f_s,f_{N_s})}{P(f'_s,f_{N_s})} \right)$$

$$N(f_s,f_{N_s}) = \begin{bmatrix} \delta_{f_s,0}, \\ \delta_{f_s,1}, \\ \gamma(f_s,f_{we}) + \gamma(f_s,f_{ea}) \\ \gamma(f_s,f_{no}) + \gamma(f_s,f_{so}) \\ \gamma(f_s,f_{ne}) + \gamma(f_s,f_{sw}) \\ \gamma(f_s,f_{nw}) + \gamma(f_s,f_{se}) \end{bmatrix}^T$$

Least squares solution:

$$N\theta_p = p \quad \theta_p = N^+ p \quad N^+ = (N^T N)^{-1} N^T$$
Estimation of the MRF parameters

Least squares method [Derin et al., 1987]. Input data: pairs of different labels $f_s$ and $f'_s$ with the same neighborhood $N_s$.

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

$$
N(f_s, f_{N_s}) = \begin{bmatrix}
\delta_{f_s,0}, \\
\delta_{f_s,1}, \\
\gamma(f_s, f_{we}) + \gamma(f_s, f_{ea}), \\
\gamma(f_s, f_{no}) + \gamma(f_s, f_{so}), \\
\gamma(f_s, f_{ne}) + \gamma(f_s, f_{sw}), \\
\gamma(f_s, f_{nw}) + \gamma(f_s, f_{se})
\end{bmatrix}^T
$$

Least squares solution:

$$
N\theta_p = p \quad \theta_p = N^+ p \quad N^+ = (N^T N)^{-1} N^T
$$
Least squares method [Derin et al., 1987]. Input data: pairs of different labels $f_s$ and $f'_s$ with the same neighborhood $N_s$

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

$$N(f_s, f_{N_s}) = \begin{bmatrix} \delta_{f_s,0} \\ \delta_{f_s,1} \\ \gamma(f_s, f_{we}) + \gamma(f_s, f_{ea}) \\ \gamma(f_s, f_{no}) + \gamma(f_s, f_{so}) \\ \gamma(f_s, f_{ne}) + \gamma(f_s, f_{sw}) \\ \gamma(f_s, f_{nw}) + \gamma(f_s, f_{se}) \end{bmatrix}^T$$

Least squares solution:

$$N\theta_p = p \quad \implies \quad \theta_p = N^+ p \quad \implies \quad N^+ = (N^TN)^{-1}N^T$$
Assumption

- Most space on the document page is occupied by background.
- The ink is 100% opaque and therefore a recto text pixel completely covers a verso pixel.
Replacement of the verso pixels

Leaves:

\[
M_s = \begin{cases} 
1 & f_s^1 = \text{backgr.} \\
0 & \text{else}
\end{cases}
\]

\[
O_s = d_s
\]

Parent nodes:

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

\[
O_s = \frac{\sum_{s' \in s_-} M_{s'} O_{s'}}{\left(\sum_{s' \in s_-} M_{s'}\right)}
\]

\[s_- : \text{children of site } s\]
Plan

1. Introduction
2. The prior model
3. The observation model
4. Iterative graph cuts - the posterior probability and its maximization
5. Pre- and postprocessing
6. Results
Experiments

3 types of experiments:
- Synthetic images: recall/precision on pixel level
- Real images, evaluation by humans: ranking of methods
- Real images, evaluation by OCR: Levenshtein cost and recall/precision on character level

2 types of optimization:
- Simulated annealing
- Boykov and Kolmogorov’s implementation of minimum cut/maximum flow
### Experiments on synthetic images

<table>
<thead>
<tr>
<th>Noise level</th>
<th>Nr. of classes</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma = 10$</td>
<td>K-Means (k=3)</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Single MRF</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Double MRF</td>
<td><strong>0.01</strong></td>
</tr>
<tr>
<td>$\sigma = 15$</td>
<td>K-Means (k=3)</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>Single MRF</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Double MRF</td>
<td><strong>0.08</strong></td>
</tr>
<tr>
<td>$\sigma = 15$</td>
<td>K-Means (k=3)</td>
<td>3.56</td>
</tr>
<tr>
<td></td>
<td>Single MRF</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Double MRF</td>
<td><strong>0.31</strong></td>
</tr>
</tbody>
</table>

Inference with simulated annealing
Experiments on real images

Input

Single MRF

K-means

Double MRF
Experiments on real images

Input

K-means

Single MRF

Double MRF
Experiments on real images

Input

K-means

Single MRF

Double MRF

Christian Wolf - http://liris.cnrs.fr/christian.wolf

Restore the recto by estimating the verso
Experiments on real images

Input

K-means

Single MRF

Double MRF

Christian Wolf - http://liris.cnrs.fr/christian.wolf

Restore the recto by estimating the verso
## Experiments on real images

<table>
<thead>
<tr>
<th>Method</th>
<th>Ranked 1</th>
<th>Ranked 2</th>
<th>Ranked 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Means</td>
<td>18</td>
<td>10</td>
<td>36</td>
</tr>
<tr>
<td>Single-MRF</td>
<td>13</td>
<td>39</td>
<td>12</td>
</tr>
<tr>
<td>Double-MRF</td>
<td>33</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
</tbody>
</table>

Binomial-test: P-value of 0.00197. Significance level of $\alpha = 0.025 \implies$ reject of $H_0$. 
Real images: input and k-means (graph cut)
bras de la Compagnie: Il étoit accompagné de son Oncle, Frère de sa Mère. Le jeune Prince avait pour motif de la fuite le désir de trouver les moyens de retirer son Père du triste état, où il étoit réduit.

bras de l:iCompagnie: Il étoit accompagné dé (on _
Oncle, Frère de fa Mere. Le ierzne Prince avoit pour motif de la fuite le delîr de trouver les mbvens 'dë`r: iter *);}>; Père du ttíftc état,. ou il. étoit réduitr;

bras de l:iCompagnie: Il étoit accompagné dé (on _
Oncle, Frère de sa Mère. Le jeune Prince avoit pour motif de la fuite le désir de trouver les moyens de retirer son Père du triste état, où il étoit réduit.
### Real images: Improvement of OCR results (graph cut)

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th>Precision</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>No restoration</td>
<td>84.25</td>
<td>60.71</td>
<td>80,512</td>
</tr>
<tr>
<td>K-Means (k=3)</td>
<td>90.19</td>
<td>73.26</td>
<td>47,611</td>
</tr>
<tr>
<td>Double MRF</td>
<td>90.45</td>
<td>75.57</td>
<td>42,314</td>
</tr>
</tbody>
</table>

### 1 page, inference with graph cuts. OCR: Google Tesseract.

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th>Precision</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>No restoration</td>
<td>91.65</td>
<td>70.72</td>
<td>502</td>
</tr>
<tr>
<td>Niblack</td>
<td>91.03</td>
<td>70.64</td>
<td>492</td>
</tr>
<tr>
<td>Sauvola</td>
<td>91.75</td>
<td>76.66</td>
<td>349</td>
</tr>
<tr>
<td>K-Means (k=3)</td>
<td>92.68</td>
<td>77.10</td>
<td>326</td>
</tr>
<tr>
<td>Source separation RGB</td>
<td>80.72</td>
<td>54.53</td>
<td>911</td>
</tr>
<tr>
<td>Single MRF</td>
<td>94.22</td>
<td>81.68</td>
<td>275</td>
</tr>
<tr>
<td>Double MRF</td>
<td>94.43</td>
<td>83.27</td>
<td>262</td>
</tr>
</tbody>
</table>
## Computational complexity

<table>
<thead>
<tr>
<th>Method</th>
<th>1026×1557</th>
<th>2436×3320</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Means (k=3)</td>
<td>3.2</td>
<td>18.9</td>
</tr>
<tr>
<td>Single MRF</td>
<td>9.1</td>
<td>40.6</td>
</tr>
<tr>
<td>Double MRF</td>
<td>12.6</td>
<td>67.5</td>
</tr>
</tbody>
</table>

Inference with graph cuts

Execution time in seconds (Pentium-M 1.86Mhz 1Gb RAM).
The double MRF model is more natural and more powerful than the single MRF model.

The double MRF model using binary labels naturally leads to an efficient graph cut implementation.

The MRF methods based on classification beat the source separation based methods in terms of OCR improvement.

Perspectives:

- Creation of a in-homogeneous “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.
- A hierarchical Markov model.
- A discriminative model taking into account texture.