Bayesian Separation of Document Images with Hidden Markov Model

Feng SU and Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes
UMR 8506 (CNRS-Supèlec-Univ Paris Sud)
Supèlec, 91192 Gif-sur-Yvette, France

International Conference on Computer Vision Theory and Applications, 2007

Outline

1. Motivation
   - The Basic Problem That We Studied
   - Previous Work
2. Our Method
   - Model Assumption and Formulation
   - Bayesian Estimation of Model Parameters
3. Experiment Results

Outline

Problem: Separate Mixed Document Images

- We consider superimposition of text patterns in document images as linearly mixed sources.
- Each source is composed of at least two object types - text and background.
- Some mixture examples from historical degradation, digitization:

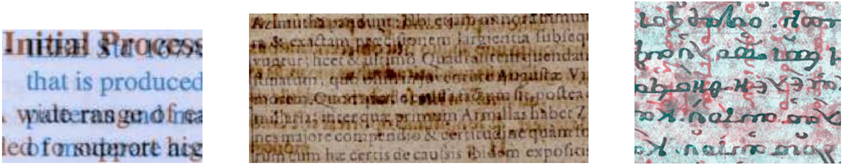
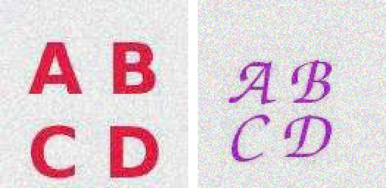


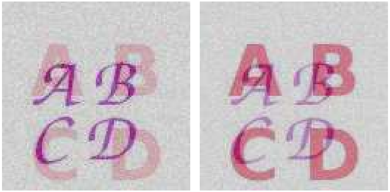
Figure: show-through, bleed-through and over-writing mixture.

Problem: From BSS Viewpoint

Objective: From M observations of mixed document images, obtain the N original source images (normally $M \geq N$).



source images $s_j(r)$



observed images $x_i(r)$

Blind Source Separation (BSS) scenario:

- Observations $\mathbf{x}$ are mixtures of unknown sources $\mathbf{s}$ with unknown mixing coefficients $\mathbf{A}$, plus additive independent white Gaussian noise ϵ with unknown variance.
- Consider linear instantaneous mixing, we have:

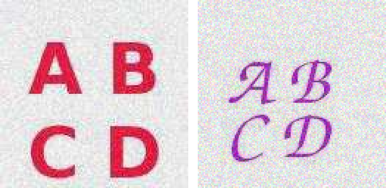
$$\mathbf{x}(r) = \mathbf{A}\mathbf{s}(r) + \epsilon(r) \qquad r \in \mathcal{R} \tag{1}$$

Previous Work

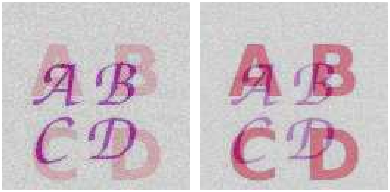
- Classic image processing methods (Non-BSS)
  - filtering (multiscale) [Sha01, TCS02], thresholding/binarization (adaptive) [NS03] ...
  - considered as a image restoration, enhancement, or segmentation problem
- BSS methods
  - PCA, ICA: [TBS04]
    - effective for detecting independent document features
    - each source is considered as a random signal sequence without internal structuring
  - Bayesian approach: [TBS06]
    - prior modeling of source: local smoothness, edge discontinuity, common classification ...
    - probability model (Markov Random Field) and numerical methods (EM, MCMC, Mean Field ...)

Outline

Mixing Model Assumption



source images $s_j(r)$



observed images $x_i(r)$

Linear instantaneous mixing model:

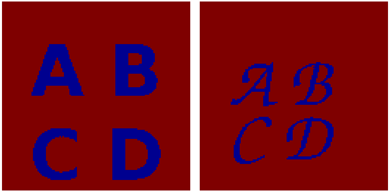
$$\mathbf{x}(r) = \mathbf{A}\mathbf{s}(r) + \epsilon(r) \qquad r \in \mathcal{R}$$

Gaussian noise model:

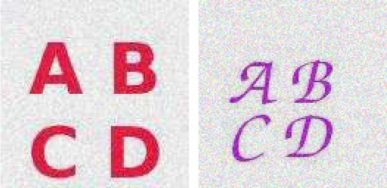
$$p(\mathbf{X}|\mathbf{S}, \mathbf{A}, \mathbf{R}_\epsilon) = \prod_r \mathcal{N}(\mathbf{A}\mathbf{s}(r), \mathbf{R}_\epsilon) \tag{2}$$

Hierachical Source Model

Double Markov process: *label process* + *intensity process*



label process $z_j(r)$



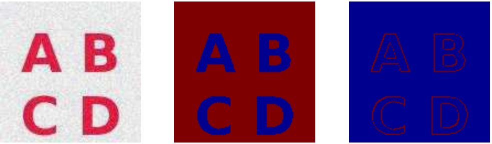
intensity process $s_j(r)|z_j(r)$

Label process $z_j(r) = \{1, \cdots, \mathcal{K}\}$:

- $\mathcal{K}$-class Potts Markov Random Field:

$$p(z_j(r), r \in \mathcal{R}) \propto \exp\left[\beta_j \sum_{r \in \mathcal{R}} \sum_{r' \in \mathcal{V}(r)} \delta(z_j(r) - z_j(r'))\right] \qquad (3)$$

Intensity Process



$s_j(r)$ $\quad$ $z_j(r)$ $\quad$ $q_j(r)$

We have *two options*:

- pixels are conditionally independent given their labels (i.i.d)

$$p(s_j(r)|z_j(r) = k) = \mathcal{N}(\mu_{jk}, \sigma^2_{jk}) \tag{4}$$

- pixels inside a region are locally dependent (Markovian)

$$p(s_j(r)|z_j(r) = k, s_j(r'), r' \in \mathcal{V}(r)) = \mathcal{N}(\bar{s}_j(r), \bar{\sigma}^2_j(r)) \tag{5}$$

with

$$\bar{s}_j(r) = q_j(r)\mu_{jk} + (1 - q_j(r))\frac{1}{|\mathcal{V}_{jk}(r)|}\sum_{r' \in \mathcal{V}_{jk}(r)} s_j(r')$$

$$\bar{\sigma}^2_j(r) = q_j(r)\sigma^2_{jk} + (1 - q_j(r))\sigma^2_j$$

Multiple-Channel Considerations

- Source:
  1. independent channels
     - separate on individual channel then merge
  2. correlated channels (for example, share a common classification per source)
     - joint demixing for all channels
- Mixing:
  1. same $\mathbf{A}_{M\times N}$ for all channels
  2. separate $\mathbf{A}_{M\times N}^{red,green,blue}$
  3. cross-channel mixing - expanded $\mathbf{A}_{ML\times NL}$ ($L = 3$ for RGB)
- Our assumptions:
  1. separate $\mathbf{A}^{r,g,b}$
  2. one common labeling Z_j for source $S_j^{(r,g,b)}$

Bayesian Estimation Framework

Joint *a posteriori* distribution of all unknowns:

$$p(\mathbf{S}, \mathbf{Z}, \theta | \mathbf{X}) \propto p(\mathbf{X} | \mathbf{S}, \mathbf{A}, \mathbf{R}_\epsilon) p(\mathbf{S} | \mathbf{Z}, \theta_s) p(\mathbf{Z}) p(\theta) \quad (6)$$

- Posterior Means (PM) estimator is chosen and approximated by MCMC Gibbs sampling algorithm

  Repeat until converge,

  1. simulate $\mathbf{Z}' \sim p(\mathbf{Z} | \mathbf{S}, \theta, \mathbf{X})$
  2. simulate $\mathbf{S}' \sim p(\mathbf{S} | \mathbf{Z}', \theta, \mathbf{X})$
  3. simulate $\theta' \sim p(\theta | \mathbf{Z}', \mathbf{S}', \mathbf{X})$

- Conjugate priors for model hyperparameters $\theta = \{\mathbf{A}, \mathbf{R}_\epsilon, \mu_{jk}, \sigma_{jk}^2\}$.

Bayesian Estimation Framework

Sampling **Z**:

$$\mathbf{Z} \sim p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) \propto p(\mathbf{X}|\mathbf{Z},\boldsymbol{\theta})p(\mathbf{Z})$$

$$\begin{aligned} p(\mathbf{X}|\mathbf{Z},\boldsymbol{\theta}) &= \prod_r p(\mathbf{x}(r)|\mathbf{z}(r),\boldsymbol{\theta}) \\ &= \prod_r \mathcal{N}(\mathbf{A}\mathbf{m}_{\mathbf{z}(r)}, \mathbf{A}\Sigma_{\mathbf{z}(r)}\mathbf{A}^t + \mathbf{R}_\epsilon) \end{aligned}$$

- $\mathbf{m}_{\mathbf{z}(r)} = [\mu_{1z_1(r)}, \ldots, \mu_{Nz_N(r)}]^t$ and $\Sigma_{\mathbf{z}(r)} = diag[\sigma^2_{1z_1(r)}, \ldots, \sigma^2_{Nz_N(r)}]$
- $p(\mathbf{Z}) = \prod_{j=1}^N p(\mathbf{z}_j)$, $p(\mathbf{z}_j)$ takes the form of Potts MRF as (3)

Bayesian Estimation Framework

Sampling $\mathbf{S} \sim p(\mathbf{S}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta})$:

$$
\begin{aligned}
p(\mathbf{S}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) &\propto p(\mathbf{X}|\mathbf{S}, \mathbf{A}, \mathbf{R}_\epsilon) p(\mathbf{S}|\mathbf{Z}, \boldsymbol{\theta}) \\
&= \prod_r \mathcal{N}(\mathbf{m}_s^{apost}(r), \mathbf{R}_s^{apost}(r))
\end{aligned}
$$

where,

$$
\begin{cases}
\mathbf{R}_s^{apost}(r) = \left[\mathbf{A}^t \mathbf{R}_\epsilon^{-1} \mathbf{A} + \Sigma_{\mathbf{z}(r)}^{-1}\right]^{-1} \\
\mathbf{m}_s^{apost}(r) = \mathbf{R}_s^{apost}(r) \left[\mathbf{A}^t \mathbf{R}_\epsilon^{-1} \mathbf{x}(r) + \Sigma_{\mathbf{z}(r)}^{-1} \mathbf{m}_{\mathbf{z}(r)}\right]
\end{cases}
$$

Bayesian Estimation Framework

Sampling $\mathbf{R}_\epsilon$:

$$p(\mathbf{R}_\epsilon|\mathbf{X},\mathbf{S},\mathbf{A}) \propto p(\mathbf{X}|\mathbf{S},\mathbf{A},\mathbf{R}_\epsilon)p(\mathbf{R}_\epsilon)$$

$p(\mathbf{R}_\epsilon)$: inverse Wishart conjugate prior.
$\mathbf{R}_\epsilon$ is *a posteriori* sampled from:

$$\begin{cases} \mathbf{R}_\epsilon^{-1} \sim \mathcal{W}_i(\alpha_\epsilon, \beta_\epsilon) \\ \alpha_\epsilon = \frac{1}{2}(|\mathcal{R}| - n), \beta_\epsilon = \frac{1}{2}|\mathcal{R}|(\mathbf{R}_{xx} - \mathbf{R}_{xs}\mathbf{R}_{ss}^{-1}\mathbf{R}_{xs}^t) \end{cases}$$

where,

- $|\mathcal{R}|$ is the number of image pixels
- $\mathbf{R}_{xx} = \frac{1}{|\mathcal{R}|}\sum_r \mathbf{x}_r\mathbf{x}_r^t$, $\mathbf{R}_{xs} = \frac{1}{|\mathcal{R}|}\sum_r \mathbf{x}_r\mathbf{s}_r^t$, $\mathbf{R}_{ss} = \frac{1}{|\mathcal{R}|}\sum_r \mathbf{s}_r\mathbf{s}_r^t$, (the sample statistics)

Bayesian Estimation Framework

Sampling $\mathbf{A} \sim p(\mathbf{A}|\mathbf{X}, \mathbf{S}, \mathbf{R}_\epsilon)$:

$$p(\mathbf{A}|\mathbf{X}, \mathbf{S}, \mathbf{R}_\epsilon) \propto p(\mathbf{X}|\mathbf{S}, \mathbf{A}, \mathbf{R}_\epsilon)p(\mathbf{A})$$

With uniform or Gaussian prior $p(\mathbf{A})$, $\mathbf{A}$ has a Gaussian posterior:

$$\begin{cases} Vec(\mathbf{A}) \sim \mathcal{N}(\mu_A, \mathbf{R}_A) \\ \mu_A = Vec(\mathbf{R}_{xs}\mathbf{R}_{ss}^{-1}), \mathbf{R}_A = \frac{1}{|\mathcal{R}|}\mathbf{R}_{ss}^{-1} \otimes \mathbf{R}_\epsilon \end{cases}$$

where,

- $\otimes$ is Kronecker product
- $Vec(.)$ is column-stacking operation

Bayesian Estimation Framework

Sampling $(\mu_{jk}, \sigma^2_{jk})$:
With conjugate Gauss prior for μ_{jk} and inverse Gamma prior for σ^2_{jk}, they are *a posteriori* sampled as:

$$\begin{cases} \mu_{jk} | \mathbf{s}_j, \mathbf{z}_j, \sigma^2_{jk} & \sim \mathcal{N}(m_{jk}, v^2_{jk}) \\ m_{jk} & = v^2_{jk} \left(\frac{\mu_{k0}}{\sigma^2_{k0}} + \frac{1}{\sigma^2_{jk}} \sum_{r \in \mathcal{R}^{(j)}_k} \mathbf{s}_j(r) \right) \\ v^2_{jk} & = \left(\frac{n^{(j)}_k}{\sigma^2_{jk}} + \frac{1}{\sigma^2_{k0}} \right)^{-1} \end{cases}$$

$$\begin{cases} \sigma^2_{jk} | \mathbf{s}_j, \mathbf{z}_j, \mu_{jk} & \sim \mathcal{IG}(\alpha_{jk}, \beta_{jk}) \\ \alpha_{jk} & = \alpha_{k0} + \frac{n^{(j)}_k}{2} \\ \beta_{jk} & = \beta_{k0} + \frac{1}{2} \sum_{r \in \mathcal{R}^{(j)}_k} (\mathbf{s}_j(r) - \mu_{jk})^2 \end{cases}$$

where, label region $\mathcal{R}^{(j)}_k = \{r : z_j(r) = k\}$ and region size $n^{(j)}_k = |\mathcal{R}^{(j)}_k|$.

Efficiency Enhancement

Parallel Gibbs Sampling [FMD05]:

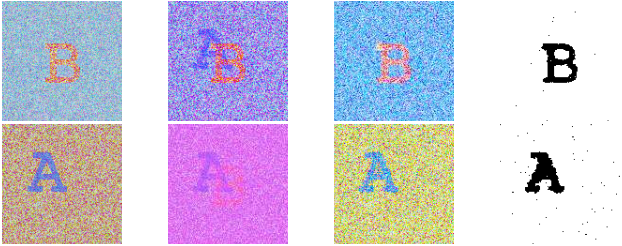
- divide pixel sites into $\mathcal{R}_B$ (black) and $\mathcal{R}_W$ (white) - chessboard pattern
- $p(\mathbf{z}_B|\mathbf{z}_W)$, $p(\mathbf{z}_W|\mathbf{z}_B)$, $p(\mathbf{s}_B|\mathbf{s}_W, \mathbf{z}_B, \mathbf{z}_W)$, $p(\mathbf{s}_W|\mathbf{s}_B, \mathbf{z}_B, \mathbf{z}_W)$ are separable by site r
- parallel simulating:

$$
\begin{aligned}
&\mathbf{z}_B^{(n)} \sim p(\mathbf{z}_B|\mathbf{z}_W^{(n-1)}, \mathbf{X}, \boldsymbol{\theta})\\
&\mathbf{s}_B^{(n)} \sim p(\mathbf{s}_B|\mathbf{z}_B^{(n)}, \mathbf{z}_W^{(n-1)}, \mathbf{X}, \boldsymbol{\theta})\\
&\mathbf{z}_W^{(n)} \sim p(\mathbf{z}_W|\mathbf{z}_B^{(n)}, \mathbf{X}, \boldsymbol{\theta})\\
&\mathbf{s}_W^{(n)} \sim p(\mathbf{s}_W|\mathbf{z}_B^{(n)}, \mathbf{z}_W^{(n)}, \mathbf{X}, \boldsymbol{\theta})\\
&\ldots
\end{aligned}
$$

Outline

Separation of Synthetic Image Mixture

Samples are synthesized from two binary (text/background) label image and the intensity and mixing model.

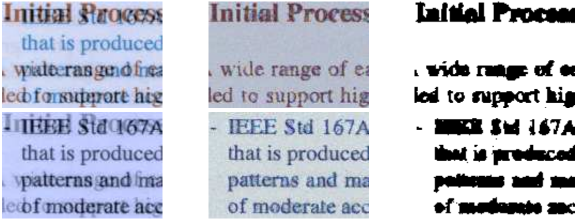


(a) (b) (c) (d)

Figure: a) original sources; b) image mixtures; c) demixed sources; d) classification labels.

Separation of Real Show-Through Image

Samples are scanned from a duplex printed paper with contrast enhanced.

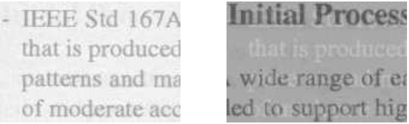


(a) (b) (c)

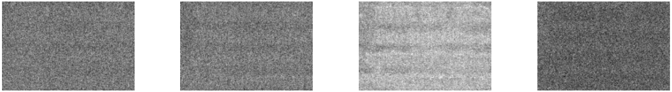
Figure: a) image mixtures; b) demixed sources; c) classification labels.

Results from ICA Algorithm

All three channels of the two observed mixtures were used simultaneously (totally 6 input components).



(a) two demixed ICs corresponding to the sources



(b) other demixed ICs containing noise-like signals

Figure: Separation results by ICA.

Due to present of cross-channel correlations, extra channel information does not necessarily improve separation.

Summary

- A hierarchical Markov model for prior modeling of document images is proposed.
- A Bayesian source and parameter estimation framework is described.
- Experiment results and comparisons are presented.
- Outlook
  - The linear and homogeneous mixing assumption may not hold precisely, especially for the complex physical degradation process in some historical documents.
  - Adaptive models for text-image mixing cases are to be explored.

For Further Reading I

Olivier Feron and Ali Mohammad-Djafari.
Image fusion and unsupervised joint segmentation using HMM and MCMC algorithms.
Journal of Electronic Imaging, 14(2), Apr. 2005.

Hirobumi Nishida and Takeshi Suzuki.
A multiscale approach to restoring scanned color document images with show-through effects.
In *Seventh International Conference on Document Analysis and Recognition*, volume 1, pages 584–588, Aug. 2003.

Gaurav Sharma.
Show-through cancellation in scans of duplex printed documents.
IEEE Transactions on Image Processing, 10(5):736–754, May 2001.

For Further Reading II

Hichem Snoussi and Ali Mohammad-Djafari.
Fast joint separation and segmentation of mixed images.
Journal of Electronic Imaging, 13:349–361, Apr. 2004.

Anna Tonazzini, Luigi Bedini, and Emanuele Salerno.
Independent component analysis for document restoration.
International Journal on Document Analysis and Recognition, 7(1):17–27, Mar. 2004.

Anna Tonazzini, Luigi Bedini, and Emanuele Salerno.
A markov model for blind image separation by a mean-field EM algorithm.
IEEE Transactions on Image Processing, 15(2):473–482, Feb. 2006.

For Further Reading III

Chew Lim Tan, Ruini Cao, and Peiyi Shen.
Restoration of archival documents using a wavelet technique.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(10):1399–1404, Oct. 2002.