Introduction to Probabilistic Image Processing and Bayesian Networks

Kazuyuki Tanaka
Graduate School of Information Sciences,
Tohoku University, Sendai, Japan
kazu@smapip.is.tohoku.ac.jp
http://www.smapip.is.tohoku.ac.jp/~kazu/
Contents

1. Introduction
2. Probabilistic Image Processing
3. Gaussian Graphical Model
4. Belief Propagation
5. Other Application
6. Concluding Remarks
Contents

1. Introduction
2. Probabilistic Image Processing
3. Gaussian Graphical Model
4. Belief Propagation
5. Other Application
6. Concluding Remarks
Markov Random Fields for Image Processing

Markov Random Fields are one of probabilistic methods for image processing.

Image Processing for Markov Random Fields (MRF)
(Simulated Annealing, Line Fields)

Image Processing in EM algorithm for Markov Random Fields (MRF) (Mean Field Methods)
In Markov Random Fields, we have to consider not only the states with high probabilities but also ones with low probabilities. In Markov Random Fields, we have to estimate not only the image but also hyperparameters in the probabilistic model. We have to perform the calculations of statistical quantities repeatedly. We need a deterministic algorithm for calculating statistical quantities. Belief Propagation
Belief Propagation has been proposed in order to achieve probabilistic inference systems (Pearl, 1988).

It has been suggested that Belief Propagation has a closed relationship to Mean Field Methods in the statistical mechanics (Kabashima and Saad 1998).

Generalized Belief Propagation has been proposed based on Advanced Mean Field Methods (Yedidia, Freeman and Weiss, 2000).

Interpretation of Generalized Belief Propagation also has been presented in terms of Information Geometry (Ikeda, T. Tanaka and Amari, 2004).
Probabilistic Model and Belief Propagation

Function consisting of a product of functions with two variables can be assigned to a graph representation.

Examples

$$\sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} A(x_1, x_2)B(x_3, x_2)C(x_4, x_2)C(x_5, x_2)$$

$$\sum_{x_1} \sum_{x_2} \sum_{x_3} A(x_1, x_2)B(x_2, x_3)C(x_3, x_1)$$

Belief Propagation can give us an exact result for the calculations of statistical quantities of probabilistic models with tree graph representations.

Generally, Belief Propagation cannot give us an exact result for the calculations of statistical quantities of probabilistic models with cycle graph representations.
Applications of belief propagation to many problems which are formulated as probabilistic models with cycle graph representations have caused to many successful results.

- Probabilistic Inference in AI (Kappen and Wiegerinck, NIPS, 2002).
Purpose of My Talk

- Review of formulation of probabilistic model for image processing by means of conventional statistical schemes.
- Review of probabilistic image processing by using Gaussian graphical model (Gaussian Markov Random Fields) as the most basic example.
- Review of how to construct a belief propagation algorithm for image processing.
Contents

1. Introduction
2. Probabilistic Image Processing
3. Gaussian Graphical Model
4. Belief Propagation
5. Other Application
6. Concluding Remarks
Digital image is defined on the set of points arranged on a square lattice. The elements of such a digital array are called pixels. We have to treat more than 100,000 pixels even in the digital cameras and the mobile phones.
Image Representation in Computer Vision

At each point, the intensity of light is represented as an integer number or a real number in the digital image data.

A monochrome digital image is then expressed as a two-dimensional light intensity function and the value is proportional to the brightness of the image at the pixel.
Noise Reduction by Conventional Filters

It is expected that probabilistic algorithms for image processing can be constructed from such aspects in the conventional signal processing.

The function of a linear filter is to take the sum of the product of the mask coefficients and the intensities of the pixels.

<table>
<thead>
<tr>
<th>192</th>
<th>202</th>
<th>190</th>
</tr>
</thead>
<tbody>
<tr>
<td>202</td>
<td>219</td>
<td>120</td>
</tr>
<tr>
<td>100</td>
<td>218</td>
<td>110</td>
</tr>
</tbody>
</table>

Average \[
\begin{bmatrix}
192 & 202 & 190 \\
202 & 219 & 120 \\
100 & 218 & 110 \\
\end{bmatrix}
\] = 173

Markov Random Fields  Algorithm  Probabilistic Image Processing
Bayes Formula and Bayesian Network

\[
\Pr\{A \mid B\} = \frac{\Pr\{B \mid A\}\Pr\{A\}}{\Pr\{B\}}
\]

**Bayes Rule**

Event $A$ is given as the observed data. Event $B$ corresponds to the original information to estimate. Thus the Bayes formula can be applied to the estimation of the original information from the given data.
Image Restoration by Probabilistic Model

Assumption 1: The degraded image is randomly generated from the original image by according to the degradation process.
Assumption 2: The original image is randomly generated by according to the prior probability.

Bayes Formula

\[
\Pr\{\text{Original Image} | \text{Degraded Image}\} = \frac{\Pr\{\text{Degraded Image} | \text{Original Image}\} \Pr\{\text{Original Image}\}}{\Pr\{\text{Degraded Image}\}}
\]
Image Restoration by Probabilistic Model

The original images and degraded images are represented by $\mathbf{f} = \{f_i\}$ and $\mathbf{g} = \{g_i\}$, respectively.

Original Image

Degraded Image

$f_i$: Light Intensity of Pixel $i$ in Original Image

$g_i$: Light Intensity of Pixel $i$ in Degraded Image

$\mathbf{r}_i = (x_i, y_i)$

Position Vector of Pixel $i$
Probabilistic Modeling of Image Restoration

Assumption 1: A given degraded image is obtained from the original image by changing the state of each pixel to another state by the same probability, independently of the other pixels.
Probabilistic Modeling of Image Restoration

Assumption 2: The original image is generated according to a prior probability. Prior Probability consists of a product of functions defined on the neighbouring pixels.

\[
\Pr\{\text{Original Image } f\} = \Pr\{F = f\} = \prod_{ij: \text{NN}} \Phi(f_i, f_j)
\]
Prior Probability for Binary Image

It is important how we should assume the function $\Phi(f_i, f_j)$ in the prior probability.

$$\Pr\{F = f\} = \prod_{i,j: \text{Nearest Neighbour}} \Phi(f_i, f_j)$$

We assume that every nearest-neighbour pair of pixels take the same state of each other in the prior probability.

$$\Phi(1,1) = \Phi(0,0) > \Phi(0,1) = \Phi(1,0)$$

Probability of Neighbouring Pixel:

- $p_{ij} = \frac{1}{2} - p$
- $p_{ji} = \frac{1}{2} - p$

$f_i = 0, 1$
Prior Probability for Binary Image

\[
\begin{align*}
\frac{1}{2} - p & > \frac{1}{2} - p \\
\end{align*}
\]

Which state should the center pixel be taken when the states of neighbouring pixels are fixed to the white states?

Prior probability prefers to the configuration with the least number of red lines.
Prior Probability for Binary Image

\[ \begin{align*}
\text{O} & = \text{\textbullet} > \text{\textbullet} = \text{O}
\end{align*} \]

Which state should the center pixel be taken when the states of neighbouring pixels are fixed as this figure?

Prior probability prefers to the configuration with the least number of red lines.
What happens for the case of large number of pixels?

Covariance between the nearest neighbour pairs of pixels

Patterns with both ordered states and disordered states are often generated near the critical point.

Sampling by Markov chain Monte Carlo

Disordered State

Critical Point (Large fluctuation)

Ordered State
Pattern near Critical Point of Prior Probability

Covariance between the nearest neighbour pairs of pixels

We regard that patterns generated near the critical point are similar to the local patterns in real world images.

\[ \text{small } p \quad \ln p \quad 0.8 \quad \text{large } p \]

similar
Bayesian Image Analysis

Prior Probability

Original Image

Degradation Process

Degraded Image

Posterior Probability

\[ \Pr\{F = f\} \rightarrow f \rightarrow \Pr\{G = g|F = f\} \rightarrow g \]

\[ \Pr\{F = f|G = g\} = \frac{\Pr\{G = g|F = f\} \Pr\{F = f\}}{\Pr\{G = g\}} \]

\[ \propto \left( \prod_{i=1}^{N} \Psi(f_i, g_i) \right) \left( \prod_{ij: \text{Nearest Neighbour}} \Phi(f_i, f_j) \right) \]

Image processing is reduced to calculations of averages, variances and co-variances in the posterior probability.
Estimation of Original Image

We have some choices to estimate the restored image from posterior probability.
In each choice, the computational time is generally exponential order of the number of pixels.

\[(1) \hat{f} = \arg\max_{z} \Pr\{F = z \mid G = g\} \quad \text{Maximum A Posteriori (MAP) estimation}\]

\[(2) \hat{f}_i = \arg\max_{z_i} \Pr\{F_i = z_i \mid G = g\} \quad \text{Maximum posterior marginal (MPM) estimation}\]

\[
\Pr\{F_i = f_i \mid G = g\} = \sum_{f \setminus f_i} \Pr\{F = f \mid G = g\}
\]

\[(3) \hat{f}_i = \arg\min_{\zeta_i} \left( \zeta_i - \sum_{z} z_i \Pr\{F = z \mid G = g\} \right)^2 \quad \text{Thresholded Posterior Mean (TPM) estimation}\]
Statistical Estimation of Hyperparameters

Hyperparameters $\alpha$, $\beta$ are determined so as to maximize the marginal likelihood $\Pr\{G=g|\alpha,\beta\}$ with respect to $\alpha$, $\beta$.

$$(\hat{\alpha}, \hat{\beta}) = \arg \max_{(\alpha, \beta)} \Pr\{G=g|\alpha, \beta\}$$

$$\Pr\{G=g|\alpha, \beta\} = \sum_{z} \Pr\{G=g|F=z, \beta\} \Pr\{F=z|\alpha\}$$

Original Image $\Omega$

$\Pr\{F=f|\alpha\} \rightarrow f \rightarrow \Pr\{G=g|F=f, \beta\} \rightarrow g$

Marginalized with respect to $F$

$\Pr\{G=g|\alpha, \beta\}$

Marginal Likelihood
Maximization of Marginal Likelihood by EM Algorithm

Marginal Likelihood:

\[
\Pr\{G = g \mid \alpha, \sigma\} = \sum_z \Pr\{G = g \mid F = z, \sigma\} \Pr\{F = z \mid \alpha\}
\]

**Q-Function**

\[
Q(\alpha, \sigma \mid \alpha', \sigma', g) = \sum_z \Pr\{F = z \mid G = g, \alpha', \sigma'\} \ln \Pr\{F = z, G = g \mid \alpha, \sigma\}
\]

EM (Expectation Maximization) Algorithm

E-step and M-Step are iterated until convergence:

**E - Step**:

\[
Q(\alpha, \sigma \mid \alpha(t), \sigma(t)) \leftarrow \sum_z \Pr\{F = z \mid G = g, \alpha(t), \sigma(t)\} \ln \Pr\{F = z, G = g \mid \alpha, \sigma\}
\]

**M - Step**:

\[
(\alpha(t+1), \sigma(t+1)) \leftarrow \arg \max_{(\alpha, \sigma)} Q(\alpha, \sigma \mid \alpha(t), \sigma(t)).
\]
Contents

1. Introduction
2. Probabilistic Image Processing
3. Gaussian Graphical Model
4. Belief Propagation
5. Other Application
6. Concluding Remarks
Bayesian Image Analysis by Gaussian Graphical Model

Prior Probability

Pr\{F = f\} \propto \exp \left(-\frac{1}{2} \alpha \sum_{ij \in B} (f_i - f_j)^2\right)

Patterns are generated by MCMC.

Markov Chain Monte Carlo Method

\(f_i \in (-\infty, +\infty)\)

\(\Omega\): Set of all the pixels

\(B\): Set of all the nearest-neighbour pairs of pixels

\(\alpha = 0.0001\)  \(\alpha = 0.0005\)  \(\alpha = 0.0030\)

Patterns generated with different values of \(\alpha\).
Bayesian Image Analysis by Gaussian Graphical Model

Degradation Process is assumed to be the additive white Gaussian noise.

\[ f_i, g_i \in (-\infty, +\infty) \]

\[
\Pr\{G = g | F = f\} = \prod_{i \in \Omega} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (f_i - g_i)^2\right)
\]

\(\Omega\): Set of all the pixels

Degraded image is obtained by adding a white Gaussian noise to the original image.

\[ g_i - f_i \sim N(0, \sigma^2) \]

Histogram of Gaussian Random Numbers
Bayesian Image Analysis by Gaussian Graphical Model

Posterior Probability

\[ P(f | g, \alpha, \sigma) \propto \exp \left( -\frac{1}{2\sigma^2} \sum_{i \in \Omega} (f_i - g_i)^2 - \frac{1}{2} \alpha \sum_{ij \in B} (f_i - f_j)^2 \right) \]

Average of the posterior probability can be calculated by using the multi-dimensional Gaussian integral Formula

\[ \int z P(z | g, \alpha, \sigma) dz = \frac{I}{I + \alpha \sigma^2 C} g \]

\[ \langle i | C | j \rangle = \begin{cases} 4 & (i = j) \\ -1 & (ij \in B) \\ 0 & \text{(otherwise)} \end{cases} \]

\( \Omega \): Set of all the pixels

\( B \): Set of all the nearest-neighbour pairs of pixels

\( N \times N \) matrix

Multi-Dimensional Gaussian Integral Formula
Bayesian Image Analysis by Gaussian Graphical Model

Iteration Procedure of EM algorithm in Gaussian Graphical Model

\[ \alpha(t) \leftarrow \left( \frac{1}{N} \text{Tr} \left( \frac{\sigma(t-1)^2 C}{I + \alpha(t-1)\sigma(t-1)^2 C} \right) + \frac{1}{N} g^T C \right)^{-1} \left( \frac{1}{N} g^T \frac{C}{I + \alpha(t-1)\sigma(t-1)^2 C} g \right) \]

\[ \sigma(t) \leftarrow \sqrt{\frac{1}{N} \text{Tr} \left( \frac{\sigma(t-1)^2 I}{I + \alpha(t-1)\sigma(t-1)^2 C} \right) + \frac{1}{N} g^T \frac{\alpha(t-1)^2 \sigma(t-1)^4 C^2}{\left( I + \alpha(t-1)\sigma(t-1)^2 C \right)^2} g} \]

\[ m(t) = \frac{I}{I + \alpha(t)\sigma(t)^2 C} g \]

\[ \hat{f}_i(t) = \arg\min_{z_i} (z_i - m_i(t))^2 \]

1 October, 2007

ALT&DS2007 (Sendai, Japan)
Image Restoration by Markov Random Field Model and Conventional Filters

Original Image | Degraded Image
---|---
![Original Image](image1.png) | ![Degraded Image](image2.png)

Restored Image

<table>
<thead>
<tr>
<th>Statistical Method</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowpass Filter (5x5)</td>
<td>413</td>
</tr>
<tr>
<td>Lowpass Filter (3x3)</td>
<td>388</td>
</tr>
<tr>
<td>Median Filter (5x5)</td>
<td>445</td>
</tr>
<tr>
<td>Median Filter (3x3)</td>
<td>486</td>
</tr>
</tbody>
</table>

\[ \text{MSE} = \frac{1}{|\Omega|} \sum_{i \in \Omega} (f_i - \hat{f}_i)^2 \]

\( \Omega \): Set of all the pixels

MRF (3x3) Lowpass (5x5) Median
Contents

1. Introduction
2. Probabilistic Image Processing
3. Gaussian Graphical Model
4. **Belief Propagation**
5. Other Application
6. Concluding Remarks
Graphical Representation for Tractable Models

**Tractable Model**

\[
\sum_{A=T,F} \sum_{B=T,F} \sum_{C=T,F} f(A,D)g(B,D)h(C,D) = \left( \sum_{A=T,F} f(A,D) \right) \left( \sum_{B=T,F} g(B,D) \right) \left( \sum_{C=T,F} h(C,D) \right)
\]

It is possible to calculate each summation independently.

**Intractable Model**

\[
\sum_{A=T,F} \sum_{B=T,F} \sum_{C=T,F} f(A,B)g(B,C)h(C,A) = \sum_{A=T,F} \sum_{B=T,F} \sum_{C=T,F} f(A,B)g(B,C)h(C,A)
\]

It is hard to calculate each summation independently.
Belief Propagation for Tree Graphical Model

After taking the summations over red nodes 3, 4 and 5, the function of nodes 1 and 2 can be expressed in terms of some messages.

\[
\sum_{f_3} \sum_{f_4} \sum_{f_5} A(f_1, f_2) B(f_2, f_3) C(f_2, f_4) D(f_2, f_5) \\
= A(f_1, f_2) \left[ \sum_{f_3} B(f_2, f_3) \right] \left[ \sum_{f_7} C(f_2, f_4) \right] \left[ \sum_{f_8} D(f_2, f_5) \right] \\
\]

\( M_{3\rightarrow2}(f_2) \quad M_{4\rightarrow2}(f_2) \quad M_{5\rightarrow2}(f_2) \)
Belief Propagation for Tree Graphical Model

By taking the summation over all the nodes except node 1, message from node 2 to node 1 can be expressed in terms of all the messages incoming to node 2 except the own message.

\[
\sum_{Z_2} M_2 \rightarrow 1 = \sum_{Z_2} M_3 \rightarrow 2 M_4 \rightarrow 2 M_5 \rightarrow 2
\]
Loopy Belief Propagation for Graphical Model in Image Processing

Graphical model for image processing is represented in terms of the square lattice. Square lattice includes a lot of cycles. Belief propagation are applied to the calculation of statistical quantities as an approximate algorithm.

Every graph consisting of a pixel and its four neighbouring pixels can be regarded as a tree graph.
Loopy Belief Propagation for Graphical Model in Image Processing

Message Passing Rule in Loopy Belief Propagation

Averages, variances and covariances of the graphical model are expressed in terms of messages.

\[
\sum \Phi_{12}(z_1, f_2) M_{3\rightarrow 1}(z_1) M_{4\rightarrow 1}(z_1) M_{5\rightarrow 1}(z_1)
\]

Message Passing Rule

\[
\sum \sum \Phi_{12}(z_1, z_2) M_{3\rightarrow 1}(z_1) M_{4\rightarrow 1}(z_1) M_{5\rightarrow 1}(z_1)
\]

Averages, variances and covariances of the graphical model are expressed in terms of messages.
Loopy Belief Propagation for Graphical Model in Image Processing

Each message passing rule includes 3 incoming messages and 1 outgoing message.

We have four kinds of message passing rules for each pixel.

Visualizations of Passing Messages
EM Algorithm for Hyperparameter Estimation

\[(\alpha(t + 1), \sigma(t + 1)) \leftarrow \arg\max_{(\alpha, \sigma)} Q(\alpha, \sigma | \alpha(t), \sigma(t), g)\]

Update Rule of Loopy BP

\[M_{2\rightarrow1} \quad M_{3\rightarrow2} \quad M_{4\rightarrow2} \quad M_{5\rightarrow2} \]

\[= \sum_{z_2} \]

EM algorithm by means of Belief Propagation
Probabilistic Image Processing by EM Algorithm and Loopy BP for Gaussian Graphical Model

\[ (\alpha(t+1), \sigma(t+1)) \leftarrow \arg \max_{(\alpha, \sigma)} Q(\alpha, \sigma | \alpha(t), \sigma(t), g). \]

\[ \hat{\alpha}_{\text{Exact}} = 37.624 \]
\[ \hat{\sigma}_{\text{Exact}} = 0.000713 \]

MSE:315

\[ \hat{\alpha}_{\text{LBP}} = 0.000600 \]
\[ \hat{\sigma}_{\text{LBP}} = 36.335 \]

MSE:327

Loopy Belief Propagation

\( f \)
Contents

1. Introduction
2. Probabilistic Image Processing
3. Gaussian Graphical Model
4. Belief Propagation
5. Other Application
6. Concluding Remarks
Digital Images Inpainting based on MRF

Markov Random Fields

Contents

1. Introduction
2. Probabilistic Image Processing
3. Gaussian Graphical Model
4. Belief Propagation
5. Other Application
6. Concluding Remarks
Summary

- Formulation of probabilistic model for image processing by means of conventional statistical schemes has been summarized.

- Probabilistic image processing by using Gaussian graphical model has been shown as the most basic example.

- It has been explained how to construct a belief propagation algorithm for image processing.
Statistical Mechanics Informatics for Probabilistic Image Processing

Original ideas of some techniques, Simulated Annealing, Mean Field Methods and Belief Propagation, is often based on the statistical mechanics.

  Image Processing for Markov Random Fields (MRF)
  (Simulated Annealing, Line Fields)

  Image Processing in EM algorithm for Markov Random Fields (MRF) (Mean Field Methods)

  Cluster Variation Method for MRF in Image Processing

Mathematical structure of Belief Propagation is equivalent to Bethe Approximation and Cluster Variation Method (Kikuchi Method) which are ones of advanced mean field methods in the statistical mechanics.
It has been suggested that statistical performance estimations for probabilistic information processing are closed to the spin glass theory. The computational techniques of spin glass theory has been applied to many problems in computer sciences.

- CDMA Multiuser Detection in Mobile Phone Communication (T. Tanaka: IEEE Information Theory, 2002).
SMAPIP Project

Member:
K. Tanaka, Y. Kabashima, H. Nishimori, T. Tanaka, M. Okada, O. Watanabe, N. Murata, ....

Webpage URL: http://www.smapip.eei.metro-u.ac.jp/

Period: 2002 –2005
Head Investigator: Kazuyuki Tanaka

MEXT Grant-in Aid for Scientific Research on Priority Areas
DEX-SMI Project

Deepening and Expansion of Statistical Mechanical Informatics

MEXT Grant-in Aid for Scientific Research on Priority Areas

Period: 2006 –2009

Head Investigator: Yoshiyuki Kabashima

http://dex-smi.sp.dis.titech.ac.jp/DEX-SMI/
References
