# Image segmentation combining Markov Random Fields and Dirichlet Processes

Jessica SODJO

IMS, Groupe Signal Image, Talence
Encadrants : A. Giremus, J.-F. Giovannelli, F. Caron, N. Dobigeon

# Plan

## Segmentation

- partition of an image in $K$ homogeneous regions called classes
- label the pixels : pixel $i \leftrightarrow z_i \in \{1, \ldots, K\}$

## Bayesian approach

- prior on the distribution of the pixels
- all the pixels in a class have the same distribution characterized by a parameter vector $U_k$
- Markov Random Fields (MRF) : exploit the similarity of pixels in the same neighbourhood

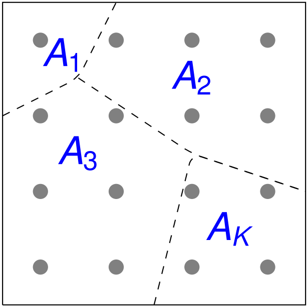
## Constraint : $K$ must be fixed a priori

## Idea : use the BNP models to directly estimate $K$

# Plan

# Notations

- $N$ is the number of pixels
- $\boldsymbol{Y}$ is the observed image
- $\boldsymbol{Z} = \{z_1, \ldots, z_N\}$
- $\Pi = \{A_1, \ldots, A_K\}$ is a partition and $\boldsymbol{m} = \{m_1, \ldots, m_K\}$ with $m_k = |A_k|$




$$m_1 = 1 \quad m_2 = 5 \quad m_3 = 6 \quad m_K = 4$$

FIGURE: Example of partition

# Markov Random Fields (MRF)

– Description of the image by a neighbouring system

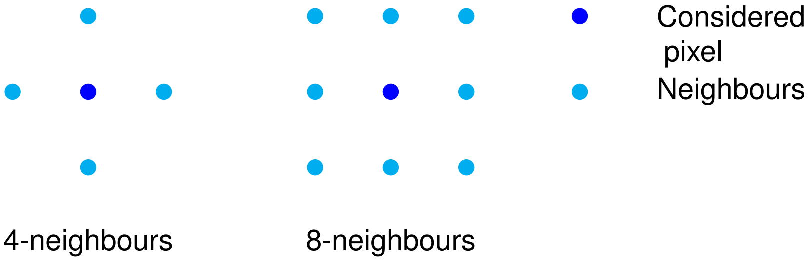



FIGURE: Examples of neighbouring system

– A clique $c$ is either a singleton either a set of pixels in the same neighbourhood

# Markov Random Fields

Let $\theta_i \in \{U_1, \ldots, U_K\}$ be the parameter vector associated to the $i$-th pixel

MRF $\Leftrightarrow p(\theta_i \mid \theta_{-i}) = p(\theta_i \mid \theta_{\mathcal{V}(i)})$
where $\mathcal{V}(i)$ is the set of neighbours of pixel $i$

Hammersley-Clifford theorem $\Rightarrow$ Gibbs field

$$p(\boldsymbol{\theta}) = \frac{1}{Z_\Phi} \exp\left(-\Phi(\boldsymbol{\theta})\right) = \frac{1}{Z_\Phi} \exp\left(-\sum_c \Phi_c(\boldsymbol{\theta}_c)\right) \qquad (1)$$

with $\Phi_c(\boldsymbol{\theta}_c)$ the local potential and $\Phi(\boldsymbol{\theta})$ the global one

Limitation : $K$ is assumed to be known

# Potts model

The Potts model is a special MRF defined by :

$$M(\Pi) \propto \exp\left(\sum_{i \leftrightarrow j} \beta_{ij} \mathbf{1}_{z_i = z_j}\right) \tag{2}$$

where

- $i \leftrightarrow j$ means that the pixels $i$ and $j$ are neighbours
- $\beta_{ij} > 0$ if $i$ and $j$ are neighbours and $\beta_{ij} = 0$ otherwise

# The DP model

$$\tau'_k \mid \gamma, H \sim \mathrm{Beta}(1, \gamma) \qquad \tau_k = \tau'_k \prod_{l=1}^{k-1}(1 - \tau'_l) \tag{3}$$

where $\mathrm{Beta}(.)$ is the Beta distribution

Let us write $\boldsymbol{\tau} \sim \mathit{Stick}(\gamma)$, $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots\}$ and $\sum_{k=1}^{\infty} \tau_k = 1$

$$\mathbb{G} \mid \gamma, H \sim DP(\gamma, H) \qquad \mathbb{G} = \sum_{k=1}^{\infty} \tau_k \delta_{U_k} \tag{4}$$

with

$$U_k \mid H \overset{\text{iid}}{\sim} H \tag{5}$$

The distribution of the observations is $f$, defined as :

$$y_i \mid \theta_i \sim f(. \mid \theta_i) \qquad \text{and} \qquad \theta_i \mid \mathbb{G} \sim \mathbb{G} \tag{6}$$

# The DP model

The Chinese Restaurant Process says,

$$\theta_i \mid \theta_{-i} \sim \sum_{k=1}^{K^{-i}} \frac{m_k^{-i}}{N-1+\gamma} \delta_{U_k} + \frac{\gamma}{N-1+\gamma} H$$

- $m_k^{-i}$ is the size of cluster $k$ if we remove pixel $i$ from the partition
- $K^{-i}$ is the number of clusters in the image with the $i$-th pixel removed
- $U_k$ is the parameter vector associated to the $k$-th cluster

Limitation : the spatial interactions are not taken into account

# Principle of the segmentation using DP models

Define a distribution on the partitions using :

- a model that allows that pixels in the same neighbourhood are likely to be in the same cluster (MRF)
- DP model to deduce automatically the number of clusters (and if needed their parameters)

# Prior distribution mixing DP and MRF

$$p(\boldsymbol{\theta}) \propto \underbrace{\frac{1}{Z_G} \exp(-\sum_i \Phi_i(\theta_i))}_{\Psi(\boldsymbol{\theta}) \rightsquigarrow DP\ model} \underbrace{\frac{1}{Z_M} \exp(-\sum_{c \in C_2} \Phi_c(\theta_c))}_{M(\boldsymbol{\theta}) \rightsquigarrow MRF\ model}$$

where

- $C_2$ means $|c| \geqslant 2$ and $|.|$ is the size.
- $\Phi_i(.)$ is defined as :

$$\Phi_i(\theta_i) = -\log \mathbb{G}(\theta_i) \text{ and } Z_G = \int \prod_{i=1}^{N} \exp(-\log \mathbb{G}(\theta_i)) \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_N$$

$$\Rightarrow \Psi(\boldsymbol{\theta}) = \prod_{i=1}^{N} \mathbb{G}(\theta_i)$$

P. Orbanz & J. M. Buhmann
*Nonparametric Bayesian image segmentation*, International Journal of Computer Vision, 2007

# Prior distribution mixing DP and MRF

We can deduce :

$$P(\theta_i \mid \theta_{-i}) \propto \sum_{k=1}^{K} M(\theta_i \mid \theta_{-i}) m_k^{-i} \delta_{U_k} + \frac{\gamma}{Z_\Phi} H \qquad (7)$$

Probability of assignment to a new cluster :

$$q_{i0} \propto \int_{\Omega_\theta} f(y_i \mid \theta) H(\theta) d\theta \qquad (8)$$

Probability of assignment to an existing cluster :

$$q_{ik} \propto m_k^{-i} \exp(-\Phi(U_k \mid \theta_{-i})) f(y_i \mid U_k) \qquad (9)$$

Parameter update :

$$U_k \sim \mathbb{G}_0(U_k) \prod_{i \mid i \in A_k} f(y_i \mid U_k) \qquad (10)$$

# Swendsen-Wang algorithm : principle

- Estimation based on the joint posterior $p(\theta, \boldsymbol{Z} \mid \boldsymbol{Y})$
- Intractable $\Rightarrow$ Markov Chain Monte Carlo (MCMC)

**Problem** : very slow convergence

**Goal** : Sample faster the partition of the image

- Introduction of a new set of latent variables $\boldsymbol{r}$ such that :

$$
\begin{aligned}
p(\Pi, \boldsymbol{r}) &= p(\Pi)p(\boldsymbol{r} \mid \Pi) \\
p(\boldsymbol{r} \mid \Pi) &= \prod_{1<i<j<N} p(r_{ij} \mid \Pi) \\
p(r_{ij} = 1 \mid \Pi) &= 1 - \exp(\beta_{ij}\delta_{ij}\mathbf{1}_{z_i=z_j})
\end{aligned}
$$

The marginal posterior $p(\theta, \boldsymbol{Z} \mid \boldsymbol{Y})$ is unchanged

- The links define the "so-called" spin-clusters

# Swendsen-Wang algorithm : principle

– Update the labels of the spin-clusters
  This operation update simultaneously the labels of all the pixels in a spin-cluster

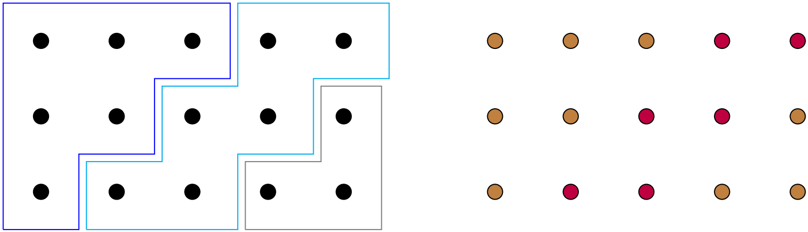


FIGURE: Example of label update for spin-clusters

# Swendsen-Wang algorithm : principle

- $r_{ij} \sim \mathrm{Ber}(1 - \exp(\beta_{ij}\delta_{ij}\mathbf{1}_{z_i=z_j}))$
  with $\mathrm{Ber}(.)$ is the Bernouilli distribution

  Let $\boldsymbol{S} = \{S_1, \ldots, S_p\}$ be the set of spin-clusters.

- While removing the spin-cluster $S_l$,

  $\Pi_{-l} = \{A_1^{-l}, \ldots, A_{K_{-l}}^{-l}\}$ is the partition obtained while removing all pixels in spin-cluster $S_l$

  $m_k^{-l} = |A_k^{-l}|$

# Swendsen-Wang algorithm : principle

For $l = 1 : p$

* The probability to assign pixels in spin-cluster $S_l$ to cluster $k$ is :

$$q_{lk} \propto \Psi(m_1^{-l}, \dots, m_k^{-l} + |S_l|, \dots, m_{K_{-l}}^{-l}) p(y_{S_l} \mid y_{A_k^{-l}}) \prod_{\{(i,j) \mid i \in S_l, r_{ij}=0\}} \exp(\beta_{ij}(1-\delta_{ij}) \mathbf{1}_{z_i = z_j})$$

* The probability to assign pixels in spin-cluster $S_l$ to a new cluster is :

$$q_{l0} = \Psi(m_1^{-l}, \dots, m_{K_{-l}}^{-l}, |S_l|) p(y_{S_l})$$

with $p(y_{A_k}) = \int \prod_{i \in A_k} f(y_i \mid U_k) H(U_k) \mathrm{d}U_k$

# Plan

# Proposed idea

- Different levels of classification can be considered
- Coarse categories : urban, sub-urban, forest, etc.
- Sub-classes shared between the categories : trees, roads, buildings

Taking into account the fact that the classes are shared between different categories can help estimating their parameters and thereby improve the segmentation

## Solution : Hierarchical DP

Let $J$ be the number of categories

$$\mathbb{G}_0 \mid \gamma, H \sim DP(\gamma, H)$$

$$\mathbb{G}_j \mid \alpha_0, \mathbb{G}_0 \sim DP(\alpha_0, \mathbb{G}_0) \quad \text{for } j = 1, \ldots, J$$

$$\alpha_0 \in \mathbb{R}_+^*$$

$\mathbb{G}_0$ is a discrete distribution

Discreteness of $\mathbb{G}_0 \Rightarrow$ clusters shared among categories

$$\mathbb{G}_0 = \sum_{k=1}^{\infty} \tau_k \delta_{U_k} \quad (11)$$

where $\boldsymbol{\tau}|\gamma \sim Stick(\gamma)$, $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots\}$ and $U_k \mid H \sim H$

$$\mathbb{G}_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{U_k} \quad (12)$$

with $\boldsymbol{\pi}_j \mid \alpha_0, \boldsymbol{\tau} \sim DP(\alpha_0, \boldsymbol{\tau})$ and $\boldsymbol{\pi}_j = \{\pi_{j1}, \pi_{j2}, \ldots\}$

$$\varphi_{ji} \mid \mathbb{G}_j \sim \mathbb{G}_j \quad (13)$$

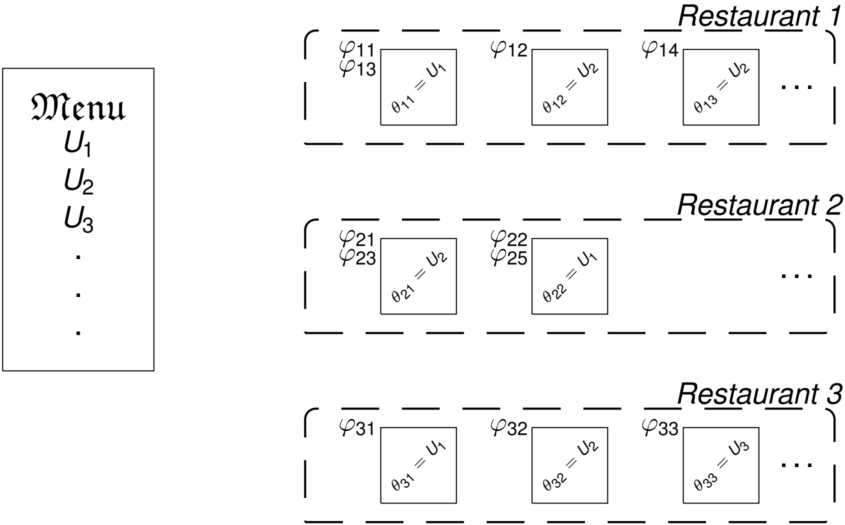
So, samples of the processes $\mathbb{G}_0$ and $\mathbb{G}_j$ can be seen as infinite countable mixtures of Dirac measures with respective coefficients $\boldsymbol{\tau}$ and $\boldsymbol{\pi}_j$.

# Principle - Chinese Restaurant Franchise

## NOTATIONS

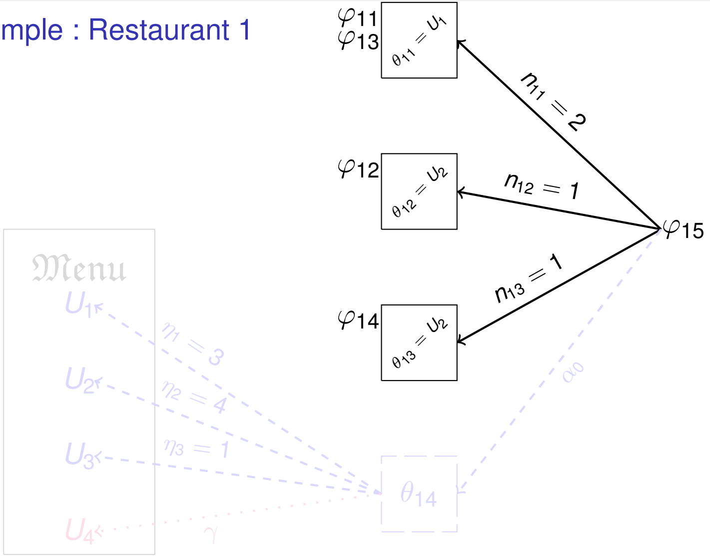
- $J$ restaurants
- Same menu for all restaurants - $U_1, U_2, \dots$
- $T_j$ is the number of tables in restaurant $j$
- $\theta_{jt}$ is the $t$-th table of restaurant $j$
- $\varphi_{ji}$ is the $i$-th client in restaurant $j$
- $n_{jt}$ is the number of clients at a table $t$
- $\eta_{jk}$ is the number of tables in restaurant $j$ which have chosen dish $U_k$ and $\eta_k = \sum_k \eta_{jk}$

# Principle - Chinese Restaurant Franchise



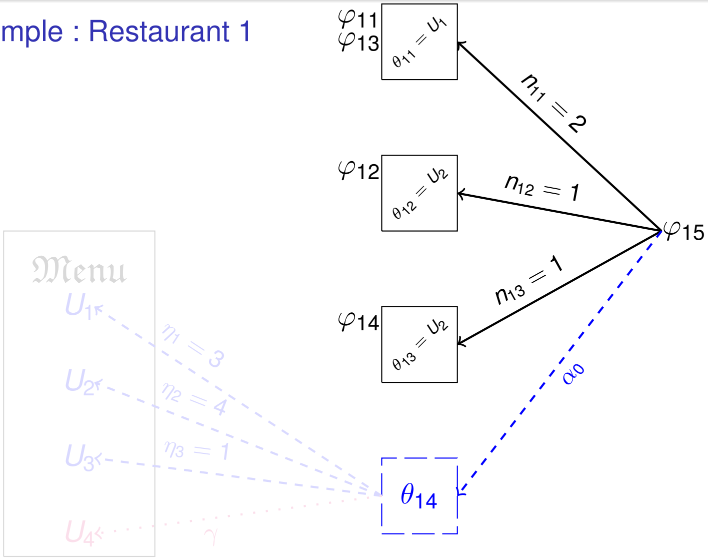
# Principle - Chinese Restaurant Franchise

## Exemple : Restaurant 1



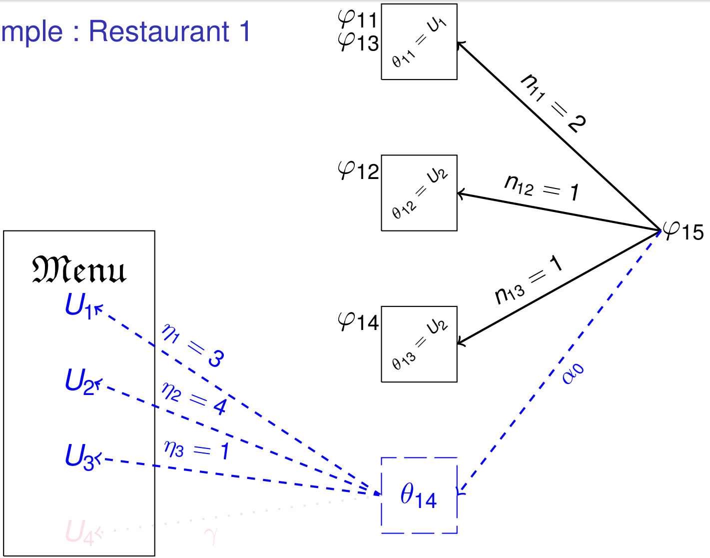
# Principle - Chinese Restaurant Franchise

## Exemple : Restaurant 1



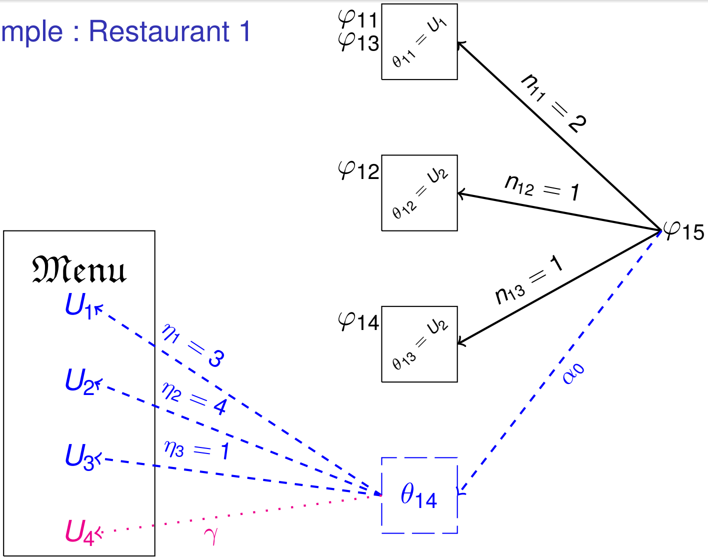
# Principle - Chinese Restaurant Franchise

## Exemple : Restaurant 1



# Principle - Chinese Restaurant Franchise

## Exemple : Restaurant 1



# Chinese Restaurant Franchise

$$\varphi_{ji} \mid \varphi_{j1}, \ldots, \varphi_{ji-1}, \alpha_0, \mathbb{G}_0 \sim \sum_{t=1}^{T_j} \frac{n_{jt}}{i-1+\alpha_0}\delta_{\theta_{jt}} + \frac{\alpha_0}{i-1+\alpha_0}\mathbb{G}_0 \tag{14}$$

$$\theta_{jt} \mid \theta_{j1}, \ldots, \theta_{21}, \ldots, \theta_{jt-1}, \gamma, H \sim \sum_{k=1}^{K} \frac{\eta_k}{\sum_k \eta_k + \gamma}\delta_{U_k} + \frac{\gamma}{\sum_k \eta_k + \gamma}H \tag{15}$$

Y. W. Teh, M. I. Jordan, M. J. Beal & D. M. Blei
*Hierarchical Dirichlet Processes*, JASA, 2006

# Plan

## Conclusion

- Spatial constraints : Potts model
- Flexibility : DP model
- Rapidity : Swendsen-Wang algorithm
- Sharing : HDP

## Perspective

- Efficient sampling algorithm

Thank you for your attention