Unsupervised Image Segmentation with Spatial Triplet Markov Trees

Hugo Gangloff[1,2,3] (hugo_gangloff@telecom-sudparis.eu),
Jean-Baptiste Courbot[4], Emmanuel Monfrini[1], Christophe Collet[2]

[1]Samovar, Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France
[2]ICube, Université de Strasbourg, CNRS UMR 7357, Strasbourg, France
[3]GEPROVAS, Strasbourg, France
[4]IRIMAS UR 7499, Université de Haute-Alsace, Mulhouse, France

International Conference on Acoustics, Speech and Signal Processing 2021

June 06-11, 2021 - Toronto, Canada (virtual conference)

Outline

Probabilistic graphical models

- Very **active research topic**

Probabilistic graphical models

- Very **active research topic**
- Often used for **unsupervised problems**

Probabilistic graphical models

- Very **active research topic**
- Often used for **unsupervised problems**
- **Sparse models** $\rightarrow$ fast and exact computations $\rightarrow$ hugeness some image data (e.g. satellite images)

Probabilistic graphical models

- Very **active research topic**
- Often used for **unsupervised problems**
- **Sparse models** $\rightarrow$ fast and exact computations $\rightarrow$ hugeness some image data (e.g. satellite images)
- **Dense models** $\rightarrow$ approximating methods $\rightarrow$ model very complex phenomena (e.g. artifacts on medical images)

Probabilistic graphical models

- Very **active research topic**
- Often used for **unsupervised problems**
- **Sparse models** → fast and exact computations → hugeness some image data (e.g. satellite images)
- **Dense models** → approximating methods → model very complex phenomena (e.g. artifacts on medical images)
- Combined with deep learning → many **top current results** in image processing (e.g. Variational Autoencoders)

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x
- Random variables: X; their realizations: x

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x
- Random variables: X; their realizations: x
- For a discrete random variable $X \rightarrow p(\{X = x\})$ is denoted $p(X = x)$, or $p(x)$

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x
- Random variables: X; their realizations: x
- For a discrete random variable $X \to p(\{X = x\})$ is denoted $p(X = x)$, or $p(x)$
- For a continuous random variable Y, $p(y)$ is the density function of Y

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x
- Random variables: X; their realizations: x
- For a discrete random variable $X \rightarrow p(\{X = x\})$ is denoted $p(X = x)$, or $p(x)$
- For a continuous random variable Y, $p(y)$ is the density function of Y

Context of **Bayesian segmentation**:

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x
- Random variables: X; their realizations: x
- For a discrete random variable $X \rightarrow p(\{X = x\})$ is denoted $p(X = x)$, or $p(x)$
- For a continuous random variable Y, $p(y)$ is the density function of Y

Context of **Bayesian segmentation**:

- Segment an image with values in $\mathbb{R}$ into K classes $\{\omega_k\}_{k\in\{1,\dots,K\}} \triangleq \Omega$

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x
- Random variables: X; their realizations: x
- For a discrete random variable $X \rightarrow p(\{X = x\})$ is denoted $p(X = x)$, or $p(x)$
- For a continuous random variable Y, $p(y)$ is the density function of Y

Context of **Bayesian segmentation**:

- Segment an image with values in $\mathbb{R}$ into K classes $\{\omega_k\}_{k\in\{1,\ldots,K\}} \triangleq \Omega$
- $\boldsymbol{X} = (X_s)_{s\in\mathcal{S}}$ with value in $\Omega^{|\mathcal{S}|} \rightarrow$ the **hidden variables**.

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x
- Random variables: X; their realizations: x
- For a discrete random variable $X \rightarrow p(\{X = x\})$ is denoted $p(X = x)$, or $p(x)$
- For a continuous random variable Y, $p(y)$ is the density function of Y

Context of **Bayesian segmentation**:

- Segment an image with values in $\mathbb{R}$ into K classes $\{\omega_k\}_{k\in\{1,\dots,K\}} \triangleq \Omega$
- $\boldsymbol{X} = (X_s)_{s\in\mathcal{S}}$ with value in $\Omega^{|\mathcal{S}|} \rightarrow$ the **hidden variables**.
- $\boldsymbol{Y} = (Y_s)_{s\in\mathcal{S}}$ with value in $\mathbb{R}^{|\mathcal{S}|} \rightarrow$ the **observed variables**.

Probabilistic graphical models: probabilistic setting

Notations:

- Vectors: $\boldsymbol{x}$ / Scalars: x
- Random variables: X; their realizations: x
- For a discrete random variable $X \rightarrow p(\{X = x\})$ is denoted $p(X = x)$, or $p(x)$
- For a continuous random variable Y, $p(y)$ is the density function of Y

Context of **Bayesian segmentation**:

- Segment an image with values in $\mathbb{R}$ into K classes $\{\omega_k\}_{k \in \{1,\dots,K\}} \triangleq \Omega$
- $\boldsymbol{X} = (X_s)_{s \in \mathcal{S}}$ with value in $\Omega^{|\mathcal{S}|} \rightarrow$ the **hidden variables**.
- $\boldsymbol{Y} = (Y_s)_{s \in \mathcal{S}}$ with value in $\mathbb{R}^{|\mathcal{S}|} \rightarrow$ the **observed variables**.
- Segmentation criteria: $\rightarrow$ Maximum A Posteriori (MAP)
  $\rightarrow$ Maximum Posterior Mode (MPM)

Hidden Markov Models (HMM)

- Hidden Markov Models (HMM) (Baum and Petrie 1966) are the most popular type of probabilistic models

Hidden Markov Models (HMM)

- Hidden Markov Models (HMM) (Baum and Petrie 1966) are the most popular type of probabilistic models
- Applications in many contexts: image segmentation, speech processing, stock index forecasting, gene prediction, ...

Hidden Markov Models (HMM)

- Hidden Markov Models (HMM) (Baum and Petrie 1966) are the most popular type of probabilistic models
- Applications in many contexts: image segmentation, speech processing, stock index forecasting, gene prediction, ...
- Different models belong to the HMM family:

Hidden Markov Models (HMM)

- Hidden Markov Models (HMM) (Baum and Petrie 1966) are the most popular type of probabilistic models
- Applications in many contexts: image segmentation, speech processing, stock index forecasting, gene prediction, ...
- Different models belong to the HMM family:
  - Hidden and observed random variables

Hidden Markov Models (HMM)

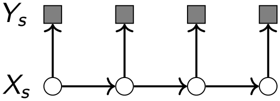
- Hidden Markov Models (HMM) (Baum and Petrie 1966) are the most popular type of probabilistic models
- Applications in many contexts: image segmentation, speech processing, stock index forecasting, gene prediction, ...
- Different models belong to the HMM family:
  - Hidden and observed random variables
  - Generative models $\rightarrow p(\boldsymbol{x}, \boldsymbol{y})$ is modeled

Hidden Markov Models (HMM)

- Hidden Markov Models (HMM) (Baum and Petrie 1966) are the most popular type of probabilistic models
- Applications in many contexts: image segmentation, speech processing, stock index forecasting, gene prediction, ...
- Different models belong to the HMM family:
  - Hidden and observed random variables
  - Generative models $\rightarrow p(\boldsymbol{x}, \boldsymbol{y})$ is modeled
  - $\boldsymbol{X}$ is a Markovian process and $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(y_s|x_s)$

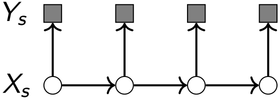
Hidden Markov Models: classical models

Hidden Markov Chain (HMC) (Baum, Petrie, et al. 1970)



Hidden Markov Models: classical models

Hidden Markov Chain (HMC) (Baum, Petrie, et al. 1970)

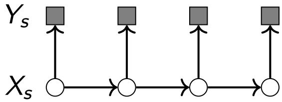


$p(\boldsymbol{x})$ is a Markov chain

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(x_r)p(y_r|x_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s|x_{s^-})p(y_s|x_s)$$

Hidden Markov Models: classical models

Hidden Markov Chain (HMC) (Baum, Petrie, et al. 1970)

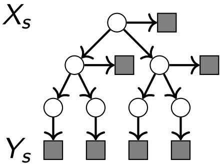


$p(\boldsymbol{x})$ is a Markov chain

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(x_r)p(y_r|x_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s|x_{s^-})p(y_s|x_s)$$

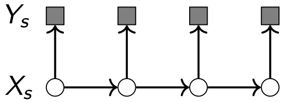
Hidden Markov Tree (HMT) (Laferté et al. 2000)

$\rightarrow$ a generalization of HMCs



Hidden Markov Models: classical models

Hidden Markov Chain (HMC) (Baum, Petrie, et al. 1970)

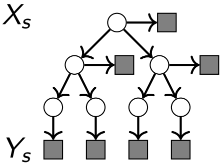


$p(\boldsymbol{x})$ is a Markov chain

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(x_r)p(y_r|x_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s|x_{s^-})p(y_s|x_s)$$

Hidden Markov Tree (HMT) (Laferté et al. 2000)

→ a generalization of HMCs

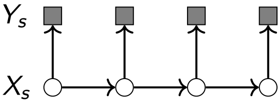


$p(\boldsymbol{x})$ is a Markov tree

$p(\boldsymbol{x}, \boldsymbol{y}) \rightarrow$ same as an HMC

Hidden Markov Models: classical models

Hidden Markov Chain (HMC) (Baum, Petrie, et al. 1970)

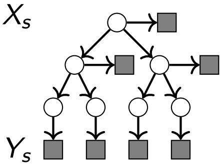


$p(\boldsymbol{x})$ is a Markov chain

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(x_r)p(y_r|x_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s|x_{s^-})p(y_s|x_s)$$

Hidden Markov Tree (HMT) (Laferté et al. 2000)

$\rightarrow$ a generalization of HMCs



$p(\boldsymbol{x})$ is a Markov tree

$p(\boldsymbol{x}, \boldsymbol{y}) \rightarrow$ same as an HMC

inference $\rightarrow$ **direct, exact computations** with Forward-Backward based algorithms

Outline

Motivations

How could we introduce richer direct dependencies in HMMs ?

Motivations

How could we introduce richer direct dependencies in HMMs ?

$\rightarrow$ Strong restrictions classically made in HMMs:

Motivations

How could we introduce richer direct dependencies in HMMs ?

→ Strong restrictions classically made in HMMs:

- X is constrained to be a Markov chain / tree

Motivations

How could we introduce richer direct dependencies in HMMs ?

→ Strong restrictions classically made in HMMs:

- $\boldsymbol{X}$ is constrained to be a Markov chain / tree
- The independent noise assumption $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(y_s|x_s)$

Motivations

How could we introduce richer direct dependencies in HMMs ?

→ Strong restrictions classically made in HMMs:

- $\boldsymbol{X}$ is constrained to be a Markov chain / tree
- The independent noise assumption $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(y_s|x_s)$
- More complex noise models: special cases of pairwise and triplet models

Motivations

How could we introduce richer direct dependencies in HMMs ?

→ Strong restrictions classically made in HMMs:

- $\boldsymbol{X}$ is constrained to be a Markov chain / tree
- The independent noise assumption $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(y_s|x_s)$
- More complex noise models: special cases of pairwise and triplet models

→ Pairwise and Triplet (Hidden) Markov Models are richer models:

Motivations

How could we introduce richer direct dependencies in HMMs ?

→ Strong restrictions classically made in HMMs:

- $\boldsymbol{X}$ is constrained to be a Markov chain / tree
- The independent noise assumption $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(y_s|x_s)$
- More complex noise models: special cases of pairwise and triplet models

→ Pairwise and Triplet (Hidden) Markov Models are richer models:

- Strict generalizations of HMMs (Gorynin et al. 2018)

Motivations

How could we introduce richer direct dependencies in HMMs ?

→ Strong restrictions classically made in HMMs:

- $\boldsymbol{X}$ is constrained to be a Markov chain / tree
- The independent noise assumption $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{s \in \mathcal{S}} p(y_s|x_s)$
- More complex noise models: special cases of pairwise and triplet models

→ Pairwise and Triplet (Hidden) Markov Models are richer models:

- Strict generalizations of HMMs (Gorynin et al. 2018)
- Conservation of the good properties of inference

Motivations

How could we introduce richer direct dependencies in HMMs ?

→ Strong restrictions classically made in HMMs:

- $\boldsymbol{X}$ is constrained to be a Markov chain / tree
- The independent noise assumption $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{s\in\mathcal{S}} p(y_s|x_s)$
- More complex noise models: special cases of pairwise and triplet models

→ Pairwise and Triplet (Hidden) Markov Models are richer models:

- Strict generalizations of HMMs (Gorynin et al. 2018)
- Conservation of the good properties of inference
- Naturally encompass extended HMMs models from the literature

Motivations

How could we introduce richer direct dependencies in HMMs ?

$\rightarrow$ Strong restrictions classically made in HMMs:

- $\boldsymbol{X}$ is constrained to be a Markov chain / tree
- The independent noise assumption $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{s\in\mathcal{S}} p(y_s|x_s)$
- More complex noise models: special cases of pairwise and triplet models

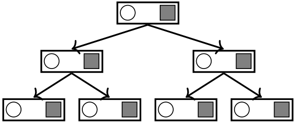
$\rightarrow$ Pairwise and Triplet (Hidden) Markov Models are richer models:

- Strict generalizations of HMMs (Gorynin et al. 2018)
- Conservation of the good properties of inference
- Naturally encompass extended HMMs models from the literature
- Triplet models integrate **auxiliary random variables** $\rightarrow$ link with deep learning models

Pairwise and triplet assumptions

Pairwise Markov Tree (Pieczynski 2002)

X_s, Y_s

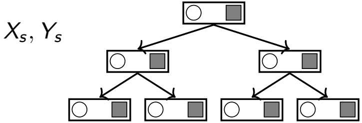


$p(\boldsymbol{x}, \boldsymbol{y})$ is a Markov tree

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(x_r, y_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s, y_s | x_{s^-}, y_{s^-})$$

Pairwise and triplet assumptions

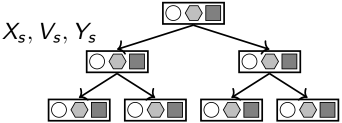
Pairwise Markov Tree (Pieczynski 2002)



$p(\boldsymbol{x}, \boldsymbol{y})$ is a Markov tree

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(x_r, y_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s, y_s | x_{s^-}, y_{s^-})$$

Triplet Markov Tree (TMT) (Courbot et al. 2018)



$p(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{y})$ is a Markov tree

$$p(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{y}) = p(x_r, v_r, y_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s, v_s, y_s | x_{s^-}, v_{s^-}, y_{s^-})$$

Triplet assumptions

- Neither $p(\boldsymbol{x})$, $p(\boldsymbol{y})$, $p(\boldsymbol{v})$, $p(\boldsymbol{x}, \boldsymbol{y})$, $p(\boldsymbol{y}, \boldsymbol{v})$, nor $p(\boldsymbol{x}, \boldsymbol{v})$ are necessarily Markovian distributions

Triplet assumptions

- Neither $p(\boldsymbol{x})$, $p(\boldsymbol{y})$, $p(\boldsymbol{v})$, $p(\boldsymbol{x},\boldsymbol{y})$, $p(\boldsymbol{y},\boldsymbol{v})$, nor $p(\boldsymbol{x},\boldsymbol{v})$ are necessarily Markovian distributions
- But $p(\boldsymbol{x},\boldsymbol{v}|\boldsymbol{y})$ (and the others...) are Markovian distributions
  $\rightarrow$ Inference can be done as in classical HMMs
  $\rightarrow$ Original hidden states:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \sum_{\boldsymbol{v}} p(\boldsymbol{x},\boldsymbol{v}|\boldsymbol{y})$$

Triplet assumptions

- Neither $p(\boldsymbol{x})$, $p(\boldsymbol{y})$, $p(\boldsymbol{v})$, $p(\boldsymbol{x}, \boldsymbol{y})$, $p(\boldsymbol{y}, \boldsymbol{v})$, nor $p(\boldsymbol{x}, \boldsymbol{v})$ are necessarily Markovian distributions
- But $p(\boldsymbol{x}, \boldsymbol{v}|\boldsymbol{y})$ (and the others...) are Markovian distributions
  $\rightarrow$ Inference can be done as in classical HMMs
  $\rightarrow$ Original hidden states:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \sum_{\boldsymbol{v}} p(\boldsymbol{x}, \boldsymbol{v}|\boldsymbol{y})$$

We now present the Spatial Triplet Markov Tree model:

$\rightarrow$ It takes advantage of the increased modeling possibilities
$\rightarrow$ It enhances the spatial correlations between random variables

Outline

Spatial Triplet Markov Trees (STMTs)

Distribution of STMTs (Gangloff et al. 2020) **(this paper)**:

$$p(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{y}) = p(x_r, \boldsymbol{v}_r, y_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s, \boldsymbol{v}_s, y_s | x_{s^-}, \boldsymbol{v}_{s^-}, y_{s^-})$$

Spatial Triplet Markov Trees (STMTs)

Distribution of STMTs (Gangloff et al. 2020) **(this paper)**:

$$p(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{y}) = p(x_r, \boldsymbol{v}_r, y_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s, \boldsymbol{v}_s, y_s | x_{s^-}, \boldsymbol{v}_{s^-}, y_{s^-})$$

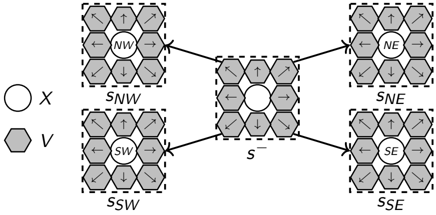
- Special design of $\boldsymbol{V}$ to improve spatial correlations in the classical HMT model. $\forall s \in \mathcal{S} : \boldsymbol{V}_s = (V^{\leftarrow}, V^{\nwarrow}, V^{\uparrow}, V^{\nearrow}, V^{\rightarrow}, V^{\searrow}, V^{\downarrow}, V^{\swarrow})$

Spatial Triplet Markov Trees (STMTs)

Distribution of STMTs (Gangloff et al. 2020) **(this paper)**:

$$p(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{y}) = p(x_r, \boldsymbol{v}_r, y_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s, \boldsymbol{v}_s, y_s | x_{s^-}, \boldsymbol{v}_{s^-}, y_{s^-})$$

- Special design of $\boldsymbol{V}$ to improve spatial correlations in the classical HMT model. $\forall s \in \mathcal{S} : \boldsymbol{V}_s = (V^{\leftarrow}, V^{\nwarrow}, V^{\uparrow}, V^{\nearrow}, V^{\rightarrow}, V^{\searrow}, V^{\downarrow}, V^{\swarrow})$
- Quadtrees: each site s^- has four sons $(s^{NW}, s^{NE}, s^{SE}, s^{SW})$ (except for last layer):

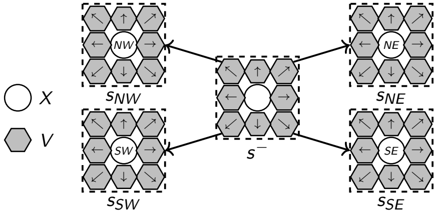


Spatial Triplet Markov Trees (STMTs)

Distribution of STMTs (Gangloff et al. 2020) **(this paper)**:

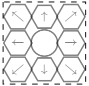
$$p(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{y}) = p(x_r, \boldsymbol{v}_r, y_r) \prod_{s \in \bar{\mathcal{S}}} p(x_s, \boldsymbol{v}_s, y_s | x_{s^-}, \boldsymbol{v}_{s^-}, y_{s^-})$$

- Special design of $\boldsymbol{V}$ to improve spatial correlations in the classical HMT model. $\forall s \in \mathcal{S} : \boldsymbol{V}_s = (V^{\leftarrow}, V^{\nwarrow}, V^{\uparrow}, V^{\nearrow}, V^{\rightarrow}, V^{\searrow}, V^{\downarrow}, V^{\swarrow})$
- Quadtrees: each site s^-has four sons $(s^{NW}, s^{NE}, s^{SE}, s^{SW})$ (except for last layer):



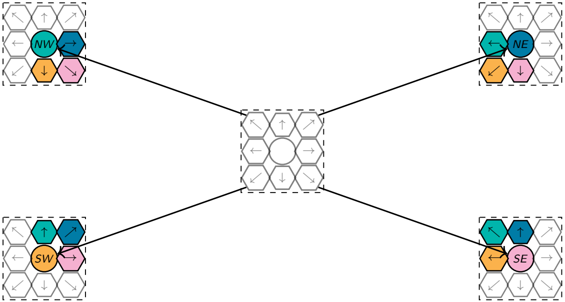
- We consider only observations Y_s at the finer resolution.

Designing the auxiliary process in STMTs



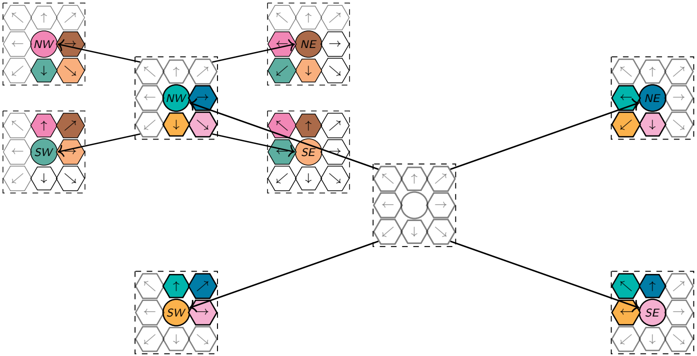
Propagation of spatial information: the same color indicates the same probability law

Designing the auxiliary process in STMTs



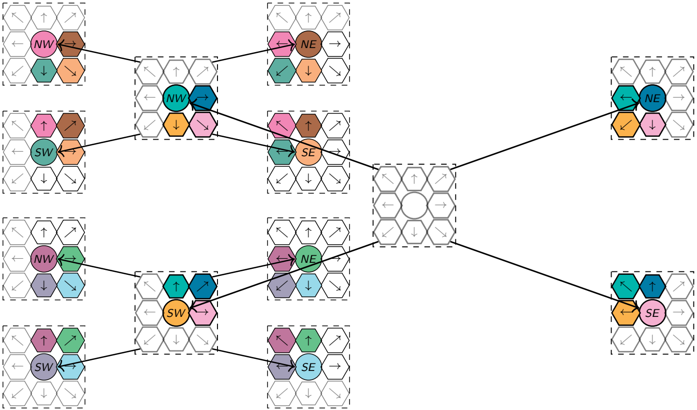
Propagation of spatial information: the same color indicates the same probability law

Designing the auxiliary process in STMTs



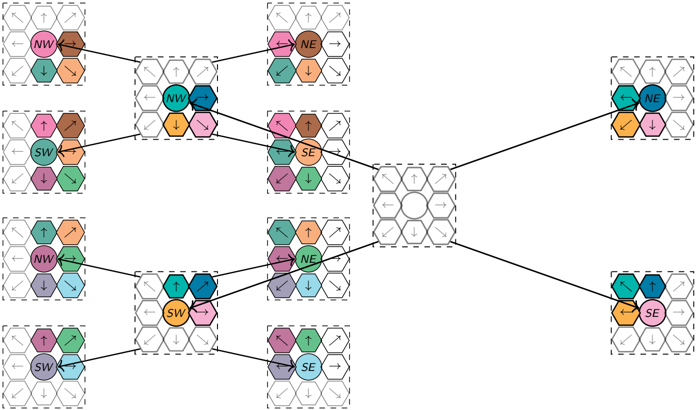
Propagation of spatial information: the same color indicates the same probability law

Designing the auxiliary process in STMTs



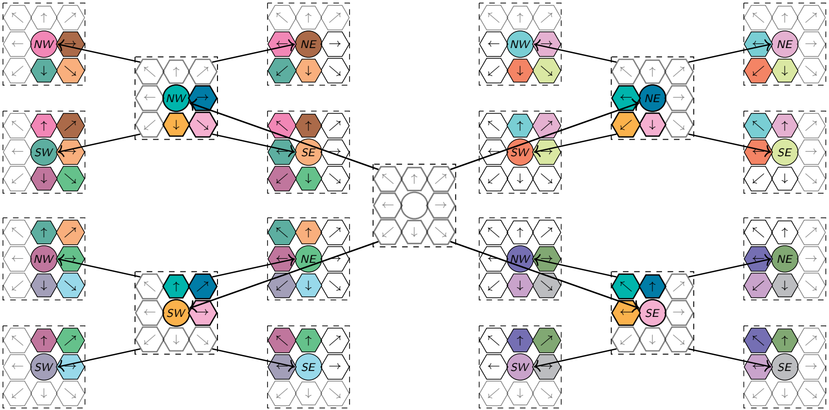
Propagation of spatial information: the same color indicates the same probability law

Designing the auxiliary process in STMTs



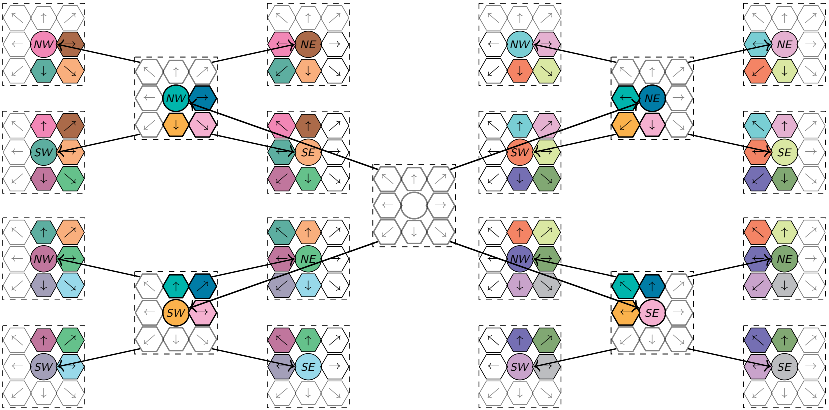
Propagation of spatial information: the same color indicates the same probability law

Designing the auxiliary process in STMTs



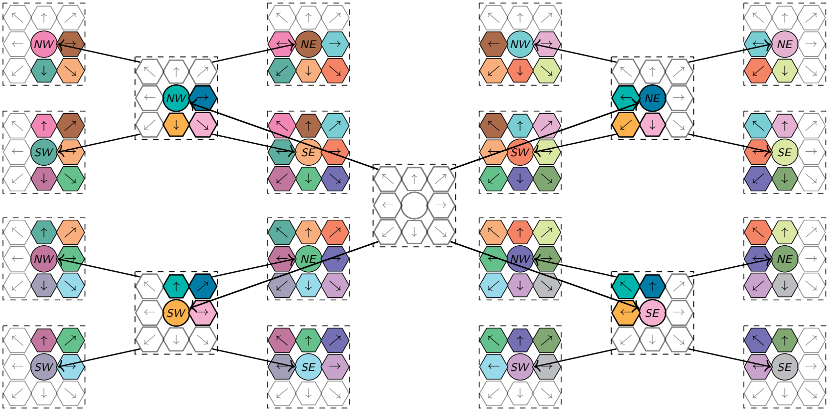
Propagation of spatial information: the same color indicates the same probability law

Designing the auxiliary process in STMTs



Propagation of spatial information: the same color indicates the same probability law

Designing the auxiliary process in STMTs



Propagation of spatial information: the same color indicates the same probability law

STMTs: numerical applications

- Ising-like transitions to propagate spatial homogeneity:

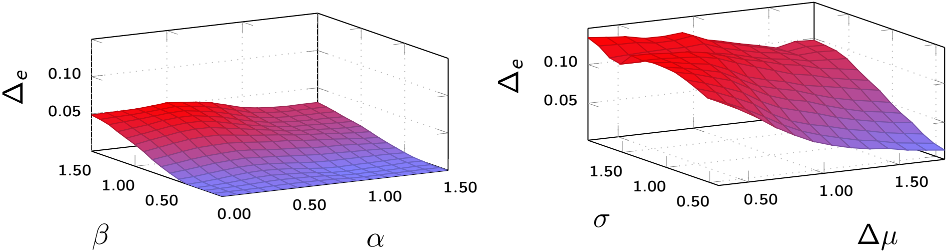
$$p(x_s|x_{s^-}, \boldsymbol{v}_{s^-}) = \frac{1}{Z}\exp\left(\alpha\delta_{x_s}^{x_{s^-}} + \sum_{v_{s^-}\in\boldsymbol{v}_{s^-}} \beta\delta_{x_s}^{v_{s^-}}\right), \text{ with } (\alpha, \beta) \in \mathbb{R}_+^2.$$

STMTs: numerical applications

- Ising-like transitions to propagate spatial homogeneity:

$$p(x_s|x_{s^-}, \boldsymbol{v}_{s^-}) = \frac{1}{Z}\exp\left(\alpha\delta_{x_s}^{x_{s^-}} + \sum_{v_{s^-}\in\boldsymbol{v}_{s^-}} \beta\delta_{x_s}^{v_{s^-}}\right), \text{ with } (\alpha,\beta)\in\mathbb{R}_+^2.$$

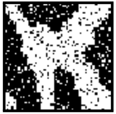
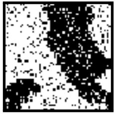
- Simulate STMTs and restore with STMTs and HMTs (known α, β parameters)



With $\Delta_e = err_{HMT} - err_{STMT}$: relative error rate

$\rightarrow$ For all model parameters and noise levels STMT performs better than HMT

Clamped sampling experiment

MF (Potts potential)		...		Average difference rate to MRF (100 simulations)
STMT (Potts-like potential)		...		6.8%
MT (Potts-like potential)		...		14.2%

Clamped samplings of the original $\boldsymbol{X}$ *process (last layer only)*

→ STMTs seem to capture the best the spatial context created by Markov Fields (MFs)

Outline

Unsupervised parameter estimation

Case of independent Gaussian observations

We introduce Linear Least Square (LLS) estimator for α^t and β^t

$$[\alpha^t, \beta^t] = (B^T B)^{-1} B^T A$$

where the generic term of vector A is, $\forall s \in \mathcal{S}^L$, $\forall (x_s, x_s') \in \Omega^2$,

$$A_s = \ln \frac{p(x_s, x_{s^-}, \boldsymbol{v}_{s^-}|\boldsymbol{y})}{p(x_s', x_{s^-}, \boldsymbol{v}_{s^-}|\boldsymbol{y})} - \ln \frac{1}{\sqrt{2\pi\sigma_{x_s}^2}} + \frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} + \ln \frac{1}{\sqrt{2\pi\sigma_{x_s'}^2}} + \frac{(y_s - \mu_{x_s'})^2}{2\sigma_{x_s}^2}$$

and the generic line of matrix B is

$$B_{s,:} = \left[\delta_{x_s}^{x_{s^-}} - \delta_{x_s'}^{x_{s^-}}, \sum_{v_{s^-} \in \boldsymbol{v}_{s^-}} \left(\delta_{x_s}^{v_{s^-}} - \delta_{x_s'}^{v_{s^-}} \right) \right]$$

Unsupervised parameter estimation

Case of independent Gaussian observations

Maximum Likelihood (ML) estimator for $\boldsymbol{\mu}^t$ and $\boldsymbol{\sigma}^t$, $\forall \omega \in \Omega$:

$$\mu_\omega^t = \frac{1}{\sum_{s\in\mathcal{S}^L} \mathbb{1}_{\{x_s^t=\omega\}}} \sum_{s\in\mathcal{S}^L} y_s \mathbb{1}_{\{x_s^t=\omega\}}$$

and

$$\sigma_\omega^t = \left(\frac{1}{\sum_{s\in\mathcal{S}^L} \mathbb{1}_{\{x_s^t=\omega\}}} \sum_{s\in\mathcal{S}^L} \left(y_s - \mu_\omega^t\right)^2 \mathbb{1}_{\{x_s^t=\omega\}} \right)^{\frac{1}{2}}$$

Unsupervised parameter estimation

Algorithm 1: Iterative Parameter Estimation for Trees for STMTs.
Case $\Omega = \{\omega_1, \omega_2\}$

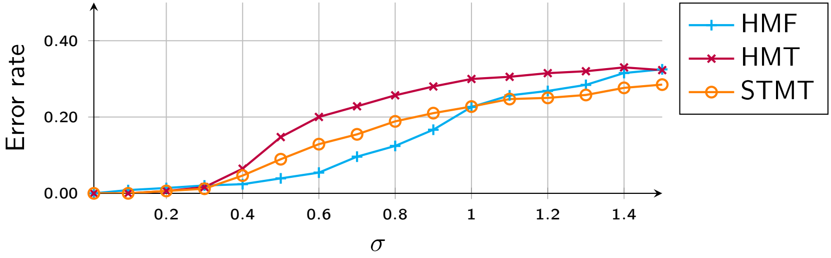
Data: $\boldsymbol{\theta}^0 = \{\alpha^0, \beta^0, \mu_0^0, \mu_1^0, \sigma_0^0, \sigma_1^0\}$, an initial set of parameters,
$\boldsymbol{y}$, the observations.

Result: $\boldsymbol{\theta}^* = \{\alpha^*, \beta^*, \mu_0^*, \mu_1^*, \sigma_0^*, \sigma_1^*\}$, the estimated parameters.

```
t ← 1
while convergence is not attained do
    1. MPM estimation:
    x̂_s^{MPM,t} = argmax_{x_s} p(x_s | y, θ^{t−1}), ∀s ∈ S
    2. Estimation with the complete data (x̂^{MPM,t}, y):
    • LLS estimator for α^t and β^t
    • ML estimator for μ_0^t and μ_1^t
    • ML estimator for σ_0^t and σ_1^t
    θ^t ←= {α^t, β^t, μ_0^t, μ_1^t, σ_0^t, σ_1^t}
    t ← t + 1
```

Unsupervised image segmentation experiment

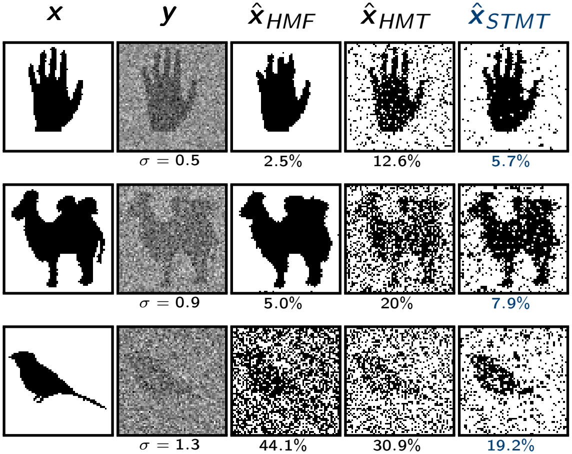
Comparing Hidden Markov Fields (HMFs), HMTs and STMTs in unsupervised segmentation:



Error rate in unsupervised segmentation function of the noise level

→ STMTs greatly improve HMT results
→ STMTs closer to HMF results

Unsupervised image segmentation experiment



Unsupervised segmentation HMF, HMT and STMT models

Outline

Conclusion

Summary

- STMTs are generalizations of HMTs
  $\rightarrow$ increased modeling possibilities
  $\rightarrow$ strengthened spatial correlations
- Inference remains exact and deterministic

Conclusion

Summary

- STMTs are generalizations of HMTs
  $\rightarrow$ increased modeling possibilities
  $\rightarrow$ strengthened spatial correlations
- Inference remains exact and deterministic

Perspectives

- *Spatial* correlations induced $\rightarrow$ theoretical links between STMTs and HMFs ?
- STMTs as the variational distribution for variational inference in trees with semi-cycles as in (Gangloff et al. 2020) for 2D segmentation

References I

[1] L. E. Baum and T. Petrie. "Statistical inference for probabilistic functions of finite state Markov chains". In: *The annals of mathematical statistics* 37.6 (1966), pp. 1554–1563.

[2] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains". In: *The annals of mathematical statistics* 41.1 (1970), pp. 164–171.

[3] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[4] J.-B. Courbot, E. Monfrini, V. Mazet, and C. Collet. "Triplet markov trees for image segmentation". In: *SSP 2018: IEEE Workshop on Statistical Signal Processing*. IEEE Computer Society, 2018, pp. 233–237.

[5] H. Gangloff, J.-B. Courbot, E. Monfrini, and C. Collet. "Spatial Triplet Markov Trees for auxiliary variational inference in Spatial Bayes Networks". In: *Stochastic Modeling Techniques and Data Analysis international conference (SMTDA'20)*. 2020. In press.

[6] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images". In: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1984), pp. 721–741.

[7] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT Press, 2016.

[8] I. Gorynin, H. Gangloff, E. Monfrini, and W. Pieczynski. "Assessing the segmentation performance of pairwise and triplet Markov models". In: *Signal Processing* 145 (2018), pp. 183–192.

[9] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT Press, 2009.

[10] J.-M. Laferté, P. Pérez, and F. Heitz. "Discrete Markov image modeling and inference on the quadtree". In: *IEEE Transactions on image processing* 9.3 (2000), pp. 390–404.

[11] K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT Press, 2012.

[12] W. Pieczynski. "Arbres de Markov couple". In: *Comptes Rendus Mathématique* 335.1 (2002), pp. 79–82.