Triplet markov trees for image segmentation
Jean-Baptiste Courbot, Emmanuel Monfrini, Vincent Mazet, Christophe Collet

To cite this version:
TRIPLET MARKOV TREES FOR IMAGE SEGMENTATION

Jean-Baptiste Courbot\textsuperscript{1,2}, Emmanuel Monfrini\textsuperscript{3}, Vincent Mazer\textsuperscript{1}, Christophe Collet\textsuperscript{1}

\textsuperscript{1}ICube, Universit\'e de Strasbourg - CNRS, 67412 Illkirch, France
\textsuperscript{2}Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, CRAL UMR5574, F-69230, Saint-Genis-Laval, France
\textsuperscript{3}D\'epartement CITI, CNRS, UMR 5157 SAMOVAR, 91011 \`Evry, France

ABSTRACT
This paper introduces a triplet Markov tree model designed to minimize the block effect that may be encountered while segmenting image using Hidden Markov Tree (HMT) modeling. We present the model specificities, the Bayesian Maximum Posterior Mode segmentation, and a parameter estimation strategy in the unsupervised context. Results on synthetic images show that the method greatly improves over HMT-based segmentation, and that the model is competitive with a hidden Markov field-based segmentation.

Index Terms— Triplet Markov Tree, Image Segmentation, Unsupervised segmentation

1. INTRODUCTION
Markov models have proven their efficiency in the context of Bayesian image segmentation, i.e. the estimation of the classification $X$ from the observation $Y = y$. Among those models, one can mention hidden Markov chains, hidden Markov fields and hidden Markov trees (see for instance [1, 2, 3, 4] and the references therein). The Bayesian segmentation is obtained after choosing an estimator, such as the Maximum Posterior Mode (MP) \[2\] in an appropriate context. Posterior Mode (MPM) \[5\] or the Maximum A Posteriori estimation, within a segmentation context, of a hierarchy representing the classification to obtain. The main drawback of this hierarchy is the introduction, in the most difficult cases, either of “speckles” or block effects in the segmentation result. Several models were introduced to compensate these effects. For instance, let us mention the evolutive Markov tree models \[6\], in which the parent-child transition probabilities depend on the considered resolution. One can also mention the hierarchical fields modeling, in which the prior density is Markovian both spatially (over each scale) and hierarchically \[7\].

In a similar fashion as Markov chains and Markov fields models, Markov tree models have been enriched by the introduction of triplets models \[8, 9, 10\]. In this framework, in addition to the observed process $Y$ and the classification $X$, a third auxiliary process denoted $U$ is adjoined. The objective of such model is the accurate handling of more complex phenomena, such as privileged orientation for instance \[11\].

In this paper, we introduce a Triplet Markov Tree (TMT) model, in which the auxiliary process modulates the parent-child transition probabilities, depending on the classification of the parent’s neighbors. This model will be referred as Spatial Triplet Markov Tree (STMT) and is introduced in Section 2, along with the MPM segmentation computation. We apply this model in the case of a single (high-resolution) image over a classical HMT method. We also show that this method is competitive with a combination of HMT and Hidden Markov Field (HMF) models.

2. SPATIAL TRIPLET MARKOV TREES

2.1. Model
Let $T = (T_s)_{s \in S}$ be a stochastic process, where $S$ is the set of resolutions of a quadtree: $S = \{S^0, \ldots, S^N\}$. Each $S^n$ contains $4^n$ sites : $n = 0$ represents the tree root, and $n = N$ is the finest resolution. Besides, we set $T_s = (Y, X, U)$, where $Y \in \mathbb{R}^{S^0}$ is an observation process, $X \in \Omega_{\mathbb{R}^{S^0}}$ is a class process and $U \in \Omega_{u}^{S^0}$ is an auxiliary process.
can be found in [12]. We adapted this method for the more concision, we note P. For concision, we note TP. We choose to use the MPM estimator for the segmentation. 2.2. MPM segmentation

For concision, we note TP. We choose to use the MPM estimator for the segmentation. 2.2. MPM segmentation

Formally, T is a Markov tree [3] and verifies:

\[ p(T) = p(T_r) \prod_{s \in S \setminus S^0} p(T_s|T_{s^-}); \]  

where \( s^- \) is the parent site of \( s \), and \( T_r \) is the value of \( T_s \) in the root site \( s \in S^0 \).

\( U \) is an auxiliary process tuning the distribution of \( X \) at each parent-child transition. Each \( U_s \) is a random vector ruled by the same distribution as \( (X_s)_{s \in V_s} \), where \( V_s \) is a neighborhood of \( s \) to be defined. In the remaining of this paper, we assume that \( X_s \) and \( U_s \) are independent given \( T_{s^-} \). This yields \( \forall s \in \{S^1, \ldots, S^N\} : \)

\[ p(T_s|T_{s^-}) = p(Y_u|X_s, U_s, T_{s^-})p(X_s|T_{s^-})p(U_s|T_{s^-}). \]  

The three distributions appearing in this equation are specified in Section 3. Figure 1 represents the quadtree dependency graph and the dependencies involved in the HMT, the TMT and the STMT models.

2.2. MPM segmentation

We choose to use the MPM estimator for the segmentation. For concision, we note \( p(A = a) \) as \( p(a) \) when possible. The MPM estimator requires the computation of \( p(X_s = \omega_s, U_s = \nu_s|X_{s^-} = \omega_{s^-}, U_{s^-} = \nu_{s^-}) \) in each \( s \in S \) and for all \( (\omega_s, \nu_s, \omega_{s^-}, \nu_{s^-}) \in (\Omega_x \times \Omega_u)^2 \).

An algorithm for computing the MPM in the HMT model can be found in [12]. We adapted this method for the more general case of TMT used here. It consists of the five steps described below, and requires the computation of auxiliary functions denoted \( A_n \) for \( n \in \{0, 1, \ldots, N\} \) in the following fine-to-coarse pass:

1. At the finest resolution, \( \forall s \in S^N : \)

\[ A_N(s; \omega_i, \nu_j, \omega_k, \nu_l) = p(y_s, \omega_i, \nu_j|y_{s^-}, \omega_k, \nu_l). \]  

2. On the intermediate resolutions, \( \forall s \in \{S^1, \ldots, S^{N-1}\} : \)

\[ A_n(s; \omega_i, \nu_j, \omega_k, \nu_l) = p(y_s, \omega_i, \nu_j|y_{s^-}, \omega_k, \nu_l) \times \prod_{s \in \mathcal{E}_s} \sum_{(\omega, \nu) \in \Omega_x \times \Omega_u} A_{n+1}(s^+; \omega, \nu, \omega_i, \nu_j); \]  

where \( \mathcal{E}_s \) is the children set of \( s \).

3. On the root, \( r \in S^0 : \)

\[ A_0(\omega_i, \nu_j) = p(y_0, \omega_i, \nu_j) \]

\times \prod_{s \in \mathcal{E}_0} \sum_{(\omega, \nu) \in \Omega_x \times \Omega_u} A_1(s^+; \omega, \nu, \omega_i, \nu_j); \]  

where \( \mathcal{E}_0 = S^1 \).

A triplet Markov tree benefits from the posterior Markovianity: \( p(X, U|Y = y) \) is a Markov distribution\(^1\). Hence, it is feasible to compute \( p(X_s = \omega_s, U_s = \nu_s|y) \) for \( s \in S^N \) and \( (\omega_s, \nu_s) \in \Omega_x \times \Omega_u \) in the following coarse-to-fine pass:

4. Computation of the posterior distribution in the root \( r : \)

\[ p(X_r = \omega_i, U_r = \nu_j|y) = \frac{A_0(\omega_i, \nu_j)}{\sum_{(\omega, \nu) \in \Omega_x \times \Omega_u} A_0(\omega, \nu)}. \]  

5. Computation of the posterior \( \forall s \in \{S^1, \ldots, S^N\} : \)

\[ p(X_s = \omega_i, U_s = \nu_j|X_{s^-} = \omega_{s^-}, U_{s^-} = \nu_{s^-}, y) = \frac{A_n(s; \omega_i, \nu_j, \omega_{s^-}, \nu_{s^-})}{\sum_{(\omega, \nu) \in \Omega_x \times \Omega_u} A_n(s; \omega, \nu, \omega_{s^-}, \nu_{s^-})}. \]  

Once the cascading posterior density computations are performed, one can obtain the MPM estimation for the root:

\( (\hat{x}_r, \hat{u}_r)_{MPM} = \arg \max_{(\omega_i, \nu_j) \in \Omega_x \times \Omega_u} p(X_r = \omega_i, U_r = \nu_j|y); \)  

and \( \forall s \in \{S^1, \ldots, S^N\} : \)

\( (\hat{x}_s, \hat{u}_s)_{MPM} = \arg \max_{(\omega_i, \nu_j) \in \Omega_x \times \Omega_u} p \left( X_s, U_s = (\omega_i, \nu_j) \mid (\hat{x}_{s^-}, \hat{u}_{s^-})_{MPM}, y \right); \)  

so that the MPM estimation is the set \( ((\hat{x}_s, \hat{u}_s)_{MPM})_{s \in S^*} \).

\(^1\)Let us add that \( p(X|Y = y) \) is not necessarily a Markov distribution.
3. APPLICATION TO IMAGE SEGMENTATION

3.1. Model specificities

An observed image is represented by \((Y_s)_{s \in \mathcal{S}} \in \mathbb{R}^{[\mathcal{S}]}\), and we assume that:

\[
p(Y_s|X_s, U_s, T_s \sim) = p(Y_s|X_s).
\]  

(10)

Besides, \(\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^{N-1}\}\), we formally consider that
\(p(Y_s|X_s, U_s, T_s \sim) \propto 1\) and \(p(Y_s|X_s, U_r) \propto 1\).

Concerning the auxiliary process \(U\), the considered neighborhood is the set of 8 closest neighbors at the same resolution. Hence \(U_s\) is 8-valued and is noted: \(U_s = (U_s,1, \ldots, U_s,8)\). Since \(U_s\) has the same distribution as \((X_v)_{v \in V_s}\) and since the \((X_v)_{v \in V_s}\) are independent given \(T_{s-}\), one can write:

\[
p(U_s|T_{s-}) = p((X_v)_{v \in V_s}|T_{s-}) = \prod_{v \in V_s} p(X_v|T_{s-}).
\]  

(11)

\(X_v\) may belong to the children of \(T_{s-}\) or not (see Figure 2). In the first case, we define \(p(X_v|T_{s-})\) along the lines of a Potts model:

\[
p(X_v|T_{s-}) \propto \exp\left(\beta \delta_{X_v}^{X_{s-}} + \gamma \sum_{1 \leq k \leq 8} \delta_{X_v}^{X_{s-},k}\right); \tag{12}
\]

where \(\delta_{X_v}^{X_{s-}}\) is the Kronecker product, which equals to 1 when \(a = b\) and 0 otherwise, and \(\beta\) and \(\gamma\) are model parameters.

Remark. When \(\gamma = 0\), the transition probabilities correspond to a classical HMT model.

In the second case, when \(X_v\) does not belong to the children of \(T_{s-}\), we know that \(X_v \in (X_w)_{w \in V_{s-}}\). Let us assume that \(U_{s-} = \omega_s\) models the same site as the parent of \(v\).

Then, we define:

\[
p(X_v|T_{s-}) \propto \exp(\beta \delta_{U_{s-}^{X_{s-}}} + \gamma \delta_{U_{s-}^{X_{s-},k}} + \gamma \sum_{1 \leq k \leq 8} \delta_{U_{s-}^{X_{s-},k,k'}}).
\]  

(13)

Besides, in the root and \(\forall (\omega_i, \nu_j) \in \Omega_\omega \times \Omega_\nu\), we have
\(p(T_0 = (y_0, \omega_i, \nu_j)) = \pi_i\), where the \(\pi_i\) are model parameters representing the prior on the tree root.

Remark. In practice, the 8-neighborhood is not defined on the image border. In the model implementation, borders are replicated to bypass this limit.

In the remaining of this paper, we assume a Gaussian noise model parametrized by \(\mu_i, \sigma_i\) and consider the segmentation with \(|\Omega_\omega| = 2\) classes.

3.2. Unsupervised Parameter Estimation

The model parameters are:

\[
\theta = \{\mu_0, \mu_1, \sigma_0, \sigma_1, \pi_0, \pi_1, \gamma, \beta\}. \tag{14}
\]

In an unsupervised framework, \(\theta\) must be estimated from \(Y = y\) only. At first, we present the estimators when a “complete” realization \((y, x, u)\) is available.

The estimation of \(\mu_i\) et \(\sigma_i\) is performed with the maximum likelihood estimator using \((x, y)\). The \(\pi_i\) are directly estimated from the posterior density computation (6):

\[
\hat{\pi}_i = p(X_r = \omega_k|Y = y) = \sum_{\nu \in \Omega_\nu} p(X_r = \omega_k, U_r = \nu|Y = y).
\]

(15)

We now detail the estimators for \(\beta\) and \(\gamma\), inspired by the least-squares estimator from [13], initially proposed for Markov field distributions. We have \(\forall s \in \{\mathcal{S}^1, \ldots, \mathcal{S}^{N}\}\) and for \(\omega_i \neq \omega_j\):

\[
p(X_s = \omega_i|T_{s-}) = \frac{p(X_s = \omega_i|T_{s-})}{p(X_s = \omega_j|T_{s-})} = \exp\left[\beta \left(\delta_{X_s}^{\omega_i} - \delta_{X_s}^{\omega_j}\right) + \gamma \sum_{1 \leq k \leq 8} (\delta_{U_{s-}}^{\omega_i,k} - \delta_{U_{s-}}^{\omega_j,k})\right].
\]  

(16)

The left-hand term of this equation is computed thanks to the posterior distribution (7)². We obtain the “partial” estimation of \(\beta\) for all pairs \((x_s, t_{s-})\):

\[
\hat{\beta}_{s,s-} = \frac{1}{|\mathcal{S} \backslash \mathcal{S}^0|} \sum_{s \in \mathcal{S} \backslash \mathcal{S}^0} \hat{\beta}_{s,s-}.
\]

(18)

²In practice, this estimator is more robust than the histogram estimation.
Based on the same reasoning, we obtain with (7) and for all pairs \((x_s, t_{s-})\):

\[
\hat{\gamma}_{s, s^{-}} = \log \frac{p(x_s = \omega_i | t_{s-}, Y = y)}{p(x_s = \omega_j | t_{s-}, Y = y)} - \beta \left( \delta_{\omega_i} - \delta_{\omega_j} \right) - \sum_{1 \leq k \leq 8} \left( \delta_{u_{s-}, k} - \delta_{u_{s-}, k} \right)
\]

When \(\sum_{1 \leq k \leq 8} \left( \delta_{u_{s-}, k} - \delta_{u_{s-}, k} \right) = 0\), this “partial” estimation is not defined. Denoting \(C(s^{-}, i, j)\) this value, the least-squares estimation of \(\gamma\) is:

\[
\hat{\gamma} = \frac{\sum_{s \in S, t \in O} I\{C(s^{-}, i, j) \neq 0\} \hat{\gamma}_{s, s^{-}}}{\sum_{s \in S, t \in O} I\{C(s^{-}, i, j) \neq 0\}}.
\]

The parameters \(\theta\) are estimated with the Stochastic Expectation-Maximization (SEM) method [14]. This stochastic variant of EM [15] iteratively generates realizations of \((X, U)\) thanks to the previous value of \(\theta\), and re-estimate \(\theta\) thanks to the simulated realizations with the estimators presented in this section.

## 4. NUMERICAL RESULTS

We use two synthetic images of size \(128^2\) pixels, denoted \(x_A\) and \(x_B\). The former is a realization of a STMT using \(\beta = 2.25\) et \(\gamma = 0.25\) (Figure 3a), and the latter present wide homogeneous regions with smooth boundaries (Figure 3f). Besides, we set \(\mu_0 = 0, \mu_1 = 1\) and \(\sigma_0 = \sigma_1 = \sigma\) so that \(\sigma\) tunes the Signal-to-Noise Ratio (SNR) defined as \(20 \log_{10}(\bar{x}/\sigma)\), \(\bar{x}\) being the averaged value of \(x\). In this setting, we evaluate three methods:

- a “mixed” HMT/HMF method: the segmentation is made within a classical HMF model, whose Gibbs sampling are initialized using the HMT result. This model is expected to handle well both spatial and hierarchical features in images;
- the STMT-based segmentation introduced in this paper.

Figures 3b-3d and 3g-3i illustrates the segmentation results, and the error rate evaluated on 100 simulations are displayed in Figures 3e and 3j.

These results first show that, in any cases, the HMT-based segmentation is outperformed by its alternatives. Besides, the HMT/HMF and STMT model yields close error rates, and the achievement of the best average error rate depends on the image to process. This can be visually interpreted as a consequence of the “smoothness” of the image \(x_B\), which favors a Markov field-based method. As a final remark, let us add that our numerical experiments showed that the computation of the STMT-based segmentation method is always faster than the HMT/HMF-based segmentation (up to 10 times faster on \(512^2\) images) due to the ability to compute exactly the posterior densities.

## 5. CONCLUSION

This paper introduced the STMT model, as well as the computation of the MPM criterion and the segmentation of images. Results showed that the model is robust and competitive with a mixed HMT/HMF model, while providing the ability to compute exactly the posterior densities.
6. REFERENCES


