



Research article

Unsupervised segmentation of images using bi-dimensional pairwise Markov chains model

A. Joumad¹, A. El Moutaouakkil¹, A. Nasroallah², O. Boutkhoum¹, Mejdil Safran^{3,*}, Sultan Alfarhood³ and Imran Ashraf^{4,*}

¹ Department of informatics, Chouaib Doukkali University, Faculty of Sciences, B. P. 299-24000, El Jadida, Morocco

² Department of mathematics, Cadi Ayyad University, Faculty of Sciences Semlalia, B. P. 2390, Marrakesh, Morocco

³ Department of Computer Science, College of Computer and Information Sciences, King Saud University, P.O.Box 51178, Riyadh 11543, Saudi Arabia

⁴ Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea

* **Correspondence:** Email: imranashraf@ynu.ac.kr, mejdil@ksu.edu.sa.

Abstract: The pair-wise Markov chain (PMC) model serves as an extension to the hidden Markov chain (HMC) model and has been widely used in unsupervised restoration tasks associated with reconstructing the hidden data. In fact, the PMC model can treat fairly complicated situations for which application of Bayesian restoration estimators such as maximum *A Posteriori* (MAP), or maximal *Posterior* mode (MPM) remains possible. The major novelty in this work is to construct a PMC model with observational data in two dimensions, and subsequently adapt the estimation algorithms, as well as, image restoration methods for that context. Often, the transformation of an image from a two-dimensional format to a one-dimensional sequence occurs via Hilbert-Peano scan (HPS), whereas in the proposed model, the second component of the observed process takes over this role to exceed the situation of pixel missing information after transformation for a to be segmented image. To reconstruct the hidden process, we used the MPM decision criterion after estimating the model's parameters with two algorithms: Stochastic expectation maximization (SEM) and iterative conditional estimation (ICE). In this study, experimental, numerical, and visual results are shown to demonstrate the superiority of the proposed model over the classical PMC for unsupervised restorations.

Keywords: Bayesian restoration; hidden Markov chain model; pairwise Markov chain model; unsupervised segmentation; Hilbert-Peano scan; ICE algorithm; SEM algorithm; bi-dimensional pairwise Markov chain

1. Introduction

The hidden Markov chain (HMC) model, as characterized in [1–5], is a doubly stochastic process, usually denoted by (\mathbf{X}, \mathbf{Y}) . The process \mathbf{X} , which models a to be segmented image, is assumed to be hidden and Markovian, and the process \mathbf{Y} , which models the noisy version of the image to be recuperated, is assumed to be observed and real. The success of the HMC model stems from the fact that, when the noise is not too complex, several Bayesian classification approaches such as maximum *A Posteriori* (MAP) or maximal *A Posteriori* (MPM) estimators can be used to determine the image $\mathbf{X} = \mathbf{x}$ from the observed version $\mathbf{Y} = \mathbf{y}$. In addition, the spatial regularity of pixels may hinder the effectiveness of the HMC model in image segmentation.

We consider the problem of segmenting a satellite image into two classes, “water” and “trees”. The hidden image represents all of the classes, whereas the observed image represents all of the gray levels on each pixel. In general, all neighboring pixels are considered to have a higher probability of the same class than pixels situated further from each other. Nevertheless, in this situation, we can detect pixels that are among a set of pixels in identical groups or near edges but have distinct appearances. According to the classical hypothesis, this attribute cannot be accommodated in the HMC model. For this reason, the HMC model has been augmented to include the pair-wise Markov chain (PMC) model. The PMC model is more general than the HMC model, given the process \mathbf{X} is not necessarily a Markov chain and the pair (\mathbf{X}, \mathbf{Y}) is directly Markovian. Another advantage of PMC against HMC is its capacity to easily take into account the noise correlation. Furthermore, the PMC model, like HMC, is capable of implementing Bayesian MAP and MPM restorations.

Another major problem is that when a one-dimensional sequence is generated from a two-dimensional image using any reading process, such as human preference score (HPS), some locality information of the pixels is lost. What prompted us in this study was the success of adding additional components to the observations in the studies [6–8] in the framework of the HMC model by taking advantage of a neighborhood of each pixel (for example, four nearest neighbors), which culminated in an impressive model on a range of synthetic images. In this work, we apply the same concept to the PMC model in order to tackle the main issue of spatial regularity of pixels, as well as the problem of pixel information loss in an image, which is the objective of the segmentation process. As a result, we formulate a PMC model with bidimensional observed data (BPMC), whose most important task is to segment images while taking into account the above-mentioned last two difficulties. The focus of this study is to look into the potential benefits of BPMC over PMC in terms of unsupervised segmentation. Initially, the unidentified parameters of the model must be estimated. In the case of PMC, we cannot simply maximize the likelihood analytically, which is the case for the expectation-maximization (EM) algorithm [9, 10], at each iteration, so we need to apply iterative conditional estimation (ICE) and SEM [11–13]. In general, the ICE and structural equation models (SEM) algorithms are very powerful when used with Gaussian estimation and generalized mixture. In this study, we adapt the ICE and SEM algorithms to the new model. Furthermore, SEM and ICE algorithms are examined in the presence of Gaussian noise for both PMC and BPMC, and it is discovered that SEM and ICE algorithms are more

adaptable because likelihood is not always required. In addition, the estimators beyond the maximum likelihood can also be utilized. Therefore, we need to answer the following questions.

- Does the use of BPMC significantly improve restoration results, compared to the PMC model, and how do ICE and SEM algorithms work in the BPMC context?
- Does the use of BPMC compared to PMC permit the image to be converted into a one-dimensional sequence without losing pixel locality information?

The last point ought to be considered since all kinds of hidden Markov chain models based on HPS have shown potential for image segmentation, and in certain situations, they may even be competing with the hidden Markov field (HMF) model in classification [14–16]. On top of that, the HMF model allows for a more precise and more intuitive modeling of the spatial relationships between pixels [17], while the PMC model requires HPS to account for the spatial details of the pixels. Conversely, the BPMC model takes into account temporal and spatial pixel information by incorporating an additional element right into the observed process, allowing it to compete with the PMC model in this regard.

The rest of this paper is structured into seven sections. Section 2 presents studies relevant to the current study. In Section 3, we provide the reader with a mathematical background of the BPMC model. Section 4 presents the two models discussed in this study and their properties. Section 5 is committed to the simulation of PMC and BPMC models in a supervised way using MPM classifier-based classifications. Section 6 deals with estimating parameter algorithms; here, we describe the main steps of the ICE and SEM algorithms in the case of the two models, PMC and BPMC. In Section 7, with numerical experiments for various noise factors, we give a study of comparison of the two estimation procedures based on the accuracy of parameter estimates for both models. Furthermore, using error rates and the peak signal-to-noise ratio (PSNR) index, we demonstrate the performance of these models on a sample of synthetic images. Finally, conclusions and some perspectives are discussed in Section 8.

2. Literature review

The pair-wise terminology was first attached to the hidden Markov model in the works [18–20], where the main motivation was to solve the problem of correlation between noise and pixels located on the image boundary. In fact, this difficulty has been emphasized for the first time in the pair-wise Markov field (PMF) model and the pair-wise Markov tree (PMT) model. The PMF model has been successfully used for textured image segmentation [20], real synthetic-aperture radar (SAR) images [21], and hyper-spectral images [22]. The PMT model has also been used for image and signal restoration tasks [19]. Recently, the PMC model [18] has been used for applications including signal and image processing. The PMC model has specific applications in image processing, such as textured and real images segmentation [14, 23–27] and text segmentation [28, 29], where the authors used an HPS for converting the image and copula to model conditional densities of class. Images with multi-scale can be segmented using PMC [30]. Other uses include fuzzy PMC to treat spatially correlated fuzzy classes between observations [31, 32].

Several other methods based on hidden Markov models have been attempted in order to solve the question of the relationship between pixels in the chain after the transformation of a bidimensional image. Among these, an approach suggested by [15] is based on a combination of the HMC and HMF models. The first model is employed in the estimation phase and the second is exploited in the final

classification to determine the \mathbf{X} configuration. Besides, another strategy in [33, 34] is centered on the integration of spectral and contextual information into the HMC model, where the structure of the gathered data is no more Markov when modeling an image using HPS. Additionally, the study [35] entails another suggestion for an algorithm that incorporates fuzzy C-means combined with HPS. In [36], a new HMC is proposed for representing the semantic and spatial interactions between pixels.

A new segmentation method in [7] has recently been introduced that tackles the issues of spatial and temporal pixels via contextual HPS. This method adds an additional element to the observation process in the HMC model so that each pixel recovered from the HPS is linked to another pair of pixels next to it in the image but not in the chain. Similarly, in [8], a second component is introduced in the process \mathbf{Y} and a two-dimensional model is generated with estimation and segmentation techniques adapted to this scenario. Table 1 summarizes all of the methods covered here.

Table 1. A brief overview of research into noise correlation and pixel misplacement in image segmentation.

Models	Image used	Technique suggested	Year	Reference
PMF	Textured image	Pixel neighborhood	2001	[20]
PMT	Synthetic image	Partition of pixels set	2003	[19]
PMC	Synthetic and SAR images	HPS	2004	[14]
PMC	Textured and SAR images	HPS and copula	2013	[26]
PMC	Scanned document	HPS	2011	[28]
FPMC	Fuzzy and Astronomical images	HPS	2008	[31]
HMC	Synthetic image	Spectral and spatial pixel information	2005	[33]
Fuzzy segmentation	multi-spectral image	FCM algorithm and HPS	2005	[35]
HMM	Medical image	Adapted Viterbi algorithm	2020	[36]
HMC	Textured image	Contextual HPS	2021	[7]
Bi-dimensional HMC	Synthetic and mammography image	Bi-dimensional observation	2023	[8]

3. Mathematical context

In this part, we describe a brief mathematical basis for the BPMC model given in this study. We consider a couple of stochastic processes $(\mathbf{X}, \mathbf{Y}) = ((X_n)_{1 \leq n \leq N}, (Y_n)_{1 \leq n \leq N})$, where N is the number of pixels. As previously explained, in the segmentation of images with the hidden Markov model (HMM), we need to estimate the latent process $\mathbf{X} = \mathbf{x}$ via the observed process $\mathbf{Y} = \mathbf{y}$.

In the context of the PMC model with two-dimensional observed data, we consider another realization of a random variable $\bar{\mathbf{Y}} = \bar{\mathbf{y}}$ modeling information about each pixel based on its neighborhood in the image, but not in the chain, in such a way that the observed process is two-dimensional and is denoted by $(\mathbf{Y}, \bar{\mathbf{Y}})$. Then, we intend to estimate the parameters of the joint

probability $p(\mathbf{x}, \mathbf{y}, \bar{\mathbf{y}})$ by using data unsupervisedly. Based on the observation, Bayesian estimators are used to restructure the hidden process. In the present model's segmentation problem, it considers two laws: An *a priori* law $p(\mathbf{x})$ and a two-dimensional conditional density $p(\mathbf{y}, \bar{\mathbf{y}}|\mathbf{x})$.

To take on the above-mentioned segmentation problem, the *a posteriori* distribution must be identified, which contains all of the information on \mathbf{X} that is available at $(\mathbf{Y}, \bar{\mathbf{Y}})$. This distribution was supplied by: $p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}}) = \frac{p(\mathbf{x}, \mathbf{y}, \bar{\mathbf{y}})}{\sum_{\mathbf{x}} p(\mathbf{x}, \mathbf{y}, \bar{\mathbf{y}})}$, where $p(\mathbf{x}, \mathbf{y}, \bar{\mathbf{y}}) = p(\mathbf{x}) \times p(\mathbf{y}, \bar{\mathbf{y}}|\mathbf{x})$.

Let L be a bivariate loss function and $\hat{\mathbf{x}}$ be an approximation of the hidden realization, where the latter is obtained by this method:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}'} \sum_{\mathbf{x}} L(\mathbf{x}, \mathbf{x}') p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}}). \quad (3.1)$$

In Bayesian statistics, the two best-known estimators are MAP and MPM, which are associated respectively with the two loss functions $L_{MAP}(\mathbf{x}, \mathbf{x}') = L(\mathbf{x}, \mathbf{x}')$ and $L_{MPM}(\mathbf{x}, \mathbf{x}') = \sum_{n=1}^N L(x_n, x'_n)$, with L here representing the Kronecker symbol.

The two estimators are consequently and respectively defined by:

$$\begin{aligned} \hat{\mathbf{x}}_{MAP} &= \arg \min_{\mathbf{x}'} \sum_{\mathbf{x}} L_{MAP}(\mathbf{x}, \mathbf{x}') p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}}) \\ &= \arg \min_{\mathbf{x}'} \sum_{\mathbf{x} \neq \mathbf{x}'} p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}}) \\ &= \arg \min_{\mathbf{x}'} (1 - p(\mathbf{x}'|\mathbf{y}, \bar{\mathbf{y}})) \\ &= \arg \max_{\mathbf{x}'} p(\mathbf{x}'|\mathbf{y}, \bar{\mathbf{y}}), \end{aligned}$$

and

$$\begin{aligned} \hat{\mathbf{x}}_{MPM} &= \arg \min_{\mathbf{x}'} \sum_{\mathbf{x}} L_{MPM}(\mathbf{x}, \mathbf{x}') p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}}) \\ &= \arg \min_{\mathbf{x}'} \sum_{\mathbf{x}} \sum_{n=1}^N L(x_n, x'_n) p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}}) \\ &= \arg \min_{\mathbf{x}'} \sum_{n=1}^N \sum_{x_n \neq x'_n} p(x_n|\mathbf{y}, \bar{\mathbf{y}}) \\ &= \arg \min_{\mathbf{x}'} \sum_{n=1}^N (1 - p(x'_n|\mathbf{y}, \bar{\mathbf{y}})) \\ &= \arg \max_{\mathbf{x}'} \sum_{n=1}^N p(x'_n|\mathbf{y}, \bar{\mathbf{y}}). \end{aligned}$$

When comparing these two estimators, MPM maximizes the *a posteriori* marginal probability for each x_n , while MAP maximizes the *a posteriori* probability $p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}})$ in a direct way, and allows us to estimate the whole \mathbf{x} sequence. Next, we only utilize the MPM estimator since it is straightforward to use. We use iterative strategies to estimate the joint distribution of the unobserved process using the data available $(\mathbf{y}, \bar{\mathbf{y}})$. For calculating the joint distribution, we often use a conditional expectation approximation in the case of the ICE method or stochastic generators to estimate a sequence of model parameters from the dataset in the case of the SEM algorithm.

4. Models

Let's consider two stochastic processes $\mathbf{X} = (X_n)_{1 \leq n \leq N}$ and $\mathbf{Y} = (Y_n)_{1 \leq n \leq N}$, where N is the number of pixels of an image. The process \mathbf{X} is the unknown image whereas \mathbf{Y} is the observed one. Each random variable X_n has the values from a finite set of K classes $\Omega = \{\omega_1, \dots, \omega_K\}$, and each Y_n has attributed the values from the set of real numbers R . Realizations of \mathbf{X} and \mathbf{Y} are denoted by $\mathbf{x} = (x_n)_{1 \leq n \leq N}$ and $\mathbf{y} = (y_n)_{1 \leq n \leq N}$, respectively.

We suppose that the process \mathbf{X} is stationary and that the random variables $\mathbf{Y} = (Y_n)_{1 \leq n \leq N}$ are correlated and conditionally depend on \mathbf{X} . As in all Markov models, the difficulty of classification is estimating the latent process $\mathbf{X} = \mathbf{x}$ based on the observed process $\mathbf{Y} = \mathbf{y}$.

4.1. PMC model

This study uses the PMC model and proposes updating the classical pair-wise model. The classical PMC model is a doubly stochastic process, usually noted (X, Y) , where Y (one-dimensional) represents the pixels of a hidden image X . The same concept for the bidimensional PMC model is true, except for the process Y . In the classical model, a direct Hilbert Peano scan is used to transform the image from its two-dimensional version to a one-dimensional chain, followed by an inverse scan to reconstruct the final image, but in the suggested model, the second component of Y provides for this goal. The visual results revealed a significant advantage of the proposed model over the classical model, especially for isolated pixels in the to-be-segmented image.

An image of a bidimensional form can be transformed into a one-dimensional form inevitably by combining the image column by column or line by line. Moreover, we can use the classical HPS for an image $2^p \times 2^p$ [2], or the generalized HPS for any given image [37]. In a moderately different sense than above, now let's consider the pair-wise process $\mathbf{Z} = (Z_1, \dots, Z_n, \dots, Z_N)$ corresponding to \mathbf{X} the hidden process and \mathbf{Y} the observed process, where $Z_n = (X_n, Y_n)$ for $n = 1, \dots, N$. Lowercase letters are used to identify the realizations of such variables and processes. To keep things straightforward and brief, we use the letters p and f , respectively, to describe various distributions on Ω and densities on R in the following text. $p(z_n)$ is also employed for $p(Z_n = z_n)$. Referring to the previous, the pair \mathbf{Z}

is a PMC model when its joint distribution can be given as follows $p(\mathbf{z}) = p(z_1) \prod_{n=2}^N p(z_n|z_{n-1})$, with

$$\begin{aligned} p(z_1) &= p(x_1, y_1) \\ &= \sum_{x_2} \int_R p(x_1, x_2, y_1, y_2) dy_2 \end{aligned}$$

$$\begin{aligned}
&= \sum_{x_2} \int_R p(x_1, x_2) p(y_1, y_2 | x_1, x_2) dy_2 \\
&= \sum_{x_2} p(x_1, x_2) \int_R p(y_1, y_2 | x_1, x_2) dy_2 \\
&= \sum_{x_2} p(x_1, x_2) p(y_1 | x_1, x_2) \\
&= \sum_{x_2} p(x_1, x_2) f_{x_1, x_2}(y_1),
\end{aligned}$$

and

$$p(z_n | z_{n-1}) = p(x_n, y_n | x_{n-1}, y_{n-1}) = p(y_n | x_{n-1}, y_{n-1}, x_n) p(x_n | x_{n-1}, y_{n-1}) \text{ for } n = 2, \dots, N. \quad (4.1)$$

Based upon whether the process \mathbf{X} is stationary, we can also define the \mathbf{Z} distribution only as $p(z_1)p(z_2|z_1) = p(z_1, z_2)$. Thus the following proposition is inspired by the following works [4, 6, 38].

Proposition 4.1. *For \mathbf{Z} a stationary PMC model related to processes \mathbf{X} and \mathbf{Y} , and for an n from $\{2, \dots, N\}$, the following results are obtained.*

- (1) *Both the distributions of \mathbf{X} conditionally to $\mathbf{Y} = \mathbf{y}$ and of \mathbf{Y} conditionally to $\mathbf{X} = \mathbf{x}$ are Markovians.*
- (2) *If $p(x_n | x_{n-1}, y_{n-1}) = p(x_n | x_{n-1})$ and $p(y_n | x_n, x_{n-1}, y_{n-1}) = p(y_n | x_n, y_{n-1})$, then \mathbf{Z} is an HMC with dependent noise.*
- (3) *If $p(x_n | x_{n-1}, y_{n-1}) = p(x_n | x_{n-1})$ and $p(y_n | x_n, x_{n-1}, y_{n-1}) = p(y_n | x_n)$, then \mathbf{Z} is an HMC with independent noise.*
- (4) *If $p(y_{n-1} | x_{n-1}, x_n) = p(y_{n-1} | x_{n-1})$, then \mathbf{Z} is an HMC with dependent noise.*

Proof. (1) Let $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ be a PMC model. Then, n belongs to $\{2, \dots, N\}$, and we have

$$\begin{aligned}
p(x_n | x_1, \dots, x_{n-1}, \mathbf{y}) &= \frac{p(x_1, \dots, x_{n-1}, x_n, \mathbf{y})}{p(x_1, \dots, x_{n-1}, \mathbf{y})} = \frac{\sum_{x_{n+1}, \dots, x_N} p(\mathbf{x}, \mathbf{y})}{\sum_{x_n, \dots, x_N} p(\mathbf{x}, \mathbf{y})} \\
&= \frac{\sum_{x_{n+1}, \dots, x_N} p(x_1, y_1) p(x_2, y_2 | x_1, y_1) \dots p(x_N, y_N | x_{N-1}, y_{N-1})}{\sum_{x_n, \dots, x_N} p(x_1, y_1) p(x_2, y_2 | x_1, y_1) \dots p(x_N, y_N | x_{N-1}, y_{N-1})} \\
&= \frac{p(x_1, y_1) \dots p(x_n, y_n | x_{n-1}, y_{n-1}) \sum_{x_{n+1}, \dots, x_N} p(x_{n+1}, y_{n+1} | x_n, y_n) \dots p(x_N, y_N | x_{N-1}, y_{N-1})}{p(x_1, y_1) \dots p(x_{n-1}, y_{n-1} | x_{n-2}, y_{n-2}) \sum_{x_n, x_{n+1}, \dots, x_N} p(x_n, y_n | x_{n-1}, y_{n-1}) p(x_{n+1}, y_{n+1} | x_n, y_n) \dots p(x_N, y_N | x_{N-1}, y_{N-1})} \\
&= \frac{p(x_n, y_n | x_{n-1}, y_{n-1}) \sum_{x_{n+1}, \dots, x_N} p(x_{n+1}, y_{n+1} | x_n, y_n) \dots p(x_N, y_N | x_{N-1}, y_{N-1})}{\sum_{x_n, x_{n+1}, \dots, x_N} p(x_n, y_n | x_{n-1}, y_{n-1}) p(x_{n+1}, y_{n+1} | x_n, y_n) \dots p(x_N, y_N | x_{N-1}, y_{N-1})}.
\end{aligned}$$

We also have

$$\begin{aligned}
 p(x_n|x_{n-1}, \mathbf{y}) &= \frac{p(x_{n-1}, x_n, \mathbf{y})}{p(x_{n-1}, \mathbf{y})} = \frac{\sum_{\{x_1, \dots, x_N\} \setminus \{x_{n-1}, x_n\}} p(\mathbf{x}, \mathbf{y})}{\sum_{\{x_1, \dots, x_N\} \setminus \{x_{n-1}\}} p(\mathbf{x}, \mathbf{y})} \\
 &= \frac{\sum_{\{x_1, \dots, x_N\} \setminus \{x_{n-1}, x_n\}} p(x_1, y_1)p(x_2, y_2|x_1, y_1) \dots p(x_N, y_N|x_{N-1}, y_{N-1})}{\sum_{\{x_1, \dots, x_N\} \setminus \{x_{n-1}\}} p(x_1, y_1)p(x_2, y_2|x_1, y_1) \dots p(x_N, y_N|x_{N-1}, y_{N-1})} \\
 &= \frac{p(x_n, y_n|x_{n-1}, y_{n-1}) \sum_{x_1, \dots, x_{n-2}} p(x_1, y_1)p(x_2, y_2|x_1, y_1) \dots p(x_{n-1}, y_{n-1}|x_{n-2}, y_{n-2}) \sum_{x_{n+1}, \dots, x_N} p(x_{n+1}, y_{n+1}|x_n, y_n) \dots p(x_N, y_N|x_{N-1}, y_{N-1})}{\sum_{x_1, \dots, x_{n-2}} p(x_1, y_1)p(x_2, y_2|x_1, y_1) \dots p(x_{n-1}, y_{n-1}|x_{n-2}, y_{n-2}) \sum_{x_n, x_{n+1}, \dots, x_N} p(x_n, y_n|x_{n-1}, y_{n-1}) \dots p(x_N, y_N|x_{N-1}, y_{N-1})} \\
 &= \frac{p(x_n, y_n|x_{n-1}, y_{n-1}) \sum_{x_{n+1}, \dots, x_N} p(x_{n+1}, y_{n+1}|x_n, y_n) \dots p(x_N, y_N|x_{N-1}, y_{N-1})}{\sum_{x_n, x_{n+1}, \dots, x_N} p(x_n, y_n|x_{n-1}, y_{n-1}) \dots p(x_N, y_N|x_{N-1}, y_{N-1})}.
 \end{aligned}$$

This gives $p(x_n|x_1, \dots, x_{n-1}, \mathbf{y}) = p(x_n|x_{n-1}, \mathbf{y})$, therefore \mathbf{X} conditionally to $\mathbf{Y} = \mathbf{y}$ is a Markov chain. In the same way, we can show that \mathbf{Y} conditionally to $\mathbf{X} = \mathbf{x}$ is a Markov chain, and \mathbf{x} and \mathbf{y} do a symmetrical function here.

(2) \mathbf{Z} is a PMC model, then

$$\begin{aligned}
 p(z) &= p(x_1, y_1) \prod_{n=2}^N p(y_n|x_{n-1}, y_{n-1}, x_n)p(x_n|x_{n-1}, y_{n-1}) \\
 &= p(x_1)p(y_1|x_1) \prod_{n=2}^N p(y_n|x_n, y_{n-1})p(x_n|x_{n-1}),
 \end{aligned}$$

from which \mathbf{Z} is an HMC with dependent noise, identified by the initial distribution $p(z_1)$ by

$$p(z_1) = p(x_1)p(y_1|x_1),$$

and the transition matrix $p(z_n|z_{n-1})$ given by

$$p(z_n|z_{n-1}) = p(x_n|x_{n-1})p(y_n|x_n, y_{n-1}).$$

(3) \mathbf{Z} is a PMC model, then

$$\begin{aligned}
 p(z) &= p(x_1, y_1) \prod_{n=2}^N p(y_n|x_{n-1}, y_{n-1}, x_n)p(x_n|x_{n-1}, y_{n-1}) \\
 &= p(x_1)p(y_1|x_1) \prod_{n=2}^N p(y_n|x_n, y_{n-1})p(x_n|x_{n-1})
 \end{aligned}$$

$$\begin{aligned}
&= p(x_1) \prod_{n=2}^N p(x_n|x_{n-1}) p(y_1|x_1) \prod_{n=2}^N p(y_n|x_n, y_{n-1}) \\
&= p(x_1) \prod_{n=2}^N p(x_n|x_{n-1}) \prod_{n=1}^N p(y_n|x_n).
\end{aligned}$$

From which \mathbf{Z} is an HMC with independent noise.

(4) Let us show here that the process \mathbf{X} is a Markov chain. We have, for n from $\{2, \dots, N\}$,

$$\begin{aligned}
p(x_n|x_1, \dots, x_{n-1}) &= \frac{p(x_1, \dots, x_{n-1}, x_n)}{p(x_1, \dots, x_{n-1})} = \frac{\sum_{x_{n+1}, \dots, x_N} \int_{\mathcal{R}^N} p(\mathbf{x}, \mathbf{y}) dy_1 \dots dy_N}{\sum_{x_n, \dots, x_N} \int_{\mathcal{R}^N} p(\mathbf{x}, \mathbf{y}) dy_1 \dots dy_N} \\
&= \frac{\sum_{x_{n+1}, \dots, x_N} \int_{\mathcal{R}^N} p(x_1, y_1) p(x_2, y_2|x_1, y_1) \dots p(x_N, y_N|x_{N-1}, y_{N-1}) dy_1 \dots dy_N}{\sum_{x_n, \dots, x_N} \int_{\mathcal{R}^N} p(x_1, y_1) p(x_2, y_2|x_1, y_1) \dots p(x_N, y_N|x_{N-1}, y_{N-1}) dy_1 \dots dy_N} \\
&= \frac{\int_{\mathcal{R}} p(x_n, y_n|x_{n-1}, y_{n-1}) dy_n \sum_{x_{n+1}, \dots, x_N} \int_{\mathcal{R}^{N-n-1}} p(x_{n+1}, y_{n+1}|x_n, y_n) \dots p(x_N, y_N|x_{N-1}, y_{N-1}) dy_{n+1} \dots dy_{N-n-2}}{\sum_{x_n} \int_{\mathcal{R}} p(x_n, y_n|x_{n-1}, y_{n-1}) dy_n \sum_{x_{n+1}, \dots, x_N} \int_{\mathcal{R}^{N-n-2}} p(x_{n+1}, y_{n+1}|x_n, y_n) \dots p(x_N, y_N|x_{N-1}, y_{N-1}) dy_{n+1} \dots dy_{N-n-2}} \\
&= \frac{\int_{\mathcal{R}} \frac{p(x_{n-1}, x_n) p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(x_{n-1}) p(y_{n-1}|x_{n-1})} dy_n}{\sum_{x_n} \int_{\mathcal{R}} \frac{p(x_{n-1}, x_n) p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(x_{n-1}) p(y_{n-1}|x_{n-1})} dy_n} \\
&= \frac{\frac{p(x_{n-1}, x_n) p(y_{n-1}|x_{n-1}, x_n)}{p(x_{n-1}) p(y_{n-1}|x_{n-1})}}{\sum_{x_n} \frac{p(x_{n-1}, x_n) p(y_{n-1}|x_{n-1}, x_n)}{p(x_{n-1}) p(y_{n-1}|x_{n-1})}}.
\end{aligned}$$

Taking advantage that $p(y_{n-1}|x_{n-1}, x_n) = p(y_{n-1}|x_{n-1})$, we thus have $p(x_n|x_1, \dots, x_{n-1}) = p(x_n|x_{n-1})$, which completes the proof. \square

Remark 4.1. *Properties 2 and 3 in Proposition 4.1 demonstrate the strict generality of the PMC model compared with the HMC model.*

4.2. BPMC model

In this model, we consider another observed process noted $\bar{\mathbf{Y}}$ given by $\bar{\mathbf{Y}} = (Y_1, \bar{Y}_1, \dots, Y_n, \bar{Y}_n, \dots, Y_N, \bar{Y}_N)$. Consequently, for each $n = 1, \dots, N$, the random variable X_n is associated with the pair (Y_n, \bar{Y}_n) , where \bar{Y}_n is related only to the observation Y_n as defined below. The problem remains unchanged; determining \mathbf{X} based on \mathbf{Y} . To establish this new model, we may then state the following conditions (Figure1).

- (1) The process $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ is a PMC model.
- (2) The random variables $(\bar{Y}_n)_{1 \leq n \leq N}$ are conditionally independent to \mathbf{Z} .
- (3) Each \bar{Y}_n is conditionally distributed to \mathbf{Z} and its distribution is conditionally to Z_n .

The model to be studied $(\mathbf{Z}, \bar{\mathbf{Y}})$, called the BPMC model, could be interpreted as a semi-hidden Markov chain model [4]. Its distribution can be stated as follows.

$$\begin{aligned}
 p(\mathbf{z}, \bar{\mathbf{y}}) &= p(\mathbf{z}, \bar{y}_1, \dots, \bar{y}_N) \\
 &= p(\mathbf{z})p(\bar{y}_1, \dots, \bar{y}_N|\mathbf{z}) \\
 &= p(z_1) \prod_{n=1}^{N-1} p(z_{n+1}|z_n) \prod_{n=1}^N p(\bar{y}_n|z_n) \\
 &= p(x_1, y_1, \bar{y}_1) \prod_{n=1}^{N-1} p(x_{n+1}, y_{n+1}, \bar{y}_{n+1}|x_n, y_n).
 \end{aligned}$$

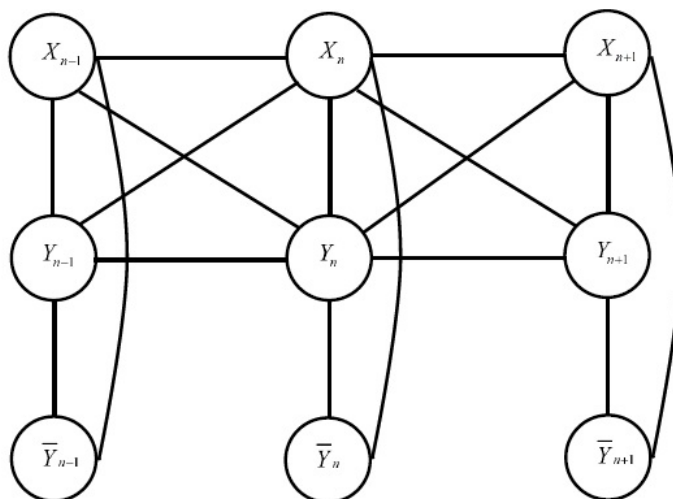


Figure 1. Probability graph of BPMC model for three successive variables.

Let's emphasize that the BPMC model is particularly interesting because, given the pair $(\mathbf{Y}, \bar{\mathbf{Y}}) = (\mathbf{y}, \bar{\mathbf{y}})$, the distribution of X is a Markov. Using the same logic as the demonstration of the first point in *Proposition 4.1*, we get the following proposition.

Proposition 4.2. For $(\mathbf{Z}, \bar{\mathbf{Y}})$, a stationary BPMC model related to processes X and Y , the following result is obtained.

The distributions of X conditional on $(\mathbf{Y}, \bar{\mathbf{Y}}) = (\mathbf{y}, \bar{\mathbf{y}})$ is a Markov chain.

5. Restoration problem of simulated models

The reason behind the achievements of the HMC model can be attributed to the fact that the process \mathbf{X} conditionally on $\mathbf{Y} = \mathbf{y}$ is a Markov chain and its transitions can be computed. This observation holds true in the case of the PMC model. However, when it comes to the BPMC model, this conditional distribution exhibits a complex structure. That is precisely why we employ the distribution $p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}})$.

Both models, PMC and BPMC, are assumed to be stationary, such that their distributions are given by $p(x_1, y_1, x_2, y_2)$ for PMC and $p(x_1, y_1, \bar{y}_1, x_2, y_2, \bar{y}_2)$ for BPMC.

In this section, we contrast the two models using the MPM restoration algorithm in the Gaussian context. Both models utilize Gaussian densities, which implies that $p(y_{n-1}, y_n | x_{n-1}, x_n)$ and $p(\bar{y}_{n-1} | x_{n-1})$, for each n , are Gaussian (in the BPMC model case, we choose $p(\bar{y}_{n-1} | x_{n-1})$ to be equal to $p(\bar{y}_n | x_{n-1}, x_n)$). In this scenario, we have K^2 probabilities $p(x_{n-1} = \omega_i, x_n = \omega_j)$, K^2 mean vectors $\mu_{ij} = \begin{pmatrix} \mu_{ij}^1 \\ \mu_{ij}^2 \end{pmatrix}$, and K^2 variance-covariance matrix $\Sigma_{ij} = \begin{pmatrix} \sigma_{ij}^1 & \sigma_{ij}^{12} \\ \sigma_{ij}^{12} & \sigma_{ij}^2 \end{pmatrix}$ of the bidimensional densities $f_{ij}(y_{n-1}, y_n)$ to represent the PMC model. Additionally, the BPMC model is represented by the parameters μ_{ij}^1 , μ_{ij}^2 , σ_{ij}^1 , and σ_{ij}^2 of the K^2 mono-dimensional densities $f_{ij}(\bar{y}_{n-1})$. In the following subsection, we will simulate the two processes \mathbf{X} and \mathbf{Y} assuming that all the aforementioned parameters are initially known. Furthermore, we will reconstruct the process \mathbf{X} only based on the process \mathbf{Y} using the PMC model, while considering both \mathbf{Y} and $\bar{\mathbf{Y}}$ simultaneously using the BPMC model.

5.1. Simulation of processes in models

Here, we evoke the previously defined distributions of the PMC model and BPMC model. The distribution of the PMC model is given by $p(x_1, y_1) = p(x_1)p(y_1|x_1)$ and $p(x_n, y_n | x_{n-1}, y_{n-1}) = p(y_n | x_{n-1}, y_{n-1}, x_n)p(x_n | x_{n-1}, y_{n-1})$, while the distribution of the BPMC model is represented by $p(x_1, y_1, \bar{y}_1) = p(x_1)p(y_1|x_1)p(\bar{y}_1|x_1, y_1)$, $p(x_n, y_n | x_{n-1}, y_{n-1})$, and $p(\bar{y}_n | x_n, y_n)$.

We can simulate the processes \mathbf{X} , \mathbf{Y} and $\bar{\mathbf{Y}}$ of models concurrently and interchangeably using the following approach.

- For the PMC model, x_1, y_1, x_n , and y_n , for each $n = 2, \dots, N$, are simulated with the following probabilities:

$$p(x_1) = \sum_{x_2} p(x_1, x_2). \quad (5.1)$$

$$p(y_1 | x_1) = \frac{p(x_1, y_1)}{p(x_1)} = \frac{\sum_{x_2} p(x_1, x_2) f_{x_1, x_2}(y_1)}{\sum_{x_2} p(x_1, x_2)}. \quad (5.2)$$

$$p(x_n | x_{n-1}, y_{n-1}) = \frac{p(x_n, x_{n-1}, y_{n-1})}{p(x_{n-1}, y_{n-1})} = \frac{p(x_{n-1}, x_n) f_{x_{n-1}, x_n}(y_{n-1})}{\sum_{x_n} p(x_{n-1}, x_n) f_{x_{n-1}, x_n}(y_{n-1})}. \quad (5.3)$$

$$p(y_n | x_n, x_{n-1}, y_{n-1}) = \frac{f_{x_{n-1}, x_n}(y_{n-1}, y_n)}{f_{x_{n-1}, x_n}(y_{n-1})}. \quad (5.4)$$

- For the BPMC model, x_1, y_1, x_n , and y_n are simulated in the same way as above. Additionally, we have simulated \bar{y}_1 and \bar{y}_n , for each $n = 2, \dots, N$, with these probabilities:

$$p(\bar{y}_1 | x_1, y_1) = f_{x_1, x_2}(\bar{y}_1). \quad (5.5)$$

$$p(\bar{y}_n | x_n, y_n) = f_{x_{n-1}, x_n}(\bar{y}_n). \quad (5.6)$$

Remark 5.1. *It can be seen here that (to dig deeper, see [14, 39])*

- $f_{x_1, x_2}(y_1)$ is a mono-dimensional Gaussian density $\mathcal{N}\left(\mu_{ij}^1, \sqrt{\sigma_{ij}^1}\right)$.
- $f_{x_{n-1}, x_n}(y_{n-1})$ is a mono-dimensional Gaussian density $\mathcal{N}\left(\mu_{ij}^1, \sqrt{\sigma_{ij}^1}\right)$.
- it can be shown that y_n is generated according to a Gaussian drawing, where $p(y_n|x_n, x_{n-1}, y_{n-1})$ is a mono-dimensional Gaussian density of mean $\mu_{ij}^2 + \frac{\sigma_{ij}^{12}}{\sigma_{ij}^1}(y_{n-1} - \mu_{ij}^1)$ and standard deviation $\sqrt{\frac{\sigma_{ij}^1 \sigma_{ij}^2 - \sigma_{ij}^{12}}{\sigma_{ij}^1}}$.
- $f_{x_1, x_2}(\bar{y}_1)$ is a mono-dimensional Gaussian density $\mathcal{N}\left(\mu_{ij}^1, \sqrt{\sigma_{ij}^1}\right)$.
- $f_{x_{n-1}, x_n}(\bar{y}_n)$ is a mono-dimensional Gaussian density $\mathcal{N}\left(\mu_{ij}^2, \sqrt{\sigma_{ij}^2}\right)$.

5.2. Restoration of hidden process in models

In this work, we look at the MPM approach for reconstructing the hidden sequence $\mathbf{X} = \mathbf{x}$.

The MPM estimator for the PMC model consists of maximizing *a posteriori* marginal probability for each x_n as follows:

$$\hat{x}_n = \arg \max_{\omega_i} \chi_n(\omega_i), \quad (5.7)$$

where $\chi_n(\omega_i) = p(x_n|\mathbf{y})$. Formally, this probability is calculated using the “conditional forward probabilities” $\alpha_n(\omega_i) = p(x_n = \omega_i|y_1, \dots, y_n)$ and the “conditional backward probabilities” $\beta_n(\omega_i) = \frac{p(y_{n+1}, \dots, y_N|x_n = \omega_i, y_n)}{p(y_{n+1}, \dots, y_N|y_1, \dots, y_n)}$ to avoid the numerical problems encountered by the same probabilities in the classical case. These conditional probabilities can be given using the following recursion:

- Initiation phase: For $i = 1, \dots, K$.

(1) Forward

$$\alpha_1(\omega_i) = \frac{p(x_1 = \omega_i, y_1)}{\sum_{j=1}^K p(x_1 = \omega_j, y_1)}. \quad (5.8)$$

(2) Backward

$$\beta_N(\omega_i) = 1. \quad (5.9)$$

- Induction phase: For $i = 1, \dots, K$ and $n = 1, \dots, N - 1$.

(1) Forward

$$\alpha_{n+1}(\omega_i) = \frac{\sum_{j=1}^K \alpha_n(\omega_j) p(x_{n+1} = \omega_i, y_{n+1}|x_n = \omega_j, y_n)}{\sum_{j=1}^K \sum_{k=1}^K \alpha_n(\omega_j) p(x_{n+1} = \omega_k, y_{n+1}|x_n = \omega_j, y_n)}. \quad (5.10)$$

(2) Backward

$$\beta_n(\omega_i) = \frac{\sum_{j=1}^K \beta_{n+1}(\omega_j) p(x_{n+1} = \omega_j, y_{n+1} | x_n = \omega_i, y_n)}{\sum_{j=1}^K \sum_{k=1}^K \alpha_n(\omega_k) p(x_{n+1} = \omega_j, y_{n+1} | x_n = \omega_k, y_n)}, \quad (5.11)$$

$$\text{with } p(x_{n+1}, y_{n+1} | x_n, y_n) = \frac{p(x_n, y_n, x_{n+1}, y_{n+1})}{p(x_n, y_n)} = \frac{p(x_n, x_{n+1}) f_{x_n, x_{n+1}}(y_n, y_{n+1})}{\sum_{x_{n+1}} p(x_n, x_{n+1}) f_{x_n, x_{n+1}}(y_n)}.$$

It is easy to demonstrate that

$$\chi_n(\omega_i) = \alpha_n(\omega_i) \beta_n(\omega_i). \quad (5.12)$$

Besides this, the MPM estimator under the BPMC model can be calculated as follows.

$$\hat{x}_n = \arg \max_{\omega_i} \chi_n^*(\omega_i), \quad (5.13)$$

where $\chi_n^*(\omega_i) = p(x_n | \mathbf{y}, \bar{\mathbf{y}})$. This probability can be achieved using the following conditional forward and backward probabilities $\alpha_n^*(\omega_i) = p(x_n = \omega_i | y_1, \bar{y}_1, \dots, y_n, \bar{y}_n)$ and $\beta_n^*(\omega_i) = \frac{p(y_{n+1}, \bar{y}_{n+1}, \dots, y_N, \bar{y}_N | x_n = \omega_i, y_n, \bar{y}_n)}{p(y_{n+1}, \bar{y}_{n+1}, \dots, y_N, \bar{y}_N | y_1, \bar{y}_1, \dots, y_n, \bar{y}_n)}$. These conditional probabilities can be given using the following recursion:

- Initiation phase: For $i = 1, \dots, K$.

(1) Forward

$$\alpha_1^*(\omega_i) = \frac{p(x_1 = \omega_i, y_1, \bar{y}_1)}{\sum_{j=1}^K p(x_1 = \omega_j, y_1, \bar{y}_1)}. \quad (5.14)$$

(2) Backward

$$\beta_N^*(\omega_i) = 1. \quad (5.15)$$

- Induction phase: For $i = 1, \dots, K$ and $n = 1, \dots, N - 1$.

(1) Forward

$$\alpha_{n+1}^*(\omega_i) = \frac{\sum_{j=1}^K \alpha_n^*(\omega_j) p(x_{n+1} = \omega_i, y_{n+1}, \bar{y}_{n+1} | x_n = \omega_j, y_n, \bar{y}_n)}{\sum_{j=1}^K \sum_{k=1}^K \alpha_n^*(\omega_k) p(x_{n+1} = \omega_k, y_{n+1}, \bar{y}_{n+1} | x_n = \omega_j, y_n, \bar{y}_n)}. \quad (5.16)$$

(2) Backward

$$\beta_n^*(\omega_i) = \frac{\sum_{j=1}^K \beta_{n+1}^*(\omega_j) p(x_{n+1} = \omega_j, y_{n+1}, \bar{y}_{n+1} | x_n = \omega_i, y_n, \bar{y}_n)}{\sum_{j=1}^K \sum_{k=1}^K \alpha_n^*(\omega_k) p(x_{n+1} = \omega_j, y_{n+1}, \bar{y}_{n+1} | x_n = \omega_k, y_n, \bar{y}_n)}, \quad (5.17)$$

$$\text{with } p(x_{n+1}, y_{n+1}, \bar{y}_{n+1} | x_n, y_n, \bar{y}_n) = \frac{p(x_n, x_{n+1}) f_{x_n, x_{n+1}}(y_n, y_{n+1}) f_{x_n, x_{n+1}}(\bar{y}_{n+1})}{\sum_{x_{n+1}} p(x_n, x_{n+1}) f_{x_n, x_{n+1}}(y_n)}$$

Considering the foregoing considerations, the distribution of X_n conditionally on $(\mathbf{Y} = \mathbf{y}, \bar{\mathbf{Y}} = \bar{\mathbf{y}})$ is given by

$$\begin{aligned} p(x_n | \mathbf{y}, \bar{\mathbf{y}}) &= \frac{p(x_n, \mathbf{y}, \bar{\mathbf{y}})}{p(\mathbf{y}, \bar{\mathbf{y}})} \\ &= \frac{p(x_n, y_1, \bar{y}_1, \dots, y_n, \bar{y}_n) p(y_{n+1}, \bar{y}_{n+1}, \dots, y_N, \bar{y}_N | x_n, y_1, \bar{y}_1, \dots, y_n, \bar{y}_n)}{p(y_{n+1}, \bar{y}_{n+1}, \dots, y_N, \bar{y}_N | y_1, \bar{y}_1, \dots, y_n, \bar{y}_n) p(y_1, \bar{y}_1, \dots, y_n, \bar{y}_n)} \\ &= \frac{p(x_n | y_1, \bar{y}_1, \dots, y_n, \bar{y}_n) p(y_{n+1}, \bar{y}_{n+1}, \dots, y_N, \bar{y}_N | x_n, y_1, \bar{y}_1, \dots, y_n, \bar{y}_n)}{p(y_{n+1}, \bar{y}_{n+1}, \dots, y_N, \bar{y}_N | y_1, \bar{y}_1, \dots, y_n, \bar{y}_n)}. \end{aligned}$$

This gives

$$\chi_n^*(\omega_i) = \alpha_n^*(\omega_i) \beta_n^*(\omega_i). \quad (5.18)$$

5.3. Restoration problem results

In this section, we attempt to show the importance of the BPMC model compared to the PMC model. For that, we treat an example by considering $N = 100000$ and $K = 2$ (i.e., $\Omega = \{\omega_1, \omega_2\}$). We shall investigate the models' performance by taking two factors into account, two different chain homogeneities, noted H_1 and H_2 , and different noise parameters:

- Factor chain:

- Case H_1 : $p(\omega_1, \omega_1) = p(\omega_2, \omega_2) = 0.48$ and $p(\omega_1, \omega_2) = p(\omega_2, \omega_1) = 0.02$.
- Case H_2 : $p(\omega_1, \omega_1) = p(\omega_2, \omega_2) = 0.25$ and $p(\omega_1, \omega_2) = p(\omega_2, \omega_1) = 0.25$.

- Factor noise:

- Parameters of density f_{ω_1, ω_1} : $\mu_{\omega_1, \omega_1}^1 = -3$, $\mu_{\omega_1, \omega_1}^2 = -3$, $\sigma_{\omega_1, \omega_1}^1 = 14$, $\sigma_{\omega_1, \omega_1}^2 = 14$, and $\sigma_{\omega_1, \omega_1}^{12} = 0.1$.
- Parameters of density f_{ω_1, ω_2} : $\mu_{\omega_1, \omega_2}^1 = 5$, $\mu_{\omega_1, \omega_2}^2 = -5$, $\sigma_{\omega_1, \omega_2}^1 = 11$, $\sigma_{\omega_1, \omega_2}^2 = 9$, and $\sigma_{\omega_1, \omega_2}^{12} = 0.9$.
- Parameters of density f_{ω_2, ω_1} : $\mu_{\omega_2, \omega_1}^1 = -3$, $\mu_{\omega_2, \omega_1}^2 = 3$, $\sigma_{\omega_2, \omega_1}^1 = 9$, $\sigma_{\omega_2, \omega_1}^2 = 11$, and $\sigma_{\omega_2, \omega_1}^{12} = 0.1$.
- Parameters of density f_{ω_2, ω_2} : $\mu_{\omega_2, \omega_2}^1 = 5$, $\mu_{\omega_2, \omega_2}^2 = 5$, $\sigma_{\omega_2, \omega_2}^1 = 18$, $\sigma_{\omega_2, \omega_2}^2 = 18$, and $\sigma_{\omega_2, \omega_2}^{12} = 0.1$.

The following parameter is used to determine noise correlation influence on the error rate

$$\rho = \sum_{i,j=1}^K \left(\frac{\rho_{\omega_i, \omega_j}}{\rho_{\omega'_i, \omega'_j}} \right)^2 \text{ with } \rho_{\omega_i, \omega_j} > \rho_{\omega'_i, \omega'_j}, \text{ where } \rho_{\omega_i, \omega_j} = \frac{\sigma_{\omega_i, \omega_j}^{12}}{\sqrt{\sigma_{\omega_i, \omega_j}^1} \sqrt{\sigma_{\omega_i, \omega_j}^2}}, \text{ for } i, j = 1, \dots, K.$$

We consider four sets of ρ and the covariance values associated with these four simulations are shown in Table 2.

Table 2. Covariance values associated to parameter ρ used in the four experiments.

Experiments	$\sigma_{\omega_1, \omega_1}^{12}$	$\sigma_{\omega_1, \omega_2}^{12}$	$\sigma_{\omega_2, \omega_1}^{12}$	$\sigma_{\omega_2, \omega_2}^{12}$
Experiment 1 $\rho = 190$	0.1	0.9	0.1	0.1
Experiment 2 $\rho = 45$	0.2	0.9	0.3	0.2
Experiment 3 $\rho = 17$	0.3	0.1	0.3	0.2
Experiment 4 $\rho = 5$	0.3	0.4	0.4	0.9

We specify the misclassification rates computed by the MPM method in Table 3 after calculating the means from 300 independent experiments.

Table 3. Incorrectly classified (%) pixels given by PMC and BPMC models-based Bayesian MPM for two cases of chain.

Experiments	Case H_1		Case H_2	
	PMC	BPMC	PMC	BPMC
Experiment 1	13.66 %	9.712 %	32.08 %	35.86 %
Experiment 2	15.74 %	10.42 %	33.31 %	34.57 %
Experiment 3	16.87 %	12.65 %	30.18 %	32.73 %
Experiment 4	44.29 %	37.12 %	43.62 %	48.03 %

According to these results, for chain in Case H_1 , the misclassification rates are more important when the coefficient ρ is small (Experiment 4), and becomes progressively smaller when the coefficient ρ decreases. It can be seen from all experiments that the BPMC model-based MPM restorations often work better than the PMC model-based ones. For the case H_2 of chain, the misclassification rates are consistently more important for both models. In this situation, the PMC model is preferable to the BPMC model. Other simulations have been done, showing that chain homogeneity is an important element and it can be used during the restoration phase when the data is complete.

6. PMC and BPMC parameters estimation

The parameter estimation problem from incomplete data for the PMC is considerably more complex than for the HMM. Considering the fact that we are dealing with unsupervised segmentation, all of the parameters noted $\theta = (\theta_1, \dots, \theta_l)$ must be estimated. We are interested, in this section, in estimating K^2 probabilities $p(\omega_i, \omega_j)$, $2K^2$ parameters of μ_{ij} and $3K^2$ parameters of Σ_{ij} . Considering that the log-likelihood of the models under examination cannot be maximized analytically, the methods follow the simulation according to posterior distributions used for parameter estimation. We use here ICE and SEM algorithms.

6.1. Estimation using ICE algorithm

6.1.1. ICE algorithm according to PMC model

The ICE algorithm under the PMC model is an iterative procedure, working as follows:

- initialize θ_i^0 ;

- θ_i^{q+1} is calculated using $\theta_i^{q+1} = E_{\theta^q}[\hat{\theta}_i(\mathbf{X}, \mathbf{Y})|\mathbf{Y} = \mathbf{y}]$ if this expectation is computable, or $\theta_i^{q+1} = \frac{1}{T} \sum_{t=1}^T \hat{\theta}_i(\mathbf{x}^t, \mathbf{y})$ if the expectation above is not computable, where $\mathbf{x}^1, \dots, \mathbf{x}^T$ are simulated according to $p(\mathbf{x}|\mathbf{y}, \theta^q)$.

Returning to our problem, we can take the following estimator for $p(\omega_i, \omega_j)$:

$$\hat{p}(\omega_i, \omega_j)(\mathbf{z}) = \frac{1}{N-1} \sum_{n=1}^{N-1} \mathbb{1}_{\{x_n=\omega_i, x_{n+1}=\omega_j\}}. \quad (6.1)$$

The conditional expectation of this estimator, at iteration $(q+1)$, can be calculated and gives

$$p^{(q+1)}(\omega_i, \omega_j) = E[\hat{p}(\omega_i, \omega_j)(\mathbf{z})|\mathbf{Y} = \mathbf{y}] = \frac{1}{N-1} \sum_{n=1}^{N-1} \psi_n^{(q)}(\omega_i, \omega_j), \quad (6.2)$$

where $\psi_n^{(q)}(\omega_i, \omega_j) = p(x_n = \omega_i, x_{n+1} = \omega_j|\mathbf{y})$ is a joint *a posteriori* probability of two consecutive classes, and can be calculated using

$$\psi_n^{(q)}(\omega_i, \omega_j) = \frac{\alpha_n(\omega_i)p(x_{n+1} = \omega_j, y_{n+1}|x_n = \omega_i, y_n)\beta_{n+1}(\omega_j)}{\sum_{l=1}^K \sum_{k=1}^K \alpha_n(\omega_l)p(x_{n+1} = \omega_k, y_{n+1}|x_n = \omega_l, y_n)\beta_{n+1}(\omega_k)}. \quad (6.3)$$

For the parameters μ_{ij} and Σ_{ij} , we can choose the following estimators:

$$\hat{\mu}_{ij}(\mathbf{z}) = \frac{\sum_{n=1}^{N-1} \binom{y_n}{y_{n+1}} \mathbb{1}_{\{x_n=\omega_i, x_{n+1}=\omega_j\}}}{\sum_{n=1}^{N-1} \mathbb{1}_{\{x_n=\omega_i, x_{n+1}=\omega_j\}}}. \quad (6.4)$$

$$\hat{\Sigma}_{ij}(\mathbf{z}) = \frac{\sum_{n=1}^{N-1} \left(\binom{y_n}{y_{n+1}} - \hat{\mu}_{ij}(\mathbf{z}) \right) \left(\binom{y_n}{y_{n+1}} - \hat{\mu}_{ij}(\mathbf{z}) \right)^t \mathbb{1}_{\{x_n=\omega_i, x_{n+1}=\omega_j\}}}{\sum_{n=1}^{N-1} \mathbb{1}_{\{x_n=\omega_i, x_{n+1}=\omega_j\}}}. \quad (6.5)$$

However, we cannot calculate the conditional expectation of these estimators. In this situation, we employ a stochastic strategy based on the simulation of \mathbf{X} according to its *a posteriori* distribution, which is a nonstationary Markov chain, as we demonstrated previously:

$$p(\mathbf{x}|\mathbf{y}) = p(x_1|\mathbf{y}) \prod_{n=1}^{N-1} p(x_{n+1}|x_n, \mathbf{y}), \quad (6.6)$$

with an initial distribution $\chi_1(\omega_i)$ and a transition matrix

$$p(x_{n+1} = \omega_j | x_n = \omega_i, \mathbf{y}) = \frac{\psi_n(\omega_i, \omega_j)}{\chi_n(\omega_i)}. \quad (6.7)$$

The process for applying the ICE algorithm to the PMC model is then presented in the following manner.

- Iteration ($q = 0$), for $i, j = 1, \dots, K$.
 - (1) Initialization of the algorithm with $p^{(q)}(\omega_i, \omega_j), \mu_{ij}^{(q)}, \Sigma_{ij}^{(q)}$.
 - (2) Calculation of probabilities $\alpha_n^{(q)}(\omega_i)$ according to (5.8) and (5.10).
 - (3) Calculation of probabilities $\beta_n^{(q)}(\omega_i)$ according to (5.9) and (5.11).
 - (4) Deduction of probabilities $\chi_n^{(q)}(\omega_i)$ according to (5.12).
 - (5) Deduction of probabilities $\psi_n^{(q)}(\omega_i, \omega_j)$ according to (6.3).
 - (6) Simulation of $\mathbf{x}^{(q),1}, \dots, \mathbf{x}^{(q),T}$ according to (6.6).
- Iteration ($q + 1$), for $i, j = 1, \dots, K$.
 - (1) Calculation of $p^{(q+1)}(\omega_i, \omega_j)$ according to (6.2).
 - (2) Calculation of $\mu_{ij}^{(q+1)}$ and $\Sigma_{ij}^{(q+1)}$ by

$$\mu_{ij}^{(q+1)} = \frac{1}{T} \sum_{t=1}^T \mu_{ij}^t, \quad \Sigma_{ij}^{(q+1)} = \frac{1}{T} \sum_{t=1}^T \Sigma_{ij}^t, \quad (6.8)$$

where μ_{ij}^t and Σ_{ij}^t are calculated according to (6.4) and (6.5).

- Iterations are stopped if the algorithm converges; otherwise, the preceding step is repeated.

6.1.2. ICE algorithm according to BPMC model

The ICE algorithm follows this method:

- Initialize θ_i^0 ;
- θ_i^{q+1} is calculated using $\theta_i^{q+1} = E_{\theta_i^q}[\hat{\theta}_i(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{Y}}) | \mathbf{Y} = \mathbf{y}, \bar{\mathbf{Y}} = \bar{\mathbf{y}}]$ if this expectation is computable, or $\theta_i^{q+1} = \frac{1}{T} \sum_{t=1}^T \hat{\theta}_i(\mathbf{x}^t, \mathbf{y}, \bar{\mathbf{y}})$ if the expectation above is not computable, where $\mathbf{x}^1, \dots, \mathbf{x}^T$ are simulated according to $p(\mathbf{x} | \mathbf{y}, \bar{\mathbf{y}}, \theta^q)$.

Here, we take the same estimator for $p(\omega_i, \omega_j)$, as given in (6.1). The conditional expectation of this estimator, at iteration ($q + 1$), can be calculated and gives

$$p^{(q+1)}(\omega_i, \omega_j) = E[\hat{p}(\omega_i, \omega_j)(\mathbf{z}) | \mathbf{Y} = \mathbf{y}, \bar{\mathbf{Y}} = \bar{\mathbf{y}}] = \frac{1}{N-1} \sum_{n=1}^{N-1} \psi_n^{*(q)}(\omega_i, \omega_j), \quad (6.9)$$

where $\psi_n^{*(q)}(\omega_i, \omega_j) = p(X_n = \omega_i, x_{n+1} = \omega_j | \mathbf{y}, \bar{\mathbf{y}})$ is a joint *a posteriori* probability of two consecutive classes, and it can be shown that

$$\psi_n^{*(q)}(\omega_i, \omega_j) = \frac{\alpha_n^*(\omega_i) p(x_{n+1} = \omega_j, y_{n+1}, \bar{y}_{n+1} | x_n = \omega_i, y_n, \bar{y}_n) \beta_{n+1}^*(\omega_j)}{\sum_{l=1}^K \sum_{k=1}^K \alpha_n^*(\omega_l) p(x_{n+1} = \omega_k, y_{n+1}, \bar{y}_{n+1} | x_n = \omega_l, y_n, \bar{y}_n) \beta_{n+1}^*(\omega_k)}. \quad (6.10)$$

For the parameters μ_{ij} and Σ_{ij} , we choose the same estimators as in (6.4) and (6.5).

Here, we simulate the process \mathbf{X} according to its *a posteriori* distribution under BPMC model $p(\mathbf{x} | \mathbf{y}, \bar{\mathbf{y}})$, where

$$p(\mathbf{x} | \mathbf{y}, \bar{\mathbf{y}}) = p(x_1 | \mathbf{y}, \bar{\mathbf{y}}) \prod_{n=1}^{N-1} p(x_{n+1} | x_n, \mathbf{y}, \bar{\mathbf{y}}), \quad (6.11)$$

with an initial distribution $\chi_1^*(\omega_i)$ and a transition matrix

$$p(x_{n+1} = \omega_j | x_n = \omega_i, \mathbf{y}, \bar{\mathbf{y}}) = \frac{\psi_n^*(\omega_i, \omega_j)}{\chi_n^*(\omega_i)}. \quad (6.12)$$

The ICE algorithm procedure for the BPMC model is then explained as follows.

- Iteration ($q = 0$), for $i, j = 1, \dots, K$.
 - (1) Initialization of the algorithm with $p^{(q)}(\omega_i, \omega_j), \mu_{ij}^{(q)}, \Sigma_{ij}^{(q)}$.
 - (2) Calculation of probabilities $\alpha_n^{*(q)}(\omega_i)$ according to (5.14) and (5.16).
 - (3) Calculation of probabilities $\beta_n^{*(q)}(\omega_i)$ according to (5.15) and (5.17).
 - (4) Deduction of probabilities $\chi_n^{*(q)}(\omega_i)$ according to (5.18).
 - (5) Deduction of probabilities $\psi_n^{*(q)}(\omega_i, \omega_j)$ according to (6.10).
 - (6) Simulation of $\mathbf{x}^{(q),1}, \dots, \mathbf{x}^{(q),T}$ according to (6.11).
- Iteration ($q + 1$), for $i, j = 1, \dots, K$.
 - (1) Calculation of $p^{(q+1)}(\omega_i, \omega_j)$ according to (6.9).
 - (2) Calculation of $\mu_{ij}^{(q+1)}$ and $\Sigma_{ij}^{(q+1)}$ according to (6.8).
- Iterations are stopped if the algorithm converges; otherwise, the preceding step is repeated.

6.2. Estimation using SEM algorithm

The SEM algorithm is an iterative procedure that uses stochastic drawings to estimate a sequence of model parameters from observations and realizations of \mathbf{X} . In each iteration, we simulate the process \mathbf{X} according to its *a posteriori* distribution based on the parameters obtained in the current iteration. Then, the SEM algorithm for the models in this work proceeds as follows:

- We initialize the algorithm with θ_i^0 .
- At each iteration (q), we simulate just one realization \mathbf{x} of \mathbf{X} according to its *a posteriori* distribution considering the current parameters. The parameters $\theta_i^{(q+1)}$ are then calculated.

6.2.1. SEM algorithm according to PMC model

The SEM algorithm under the PMC model runs as follows:

- Iteration ($q = 0$), for $i, j = 1, \dots, K$.
 - (1) Initialization of the algorithm with $p^{(q)}(\omega_i, \omega_j), \mu_{ij}^{(q)}, \Sigma_{ij}^{(q)}$.
 - (2) Calculation of probabilities $\alpha_n^{(q)}(\omega_i)$ according to (5.8) and (5.10).
 - (3) Calculation of probabilities $\beta_n^{(q)}(\omega_i)$ according to (5.9) and (5.11).
 - (4) Deduction of probabilities $\chi_n^{(q)}(\omega_i)$ according to (5.12).
 - (5) Deduction of probabilities $\psi_n^{(q)}(\omega_i, \omega_j)$ according to (6.3).
 - (6) Simulation of $\mathbf{x}^{(q)}$ according to (6.6).
- Iteration ($q + 1$), for $i, j = 1, \dots, K$.
 - (1) Calculation of $p^{(q+1)}(\omega_i, \omega_j)$ by

$$p^{(q+1)}(\omega_i, \omega_j) = \frac{1}{N-1} \sum_{n=1}^{N-1} \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}. \quad (6.13)$$

- (2) Calculation of $\mu_{ij}^{(q+1)}$ and $\Sigma_{ij}^{(q+1)}$ by

$$\mu_{ij}^{(q+1)} = \frac{\sum_{n=1}^{N-1} \begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}}{\sum_{n=1}^{N-1} \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}}. \quad (6.14)$$

$$\Sigma_{ij}^{(q+1)} = \frac{\sum_{n=1}^{N-1} \left(\begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} - \mu_{ij}^{(q)} \right) \left(\begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} - \mu_{ij}^{(q)} \right)^t \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}}{\sum_{n=1}^{N-1} \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}}. \quad (6.15)$$

- Iterations are stopped if the algorithm converges; otherwise, the preceding step is repeated.

6.2.2. SEM algorithm according to BPMC model

The SEM algorithm under the BPMC model runs as follows:

- Iteration ($q = 0$), for $i, j = 1, \dots, K$.
- Initialization of the algorithm with $p^{(q)}(\omega_i, \omega_j), \mu_{ij}^{(q)}, \Sigma_{ij}^{(q)}$.
 - (1) Calculation of probabilities $\alpha_n^{*(q)}(\omega_i)$ according to (5.14) and (5.16).
 - (2) Calculation of probabilities $\beta_n^{*(q)}(\omega_i)$ according to (5.15) and (5.17).

- (3) Deduction of probabilities $\chi_n^{*(q)}(\omega_i)$ according to (5.18).
 (4) Deduction of probabilities $\psi_n^{*(q)}(\omega_i, \omega_j)$ according to (6.10).
 (5) Simulation of $\mathbf{x}^{(q)}$ according to (6.11).

- Iteration $(q + 1)$, for $i, j = 1, \dots, K$.

- (1) Calculation of $p^{(q+1)}(\omega_i, \omega_j)$ by

$$p^{(q+1)}(\omega_i, \omega_j) = \frac{1}{N-1} \sum_{n=1}^{N-1} \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}. \quad (6.16)$$

- (2) Calculation of $\mu_{ij}^{(q+1)}$ and $\Sigma_{ij}^{(q+1)}$ by

$$\mu_{ij}^{(q+1)} = \frac{\sum_{n=1}^{N-1} \begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}}{\sum_{n=1}^{N-1} \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}}. \quad (6.17)$$

$$\Sigma_{ij}^{(q+1)} = \frac{\sum_{n=1}^{N-1} \left(\begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} - \mu_{ij}^{(q)} \right) \left(\begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} - \mu_{ij}^{(q)} \right)^t \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}}{\sum_{n=1}^{N-1} \mathbb{1}_{\{x_n^{(q)}=\omega_i, x_{n+1}^{(q)}=\omega_j\}}}. \quad (6.18)$$

- Iterations are stopped if the algorithm converges; otherwise, the preceding step is repeated.

7. Results of images segmentation

In the present section, we show several results on the use of PMC and BPMC models for gray image segmentation in the Gaussian context. First, we present a comparative study of the results obtained from both PMC and BPMC models, concerning the misclassification rates by the Bayesian MPM algorithm, PSNR index, and the parameters estimation obtained by ICE and SEM algorithms. Second, we illustrate the segmentation results obtained on simulated images corrupted with some correlated noise. These tests give an idea about the behavior of the ICE and SEM methods for numerical and visual results.

7.1. Numerical results

We segment some noisy simulated images with correlated noise using an unsupervised method for evaluating the endurance of the BPMC model in comparison to the PMC model, as well as the consequences when the data is not PMC or HMC-suited. For PMC model segmentation methods, the modeling of the image is done via an HPS to convert the bidimensional collection of pixels as a mono-dimensional chain, and then reorganize the image for segmentation using an inverse HPS. This

transformation gives us a stochastic process with a very complex structure. This difficulty can be seen in a few papers [2, 7, 14].

In this context, the observation process representing the correlated image was obtained as follows: For each x_s , we simulate the variable N_s according to a Gaussian distribution $\mathcal{N}(\mu_{x_s}, \sigma_{x_s}^2)$, then we take

$$Y_s = \frac{1}{4a + 1} \left(N_s + a \sum_{i=1}^4 N_{s_i} \right), \text{ for each } s \in S, \quad (7.1)$$

where S is the set of pixels in the image in its bidimensional form, where s_i is a neighbor of pixel s in a neighborhood of four nearest neighbors.

We study the performance of the models presented in this work on two classes of images of different sizes, which have been noised by the previous method. However, for BPMC model segmentation methods, the sequence of pixels has been obtained via a “line by line” proceeding. We use the same form of Y_s as above, and for the second element of the observable process, we set \bar{Y}_s as an average of two observations for pixels neighboring to s in the image but not in the chain.


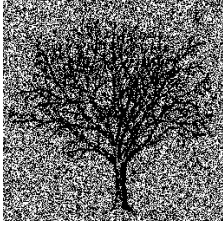
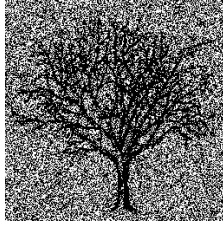
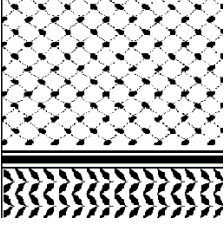
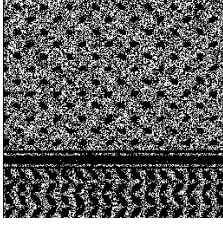
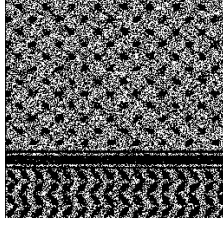

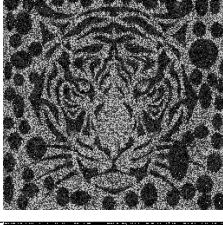
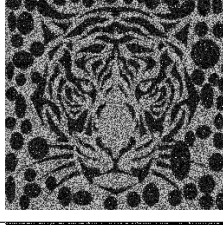
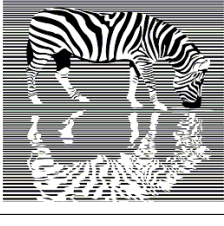

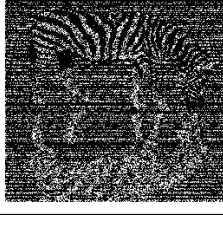
	Original images	Images with correlated noise (series 2) under PMC	Images with correlated noise (series 2) under BPMC
Image1 Size: 128×128			
Image2 Size: 256×256			
Image3 Size: 512×512			
Image4 Size: 1024×1024			

Figure 2. The synthetic images and their noisy versions used in the experiments.

The sequence of pixels collected by converting a bidimensional image to a chain of one dimension is designated by $(s_i)_{1 \leq i \leq N}$. The two realizations of the stochastic processes \mathbf{X} and \mathbf{Y} are respectively, $\mathbf{x} = (x_1, \dots, x_n, \dots, x_N)$ and $\mathbf{y} = (y_1, \dots, y_n, \dots, y_N)$, where $x_i = x_{s_i}$ and $y_i = y_{s_i}$. Under these assumptions, both processes $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ and $(\mathbf{Z}, \bar{\mathbf{Y}})$ have fairly complex structures, and we can see that the distribution of \mathbf{Y} conditionally to $\mathbf{X} = \mathbf{x}$ is not necessarily a Markov; however, we have segmented the noisy image using an unsupervised method based on MPM restoration under PMC and BPMC models, where all parameters are estimated with ICE and SEM algorithms. The question is therefore whether using the BPMC model instead of the PMC model in such a context can improve segmentation results. To give some experiences, we consider two series of parameters a , μ_{x_s} , and σ_{x_s} . Based on the parameters shown in Table 4, the original images and their noisy versions are reported in Figure 2.

We performed six segmentations, and we note that the complete data (C-D) means the original image and its noised version.

- Segmentation by MPM based on parameters of PMC model obtained from the C-D.
- Segmentation by MPM based on parameters of BPMC model obtained from the C-D.
- Segmentation by MPM based on parameters of PMC model obtained with ICE algorithm.
- Segmentation by MPM based on parameters of PMC model obtained with SEM algorithm.
- Segmentation by MPM based on parameters of BPMC model obtained with ICE algorithm.
- Segmentation by MPM based on parameters of BPMC model obtained with SEM algorithm.

Table 4. Parameters used for the segmentation experiments.

	a	μ_1	μ_2	σ_1	σ_2
series 1	0.5	16	10	1	4
series 2	0.05	1	7	1	2

We calculated some evaluated criteria to confirm the visual obtained results. The comparison of the algorithms is performed using quality measures: the error rate and the PSNR index (Table 5).

To start, we can make a number of important observations. Columns 1 and 2 can be seen as a kind of reference for the two models presented in this work, on the grounds that the complete data is used in the estimations. Comparing the results obtained by the estimation algorithms with those calculated by the complete data, one can notice that both the ICE and SEM algorithms implemented under the BPMC model shows favorable results indicating that misclassification rates and PSNR indexes are similar for both noise parameter series, except in the case of Image4 correlated by noise with the parameter of type series 1.

The first conclusion of interest in image segmentation is that the SEM algorithm behaves correctly in the situation under consideration. The comparison of BPMC and PMC models is the second conclusion. It is known that if the data did not fit either an HMC or a PMC, the PMC model-based MPM with HPS might have given better results than another model. This is not the case, since according to the findings and other experiments we have executed, regardless of the presence of strong enough

noise, BPMC model-based MPM restorations surpass PMC model-based ones. Accordingly, using MPM under the BPMC model, ICE and SEM algorithms provide strong unsupervised segmentation methods that may substantially beat PMC model-based ones with HPS.

Table 5. Incorrectly classified and PSNR index given by different MPM segmentations for two series of parameters.

Data		MPM based on BPMC and C-D		MPM based on PMC and C-D		MPM based on BPMC and ICE		MPM based on PMC and ICE		MPM based on BPMC and SEM		MPM based on PMC and SEM	
		$\tau(\%)$	PSNR	$\tau(\%)$	PSNR	$\tau(\%)$	PSNR	$\tau(\%)$	PSNR	$\tau(\%)$	PSNR	$\tau(\%)$	PSNR
Image1	series 1	3.483	40.91	8.996	38.12	4.473	40.60	12.06	40.48	5.471	42.25	9.730	37.03
	series 2	2.366	40.37	9.654	37.12	3.277	39.50	2.571	39.79	2.030	39.95	3.981	36.69
Image2	series 1	3.013	43.25	6.553	41.99	4.400	43.45	11.89	43.32	2.662	42.59	9.730	40.79
	series 2	2.655	42.84	4.801	41.94	2.500	42.87	2.900	43.52	2.465	42.51	2.767	39.84
Image3	series 1	1.437	35.62	2.740	35.52	6.786	35.46	13.36	35.05	0.618	35.69	2.752	35.52
	series 2	1.212	35.68	1.820	35.59	4.059	35.58	4.669	35.51	0.688	35.68	1.962	35.56
Image4	series 1	3.561	31.04	11.015	30.74	12.554	30.69	15.53	30.78	4.981	30.73	6.212	30.64
	series 2	2.421	31.04	2.713	31.02	3.988	31.03	5.633	30.50	3.117	30.74	4.389	30.67

Also, in Table 6, parameters are also found using complete data and the parameters estimated from ICE and SEM algorithms based on BPMC and PMC models. Regarding these estimations by both algorithms, one can perceive that most of the time the estimates are similar by comparing with those obtained by C-D for PMC and BPMC models in the majority of situations of segmented images. With respect to this, the takeaway is that the BPMC model, in its current form, provides another choice that can compete with and likely outperform nearly any other model in the circumstances at hand.

Table 6. Estimations of the parameters for Image1 with correlated noise by series1.

<i>Data</i>		Estimates from C-D under BPMC	Estimates from C-D under PMC	Estimates from ICE under BPMC	Estimates from ICE under PMC	Estimates from SEM under BPMC	Estimates from SEM under PMC
$p(\omega_i, \omega_j)$	$p(\omega_1, \omega_1)$	0.095	0.098	0.125	0.117	0.141	0.028
	$p(\omega_1, \omega_2)$	0.057	0.053	0.374	0.382	0.054	0.032
	$p(\omega_2, \omega_1)$	0.057	0.053	0.064	0.109	0.054	0.032
	$p(\omega_2, \omega_2)$	0.790	0.794	0.435	0.390	0.749	0.906
$f_{\omega_i, \omega_j}(\gamma_n, \gamma_{n+1})$	μ_{ω_1, ω_1}	(14.97,14.96)	(14.30,14.30)	(14.68,14.69)	(13.90,13.90)	(14.44,14.46)	(14.73,14.71)
	μ_{ω_1, ω_2}	(13.95,11.89)	(13.22,11.95)	(13.18,11.36)	(12.17,10.69)	(12.86,11.47)	(14.46,13.03)
	μ_{ω_2, ω_1}	(11.90,13.97)	(11.94,13.23)	(11.37,13.18)	(10.69,12.18)	(11.47,12.82)	(13.07,14.48)
	μ_{ω_2, ω_2}	(10.13,10.13)	(10.27,10.27)	(10.20,10.21)	(10.17,10.17)	(10.06,10.06)	(10.59,10.59)
	$\sigma_{\omega_1, \omega_1}^1$	1.059	1.045	1.306	1.322	1.836	0.331
	$\sigma_{\omega_1, \omega_1}^2$	1.065	1.060	1.313	1.316	1.782	0.361
	$\sigma_{\omega_1, \omega_1}^{12}$	0.092	0.258	0.510	0.602	0.294	0.078
	$\sigma_{\omega_1, \omega_2}^1$	1.022	1.161	3.272	1.713	2.120	0.806
	$\sigma_{\omega_1, \omega_2}^2$	1.405	1.428	2.427	2.113	1.237	1.099
	$\sigma_{\omega_1, \omega_2}^{12}$	0.018	0.255	1.094	0.943	0.315	0.229
	$\sigma_{\omega_2, \omega_1}^1$	1.393	1.447	2.409	2.122	1.189	1.090
	$\sigma_{\omega_2, \omega_1}^2$	1.045	1.163	3.251	1.725	2.141	0.789
	$\sigma_{\omega_2, \omega_1}^{12}$	0.047	0.266	1.077	0.966	0.351	0.227
	$\sigma_{\omega_2, \omega_2}^1$	1.084	1.180	1.441	1.242	1.001	2.031
	$\sigma_{\omega_2, \omega_2}^2$	1.085	1.179	1.442	1.242	0.997	2.039
	$\sigma_{\omega_2, \omega_2}^{12}$	0.073	0.191	0.521	0.530	0.015	0.994

7.2. Visual results

The visual results of the unsupervised segmentation using both BPMC and PMC models are discussed here. The latter have been initialized using a two-class segmentation done by the K -means algorithm to generate the initial configuration of the process X .

The images (Figure 2) to be segmented are chosen based on the homogeneity of the objects. Images 1 and 2 show extremely fine details, however Images 3 and 4 are not particularly homogeneous and missing fine details. Thus, it is easy to demonstrate a significant advantage of the BPMC model over the PMC model, with HPS in the context of image segmentation. In fact, regarding Table 5, the rating is based on the average rate of each algorithm across eight experiments. Under the BPMC model, ICE was 5.25% and SEM was 2.75%, while under the PMC model, ICE was 8.57% and SEM was 5.19%. The PSNR index average results in the following rankings: ICE (37.39), SEM (37.51) under the BPMC model and ICE (37.36), and SEM (35.84) under the PMC model.

We can also visually observe that, depending on the segmentation phase required in these experiments, the error rate is not always the most important criterion. In fact, in Image 4, the details of the zebra (Figure 3) are better restored and there are fewer spurious whites and blacks on the background in the case of the BPMC model, whereas this is not the case for the PMC model

(Figure 4). However, the error rates of the two models are very close.

Nonetheless, we obtained the following conclusions:

(1) The introduction of a second component in the observation process makes the BPMC model a good rival to the PMC model with HPS in all numerical (Table 5) and visual scenarios (Figure 3) in the context of image segmentation.

(2) The SEM algorithm performs better than the ICE algorithm, especially when working on the BPMC model (Table 5).

Remark 7.1. *The HPS is a space-filling curve used to read pixels during image processing. This sweep takes advantage of the two-dimensional locality of pixels. It should be noted that the introduction of such a scan is only conceivable in the case of $2^n \times 2^n$ images, which is why we limit our study to the PMC model. However, the advantage of the proposed model BPMC is that it is capable of handling images of any size.*



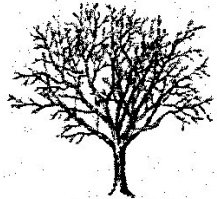
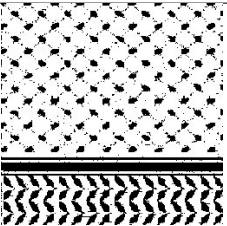
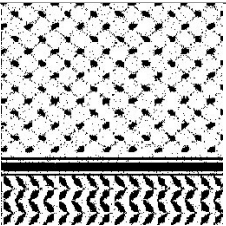
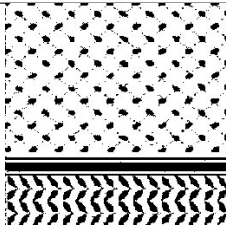

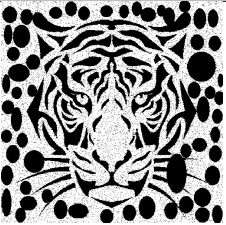

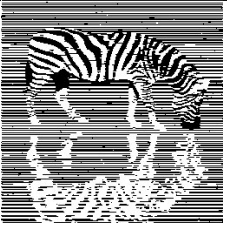
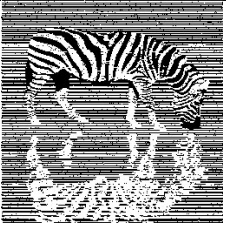

	MPM under BPMC with C-D	MPM under BPMC with ICE	MPM under BPMC with SEM
Image1 (series 1)			
Image2 (series 1)			
Image3 (series 2)			
Image4 (series 2)			

Figure 3. Segmentation results corresponding to the four synthetic images for two noises under the BPMC model.

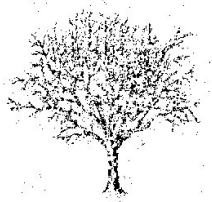
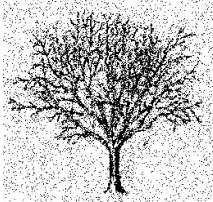
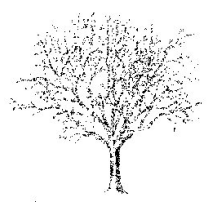
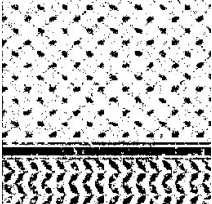
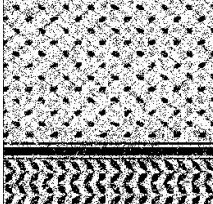
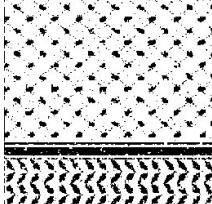

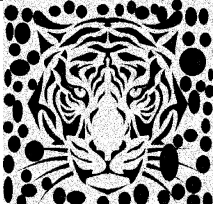



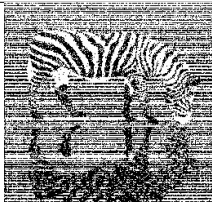
	MPM under PMC with C-D	MPM under PMC with ICE	MPM under PMC with SEM
Image1 (series 1)			
Image2 (series 1)			
Image3 (series 2)			
Image4 (series 2)			

Figure 4. Segmentation results corresponding to the four synthetic images for two noises under PMC model.

7.3. Practical implications of proposed approach

An unsupervised approach is particularly useful when a smaller amount of or no labeled data is available for training. The proposed approach has shown superior results compared to the classical PMC model and can be potentially utilized for various applications. Particularly, the suggested model can be used for the segmentation of radar images, textured images, and medical images, because these types of images have an important correlation of noise.

8. Conclusions and perspectives

We proposed an unsupervised approach for restoring hidden data by applying the PMC model. This study contributes to provide a 2D observed process for the latter model. The novel BPMC model is proposed first to compete with the classical PMC model for image segmentation and then to solve the problem of correlation between noise and the difficulty of two-dimensional pixel locality during one-dimensional image transformation. The parameter estimation methods described for this model are applicable to Gaussian and possibly correlated noise. We initially found that when the

data follows a PMC structure, Bayesian restorations according to the BPMC model beat those based on the PMC model. Likewise, the parameter estimating methods described for the BPMC model are extremely efficient, as restorations based on complete parameters are comparable to restorations using approximated values. Next, we executed a series of tests combining the use of unsupervised segmentation of images via an HPS in the case of the PMC model and the insertion of a second element into the observed process in the case of the BPMC model. The assessment of interest in this study is based on two points. In the first step, we show that the BPMC model works best when the stochastic process from a noisy class image with correlated noise is very complex (which is neither PMC nor HMC). Several experiments in this study show that the BPMC model always gives better results compared to the PMC model, and the difference can even be quite significant in some situations. So, one way to generate complex data is to use a noisy image in a correlative version and an HPS or line by line technique. In the second step, we proved that estimation methods corresponding to the proposed model can compete with classical methods based on the PMC model with high efficiency and are therefore of interest for the image segmentation problem.

As a general conclusion, we can confirm that the proposed model, where we introduced a second component in the observed process, offers an interesting opportunity compared to the PMC model with an HPS. Basically, the composition of the second component of \mathbf{Y} based on the neighborhood of four nearest neighbors of each pixel provides the same role as the HPS, while on the other hand, the BPMC-based unsupervised restorations with parameter estimation methods surpass those based on the PMC model.

As perspectives for further work, we may apply the concept used here to the triplet Markovian model described in [40, 41] in the case of Gaussian noise [4], in a way that the observed process will be two-dimensional and the resulting model will therefore be a bidimensional triplet Markov chain. Of course, the same concept could be used for the segmentation of fuzzy data. It consists of considering a fuzzy HMC model with bidimensional observations as presented in [42] and applying it to medical images. In the proposed model, both spatial correlation and locality details about pixels should be considered at the same time. Another potential direction is concerning types of noise, such as applying non-Gaussian noise to image segmentation using bidimensional Markov models.

Author contributions

A. Joumad: Conceptualization, Data curation, Writing-original draft; A. El Moutaouakkil: Conceptualization, Formal analysis, Writing-original draft; A. Nasroallah: Data curation, Formal analysis, Methodology; O. Boutkhoul: Methodology, Project administration, Software; Mejdil Safran: Funding acquisition, Investigation, Visualization; Sultan Alfarhood: Investigation, Project administration, Visualization; Imran Ashraf: Supervision, Validation, Writing-review & editing. All authors have read and approved the final version of the manuscript for publication.

Acknowledgements

The authors extend their appreciation to King Saud University for funding this research through Researchers Supporting Project Number (RSPD2024R890), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. O. Cappé, E. Moulines, T. Rydén, Inference in hidden markov models, *Proceedings of EUSFLAT Conference*, 2009, 14–16.
2. B. Benmiloud, W. Pieczynski, Estimation des paramètres dans les chaînes de markov cachées et segmentation d'images, *Traitement du signal*, **12** (1995), 433–454.
3. J. B. M. Robert, J. Elliott, L. Aggoun, *Hidden Markov models: Estimation and control*, Science & Business Media, 1995.
4. C. Fernandes, *Chaînes de Markov triplets et segmentation non supervisée d'images*, Institut Polytechnique de Paris, 2022.
5. L. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, *Proceedings of IEEE*, **77** (1989), 257–286. <https://doi.org/10.1109/5.18626>
6. W. Pieczynski, Pairwise markov chains, *IEEE T. Pattern Anal.*, **25** (2003), 634–639. <https://doi.org/10.1109/TPAMI.2003.1195998>
7. C. Fernandes, T. Monti, E. Monfrini, W. Pieczynski, Fast image segmentation with contextual scan and Markov chains, In: *29th European Signal Processing Conference, Dublin, Ireland*, 2021, 626–630. <https://doi.org/10.23919/EUSIPCO54536.2021.9616332>
8. A. Joumad, A. E. Moutaouakkil, A. Nasroallah, O. Boutkhoum, F. Rustam, I. Ashraf, Unsupervised statistical image segmentation using bi-dimensional hidden markov chains model with application to mammography images, *J. King Saud Univ.-Com.*, **35** (2023), 101715. <https://doi.org/10.1016/j.jksuci.2023.101715>
9. A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc. B*, **39** (1977), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
10. G. Celeux, F. Forbes, N. Peyrard, EM procedures using mean field-like approximations for markov model-based image segmentation, *Pattern Recogn.*, **36** (2003), 131–144. [https://doi.org/10.1016/S0031-3203\(02\)00027-4](https://doi.org/10.1016/S0031-3203(02)00027-4)
11. W. Pieczynski, Hidden markov fields and iterative conditional estimation, *Traitement du Signal*, **11** (1994), 141–154.
12. S. Allasonnière, E. Kuhn, Convergent stochastic expectation maximization algorithm with efficient sampling in high dimension. application to deformable template model estimation, *Comput. Stat. Data An.*, **91** (2015), 4–19. <https://doi.org/10.1016/j.csda.2015.04.011>
13. S. Huda, J. Yearwood, R. Togneri, A stochastic version of expectation maximization algorithm for better estimation of hidden Markov model, *Pattern Recogn. Lett.*, **30** (2009), 1301–1309. <https://doi.org/10.1016/j.patrec.2009.06.006>
14. S. Derrode, W. Pieczynski, Signal and image segmentation using pairwise Markov chains, *IEEE T. Signal Proces.*, **52** (2004), 2477–2489. <https://doi.org/10.1109/TSP.2004.832015>

15. R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, F. Tupin, Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields, In: *IEEE Transactions on Geoscience and Remote Sensing*, **41** (2003), 675–686. <https://doi.org/10.1109/TGRS.2003.809940>
16. S. Z. Li, *Markov random field modeling in image analysis*, Springer Science & Business Media, 2009.
17. Q. Jackson, D. A. Landgrebe, Adaptive Bayesian contextual classification based on Markov random fields, In: *IEEE Transactions on Geoscience and Remote Sensing*, **40** (2002), 2454–2463. <https://doi.org/10.1109/TGRS.2002.805087>
18. K. Kuljus, J. Lember, Pairwise Markov models and hybrid segmentation approach, *Methodol. Comput. Appl. Probab.*, **25** (2023). <https://doi.org/10.1007/s11009-023-10044-z>
19. E. Monfrini, J. Lecomte, F. Desbouvries, W. Pieczynski, Image and signal restoration using pairwise Markov trees, In: *IEEE Workshop on Statistical Signal Processing, Saint Louis, MO, USA*, 2003, 174–177. <https://doi.org/10.1109/SSP.2003.1289372>
20. W. Pieczynski, A. N. Tebbache, Pairwise Markov random fields and segmentation of textured images, *Mach. Graph. Vision*, **9** (2000), 705–718.
21. M. E. Y. Boudaren, L. An, W. Pieczynski, Dempster-Shafer fusion of evidential pairwise Markov fields, *Int. J. Approx. Reason.*, **74** (2016), 13–29. <https://doi.org/10.1016/j.ijar.2016.03.006>
22. J. B. Courbot, V. Mazet, E. Monfrini, C. Collet, Pairwise Markov fields for segmentation in astronomical hyperspectral images, *Signal Process.*, **163** (2019), 41–48. <https://doi.org/10.1016/j.sigpro.2019.05.005>
23. S. Derrode, SAR image segmentation using generalized pairwise Markov chains, In: *Proceedings of SPIE-The International Society for Optical Engineering*, **4885** (2002). <https://doi.org/10.1117/12.463177>
24. S. Derrode, W. Pieczynski, Segmentation non supervisée d’images par chaîne de Markov couple, In: *Ateliers Traitement et Analyse de l’Information: Méeshodes et Applications, Hammamet, Tunisie*, 2003.
25. N. Brunel, F. Barbaresco, Doppler and polarimetric statistical segmentation for radar clutter map based on pairwise Markov chains, *Proc. of IEEE RADAR*, (2007), 8–10.
26. S. Derrode, W. Pieczynski, Unsupervised data classification using pairwise Markov chains with automatic copulas selection, *Comput. Stat. Data An.*, **63** (2013), 81–98. <https://doi.org/10.1016/j.csda.2013.01.027>
27. A. K. Atiampo, G. L. Loum, Unsupervised image segmentation with pairwise Markov chains based on nonparametric estimation of copula using orthogonal polynomials, *Int. J. Image Graph.*, **16** (2016), 2526–2541. <https://doi.org/10.1142/S0219467816500200>
28. S. Rafi, M. Castella, W. Pieczynski, Pairwise Markov model applied to unsupervised image separation, In: *IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, Innsbruck, Austria*, 2011. <https://doi.org/10.2316/P.2011.721-044>
29. E. Azeraf, E. Monfrini, E. Vignon, W. Pieczynski, Highly fast text segmentation with pairwise Markov chains, In: *6th IEEE International Congress on Information Science and Technology, Machine Learning for Natural Language Processing, Agadir-Essaouira, Morocco*, (2020), 361–366. <https://doi.org/10.1109/CiSt49399.2021.9357304>

30. I. Papila, O. Ersoy, Multiscale segmentation of remotely sensed images using pairwise Markov chains, In: *IEEE Antennas and Propagation Society Symposium*, **2** (2004), 2123–2126. <https://doi.org/10.1109/APS.2004.1330629>
31. S. L. Cam, F. Salzenstein, C. Collet, Fuzzy pairwise Markov chain to segment correlated noisy data, *Signal Proces.*, **88** (2008), 2526–2541. <https://doi.org/10.1016/j.sigpro.2008.05.003>
32. C. Carincotte, S. Derrode, S. Bourennane, Multivariate fuzzy hidden Markov chains model applied to unsupervised multiscale SAR image segmentation, In: *The 14th IEEE International Conference on Fuzzy Systems, Reno, NV, USA, 2005*, 288–293. <https://doi.org/10.1109/FUZZY.2005.1452408>
33. B. Tso, R. C. Olsen, Combining spectral and spatial information into hidden Markov models for unsupervised image classification, *Int. J. Remote Sens.*, **26** (2005), 2113–2133. <https://doi.org/10.1080/01431160512331337844>
34. J. F. Mari, F. L. Ber, Temporal and spatial data mining with second-order hidden markov models, *Soft Comput.*, **10** (2006), 406–414.
35. A. Hafiane, B. Zavidovique, S. Chaudhuri, A modified FCM with optimal Peano scans for image segmentation, In: *IEEE International Conference on Image Processing, Genova, Italy, 2005*, III–840. <https://doi.org/10.1109/ICIP.2005.1530523>
36. Y. L. Song, B. Adobah, J. F. Qu, C. M. Liu, Segmentation of ordinary images and medical images with an adaptive hidden Markov model and viterbi algorithm, *Curr. Signal Transd. T.*, **15** (2020), 109–123. <https://doi.org/10.2174/1574362413666181109113834>
37. W. Skarbek, Generalized hilbert scan in image printing, In: *Theoretical Foundations of Computer Vision, R. Klette et W. G. Kropetsh, editors, Akademik Verlag, Berlin, 1992*, 45–57.
38. N. Brunel, *Sur quelques extensions des chaînes de Markov cachées et couples. Application à la segmentation non supervisée des signaux radar*, Université Pierre et Marie Curie-Paris VI, 2005.
39. P. Lanchantin, *Chaînes de Markov triplets et segmentation non supervisée de signaux*, Evry: Institut national des télécommunications, 2006.
40. W. Pieczynski, Chaines de Markov triplet, *Comptes Rendus. Mathématiques*, **335** (2002), 275–278. [https://doi.org/10.1016/S1631-073X\(02\)02462-7](https://doi.org/10.1016/S1631-073X(02)02462-7)
41. W. Pieczynski, F. Desbouvries, On triplet Markov chains, In: *Proceeding of the International Symposium on Applied Stochastic Models and Data Analysis, Brest, France, 335* (2005).
42. A. Joumad, A. E. Moutaouakkil, A. Nasroallah, O. Boutkhoum, Unsupervised image segmentation using fuzzy hidden Markov chain with bi-dimensional data, In: *11th International Symposium on Signal, Image, Video and Communications, El Jadida, Morocco, 2022*, 1–6. <https://doi.org/10.1109/ISIVC54825.2022.9800731>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)