Non-stationary fuzzy Markov chain

F. Salzenstein a,*, C. Collet b, S. Lecam b, M. Hatt c

a Laboratoire INESS, UMR CNRS 7163, Université Strasbourg 1 (ULP), France
b LSIT UMR CNRS 7005, Université Strasbourg 1 (ULP), France
c Laboratoire LATIM, INSERM U650, Université de Brest (UBO), France

Received 18 July 2006; received in revised form 17 April 2007
Available online 17 July 2007
Communicated by J.A. Robinson

Abstract

This paper deals with a recent statistical model based on fuzzy Markov random chains for image segmentation, in the context of stationary and non-stationary data. On one hand, fuzzy scheme takes into account discrete and continuous classes through the modeling of hidden data imprecision and on the other hand, Markovian Bayesian scheme models the uncertainty on the observed data. A non-stationary fuzzy Markov chain model is proposed in an unsupervised way, based on a recent Markov triplet approach. The method is compared with the stationary fuzzy Markovian chain model. Both stationary and non-stationary methods are enriched with a parameterized joint density, which governs the attractiveness of the neighbor states. Segmentation task is processed with Bayesian tools, such as the well known MPM (Mode of Posterior Marginals) criterion. To validate both models, we perform and compare the segmentation on synthetic images and raw optical patterns which present diffuse structures.

Keywords: Fuzzy Markov chain; Triplet Markov chain; Non-stationary chain; Multispectral image segmentation

1. Introduction

The fuzzy segmentation problem consists of estimating the hidden realization \( X = (X_s)_{1 \leq s \leq N} \), for a given set of \( D \) observations \( Y = y = \{y_s \in \mathbb{R}^D\} \), where \( x_s = (e_1(s), e_2(s), \ldots, e_K(s)) \). Each component \( e_i(s) \) represents the contribution of each class \( \omega_i \) in a finite discrete set \( \Omega = \{\omega_1, \ldots, \omega_K\} \) of \( K \) hard classes. The fuzzy belonging of each pixel respects the normalization condition: \( e_1(s) + e_2(s) + \cdots + e_K(s) = 1 \). If the context of two “hard” classes, a set \( \Omega = \{0, 1\} \) yields \( x_s \in [0, 1] \). Then, all values \( x_s \in [0, 1] \) model the proportion of the class “0” in the pixel related to \( X_s \), whereas \( 1 - x_s \) corresponds to the proportion of the class “1”. The distribution at each random variable \( X_s \) is given by a density \( h_s \) with respect to a measure \( \nu \) including discrete components (Dirac functions \( \delta_0, \delta_1 \) on \( \{0, 1\} \)) and a continuous component (the Lebesgue measure \( \mu \) on \( [0, 1]\) (Caillol et al., 1993):

\[
\nu = \delta_0 + \delta_1 + \mu.
\]

The discrete components of \( \nu \) are associated with the hard classes, whereas the continuous component \( \mu \) is associated with the fuzzy feature. In this paper, we will consider the case \( D = N \) (mono-spectral context). When \( X \) is a Markov chain called “fuzzy Markov chain” (FMC) and the variable \( Y \) is independent conditionally on \( X \), it is possible to express the joint distribution \( p(x, y) \) with respect to a measure \( \nu^N \otimes \mu^N \), as follows:

\[
p(x, y) = p(x_1)p(x_2|x_1) \cdots p(x_N|x_{N-1})p(y_1|x_1) \cdots p(y_N|x_N).
\]

In particular, the posterior field \( X \) conditional on \( Y \) is Markovian. Thus, one can process the posterior realizations of
the hidden variable $X$, called Hidden Fuzzy Markov Chain (HFMC). Generally the distribution $p(x,y)$ depends on unknown parameter $\theta = (\theta_X, \theta_Y)$ where the prior parameters $\theta_X$ define the prior density of the Markov chain and the parameters $\theta_Y$ define the distribution parameters of the driven data conditional on $X$. Algorithms like “Expectation Maximization” (EM) (McLachlan and Krishnan, 1997) or its stochastic version (SEM) (Celeux and Diebolt, 1985) are efficient to estimate the hyper-parameter when $\theta_X$ does not vary locally, i.e., when the variable $X$ is stationary. Recent studies have focused on unsupervised segmentation of Markov chain in the fuzzy context (Avrachenkov and Sanchez, 2002; Mohammed and Gader, 2000; Carincolite et al., 2004). In particular, we derive $\theta_X$ from the prior joint density $p(x_s, x_{s+1})$ at each neighbored site. We present here a new model based on a parameterized joint density, which governs locally the attractiveness between two neighbored states. Unfortunately, when $\theta_X$ does not vary locally, these approaches are sometimes badly adapted. Thus one has to introduce a new fuzzy hidden Markov chain model which represents non-stationary data. In this work we model the non-stationarity by a third auxiliary process $U$, which governs the changing values of $\theta_X$ in the hidden process. A such method has been successfully applied in the hard context (Hughes et al., 1999; Lanchantin and Pieczynski, 2004), and we propose in this article to extend non-stationary Markov chain to the fuzzy case. The solution proposed in (Lanchantin and Pieczynski, 2004) is derived from a recent triplet Markov chain model (Pieczynski, 2002) which can be described in the following manner: the pairwise process $Z = (X, U)$ is assumed to be Markovian, $X$ and $U$ separately are not necessary Markovian. The triplet process $T = (X, U, Y)$ is then a particular triplet Markov chain. In Section 2, we present the stationary fuzzy Markov chain (SFMFC) with and without a parameterized joint density (P-SFMC versus NP-SFMC). We briefly introduce the stationary fuzzy Markov field (SFMF) (Salzenstein and Pieczynski, 1997), which is used in the experimental part to enrich the comparisons. In the next Section 3 we generalize the non-stationary model of Lanchantin and Pieczynski (2004) presenting a new fuzzy model in the context of non-stationary Markov chain (NSFMC) with a possibly joint parameterized density (P-NSFMC versus NP-NSFMC). We describe the noise model used (Section 4), the MPM segmentation procedure applied to the S/NS-FMC methods (Section 5) and the associated hyper-parameter estimation step (Section 6). Finally we show the efficiency of the new method though synthetic images (Section 7) and real images (Section 8).

2. The stationary fuzzy Markov chain (SFMFC)

Let us consider now a Markov chain $X = (x_s)_{1 \leq s \leq N}$ with continuous statements, i.e., $X_s \in [0,1]$. To define the distribution $\pi(x)$ of the variable $X$, we need the density $p(x_1)$ of the initial distribution, and the transition densities $p(x_s | x_{s-1})_{1 \leq s \leq N}$.

\[ p(x_1, x_2, \ldots, x_N) = p(x_1) \cdot p(x_2 | x_1) \cdots p(x_N | x_{N-1}). \]

(3)

When the chain is stationary, all prior distributions can be deduced from a joint density. The prior joint density $p(x_{s}, x_{s+1})$ is defined on the pairwise $(x_s, x_{s+1}) \in [0,1]^2$. According to a measure $\nu \otimes \nu$, the normalization condition yields

\[ \int_0^1 \int_0^1 p(u, v) d(\nu \otimes \nu)(u, v) = 1. \]

(4)

We propose a general model to define it

\[ p(e_1, e_2) = a \cdot \phi(e_1, e_2) + b \quad \text{with} \quad \phi(e_1, e_2) = \phi(e_2, e_1); \quad (a, b) \in \mathbb{R}^2. \]

(5)

The function $\phi(e_1, e_2)$, is applied when at least one label is fuzzy, i.e., $e_1$ or $e_2 \in [0, 1]$. If both labels are hard, we note $p(0,0) = \pi_{00}, p(1,1) = \pi_{11}, p(0,1) = \pi_{01}, p(1,0) = \pi_{10}$. We model a parameterized function $\phi$ as follows:

\[ \phi(e_1, e_2) = (1 - |e_1 - e_2|)^r \quad r \in \mathbb{R}. \]

(6)

When $r$ increases, the probability of having two similar neighbored pixels increases: thus, the parameter $r$ governs the homogeneity of the image, i.e., the attractiveness between the different states. Moreover, the limit conditions that we impose yield

\[ p(0,1) = p(1,0) = b = \pi_{01} = \pi_{10}. \]

(7)

Applying Eq. (4) yields to the general condition:

\[ p(0,0) + p(1,1) + p(0,1) + p(1,0) + 2 \cdot \int_{[0,1]} p(0, u) du \]

\[ + 2 \cdot \int_{[0,1]} p(1, u) du + \int_{[0,1]} \int_{[0,1]} p(u, v) du dv = 1. \]

(8)

This gives a relationship between all prior parameters $(\pi_{00}, \pi_{11}, \pi_{01}, \pi_{10}, a, b)$. When the prior joint density is defined by the function (6), we compute (8) using a quantization of the interval $[0,1]$ into $M$ equidistant values: $\{e_0 = 0, e_1 = \frac{1}{M}, \ldots, e_2 = \frac{2}{M}, \ldots, e_0 = 1\}$. Then, we derive the initial density $p(x_1)$, which corresponds to the marginal distribution (9)

\[ p(x_s) = \int_0^1 p(x_s, \nu) d\nu(\nu) \]

\[ = p(x_s, 0) + p(x_s, 1) + \int_{[0,1]} p(x_s, \nu) d\nu. \]

(9)

In this section we briefly described a new fuzzy Markov random chain model based on a parameterized joint density associated to the transition probabilities. The stationary fuzzy chain associated with a non-parameterized density is named NP-SFMC. The stationary fuzzy chain associated with a parameterized density is named P-SFMC. It is also possible to define the same way, the stationary Markovian random field (SFMF) (Salzenstein and Pieczynski, 1997) $X \in [0,1]^N$, for which the distribution $p_x$ with respect to a measure $\nu^X$ is given by Salzenstein and Pieczynski (1997).
where the fuzzy energy $U_i$ is a sum of functions $\Phi_C$ defined on neighborhoods:

$$U_i(x) = \sum_{(i, j) \in C} \Phi_C(x_i, x_j),$$

where $(x_i, x_j) \in [0, 1]^2$ represents a pairwise of neighborhood pixels (Geman and Geman, 1984) in the image (vertical, diagonal or horizontal neighborhood), and the associated functions $\Phi_C$ defined on $[0, 1]^2$. Let us notice, the SFMF set $[0,1]$ represents a pairwise of neighborhood pixels $(x_i, x_j) \in [0,1]^2$.

In (Lanchantin and Pieczynski, 2004), the couple $Z = (X, U) = ((x_1, u_1), (x_2, u_2), \ldots, (x_N, u_N))$ is supposed to be a stationary Markov chain, where $X$ is an interested non-stationary process, and $U$ models auxiliary states:

$$p(z_i = (x_i, u_i)|z_{i-1}, z_{i-2}, \ldots, z_1) = p(z_i|z_{i-1}).$$

In (Lanchantin and Pieczynski, 2004), $X$ and $U$ take their values into discrete classes. We propose to generalize this model by labeling each component $X_i$ into a continuous set $[0,1]$. The intermediate variable $U$ takes its values into a finite set, in order to define stationary partitions of the variable $X$. The chain $Z$ is defined by a prior joint density

$$p(z_i, z_{i+1}) = p(x_i, x_{i+1}, u_i, u_{i+1}),$$

according to the measure $(v + \sum_{n=1}^{K} \delta_n) \otimes (v + \sum_{n=1}^{K} \delta_n)$, the initial probability is computed by

$$P(X_1, X_{i+1} \in I_{i+1}, U_i = \lambda_i, U_{i+1} = \lambda_{i+1}) = \int_{I_{i+1}} \int_{I_{i+1}} p(c, \eta, \lambda_i, \lambda_{i+1}) d(v \otimes v)(c, \eta),$$

with $I_i \subset [0,1]$ and $I_{i+1} \subset [0,1]$. As in the stationary case, it is possible to define a parameterized and non-parameterized joint density, provided that the normalization condition (14) is established:

$$\sum_{x_i} \sum_{x_j} \int_0^1 \int_0^1 p(c, \eta, \lambda_i, \lambda_j) d(v \otimes v)(c, \eta) = 1.$$

In particular, this condition is written

$$\sum_{x_i} \sum_{x_j} P_{ij} = 1,$$

where

$$\int_0^1 \int_0^1 g(c, \eta, \lambda_i, \lambda_j) d(v \otimes v)(c, \eta) = P[\lambda_i, \lambda_j] = P_{ij}.$$

Thus, it is possible to construct a parameterized mode of $p(x_i, x_{i+1})$ by the means of the parameters $\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}$

$$p(0, 0, \lambda_i, \lambda_j) = \pi_{00}, \quad p(0, 1, \lambda_i, \lambda_j) = \pi_{01},$$

$$p(1, 0, \lambda_i, \lambda_j) = \pi_{10}, \quad p(1, 1, \lambda_i, \lambda_j) = \pi_{11}. (16)$$

When $c_1 \lor c_2 \in [0,1]$, let us express $a_{ij}, b_{ij}$ and the auxiliary function $\phi$ defined by (6)

$$a_{ij} \cdot \phi(0,1) + b_{ij} = a_{ij} \cdot \phi(1,0) + b_{ij} = \pi_{01} = \pi_{10}. (20)$$

Finally the neighborhood prior density depends on $4 \times K^2$ parameters $\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}$. The other parameters $\pi_{10}, b_{ij}$ and $a_{ij}$ are computed by the conditions (15), (16), (20). The non-stationary Markov chain based on a parameterized joint density will be named P-NSFMC whereas the model based on a non-parameterized density is named NP-NSFMC. In this section we introduced a new fuzzy non-stationary Markov random chain model. We defined the associated prior joint density, initial and transition probabilities. Let us now describe the segmentation task.

3. The non-stationary fuzzy Markov chain (NSFMC)

Authors (Lanchantin and Pieczynski, 2004) propose to add to an initial process $X$ an additional process $U$, which takes its values in a finite set $A = \{\lambda_1, \lambda_2, \ldots, \lambda_K\}$. The couple $Z = (X, U) = \{(x_1, u_1), (x_2, u_2), \ldots, (x_N, u_N)\}$ is supposed to be a stationary Markov chain, where $X$ is an interested non-stationary process, and $U$ models auxiliary states:

$$p(z_i = (x_i, u_i)|z_{i-1}, z_{i-2}, \ldots, z_1) = p(z_i|z_{i-1}).$$

In (Lanchantin and Pieczynski, 2004), $X$ and $U$ take their values into discrete classes. We propose to generalize this model by labeling each component $X_i$ into a continuous set $[0,1]$. The intermediate variable $U$ takes its values into a finite set, in order to define stationary partitions of the variable $X$. The chain $Z$ is defined by a prior joint density

$$p(z_i, z_{i+1}) = p(x_i, x_{i+1}, u_i, u_{i+1}),$$

according to the measure $(v + \sum_{n=1}^{K} \delta_n) \otimes (v + \sum_{n=1}^{K} \delta_n)$, the initial probability is computed by

$$P(X_s \in I_s, X_{s+1} \in I_{s+1}, U_s = \lambda_s, U_{s+1} = \lambda_{s+1}) = \int_{I_s} \int_{I_{s+1}} p(c, \eta, \lambda_s, \lambda_{s+1}) d(v \otimes v)(c, \eta),$$

with $I_s \subset [0,1]$ and $I_{s+1} \subset [0,1]$. As in the stationary case, it is possible to define a parameterized and non-parameterized joint density, provided that the normalization condition (14) is established:

$$\sum_{x_i} \sum_{x_j} \int_0^1 \int_0^1 p(c, \eta, \lambda_i, \lambda_j) d(v \otimes v)(c, \eta) = 1.$$
where \( \mu_s = [\mu_s^{(1)}, \ldots, \mu_s^{(D)}] \) and \( \Gamma_s \in \mathbb{R}^D \times \mathbb{R}^D \), respectively, define a mean vector and variance–covariance matrix, at each fuzzy/hard site. Let be \( (\mu_0, \mu_1) \) and \( (\Gamma_0, \Gamma_1) \) the mean vectors and variance–covariance matrix related to the hard classes “0” and “1”. For each fuzzy site \( x_s = \varepsilon_s \), the related mean vector and covariance matrix \( \mu_s \) and \( \Gamma_s \) are written

\[
\mu_s = (1 - \varepsilon_s) \cdot \mu_0 + \varepsilon_s \cdot \mu_1, \tag{25}
\]

\[
\Gamma_s = (1 - \varepsilon_s)^2 \cdot \Gamma_0 + \varepsilon_s^2 \cdot \Gamma_1. \tag{26}
\]

5. Segmentation procedure

5.1. Segmentation of the SFMC

Given the set of observations \( Y = y \), we wish to estimate one realization \( X = x \in [0, 1]^N \). It is possible to adapt the MPM criterion (Maroquin et al., 1987) to the fuzzy context (Salzenstein and Pieczynski, 1997). For such a process, the final decision process is performed as follows: given a realization \( Y = y \), the Bayesian decision \( d_s \) such that \( d_s(Y) = X_s \), will involve minimizing a conditional expectation (27) at each location \( s \), in order to obtain an optimal value of \( X_s \)

\[
\hat{x}_s^{\text{opt}} = \arg \min_{x_s} E[L_s(X_s, \hat{X}_s)|Y = y]. \tag{27}
\]

The loss function \( L_s(x_s, \hat{x}_s) \) models the severity of attributing the value \( \hat{x}_s \) instead of a true one \( x_s \) to the pixel. Although there are numerous possibilities in the choice of the loss function, the ‘absolute distance’ \( L_s(x_s, \hat{x}_s) = |x_s - \hat{x}_s| \), gives efficient results for segmentation tasks. For a stationary variable \( X \), the error rate is approximated by

\[
E[L_s(X_s, \hat{X}_s)] \approx \frac{1}{N} \sum_{s=1}^{N} L_s(x_s, \hat{x}_s). \tag{28}
\]

The calculus of (27) requires the knowledge of the posterior distribution \( p_{\hat{X}_s} \) at each \( X_s \):

\[
E[L_s(X_s, \hat{X}_s)|y] = p_{\hat{X}_s}(0) \cdot L_s(0, \hat{X}_s) + p_{\hat{X}_s}(1) L_s(1, \hat{X}_s) + \int_{[0,1]} p_{\hat{X}_s} \cdot L_s(t, \hat{X}_s) dt. \tag{29}
\]

Segmentation is performed by affecting to each pixel a value \( \hat{X}_s \in [0, 1] \) which minimizes (29). When the hidden process \( X \) is a Markov chain, one can compute the posterior density using the forward/backward procedure (Devijver, 1985) in the fuzzy context. The forward and backward densities \( \alpha_s(x_s), \beta_s(x_s) \) are defined by

\[
\alpha_s(x_s) = p(x_s, y_1, \ldots, y_s), \tag{30}
\]

\[
\beta_s(x_s) = \frac{p(y_{s+1}, \ldots, y_N|x_s)}{p(Y_{s+1}, \ldots, Y_N|y_1, \ldots, y_s)}. \tag{31}
\]

The recurrence formula providing these quantities, are analogous to the hard segmentation process

\[
\alpha_s(x_s) \propto p(y_s|x_s) \int_0^1 \alpha_{s-1}(u) \cdot p(x_s|u) dv(u), \tag{32}
\]

\[
\beta_s(x_s) \propto \int_0^1 \beta_{s+1}(u) p(y_{s+1}|u)p(u|x_s) dv(u). \tag{33}
\]

The relationship \( \alpha_s(x_s) \cdot \beta_s(x_s) = p_{\hat{X}_s} \) gives immediately the minimization of (29). Moreover, as in the hard context, it is possible to simulate posterior realizations of the variable \( X \) using the posterior transition \( p_{\hat{X}_s}^{s+1}(x_{s+1}|x_s) \) and initial densities \( p_{\hat{X}_s}^0(x_s) \) i.e., for any \( (x_s, x_{s+1}) \in [0,1]^2 \):

\[
p_{\hat{X}_s}^0(x_s) = \alpha_s(x_s) \cdot \beta_s(x_s), \tag{34}
\]

\[
p_{\hat{X}_s}^{s+1}(x_{s+1}|x_s) = \frac{p(x_{s+1}|x_s) \cdot p(y_{s+1}|x_{s+1}) \cdot \beta_{s+1}(x_{s+1})}{\int_0^1 p(x|x_s)p(y_{s+1}|x) \cdot \beta_{s+1}(x) dv(x)}. \tag{35}
\]

5.2. Segmentation of the NSFMC

The segmentation problem consists then in estimating a realization of an hidden process \( X = x \), which is not necessary Markovian, given a set \( Y = y \). In order to estimate the fuzzy process \( X \), one has to minimize the conditional expectation (27). In order to evaluate the hidden auxiliary process \( U \), we apply the classic decision task (36) corresponding to the local “0–1” loss function:

\[
\hat{U}_{x_s}^{\text{opt}} = \arg \min_{U} P[U|Y = y]. \tag{36}
\]

Thus, it is necessary to compute the posterior densities \( p(X_s|Y = y) \) and \( p(U|Y = y) \) in order to perform the decision processes (27) and (36). They correspond to the marginalization of the distribution \( p(Z|y) \):

\[
p(x_s|y) = \sum_{a} p(z_s|a), \tag{37}
\]

\[
p(u_s|y) = \int_0^1 p(z_s|y) dv, \tag{38}
\]

\[
Z = (X, U) \text{ being a pairwise Markov chain, it is possible to compute the posterior distribution } \ p(z_s|y) \text{ by the means of the forward–backward procedure, extended to the fuzzy context:}
\]

\[
\alpha_s(z_s) \propto p(y_s|z_s) \sum_{\lambda \in A} \int_0^1 \alpha_{s-1}(\varepsilon, \lambda) \cdot p(z_s|\varepsilon, \lambda) dv(\varepsilon), \tag{39}
\]

\[
\beta_s(z_s) \propto \sum_{\lambda \in A} \int_0^1 \beta_{s+1}(\varepsilon, \lambda)p(y_{s+1}|\varepsilon, \lambda)p(\varepsilon, \lambda|z_s) dv(\varepsilon). \tag{40}
\]

Using the hypothesis related to the observed data, we simplify this procedure in the following manner:

\[
\alpha_s(z_s) \propto p(y_s|x_s) \sum_{\lambda \in A} \int_0^1 \alpha_{s-1}(\varepsilon, \lambda) \cdot p(z_s|\varepsilon) dv(\varepsilon), \tag{41}
\]

\[
\beta_s(z_s) \propto \sum_{\lambda \in A} \int_0^1 \beta_{s+1}(\varepsilon, \lambda)p(y_{s+1}|\varepsilon)p(\varepsilon, \lambda|z_s) dv(\varepsilon). \tag{42}
\]

At each step, we compute the posterior distribution

\[
p(x_s, u_s|y) = \alpha_s(x_s, u_s) \cdot \beta_s(x_s, u_s). \tag{43}
\]
In the same manner as we have seen in Section 5.1, it is possible to simulate hidden realizations of the variable \( Z = (X, U) \) according to the posterior initial and transition densities. This allows us to estimate hyper-parameters by using the well known SEM procedure.

6. Hyper-parameter estimation

We focus in this section on the estimation of the parameter \( \theta \) in the context of a non-stationary variable. Actually, the stationary context is a particular case, for which \( U \) owns one discrete state i.e., \( \text{Card} A = 1 \). The final segmentation step requires the parameter set \( \theta = (\theta_X, \theta_Y) \) where the prior parameters \( \theta_Z \) define the prior density of the Markov chain \( Z \), which could be the set of parameters \((\pi_{0,0}, \pi_{0,1}, \pi_{1,0}, \pi_{1,1}, a_{ij}, b_{ij})\) for the P-SFMC and P-NSFMC approaches. The parameters \( \theta_Y = ((\mu_{0}, \mu_{1});(\sigma_{0}, \sigma_{1})) \) define the distribution of the data driven conditional on \( X \). For each posterior realization of the field \( Z = (X, U) \), an SEM estimator (empirical frequencies and moments) is used to estimate the hyper-parameters. When the sequence \( \theta^{[p]} \) approaches steady state – for example 1\% of the relative change in the values – we stop the procedure. Let us consider now the problem in estimating \( \theta_Z \) and \( \theta_Y \) separately.

6.1. Data driven parameter estimation

Let us suppose now, we observe a realization \((x, y)\) of the pairwise \((X, Y)\). In a hard classification the empirical moment estimator \( \theta_Y \) of \( \theta_Y \) corresponds to the maximum likelihood under conditional Gaussian laws assumption. When we use a fuzzy classification, it is enough to estimate the parameters dealing with hard classes. We generalize the method proposed in (Salzenstein and Pieczynski, 1997) applying the empirical moments to the hard pixels. Let be \( Q_p = \{ s \in S : X_s = p \}, p = 0, 1 \) the sets of pixels which belong to the hard classes. Our aim is to estimate the set of parameters \( \theta_Y = (\pi_0, \pi_1; \mu_0, \mu_1) \) where \( 1 < i,j < D \). Applying the empirical moment method on the hard pixels yields

\[
\hat{\mu}^{(i)}_{p} = \frac{\sum_{s \in Q_p} x_{s} \cdot \delta(x, p)}{\sum_{s \in Q_p} \delta(x, p)},
\]

\[
\hat{\pi}^{(i)}_{p} = \frac{\sum_{s \in Q_p} (y_{s}^{(i)} - \hat{\mu}^{(i)}_{p}) \cdot (y_{s}^{(i)} - \hat{\mu}^{(i)}_{p}) \delta(x, p)}{\sum_{s \in Q_p} \delta(x, p)}.
\]

6.2. Prior parameter estimation

Let us consider an hidden chain \( Z = (X, U) \) simulated by its posterior distribution, according to the procedure described in Section 5.1. The prior parameter \( \theta_Z \) corresponds to the initial and transition densities. They can be deduced from the joint density, as seen in Section 3. We consider two hypothesis: (i) the joint density is a non-parameterized density. (ii) It depends on a parameterized function \( \phi \).

(i) NP-NSFMC: We compute the empirical \( M^2 \times K^2 \) joint probabilities (45) according to different neighborhood configurations

\[
P[X_s \in I_i, X_{s+1} \in I_j, U_s = \lambda_p, U_{s+1} = \lambda_q] = \frac{1}{M^2} \cdot P[X_s \in I_i, X_{s+1} \in I_j, U_s = \lambda_p, U_{s+1} = \lambda_q].
\]

We deduce the joint density from these probabilities. For instance, when \( I_i = \left[ \frac{i}{M}, \frac{i+1}{M} \right] \) and \( I_j = \left[ \frac{j}{M}, \frac{j+1}{M} \right] \) corresponding to a discretization of \([0,1]\]

\[
p\left( \gamma_i = \frac{i}{M}, \gamma_j = \frac{j}{M}, \lambda_p, \lambda_q \right) \approx \frac{1}{M^2} \cdot P[X_s \in I_i, X_{s+1} \in I_j, U_s = \lambda_p, U_{s+1} = \lambda_q].
\]

(ii) P-NSFMC: when the joint density is parameterized, we have to estimate the quantities \( \pi_{0,0}^{[i]}, \pi_{0,1}^{[i]}, \pi_{1,1}^{[i]} \) and \( \pi_{ij} \) according to the empirical frequencies

\[
\hat{\pi}_{00}^{[i]} = \sum_{j=1}^{N-1} \frac{1}{N} \cdot \frac{1}{N} \cdot \sum_{k=1}^{N-1} \frac{(x_{s+k}^{(i)} = 0, j) \cdot (x_{s+k+1}^{(i)} = 0, j)}{N},
\]

\[
\hat{\pi}_{01}^{[i]} = \sum_{j=1}^{N-1} \frac{1}{N} \cdot \frac{1}{N} \cdot \sum_{k=1}^{N-1} \frac{(x_{s+k}^{(i)} = 0, j) \cdot (x_{s+k+1}^{(i)} = 1, j)}{N},
\]

\[
\hat{\pi}_{11}^{[i]} = \sum_{j=1}^{N-1} \frac{1}{N} \cdot \frac{1}{N} \cdot \sum_{k=1}^{N-1} \frac{(x_{s+k}^{(i)} = 1, j) \cdot (x_{s+k+1}^{(i)} = 1, j)}{N},
\]

\[
\hat{\pi}_{ij} = \sum_{j=1}^{N-1} \frac{1}{N} \cdot \frac{1}{N} \cdot \sum_{k=1}^{N-1} \frac{(x_{s+k}^{(i)} = i, j) \cdot (x_{s+k+1}^{(i)} = j)}{N},
\]

\[
\pi_{ij}^{[i]} ñ a_{ij} \text{ and } b_{ij} \text{ are deduced by conditions (15), (16), (20).}
\]

Fig. 1. (a) A non-stationary fuzzy Markov chain \( X \); (b) its related states \( U \); (c) its noisy version \( Y \).
7. Results on synthetic images

We simulated a non-stationary fuzzy Markov chain on \( M = 10 \) discrete fuzzy levels, with two homogeneous states (\( \text{Card}\, A = 2 \)) and \( r = 1 \). The variables \( X \) and \( U \) are represented in Fig. 1a and b. The class “0” (in black) of \( U \) corresponds to an hard-dominating area in \( X \) \( (\pi_{00}^{\text{H}} = \pi_{11}^{\text{H}} = 0.2) \), where as the class “1” (in white) corresponds to a fuzzy area in \( X \) \( (\pi_{00}^{\text{F}} = \pi_{11}^{\text{F}} = 0.05) \). A noisy version is represented in Fig. 1c. We give below the following corresponding prior and data driven parameters:

\[
\begin{align*}
\pi_{00}^{\text{F}} &= \pi_{11}^{\text{F}} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.05 \end{pmatrix}, \\
\pi_{00}^{\text{H}} &= \pi_{11}^{\text{H}} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \\
(P_{ij}) &= \begin{pmatrix} 0.4995 & 0.005 \\ 0.005 & 0.4995 \end{pmatrix}, \\
(\mu_0, \mu_1) &= (120, 142), \\
(\sigma_0, \sigma_1) &= (4, 4).
\end{align*}
\]

Fig. 2a and b gives the segmented fields \( X \) corresponding to the non-stationary (\( \text{Card}\, A = 2 \)) case when one considers respectively the parameterized (P-NSFMC with \( r = 1 \)) and non-parameterized (NP-NSFMC) approaches. Fig. 2c–e corresponds to the stationary algorithm (\( \text{Card}\, A = 1 \)) applied to the image, respectively by the means of a parameterized (P-SFMC), non-parameterized (NP-SFMC) fuzzy Markov chain. Another stationary method used is based on a fuzzy Markov field (SFMF), briefly presented in Section 2. For the non-stationary approaches with two states, the segmented field \( U \) is represented in Fig. 3a and b. Moreover, the estimated prior parameters are indicated in Table 1, where as the noise parameters for NSFMC, SFMC and SFMF are given, respectively, in Tables 2 and 3. Finally the error rates computed by (28) are given for all NP/P-(N)SFMC and SFMF procedures in Table 4. The stationary method based on FMC or FMF give higher rates of error than the non-stationary methods. The highest rate is given by the SFMF.
The SFMC are staying competitive but do not provide any accuracy information concerning the homogeneity of the image. Actually, the non-stationary procedure estimates correctly the homogeneous areas (see Fig. 3 a and b). Let us notice the method using a non-parameterized neighborhood density stay competitive facing the parameterized assumption. Further studies must be performed in order to measure the influence of the parameter \( r \).

8. Results on real images

We wish to identify different homogeneous regions inside an image. We processed here our images in the mono-spectral context. We present in Fig. 4 a and b two images of Oakland typically exhibiting a such situation.

![Figures](image1.png)

Fig. 4. (a, b) The observation; (c, d) segmented images using SFMC; (e, f) segmented images using SFMF.

**Table 4**

<table>
<thead>
<tr>
<th>Procedure</th>
<th>P-SFMC</th>
<th>NP-SFMC</th>
<th>P-NSFMC</th>
<th>NP-NSFMC</th>
<th>SFMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rates</td>
<td>5.85%</td>
<td>6.01%</td>
<td>5.18%</td>
<td>5.35%</td>
<td>6.79%</td>
</tr>
</tbody>
</table>

Fig. 4a contains a sea area and the city, which appear to be inhomogeneous on the picture, noticing that the distribution of this part of the image behaves differently from the distribution of the sea part. Fig. 4b contains a cloudy area and a town area. We processed first all images Fig. 4a and b using a stationary approach. The results are given in Fig. 4c–f, respectively by the FMC and the FMF algorithm. These methods do not provide the fine details and give comparable results. It is then necessary to take into account the stationary information, initializing the algorithm with more than two stationary states (\( K > 2 \)). The segmented fields \( X \) and \( U \) are represented, respectively, in Fig. 5a–d. In particular, the class “0” of \( U \) indicates the high density region i.e., the city part. Our procedure ensures a convergence of the stationary field towards \( K = 2 \) classes, which suits the initial hypothesis of both groundtruth models i.e., city/sea areas for Fig. 4a and city/cloud areas for Fig. 4b. The results given by the hidden realizations \( X \) are less convenient because the method tends to lose some details, concerning the lower homogeneous regions. In order to enrich the information providing by this field, we propose to combine the NP-FMC and NP-NSFMC approaches into an algorithm as follows:

- (i) Perform the NP-NSFMC method to the observed data.
- (ii) For each stationary state \( U_i, i = 1, 2, \ldots, K \), apply a stationary NP-FMC method. For each state \( i \), the data corresponding to \( U_j, j \neq i \) are processed as missing data: one has to suppress them in the Hilbert-Peano path (Salzenstein and Collet, 2006).

![Figures](image2.png)

Fig. 5. Segmented images of Fig. 4a and b, using a non-stationary method. (a, b) Realizations of \( X \); (c, d) associated \( U \) containing two states \( K = 2 \).
Consequently, we applied the method separately for each stationary parts of Fig. 4a and b. The segmented data corresponding are given in Fig. 6. Although the interpretation is more delicate, requiring separate graylevels/pictures, the final results provide more detailed groundtruth.

9. Conclusion

We presented in this paper a new fuzzy Markov chain model based on a non-stationary approach. On one hand we modeled the prior parameters of a stationary chain using a parameterized joint density defined on numbered sites. On the other hand, we used an intermediate field $U$ in order to govern the switching in the distribution of $X$. Here the classes in $U$ are discrete while the classes in $X$ are continuous. The proposed method merges the fuzzy processing technique and a recent technique that has been used to describe non-stationary images for hard classification. This model is more flexible than the single chain based procedure, in the following manner: (i) the pairwise $(X, U)$ is a Markov chain, but $X$ is not necessary Markovian; (ii) the stationary model is a particular case with one discrete state in $U$. The fuzzy context should be better adapted than the hard approach when the scene owns diffuse structure. In order to take into account the complexity of multiple real situations, it worthy to extend this model to the multi-sensors context and non-Gaussian data driven distributions.

References