



Hidden fuzzy Markov chain model with K discrete classes

Ahmed Gamal Eldin, Fabien Salzenstein, Christophe Collet

► **To cite this version:**

Ahmed Gamal Eldin, Fabien Salzenstein, Christophe Collet. Hidden fuzzy Markov chain model with K discrete classes. Information Sciences Signal Processing and their Applications (ISSPA), May 2010, Kuala Lumpur, Malaysia. 2010. <hal-00616372>

HAL Id: hal-00616372

<https://hal.archives-ouvertes.fr/hal-00616372>

Submitted on 22 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HIDDEN FUZZY MARKOV CHAIN MODEL WITH K DISCRETE CLASSES

Ahmed Gamal-Eldin^{1,3}, Fabien Salzenstein², Christophe Collet³

¹ INRIA Sophia Antipolis 2004 - France ² Laboratoire InESS, UMR CNRS 7163 Université de Strasbourg - France

³ Université Strasbourg, LSIIT - UMR CNRS 7005 - France

¹ agamalel@sophia.inria.fr, ² salzenst@iness.c-strasbourg.fr, ³ collet@lsiit.u-strasbg.fr

ABSTRACT

This paper deals with a new unsupervised fuzzy Bayesian segmentation method based on the hidden Markov chain model, in order to separate continuous from discrete components in the hidden data. We present a new F-HMC (fuzzy hidden Markov chain) related to three hard classes, based on a general extension of the previously algorithms proposed. For a given observation, the hidden variable owns a density according to a measure containing Dirac and Lebesgue components. We have performed our approach in the multispectral context. The hyper-parameters are estimated using a Stochastic Expectation Maximization (SEM) algorithm. We present synthetic simulations and also segmentation results related to real multi-band data.

Index terms: Hidden fuzzy Markov chain, multispectral image segmentation, parameterized joint density.

1. INTRODUCTION

In many applications a classification which labels the original scene into finite discrete sets artificially eliminates the imprecision related to the resolution level. Starting from the fuzzy set theory [1], some authors combined the imprecision of the hidden data and the uncertainty of the observed data using probabilistic methods. A contextual approach has been proposed [2], using a measure of density including components related to the discrete information (Dirac function) and components related to the imprecise information (Lebesgue measure). Hidden Fuzzy Markov field (F-HMF) [3, 4, 5] or fuzzy Markov chain (F-HMC) [6, 7, 8] model have been then developed. However, due to the complexity of these models and the intrinsic properties of the processed images, the previously methods focused on two hard classes segmentation. In [9], authors have proposed a histogram based method for mono-band data (which leads to an underdetermined problem) and did not provide a specific measure in order to separate the hard from the fuzzy components. In [10] the authors combine discrete and continuous classes, but do not rely the continuous and discrete information. We propose here a multi-band extension of the F-HMC model, dealing with three hard classes. Both prior and data driven parameters are estimated using a SEM algorithm. We perform an MPM

procedure (Maximum Posterior Marginal) [11] as a final decision process using an euclidian distance.

2. FUZZY MARKOV CHAIN MODEL (F-HMC)

2.1. Introduction

The fuzzy segmentation problem consists in estimating the hidden realization $x = (x_s)_{1 \leq s \leq N}$, for a given set of D observations $\mathbf{y} = \{\mathbf{y}_s\} \in \mathbb{R}^D$, where

$$x_s = (\varepsilon_0(s), \dots, \varepsilon_{K-1}(s))$$

Each $\varepsilon_i(s)$ represents the contribution of the class ω_i which belongs to a finite discrete set $\Omega = \{\omega_0, \dots, \omega_{K-1}\}$ of K hard classes ω_i . The normalization condition yields:

$$\varepsilon_0(s) + \varepsilon_1(s) + \dots + \varepsilon_{K-1}(s) = 1 \quad (1)$$

For instance a site belonging to the class ω_i is associated locally with $x_s = \{0, 0, \dots, 1, 0, 0, \dots, 0\}$. According to our assumption, the variable X is a Markov chain called "fuzzy Markov chain" (FMC) with respect to a measure ν^N defined in section 2.2.

$$\pi(x) = p(x_1)p(x_2|x_1)\dots p(x_N|x_{N-1}) \quad (2)$$

Thus, our model is able to classify the data containing simultaneously discrete and continuous components. In particular, we wish to identify better the areas at the intersection between two or more homogeneous classes, called 'hard' pixels represented by the variables ω_i .

2.2. Associated measure of the local density

The distribution at each random variable X_s is given by a density h_s with respect to a measure ν including discrete components (Dirac functions $\delta_0, \delta_1, \delta_2$) and continuous components, which represent the mixture of the 'hard' classes. The figure 1 represents the 3D space associated with three hard classes. The hyperplane (triangle) corresponds to the relation (1), where $\varepsilon_i(s) \in [0, 1]$. The extremities of the triangle represent the hard components, each segment a, b or c represents the variable produced by a mixture of two hard classes and the surface of the triangle represents any variable produced by a mixture of three hard classes. In a particular context of our model, we use the following measure ν :

$$\nu = \delta_{\omega_0} + \delta_{\omega_1} + \delta_{\omega_2} + \mu_a + \mu_b + \mu_c \quad (3)$$

The authors would like to thank the laboratory INRIA Sophia-Antipolis (France) for the financial support.

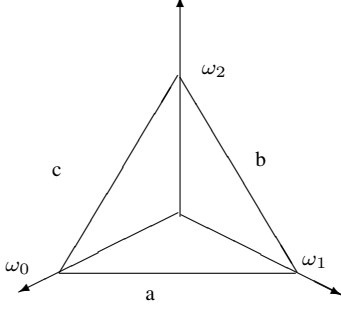


Fig. 1: Measure density for the fuzzy Markov model with three hard classes ω_0 , ω_1 and ω_2 which are presented by three δ_i and the Lebesgue measure for the three segments a,b,c.

Where μ_a, μ_b, μ_c are respectively the Lebesgue measure on $[0, 1]$ i.e, related to the points located on the segments a, b and c. According to Eq. (3), we do not take into account the sites belonging to a mixture of three hard classes (the surface of the hyperplane). We made the realistic assumption, that in a three classes images, the mixed pixels are only between a pair of homogeneous classes simultaneously. In the following sections, let us call Δ the triangle excluding its surface.

2.3. Joint and initial densities

To define the prior density $\pi(x)$ of the variable X , we need the density $p(x_1)$ of the initial distribution, and the transition densities $p(x_s|x_{s-1})_{1 \leq s \leq N}$:

$$p(x_1, x_2, \dots, x_N) = p(x_1) \cdot p(x_2|x_1) \dots p(x_N|x_{N-1})$$

When the chain is stationary, all prior distributions can be deduced from a prior joint density. This density $g(x_s, x_{s+1})$ is defined on the pairwise $(x_s, x_{s+1}) \in \Delta^2$ as follows:

$$g(\omega_i, \omega_j) = \begin{cases} \alpha_{ij}, & \text{for } i = j \\ \gamma_{ij}, & \text{otherwise.} \end{cases} \quad (4)$$

if both states are *hard*. $(i, j) \in \{0, 1, 2\}^2$. We set $\gamma_{ij} = 0$, which means that there is no abrupt transition between the hard classes. If at least one *fuzzy* membership is present:

$$g(x_s, x_{s+1}) = \beta \cdot \phi(x_s, x_{s+1}) \quad (5)$$

where, $\phi(x_s, x_{s+1}) = \left(\frac{2 - \|x_s - x_{s+1}\|}{2} \right)^r$ and $\|x_s - x_{s+1}\|$ being the quadratic norm between two points. This expression is adapted for the three classes model and it generalizes the model introduced in [6]. The parameter r has an effect on the image homogeneity: when it increases, the probability of having two similar neighbored pixels increases. Typically $r \in [1, 50]$. Moreover, the normalization condition of the joint density is computed by an integration along Δ as follows:

$$\oint_{\Delta} \oint_{\Delta} g(x_t, x_{t+1}) d(\nu \otimes \nu)(x_t, x_{t+1}) = 1 \quad (6)$$

Finally, the measure ν and the parameterized form of the joint density yields:

$$\sum_{i=0}^2 \alpha_{ii} + (\beta_a + \beta_b + \beta_c) \int_0^1 \int_0^1 (1 - |u - v|)^r du dv + 2(\beta_a + \beta_b + \beta_c) \cdot \int_0^1 (1 - |\varepsilon|)^r + (1 - |1 - \varepsilon|)^r d\varepsilon$$

where β_i corresponds to the relation (5) for each segment a, b, c. In a stationary context, the joint density $g(x_s, x_{s+1})$ provides the initial density $h(x_s)$: $h(x) = \oint_{\Delta} g(x, y) d\nu(y)$ All transition densities are then simply defined by the Bayes formulae.

2.4. Multi-spectral observation

The detection problem is the following one: one has to estimate hidden data $X = x \in \Delta^N \subset R^{(K-1)N}$ (Δ being an hyperplane of the set R^{KN}) given an observed multi-spectral realization $\mathbf{Y} = \mathbf{y} \in R^{DN}$. In order to avoid a miss classification, we need to process at least $D = K - 1$ bands. Let be D realizations $\left\{ y^{(i)} = (y_s^{(i)})_{s \in S} \right\}_{1 \leq i \leq D}$ of the random vector $\mathbf{Y} = (Y^{(1)}, Y^{(2)}, \dots, Y^{(D)}) \in R^{ND}$ provided by different sensors, we wish to estimate one realization $X = x \in \Delta^N \subset R^{3N}$. The variables $\mathbf{Y}_s|_{s \in S}$ are supposed to be independent conditionally on X , and $P[\mathbf{Y}_s|X] = P[\mathbf{Y}_s|X_s]$. In a spectral correlated context, the distribution of $\mathbf{y}_s = [y_s^{(1)}, y_s^{(2)}, \dots, y_s^{(D)}]^t$ according to $X_s = x_s \in \Delta$ is Gaussian:

$$f_{x_s}(\mathbf{y}_s) = \frac{1}{2\pi^{D/2} (\det \Gamma_{x_s})^{1/2}} \exp\left(\frac{-1}{2} (\mathbf{y}_s - \mu_{x_s})^t \Gamma_{x_s}^{-1} (\mathbf{y}_s - \mu_{x_s})\right) \quad (7)$$

where $\mu_{x_s} = [\mu_{x_s}^{(1)}, \dots, \mu_{x_s}^{(D)}]^T$ and $\Gamma_{x_s} \in R^D \times R^D$ respectively define a mean vector and variance-covariance matrix, at each fuzzy/hard site. Let be (μ_0, μ_1, μ_2) and $(\Gamma_0, \Gamma_1, \Gamma_2)$ the mean vectors and variance-covariance matrix related to the hard classes $\omega_0, \omega_1, \omega_2$. For each fuzzy site $X_s = (\varepsilon_0(s), 1 - \varepsilon_0(s), 0)$ defined on the segment a, by a mixture of classes ω_0 and ω_1 , the related mean vector and matrix μ_{x_s} et Γ_{x_s} are written:

$$\begin{aligned} \mu_{x_s} &= \varepsilon_0(s) \cdot \mu_0 + (1 - \varepsilon_0(s)) \cdot \mu_1 \\ \Gamma_{x_s} &= \varepsilon_0(s)^2 \cdot \Gamma_0 + (1 - \varepsilon_0(s))^2 \cdot \Gamma_1 \end{aligned} \quad (8)$$

The average and variance-covariance matrixes related to the segment b and c are defined in the same manner. It is known that the density $f_{X, \mathbf{Y}}(x, \mathbf{y})$ of the distribution of (X, \mathbf{Y}) is Markovian, with respect to a measure $\nu^N \otimes (\mu^N \otimes \mu^N \otimes \dots \otimes \mu^N)$. In particular, the *posterior* variable X conditional on \mathbf{Y} is Markovian. Thus, one can process the posterior realizations of the hidden variable X , which allows us to perform the segmentation in an unsupervised way (see section 4).

3. SEGMENTATION RULE

The segmentation will be performed by the MPM criterion [11], adapted to the fuzzy context: a loss function $L(x, \hat{x}) = \sum_t L_s(x_s, \hat{x}_s)$ is introduced, where $L_s(\hat{x}_s, x_s)$ models the severity of attributing the value \hat{x}_s instead of a

true one x_s at each site s . We chose the *euclidian distance* i.e, the usual quadratic norm $L_s(x_s, \hat{x}_s) = \|x_s - \hat{x}_s\|$ between two points x_s and \hat{x}_s belonging to the triangle Δ . A decision \hat{d}_s will then involve minimizing the conditional risk $E[L_s(X_s, \hat{X}_s)|Y = y]$ at each location s .

$$\hat{x}_s^{opt} = \arg \min_{\hat{X}_s = \hat{x}_s} E[L_s(X_s, \hat{X}_s)|Y = y] \quad (9)$$

Let us express precisely the quantity (9)

$$\begin{aligned} E[L_s(X_s, \hat{X}_s)|Y = y] &= \oint_{\Delta} L_s(x_s, \hat{X}_s) p^y(x_s) d\nu(x_s) \\ &= \sum_{i=0}^{K-1} p_{x_s}^y(\omega_i) \cdot L_s(\omega_i, \hat{x}_s) \\ &\quad + \oint_{t \in \Delta \setminus \{\omega_0, \omega_1, \omega_2\}} p_{x_s}^y(x_s) L_s(x_s, \hat{X}_s) ds \end{aligned} \quad (10)$$

The minimization of (10) requires the posterior density $p_{X_s}^y$, which is performed by the forward/backward procedure. The forward and backward densities $\alpha_s(x_s), \beta_s(x_s)$ are deduced by:

$$\begin{aligned} \alpha_s(x_s) &= p(x_s, \mathbf{y}_1, \dots, \mathbf{y}_s) \\ \beta_s(x_s) &= \frac{p(\mathbf{y}_{s+1}, \dots, \mathbf{y}_N | x_s)}{p(\mathbf{y}_{s+1}, \dots, \mathbf{y}_N | \mathbf{y}_1, \dots, \mathbf{y}_s)} \end{aligned}$$

The recurrence formulae providing these quantities, are analogous to the hard segmentation processing:

$$\begin{aligned} \alpha_s(x_s) &= p(\mathbf{y}_s | x_s) \oint_{\Delta} \alpha_{s-1}(u) \cdot p(x_s | u) d\nu(u) \\ \beta_s(x_s) &= \oint_{\Delta} \beta_{s+1}(u) p(\mathbf{y}_{s+1} | u) p(u | x_s) d\nu(u) \end{aligned}$$

This leads to the posterior density:

$$p(x_s | Y) = \alpha_s(x_s) \beta_s(x_s) \quad (11)$$

4. HYPERPARAMETER ESTIMATION

The Markov chain requires two sets of parameters that needs to be estimated: the prior densities and the data driven parameters. We process the classic Stochastic Expectation Maximization (SEM) algorithm because of its simplicity in both mathematical derivation and implementation. It also prevents local minima more efficiently than the ordinary EM algorithm.

4.1. Prior parameters

The prior parameter corresponds to the initial and transition densities. For a stationary chain, they can be deduced from the joint density defined in section 2.3. Let us suppose we observe a posterior realization x of X conditioned on the observation \mathbf{Y} . The initial density yields:

$$\pi_i^{[q]} = \frac{1}{N} \sum_{s=1}^N \delta(x_1^{[q]}(s), \omega_i) \quad (12)$$

The transition matrix is given by:

$$a_{ij}^{[q]} = \frac{\sum_{t=1}^{\tau} \sum_{s=2}^N \delta(x_{s-1}^{[q]}(t), \omega_i) \cdot \delta(x_s^{[q]}(t), \omega_j)}{\sum_{t=1}^{\tau} \sum_{s=2}^N \delta(x_{s-1}^{[q]}(t), \omega_i)} \quad (13)$$

where i and j represents all existing classes, $[q]$ represent the iteration number in the SEM algorithm and t represent the number of simulations, since we use a stochastic method.

4.2. Data driven parameters

Let us suppose now, we observe a realization (x, \mathbf{y}) of the pairwise (X, \mathbf{Y}) . The average and variance-covariance matrixes related to the fuzzy components depending linearly on the hard classes, it is enough to estimate the parameters dealing with hard classes. Let $Q_p = \{s \in S / X_s = p\}$, $p = 0, 1, 2$ be the sets of pixels which belong to the hard classes. Our aim is to estimate the set of parameters $(\mu_p, \Gamma_p)_{p=0,1,2} = (\mu_p^{(i)}, \Gamma_p^{(i,j)})_{p=0,1,2}$, where $1 < i, j < D$. Applying the empirical moment method on the hard pixels yields:

$$\hat{\mu}_p^{(i)} = \frac{\sum_{s \in Q_p} y_s^{(i)} \cdot \delta(x_s, p)}{\sum_{s \in Q_p} \delta(x_s, p)} \quad (14)$$

$$\hat{\Gamma}_p^{(i,j)} = \frac{\sum_{s \in Q_p} (y_s^{(i)} - \hat{\mu}_p^{(i)}) \cdot (y_s^{(j)} - \hat{\mu}_p^{(j)}) \delta(x_s, p)}{\sum_{s \in Q_p} \delta(x_s, p)} \quad (15)$$

5. RESULTS AND DISCUSSIONS

In figure (2), we present a simulation of a Markov chain containing three hard classes. The image (2)(a) is provided by, $\alpha_{00} = \alpha_{11} = \alpha_{22} = 0.3$; $\beta_a = \beta_b = \beta_c =$ all have the same value, $r = 35$ (to insure an homogeneous field) with 5 fuzzy levels. In (2)(b) all parameters remain the same except $\alpha_{ii} = 0.2$, for $i = 0, 1, 2$. The relative values of α_{ii} between both images explain the fuzzy nature of (2)(b). Moreover, a color representation seems more appropriate than a gray level one, where the hard classes ω_0, ω_1 and ω_2 are respectively represented in red, green and blue. The mixture of classes are intermediate colors (see fig. (2)(c)). Figure 3 (a) presents a real three-band (RGB colored) image. The segmentation results are shown on Figure 3 (b). In particular, the tree (in red) is processed as a pure hard class (the final red area does not contain any fuzzy element), while the segmentation provides a fuzzy transition between the sky (blue hard class) and the clouds (green hard class) of the background. Thus our model is robust to the nature (discrete or fuzzy) of a class. Moreover, less than 10 iterations were necessary to insure the convergence. In 3(c) we present the results in gray levels and have compared it (3(d)) with a two hard classes based F-HMF model previously proposed [6].

6. CONCLUSION AND FUTURE WORK

In this work we have proposed a new F-HMC three hard classes based model, in order to classify continuous and

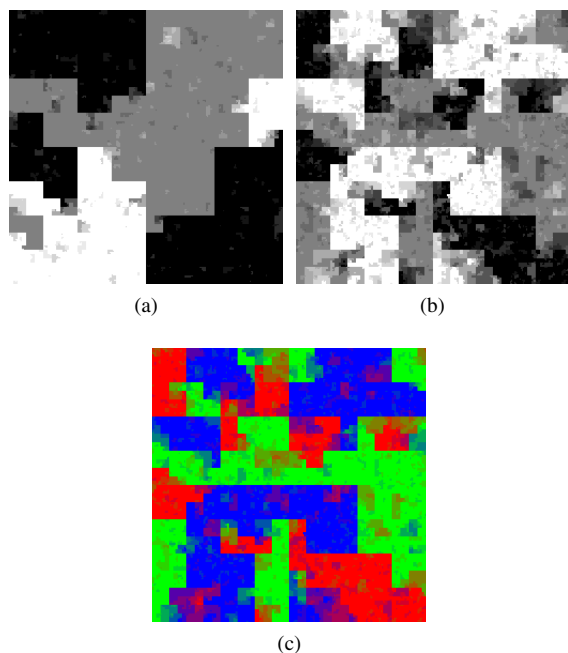


Fig. 2: (a) A simulation with almost all hard pixels. (b) A simulation with a lot of fuzzy pixels. (c) The simulation in (b) in color scheme.

discrete information. The great flexibility of the Markov chains allows us to perform either 1D, 2D or 3D signal. Although it does not include the mixtures between three hard classes, we believe that our method offers potentialities in image processing field, because a such hypothesis seems to be realistic. As a future work, we plane to develop other noise model like the Generalized Gaussian Mixture [12] and also include more complex structure as the pairwise Markov chain [8] in order to process spatially dependent observations.

7. REFERENCES

- [1] L.A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, pp. 338–352, 1965.
- [2] H. Caillol, A. Hillion, and W. Pieczynski, "Fuzzy Markov random fields and unsupervised image segmentation," *IEEE Trans. Geosci. Remote Sensing*, vol. 4, pp. 801–810, 1993.
- [3] J.T. Kent and K.V. Mardia, "Spatial classification using fuzzy membership," *IEEE, Trans. On Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, pp. 659–671, 1988.
- [4] F. Salzenstein and W. Pieczynski, "Parameter estimation in hidden fuzzy Markovian fields and image segmentation," *Graphical Models and image processing*, vol. 59, no. 4, pp. 205–220, 1997.
- [5] S. Ruan, B. Moretti, J. Fadili, and D. Bloyet, "Fuzzy Markovian segmentation in application of magnetic resonance images," *Computer Vision and Image Understanding*, vol. 85, pp. 54–69, 2002.

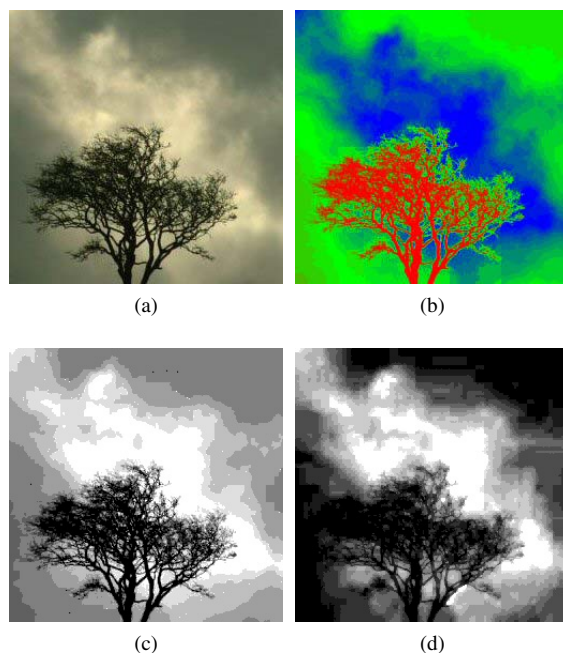


Fig. 3: (a) Real image of a tree, in the background the sky and clouds. (b) The segmentation result of image (a) in our color scheme. (c) Segmentation result in gray levels, with reduced number of fuzzy pixels. (d) Segmentation using a two hard classes model.

- [6] F. Salzenstein and C. Collet, "Fuzzy Markov random fields versus chains for multispectral image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1753 – 1767, November 2006.
- [7] F. Salzenstein, Ch Collet, S. Lecam, and M. Hatt, "Non stationary fuzzy markov chains," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2201–2208, 2007.
- [8] S. Lecam, F. Salzenstein, and C. Collet, "Fuzzy pairwise markov chain to segment correlated noisy data," *Signal Processing*, vol. 88, no. 10, pp. 2526–2541, 2008.
- [9] C. Carincotte, *Markovian Segmentation of fuzzy image, application in teledetection radar image change*, Ph.D. thesis, Université d'Aix-Marseille, 2005.
- [10] P. Boutheymy, Ch. Hardouin, G. Piriou, and J.F. Yao, "Mixed-state automodels and motion texture modeling," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 3, pp. 387–402, 2006.
- [11] J. Maroquin, S.Mitte, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of the american statistical association*, vol. 82, pp. 76–89, 1987.
- [12] F. Flitti, *Segmentation markovienne floue d'images, application en détection de changements entre images radar*, Ph.D. thesis, Université de Strasbourg, 2005.