

Unsupervised Segmentation of Random Discrete Data Hidden With Switching Noise Distributions

Mohamed El Yazid Boudaren, Emmanuel Monfrini, and Wojciech Pieczynski

Abstract—Hidden Markov models are very robust and have been widely used in a wide range of application fields; however, they can prove some limitations for data restoration under some complex situations. These latter include cases when the data to be recovered are nonstationary. The recent triplet Markov models have overcome such difficulty thanks to their rich formalism, that allows considering more complex data structures while keeping the computational complexity of the different algorithms linear to the data size. In this letter, we propose a new triplet Markov chain that allows the unsupervised restoration of random discrete data hidden with switching noise distributions. We also provide genuine parameters estimation and MPM restoration algorithms. The new model is validated through experiments conducted on synthetic data and on real images, whose results show its interest with respect to the standard hidden Markov chain.

Index Terms—Hidden Markov chains, switching noise distributions, triplet Markov chains, unsupervised segmentation.

I. INTRODUCTION

HIDDEN Markov chains (HMCs) have been extensively used to solve a wide range of inverse problems occurring in various fields including image and signal processing. Let us mention [1]–[3] as pioneering papers. When the hidden data of interest can be modeled through a finite Markov chain, and when the noise form is simple enough, these models provide satisfactory results and the hidden data can be recovered using Bayesian restoration techniques like Maximum A Posteriori (MAP) or Maximum Posterior Mode (MPM) [1], [2].

Let $X = (X_n)_{n=1}^N$ be an unobservable process that takes its values from a finite set of classes $\Omega = \{\omega_1, \dots, \omega_K\}$, and let $Y = (Y_n)_{n=1}^N$, with $Y_n \in \mathbb{R}$, be an observable process that can be seen as a noisy version of X . According to the HMC formalism, the joint probability of $Z = (X, Y)$ is given by:

$$p(x, y) = p(x_1) p(y_1|x_1) \prod_{n=2}^N p(x_n|x_{n-1}) p(y_n|x_n). \quad (1)$$

Manuscript received January 11, 2012; revised May 04, 2012; accepted June 28, 2012. Date of publication July 19, 2012; date of current version July 27, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Henry Leung.

M. E. Y. Boudaren is with the Ecole Militaire Polytechnique, BP 17, Bordj EL Bahri, Algiers 16111, Algeria and also with the Institut Mines-Telecom, TELECOM SudParis, Department of CITI, CNRS UMR 5157, Evry 91011, France and the Université des Sciences et de la Technologie Houari Boumediene, BP 32, Algiers 16111, Algeria (e-mail: Mohamed.Boudaren@it-sudparis.eu).

E. Monfrini and W. Pieczynski are with the CITI Department, Institut Mines-Telecom, Institut TELECOM SudParis, CNRS UMR 5157, Evry 91011, France (e-mail: Emmanuel.Monfrini@it-sudparis.eu; Wojciech.Pieczynski@it-sudparis.eu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2012.2209639

The Bayesian restoration can then be achieved thanks to the possibility of recursive computations of forward probabilities $\alpha_n(x_n) = p(y_1, \dots, y_n, x_n)$ and backward probabilities $\beta_n(x_n) = p(y_{n+1}, \dots, y_N|x_n)$. Moreover, these techniques may be applied in the unsupervised context thanks to parameters estimation algorithms such as Expectation-Maximization (EM) [2] and Iterative Conditional Estimation (ICE) [4].

Recently, HMCs have been generalized to pairwise Markov chains (PMCs) [5] and triplet Markov chains (TMCs) [6].

In PMC, one only assumes the Markovianity of Z . Hence, the process X is not necessarily Markovian.

In TMC, we introduce an additional underlying process U and only the triplet $T = (U, X, Y)$ is assumed to be Markovian. The new auxiliary process may have different meanings. In [7], authors used U to consider nonstationary aspect of the hidden process X . Furthermore, an interesting link has been established between TMCs and Dempster-Shafer theory of evidence to model multisensor and nonstationary data in the Markovian context [6].

Let us cite some other works that dealt with nonstationary data based on the introduction of the “time duration functions” [8]. Parameters can then be estimated via a Monte Carlo Markov chain (MCMC) approach [9]. Further extensions to hidden semi-Markov chains have also been proposed [10].

The aim of this letter is to propose a new TMC model including unknown stochastic switches of the noise distributions $p(y_n|x_n)$ in a similar manner as the TMC proposed in [7] to model the random switches of $p(x_n|x_{n-1})$. Such a model may be applied in image processing to take light environment or presence of shadow in the image into account. In finance, this model may be applied to model the fact that financial returns behave in a different way in crisis time.

The remainder of the letter is organized as follows: Section II describes the new model and provides its corresponding restoration and parameters estimation procedures. In Section III, the model validity is assessed through experiments carried on synthetic data and real images. Finally, concluding remarks and future improvements end the letter.

II. JUMPING NOISE- HIDDEN MARKOV CHAINS

In this section, we describe the proposed TMC model and its corresponding MPM restoration and parameters estimation algorithms. The proposed model will be called all along the remainder of the letter ‘jumping noise – hidden Markov chain’ abbreviated JN-HMC in contrast to the standard HMC.

A. Jumping Noise- Hidden Markov Chain Definition

Let $X = (X_n)_{n=1}^N$ be a hidden process that takes its values from a finite set of classes $\Omega = \{\omega_1, \dots, \omega_K\}$ and that is to

be estimated from an observable process $Y = (Y_n)_{n=1}^N$, with $Y_n \in \mathbb{R}$. The couple $Z = (X, Y)$ is said to be a JN-HMC if there exists an auxiliary process $U = (U_n)_{n=1}^N$ that takes its values from a finite set of auxiliary classes $\Lambda = \{\lambda_1, \dots, \lambda_M\}$ such that the triplet $T = (U, X, Y)$ is a Markov chain. Its distribution is then given by:

$$p(t) = p(t_1) \prod_{n=2}^N p(t_n | t_{n-1}) \quad (2)$$

where

$$\begin{cases} p(t_1) = p(u_1) p(x_1) p(y_1 | u_1, x_1) \\ p(t_n | t_{n-1}) = p(u_n | u_{n-1}) p(x_n | x_{n-1}) p(y_n | u_n, x_n) \end{cases} \quad (3)$$

The MPM restoration is then workable and its complexity is linear with the data size N .

B. Parameters of a Gaussian JN-HMC

Let T be a JN-HMC defined with (2) and (3). It will be said “Gaussian” if the distributions $p(y_n | u_n, x_n)$ are Gaussian. The distribution of $T = (U, Z)$ is then defined by the following parameters:

- The initial distributions $p(u_1)$ and $p(x_1)$ are given by vectors ϖ and π , respectively. Hence, $p(u_1 = \lambda_m) = \varpi(m)$ and $p(x_1 = \omega_k) = \pi(k)$;
- The stationary transition distributions $p(u_n | u_{n-1})$ and $p(x_n | x_{n-1})$ are given by the transition matrices Q and A respectively. Hence, $p(u_n = \lambda_m, | u_{n-1} = \lambda_{m'}) = q(m, m')$ and $p(x_n = \omega_k, | x_{n-1} = \omega_{k'}) = a(k, k')$, respectively.
- Since we have K hidden classes and M auxiliary cases, we need $M \times K$ Gaussian functions, and thus $M \times K$ means and $M \times K$ standard deviations. The noise distribution $p(y_n | u_n = \lambda_m, x_n = \omega_k)$ is then defined as long as the mean $\mu_{k,m}$ and standard deviation $\sigma_{k,m}$ are given.

The parameters Θ of the model being defined, let us notice that we can sample realizations of T in the classical manner: (i) sample a realization of the hidden process X by random drawings from π and A respectively; (ii) sample a realization of the auxiliary process U according to ϖ and Q respectively; and (iii) finally, sample a realization of the observed process Y by random drawing from the Gaussian density function corresponding to the hidden and auxiliary process.

C. MPM Restoration of a JN-HMC

Let $Y = y$ and let the model parameters be assumed to be known. We can achieve the MPM restoration to estimate the realizations of the hidden processes $X = \hat{x}$ and $U = \hat{u}$ in the following way. The MPM estimator is given by the formula:

$$[\hat{x} = \hat{s}_{MPM}(y)] \Leftrightarrow \left[\hat{x}_n = \underset{\omega}{\operatorname{argmax}} p(x_n = \omega | y) \right]. \quad (4)$$

Let us define the following generalized Forward functions $\alpha_n(x_n, u_n) = p(y_1, \dots, y_n, x_n, u_n)$ and Backward functions

$\beta_n(x_n) = p(y_{n+1}, \dots, y_N | x_n, u_n)$, that can be computed iteratively as follows:

$$\begin{cases} \alpha_1(x_1, u_1) = p(u_1) p(x_1) p(y_1 | u_1, x_1); \\ \alpha_n(x_n, u_n) = \sum_{x_{n-1}, u_{n-1}} \alpha_{n-1}(x_{n-1}, u_{n-1}) p(u_n | u_{n-1}) \\ \quad \cdot p(x_n | x_{n-1}) p(y_n | u_n, x_n) \end{cases} \quad (5)$$

$$\begin{cases} \beta_N(x_N, u_N) = 1; \\ \beta_n(x_n, u_n) = \sum_{x_{n+1}, u_{n+1}} \beta_{n+1}(x_{n+1}, u_{n+1}) p(u_{n+1} | u_n) \\ \quad \cdot p(x_{n+1} | x_n) p(y_{n+1} | u_{n+1}, x_{n+1}). \end{cases} \quad (6)$$

The posterior distributions $p(x_n, u_n | y)$ are then computable according to the following formula:

$$p(x_n, u_n | y) \propto \alpha_n(x_n, u_n) \beta_n(x_n, u_n), \quad (7)$$

which gives the posterior margins allowing the MPM estimation of X and U .

$$p(x_n | y) = \sum_{u_n} p(x_n, u_n | y) \quad (8)$$

$$p(u_n | y) = \sum_{x_n} p(x_n, u_n | y). \quad (9)$$

D. Parameters Estimation

When the model parameters are unknown, we propose to estimate these latter using an adapted version of the so called EM algorithm. For this purpose, let us first define the following posterior probabilities:

$$\begin{aligned} \psi_n(x_n, u_n, x_{n+1}, u_{n+1}) &= p(x_n, u_n, x_{n+1}, u_{n+1} | y) \\ &\propto \alpha_n(x_n, u_n) p(u_{n+1} | u_n) \\ &\quad \cdot p(x_{n+1} | x_n) \beta_n(x_{n+1}, u_{n+1}) \end{aligned} \quad (10)$$

$$\xi_n(x_n, u_n) = p(x_n, u_n | y). \quad (11)$$

The parameters estimation procedure is achieved in the following iterative way:

- Consider an initial parameters set $\Theta^0 = (\varpi^0, \pi^0, Q^0, A^0, (\mu_{k,m}, \sigma_{k,m})_{k,m}^0)$;
- For each iteration $i + 1$:
 - E-Step: evaluate the functions ψ_n^i and ξ_n^i using Θ^i .
 - M-Step: Derive the parameters set Θ^{i+1} as follows:

$$\varpi(m) = \sum_{x_n} \xi_1^i(x_n, u_n = \lambda_m) \quad (12)$$

$$\pi(k) = \sum_{u_n} \xi_1^i(x_n = \omega_k, u_n) \quad (13)$$

$$q_{m,m'}^{i+1} = \frac{\sum_{n=1}^{N-1} \sum_{x_n, x_{n+1}, u_n = \lambda_m, u_{n+1} = \lambda_{m'}} \psi_n^i(x_n, u_n, x_{n+1}, u_{n+1})}{\sum_{n=1}^{N-1} \sum_{x_n, u_n = \lambda_m} \xi_n^i(x_n, u_n)} \quad (14)$$

$$a_{k,k'}^{i+1} = \frac{\sum_{n=1}^{N-1} \sum_{u_n=\omega_k, x_{n+1}=\omega_{k'}} \psi_n^i(x_n, u_n, x_{n+1}, u_{n+1})}{\sum_{n=1}^{N-1} \sum_{u_n=\omega_k} \xi_n^i(x_n, u_n)} \quad (15)$$

$$\mu_{k,m}^{i+1} = \frac{\sum_{n=1}^N \sum_{x_n=\omega_k, u_n=\lambda_m} \xi_n^i(x_n, u_n) y_n}{\sum_{n=1}^N \sum_{x_n=\omega_k, u_n=\lambda_m} \xi_n^i(x_n, u_n)} \quad (16)$$

$$\sigma_{k,m}^{i+1} = \frac{\sum_{n=1}^N \sum_{x_n=\omega_k, u_n=\lambda_m} \xi_n^i(x_n, u_n) (y_n - \mu_{k,m}^{i+1})^2}{\sum_{n=1}^N \sum_{x_n=\omega_k, u_n=\lambda_m} \xi_n^i(x_n, u_n)}. \quad (17)$$

III. EXPERIMENTAL RESULTS

To assess the effectiveness of our model, we carried out three series of experiments. The first experiments set is concerned with synthetic data, whereas the last two experiments sets deal with images. To enable our “chain” model to be applied on images, these latter are converted into and from one-dimensional signals via the Hilbert-Peano scan as done in [5]. For all experiments, models parameters are initialized using a coarse K-Means clustering, which usually provides suitable initial parameters set. Notice that initialization step is crucial and a bad choice of initial parameters may lead to poor results.

A. Unsupervised Segmentation of Synthetic Data

Here we consider data sampled according to both HMC and JN-HMC formalisms.

- Experiment 1: Let $Z = (X, Y)$ be a standard HMC with $Z = (Z_n)_{n=1}^N$ where $N = 15000$, X_n takes its values from $\Omega = \{\omega_1, \omega_2\}$ and Y_n takes its values from \mathbb{R} . The first realization of the hidden process is sampled uniformly by a random draw and the next realizations are sampled by random draws according to the transition matrix $A = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}$. On the other hand, the realization of Y_n is sampled according to the distributions $p(y_n|\omega_1)$ and $p(y_n|\omega_2)$ given by the Gaussian densities whose standard deviations are all equal to 1 and whose means are equal to 0 and 1 respectively.
- Experiment 2: Let $T = (U, X, Y)$ be a stationary TMC with $T = (T_n)_{n=1}^N$ where $N = 15000$, U_n takes its values from $\Lambda = \{\lambda_1, \lambda_2\}$, X_n takes its values from $\Omega = \{\omega_1, \omega_2\}$ and Y_n takes its values from \mathbb{R} . The first realization of the hidden process is sampled uniformly by a random draw and the next realizations are sampled by random draws according to the transition matrix $A = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}$. Similarly, the realizations u_n are sampled according to the transition matrix $Q = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$. Finally, the realization of Y_n is sampled according to the distributions $p(y_n|\omega_1, \lambda_1)$, $p(y_n|\omega_1, \lambda_2)$, $p(y_n|\omega_2, \lambda_1)$ and $p(y_n|\omega_2, \lambda_2)$ given by the Gaussian densities whose standard deviations are all

TABLE I
SEGMENTATION ERROR RATIOS OF EXPERIMENTS SET 1 (%)

Experiment	$\tau_{K-Means}(\hat{x})$	$\tau_{HMC}(\hat{x})$	$\tau_{JN-HMC}(\hat{x})$	$\tau_{JN-HMC}(\hat{u})$
1	30.9	5.4	5.4	-
2	18.8	4.2	2.0	4.2

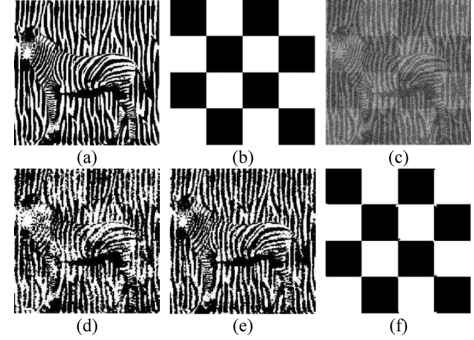


Fig. 1. Zebra image restoration according to HMC and JN-HMC formalisms. (a) Initial class-image $X = x$. (b) $U = u$ corresponding to $s = 64 \times 64$. (c) Noised image $Y = y$. (d) Image restoration considering HMC, error rate = 15.8%. (e) Image restoration considering JN-HMC, error rate = 8.5%. (f) Restoration of the underlying process, error rate = 0.1%.

equal to 1 and whose means are equal to 0, 1, 2 and 3 respectively.

For both experiments, MPM segmentation was performed according to K-Means, standard HMC ($K = 2$) and JN-HMC ($K = 2$ and $M = 2$). Average segmentation results, computed on 100 data sets per experiment, are summarized in Table I.

The obtained results show that the proposed model outperforms the standard HMC. In fact, when the data are governed by an HMC, segmentation error ratios are comparable, whereas, when the data follow the JN-HMC, the segmentation results are much better via the new model, which shows that the JN-HMC actually generalizes the HMC.

B. Unsupervised Segmentation of Noised Images

For this experiments set, we consider the “zebra” black-and-white 256×256 class-image [Fig. 1(a)]. Thus, the realization of the hidden process X , which is not necessarily Markovian, is given, where ω_1 corresponds to “black” and ω_2 to “white”. For the auxiliary process U , where each U_n takes its values from $\Lambda = \{\lambda_1, \lambda_2\}$, we consider predefined realizations in the following way: we subdivide the image into blocks of the same size (Fig. 1(b)) and these latter are assigned alternately to λ_1 (black) and λ_2 (white). The “noisy” image Y is then derived according to the same noise densities as in the second experiment of the previous sub-section. The same experiment is conducted for different values of blocks size s . The image is then restored according to K-Means, HMC and the JN-HMC formalisms. The results are summarized in Table II.

As shown in Table II, when the blocks size is relatively small ($s = 2 \times 2$ and $s = 4 \times 4$), the segmentation results provided by HMC and JN-HMC are comparable, because the noise is too strong to distinguish between dark and fair blocks. On the other hand, JN-HMC model gives significantly better results once the block’s size is larger than 8×8 . In fact, the JN-HMC permits to achieve the MPM restoration irrespective of the auxiliary class

TABLE II
SEGMENTATION ERROR RATIOS OF EXPERIMENTS SET 2 (%)

s	$\tau_{KMeans}(\hat{x})$	$\tau_{HMC}(\hat{x})$	$\tau_{JN-HMC}(\hat{x})$	$\tau_{JN-HMC}(\hat{u})$
2×2	18.9	10.3	10.3	48.3
4×4	19.1	13.9	13.9	29.3
8×8	19.1	15.6	11.7	13.3
16×16	19.1	16.0	9.1	4.1
32×32	19.0	16.3	8.9	1.0
64×64	18.9	15.8	8.5	0.1
128×128	18.9	15.8	8.5	0.0

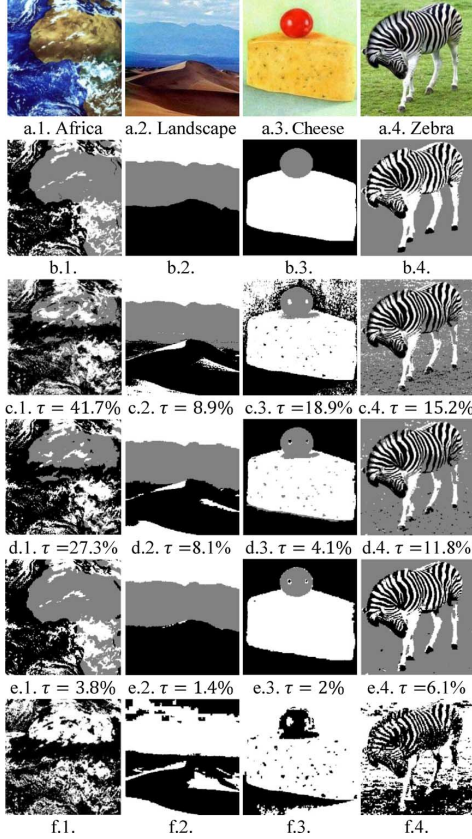


Fig. 2. Unsupervised segmentation of real color images according to HMC and JN-HMC formalisms. (a) Original color image $Y = y$. (b) Ground truth image. (c) Image segmentation into 3 classes via K-Means, error ratio $\tau_{K-Means}$. (d) Image segmentation into 3 classes according to HMC, error ratio τ_{HMC} . (e) Image segmentation into 3 classes using JN-HMC with 2 auxiliary classes, error ratio τ_{JN-HMC} . (f) Estimation of the auxiliary underlying process, “dark”: black, “fair”: white.

to which the pixel belongs to (Fig. 1), as this fact is taken into account by the modeling.

Supplementary experiments have been conducted considering different irregular predefined realizations of the auxiliary process U . The obtained results confirm the superiority of the JN-HMC model over the standard HMC.

C. Unsupervised Segmentation of Real Color Images

In this experiment, we assess the JN-HMC model against the standard HMC on a set of real color images [Fig. 2(a)]. For this purpose, ground truth images were produced manually [Fig. 2(b)]. Let us consider, for instance, the real “Africa” color image [Fig. 2(a)-1]. The aim is then to segment this latter into three classes: sea, earth and clouds. As the image presents some dark parts, it would then be interesting to achieve the segmen-

tation using the proposed JN-HMC with two auxiliary classes corresponding to the brightness of the image.

The segmentation result is then compared with the one provided by the standard HMC. Notice that the observed process takes its values here from \mathbb{R}^3 , and multivariate versions of HMC and JN-HMC are then used instead of the plain ones. As shown in Fig. 2(a), the HMC model confounds the dark part of “earth” and “sea” (global error ratio of 27.3%). This is due to the fact that pixels belonging to the dark part of “earth” class have a similar visual aspect to the “sea” class pixels. On the other hand, the JN-HMC takes this fact into account, and thus, makes it possible to distinguish between all classes, even under different brightness conditions (global error ratio of 3.8%). The auxiliary process estimate [Fig. 2(f)] can then be interpreted as the brightness along the image. We can visually check that darker pixels belong actually to the “dark” class and vice versa. Similar results were obtained on other color images as shown in Fig. 2. This proves the supremacy of the proposed model over the standard HMC.

IV. CONCLUSION

In this letter, we proposed a new TMC that extends the standard HMC in the sense that it allows one to take into account the random unknown switches of the noise distributions in a Markovian context. Such a model may be applied in situations where the observed data depend, not only on the hidden classes as in the HMC formalism, but also on a further underlying process. This latter can be used, for instance, to define the presence of shadow in an image. We showed through experiments that our model permits the MPM restoration of data in an unsupervised way using genuine EM-like parameters estimation procedure. As future improvement, we intend to use the present model to derive shadow models for computer vision applications. Another extension would be to consider cases where the auxiliary process is continuous.

REFERENCES

- [1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains,” *Ann. Math. Statist.*, vol. 41, 1970.
- [2] G. D. Fornay, “The Viterbi algorithm,” *Proc. IEEE*, vol. 61, no. 3, 1973.
- [3] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [4] J.-P. Delmas, “An equivalence of the EM and ICE algorithm for exponential family,” *IEEE Trans. Signal Process.*, vol. 45, no. 10, pp. 2613–2615, 1997.
- [5] S. Derrode and W. Pieczynski, “Signal and image segmentation using Pairwise Markov chains,” *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2477–2489, 2004.
- [6] W. Pieczynski, “Multisensor triplet Markov chains and theory of evidence,” *Int. J. Approx. Reas.*, vol. 45, no. 1, pp. 1–16, 2007.
- [7] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, “Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise,” *Signal Process.*, vol. 91, no. 2, pp. 163–175, 2011.
- [8] B. Sin and J. H. Kim, “Nonstationary hidden Markov model,” *Signal Process.*, vol. 46, pp. 31–46, 1995.
- [9] P. M. Djuric and J.-H. Chun, “An MCMC sampling approach to estimation of nonstationary hidden Markov models,” *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1113–1123, May 2002.
- [10] S.-Z. You and H. Kobayashi, “A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking,” *Signal Process.*, vol. 83, pp. 235–250, 2003.