

Arbres de Markov Triplet et théorie de l'évidence

P. LANCHANTIN et W. PIECZYNSKI

GET/INT, Dépt. CITI, CNRS UMR 5157, 9, rue Charles Fourier, 91011 Evry Cedex

Pierre.Lanchantin@int-evry.fr, Wojciech.Pieczynski@int-evry.fr

Résumé – Les chaînes de Markov Triplet (CMT) généralisent les chaînes de Markov Couple (CMCouple), ces dernières généralisant les chaînes de Markov cachées (CMC). Par ailleurs, dans une CMC la loi a posteriori du processus caché, qui est de Markov, peut être vue comme une fusion de Dempster-Shafer (fusion DS) de sa loi p avec une probabilité q définie à partir des observations. Lorsque l'on se place dans le contexte de la théorie de l'évidence en remplaçant p par une fonction de masse M , sa fusion DS avec q généralise la probabilité a posteriori. Bien que le résultat de cette fusion ne soit pas nécessairement une chaîne de Markov, il a été établi qu'il est une CMT, ce qui autorise les divers traitements d'intérêt. De plus, les résultats analogues restent valables lorsque l'on généralise les différentes chaînes de Markov aux arbres de Markov. Nous proposons d'étendre ces résultats aux chaînes et arbres de Markov Couple, dans les quels la loi du processus caché n'est pas nécessairement de Markov. Nous montrons également l'intérêt pratique de la théorie de l'évidence dans la segmentation non supervisée des chaînes de Markov non stationnaires.

Abstract – The triplet Markov chains (TMC) generalize the pairwise Markov chains (PMC), and the latter generalize the hidden Markov chains (HMC). Otherwise, in an HMC the posterior distribution of the hidden process can be viewed as a particular case of the so called "Dempster-Shafer fusion" (DS fusion) of its prior Markov distribution p with a probability q defined from the observations. When we place ourselves in the theory of evidence context by replacing p by a mass function M , the result of DS fusion of M with q generalizes the conventional posterior distribution of the hidden process. Although this result is not necessarily a Markov distribution, it has been recently shown that it is a TMC, which renders traditional restoration methods applicable. Further, these results remain valid when replacing the Markov chains with Markov trees. We propose to extend these results to Pairwise Markov chains and trees. Further, we show the practical interest if the theory of evidence in the unsupervised segmentation of non stationary hidden Markov chains.

1. Introduction

Considérons $X = (X_1, \dots, X_n, \dots)$ un processus inobservable, chaque X_i prenant ses valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_k\}$, et $Y = (Y_1, \dots, Y_n, \dots)$ un processus observé, chaque Y_i prenant ses valeurs dans R . Une modélisation classique de la loi du processus $Z = (X, Y)$, qui sera dite "chaîne de Markov cachée avec bruit indépendant" (CMC-BI), est la suivante

$$p(x, y) = p(x_1)p(x_2|x_1)\dots p(x_n|x_{n-1})p(y_1|x_1)\dots p(y_n|x_n) \quad (1)$$

Il existe trois généralisations successives du modèle classique CMC-BI. Dans la première, qui aboutit au modèle dit "chaîne de Markov cachée" (CMC), les deux processus X et Z sont de Markov. Dans la deuxième, qui est relativement récente [10] et qui aboutit au modèle dit "chaîne de Markov couple" (CMCouple), on suppose Z de Markov. Notons que CMCouple est strictement plus général que CMC ; en d'autres termes, dans un CMCouple le processus X n'est pas nécessairement de Markov. Enfin, le processus Z sera dit "chaîne de Markov triplet" (CMT) s'il existe un processus $U = (U_1, \dots, U_n, \dots)$, chaque U_i à valeurs dans un ensemble fini $\Lambda = \{\lambda_1, \dots, \lambda_m\}$, tel que $T = (X, U, Y)$ est une chaîne de Markov [13]. Le modèle CMT est alors strictement plus général que le modèle CMCouple car dans une CMT T le

processus Z n'est pas nécessairement de Markov. Les CMC-BI sont les plus anciens et bien connus, l'utilisation des CMC qui ne soient pas des CMC-BI étant plus rare. L'utilisation des CMCouple qui ne soient pas des CMC est récente et leur potentiel, qui semble prometteur, commence seulement à être exploré [5]. Notons que la généralisation des modèles de Markov cachés aux modèles de Markov Couple a débuté dans le cadre des champs aléatoires, avec pour objectif une meilleure modélisation des textures [8]. Par la suite, les champs de Markov couple ont été généralisés aux champs de Markov triplet [12]. Notons également que les processus de Markov Couple et triplet peuvent être également considérés dans le cas des processus cachés à états continus [3, 4].

Par ailleurs, dans le cas de CMC-BI on peut considérer $p(x|y)$ comme le résultat de la fusion de Dempster-Shafer (fusion DS) de la loi $p(x) = p(x_1)p(x_2|x_1)\dots p(x_n|x_{n-1})$ avec la probabilité (y est fixé) $q(x) \propto p(y_1|x_1)\dots p(y_n|x_n)$ [1, 14, 15]. Ce type de fusion est valable dans le cadre plus général, lorsque la probabilité $p(x)$ est généralisée à une « fonction de masse » M . On montre alors que le résultat de la fusion de M par q , noté $M \oplus q$ (qui devient $p(x|y)$ lorsque M devient $p(x)$), est une CMT, ce qui permet les traitements bayésiens, malgré le fait que $M \oplus q$ ne soit pas nécessairement une chaîne de Markov [13]. Cette démarche reste également valable dans le cas de CMC, où

$p(x) = p(x_1)p(x_2|x_1)\dots p(x_n|x_{n-1})$ est la même que précédemment et q définit une chaîne de Markov [11].

Les mêmes généralisations sont valables dans le cas des arbres de Markov, qui sont des extensions des chaînes de Markov. Ainsi on peut considérer des Arbres de Markov cachés avec bruit indépendant (AMC-BI [7, 16]), Arbres de Markov cachés (AMC), Arbres de Markov Couple (AMC Couple [9]), et Arbres de Markov Triplet (AMT [11]). De manière analogue à celle utilisée dans le cas des chaînes, on peut alors montrer que lorsque la loi de X devient une masse, sa fusion avec q est un AMT, ce qui permet les différents traitements d'intérêt.

L'objet de cet article est de proposer des généralisations permettant de remplacer, dans la fusion DS, les CMC et AMC par CMCouple et AMCouple. Par ailleurs, nous proposons quelques premiers exemples de l'utilisation de la fusion DS à des fins de restauration non supervisée des CMC-BI non stationnaires.

2. Fusion DS dans CMCouple

Soit $\Omega = \{\omega_1, \dots, \omega_k\}$ et $P(\Omega)$ l'ensemble des parties de Ω . Une fonction de masse (FM) est une application $M: P(\Omega) \rightarrow [0,1]$ vérifiant $M(\emptyset) = 0$ et $\sum_{A \in P(\Omega)} M(A) = 1$ (on

retrouve une probabilité classique comme cas particulier de FM nulle en dehors des singletons). On dira qu'une FM M définie sur $P(\Omega^n)$ est une chaîne de Markov évidentielle (CME) si elle est nulle en dehors de $[P(\Omega)]^n$ et si $M(A_1, \dots, A_n) = M(A_1)M(A_2|A_1)\dots M(A_n|A_{n-1})$; ainsi une chaîne de Markov (CM) classique est un cas particulier d'une CME. Nous avons alors le résultat suivant [13] :

Lemme

Soit M une CME et q une CM. Alors la fusion DS $M \oplus q$ est une CMT $T = (X, U, Y)$, où chaque U_i prend ses valeurs dans $\Lambda = P(\Omega)$, et dont la loi est donnée par

$$p(t_i) \propto 1_{[x_i \in u_i]} M(u_i) q(x_i),$$

$$p(t_{i+1}|t_i) \propto 1_{[x_i \in u_i]} 1_{[x_{i+1} \in u_{i+1}]} M(u_{i+1}|u_i) q(x_{i+1}|x_i).$$

Considérons une CMCouple $Z = (X, Y)$. Sachant que $p(z_i, z_{i+1}) = p(x_i, x_{i+1})p(y_i, y_{i+1}|x_i, x_{i+1})$, la loi de Z s'écrit

$$p(x, y) = \underbrace{p(x_1, x_2) \dots p(x_{n-1}, x_n)}_{a(x)} \underbrace{p(y_1, y_2|x_1, x_2) \dots p(y_{n-1}, y_n|x_{n-1}, x_n)}_{b(x, y)} \quad (2)$$

(2) implique (voir [10]) qu'une CNS pour que Z soit une CMC est que $a(x)$ soit la loi de X , ou encore que l'intégrale de $b(x, y)$ par rapport à y vaille 1. Le cas où Z est une CMC est étudié dans [11]; nous proposons ici de considérer

le cas plus général de CMCouple. L'écriture (2) reste valable; cependant, $a(x)$ n'est plus nécessairement la loi de X . De même, $b(x, y)$ n'est plus nécessairement la loi de Y conditionnelle à $X = x$; cependant, après normalisation et à y fixé, $b(x, y)$ est une chaîne de Markov, qui sera notée q .

Proposition 1

Soit $Z = (Z_1, \dots, Z_n)$ une CMCouple dont la loi est définie par (2). Soient q définie par $b(x, y)$ et M une CME définie par $M(A_1, A_2), \dots, M(A_{n-1}, A_n)$.

Alors la fusion DS $M \oplus q$ est une CMT $T = (X, U, Y)$, où chaque U_i prend ses valeurs dans $\Lambda = P(\Omega)$, et dont la loi est celle du lemme. Lorsque $M(A_1, A_2), \dots, M(A_{n-1}, A_n)$ ne chargent que les singletons et $M(\{x_1\}, \{x_2\}) = p(x_1, x_2), \dots, M(\{x_{n-1}\}, \{x_n\}) = p(x_{n-1}, x_n)$, la TMC $T = (X, U, Y)$ donne la CMCouple $Z = (X, Y)$.

Remarques

1. Lorsque $T = (X, U, Y)$ est stationnaire (les $p(t_i, t_{i+1})$ ne dépendent pas de i) on peut estimer les paramètres, ce qui permet la mise en place des méthodes de restauration non supervisées, par des adaptations des méthodes classiques comme la méthode « Expectation-Maximization » (EM), ou la méthode « Iterative Conditional Estimation » (ICE).
2. La généralisation au cas des r capteurs indépendants est possible en suivant une démarche analogue à celle décrite dans le cas des CMC dans [11].

3. Fusion DS et Arbres de Markov Couple

Soit S un ensemble fini d'indices, $X = (X_s)_{s \in S}$, $Y = (Y_s)_{s \in S}$ et $U = (U_s)_{s \in S}$ trois processus, avec chaque $T_s = (X_s, U_s, Y_s)$ à valeurs dans $\Omega \times \Lambda \times R$. On note $Z_s = (X_s, Y_s)$, $V_s = (X_s, U_s)$, et Z, V les processus correspondants. Soit S^1, \dots, S^n une partition de S . Posons $N = \text{Card}(S)$ et, pour $1 \leq i \leq n$, $N^i = \text{Card}(S^i)$. Supposons que $N^1 \leq N^2 \leq \dots \leq N^n$ (avec $N^1 = 1$) et pour tout $1 \leq i \leq n-1$, associons à chaque $s \in S^i$ un sous-ensemble s^+ de S^{i+1} , appelé "enfants de s ", de manière à ce que $(s^+)_{s \in S^i}$ soit une partition de S^{i+1} (on suppose $S^1 = \{s_1\}$) Pour tout $s \in S - S^1$, son unique "père" sera noté s^- . Le processus T est appelé Arbre de Markov Triplet (AMT) si sa loi s'écrit

$$p(t) = p(t_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} p(t_s | t_{s^-}).$$

C'est alors un Arbre de Markov Couple (AMC Couple) $T = (V, Y)$, ce qui permet divers traitements [9]. En particulier $p(v_s | y)$, (donc également $p(x_s | y)$) et $p(v_s | v_{s^-}, y)$ sont calculables, ce qui permet le

calcul de la restauration bayésienne MPM et la simulations de V (donc de X) selon $p(v|y)$. Par ailleurs, les AMT et AMCouple généralisent les CMT et CMCouple, ces dernières étant obtenues lorsque tous les s^+ sont des singletons.

De manière analogue à celle du paragraphe précédent, considérons AMCouple $T = (V, Y)$, dont la loi s'écrit

$$p(z) = p(z_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} p(z_s | z_{s^-}) = \underbrace{p(x_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} p(x_s | x_{s^-})}_{a(x)} \times \underbrace{p(y_{s_1} | x_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} \frac{p(y_{s^-}, y_s | x_{s^-}, x_s)}{p(y_{s^-} | x_{s^-})}}_{b(x, y)} \quad (3)$$

Comme précédemment, $a(x)$ n'est pas nécessairement la loi de X et $b(x, y)$ n'est pas nécessairement la loi de Y conditionnelle à $X = x$; cependant, après normalisation et à y fixé, $b(x, y)$ est un arbre de Markov, qui sera notée q . Nous avons le résultat suivant :

Proposition 2

Soit $Z = (Z_s)_{s \in S}$ un AMCouple dont la loi est définie par (3)

Soit q définie par $b(x, y)$ dans (3), et soit M un AME défini

$$\text{par } p(A_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} p(A_s | A_{s^-}).$$

Alors la fusion DS $M \oplus q$ est un AMT $T = (X, U, Y)$, où chaque U_i prend ses valeurs dans $\Lambda = P(\Omega)$, et dont la loi est donnée par $p(t_{s_1}) \propto 1_{[x_{s_1} \in u_{s_1}]} M(u_{s_1}) q(x_{s_1})$, et

$$p(t_s | t_{s^-}) \propto 1_{[x_{s^-} \in u_{s^-}]} 1_{[x_s \in u_s]} M(u_s | u_{s^-}) q(x_s | x_{s^-}).$$

Lorsque $M(A_{s-1}, A_s)$ ne chargent que les singletons et $M(\{x_{s^-}\}, \{x_s\}) = p(x_{s^-}, x_s)$ l'AMT $T = (X, U, Y)$ donne la l'AMCouple $T = (X, Y)$.

4. Simulations

Nous présentons dans ce paragraphe des résultats originaux montrant l'intérêt de la fusion DS dans le cadre de la segmentation non supervisée des CMC-BI. Soit (X, Y) une CMC-BI vérifiant (1), avec $\Omega = \{\omega_1, \omega_2\}$ et $n = 1000$. La chaîne de Markov $X = (X_1, \dots, X_n)$ est non stationnaire dans le sens suivant. Considérons deux matrices de transition

$$M_1 = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},$$

et posons, pour $s = 1, 2, \dots$, $X^i = (X_{(i-1)s+1}, X_{(i-1)s+2}, \dots, X_{is})$. La chaîne

$X = (X_1, \dots, X_n)$ vérifie alors : (i) la loi de X_1 est $(0.5, 0.5)$,

(ii) M_1 est la matrice de transition dans X^1, X^3, \dots , et (iii)

M_2 est la matrice de transition dans X^2, X^4, \dots . On simule

$X = x$ et $Y = y$ en utilisant (1), avec $p(y_i | x_i = \omega_1)$ gaussienne de moyenne 1 et variance 1, et $p(y_i | x_i = \omega_2)$ gaussienne de moyenne 3 et variance 1. La réalisation $X = x$ est alors estimée par la méthode bayésienne MPM de trois manières différentes :

1. La première segmentation est obtenue à partir des vrais paramètres. C'est un résultat de référence de par son optimalité ; le taux d'erreur, qui est minimal, est noté τ_{\min} .

2. La deuxième segmentation est obtenue à partir des paramètres estimés par l'algorithme « Expectation-Maximization » (EM), en supposant la chaîne stationnaire. Le taux d'erreur est noté τ_{MPM-EM} .

3. La troisième segmentation est obtenue à partir de la fusion de DS, en utilisant une fonction de masse M , dont la définition est donnée ci-après, à la place de la loi a priori de X . Le taux d'erreur est noté τ_{MPM-DS} .

La fonction de masse M est une CME stationnaire définie par $M(A_1, A_2)$ (qui définit $M(A_1)$ et les $M(A_{i+1} | A_i)$), avec les A_i dans $\{\{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}\}$. Les $M(A_1, A_2)$ sont définies à partir des probabilités $c_{ij} = p(x_1 = \omega_i, x_2 = \omega_j)$, $i, j = 1, 2$ (ces dernières étant estimées par EM et servant dans la segmentation 2.), en introduisant un paramètre ε et en procédant comme indiqué dans le tableau 1. Ensuite, la segmentation MPM est obtenue en utilisant la CMT $M \oplus q$ en suivant la procédure décrite dans [13].

TAB. 1 : Définition de $M(A_1, A_2)$ à partir des $c_{ij} = p(x_1 = \omega_i, x_2 = \omega_j)$, $i, j = 1, 2$, estimés par EM.

	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$
$\{\omega_1\}$	$c_{11}(1-\varepsilon)$	$c_{12}(1-\varepsilon)$	$\varepsilon/5$
$\{\omega_2\}$	$c_{21}(1-\varepsilon)$	$c_{22}(1-\varepsilon)$	$\varepsilon/5$
$\{\omega_1, \omega_2\}$	$\varepsilon/5$	$\varepsilon/5$	$\varepsilon/5$

TAB. 2 : Les résultats des taux d'erreurs des trois segmentations obtenues pour différentes non stationnarités (définies par s). Les taux indiqués sont obtenus pour ε optimal ε_{opt} indiqué dans la dernière colonne.

s	τ_{\min}	τ_{MPM-EM}	τ_{MPM-DS}	ε_{opt}
1	12.33	13.46	13.46	0
10	10.26	15.32	12.67	0.29
20	10.09	17.94	13.26	0.44
50	9.71	21.03	13.99	0.54
100	9.88	21.64	14.34	0.58
500	9.68	23.11	14.84	0.61

Selon les résultats présentés dans le tableau 2 – et d'autres résultats analogues obtenus par ailleurs – nous pouvons affirmer que l'utilisation de la théorie de l'évidence de

Dempster-Shafer permet d'améliorer la segmentation non supervisée des CMC-BI dans le cas de leur non stationnarité.

La segmentation non supervisée d'images constitue une des applications possibles des CMC-BI. L'ensemble bi-dimensionnel des pixels est transformé en un ensemble mono-dimensionnel via le parcours de Hilbert-Peano, ensuite le processus à temps discret obtenu est traité comme une CMC [6]. Bien entendu, le processus ainsi obtenu n'est pas nécessairement une CMC-BI; de plus, il est généralement non stationnaire. Il est alors intéressant de tester l'apport éventuel de la méthode utilisant la fusion DS par rapport à la méthode classique. L'exemple présenté sur la Figure 1 montre l'existence des situations dans lesquelles cet apport est significatif.

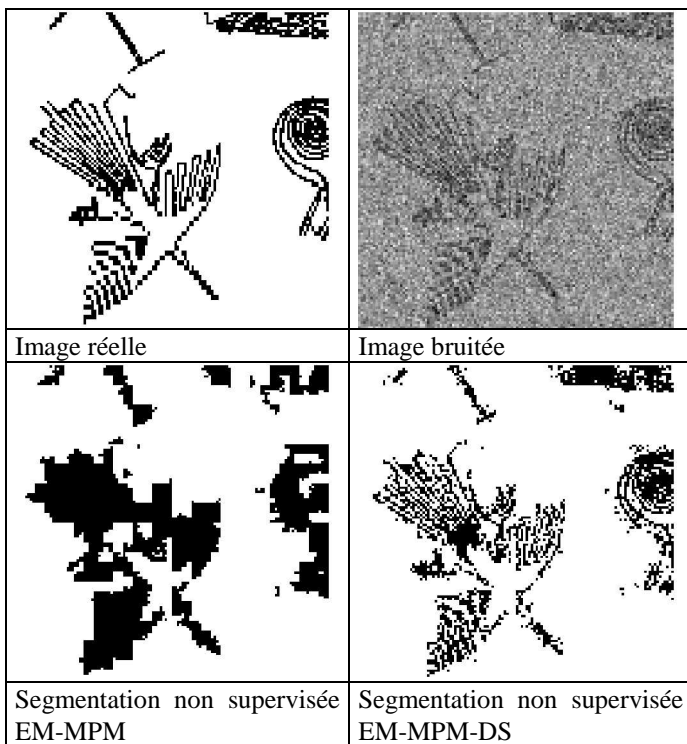


FIG. 1. : Image réelle, image bruitée par deux gaussiennes de variance commune égale à 1, et de moyennes 0, 2, résultat de la segmentation par MPM utilisant les paramètres estimés par EM, et résultat obtenu en utilisant la fusion DS.

5. Conclusions

Une des applications de la théorie de l'évidence consiste en introduction des fonctions de masse dans des modèles probabilistes lorsque ces derniers ne sont pas connus avec suffisamment de précision. Une telle situation se produit lorsque la loi a priori d'une CMC n'est pas stationnaire ; en effet, une telle loi peut être difficile à estimer avec précision. Les chaînes de Markov triplet permettent alors de traiter de telles situations et les premiers résultats, présentés dans l'article, semblent encourageant. Par ailleurs, nous avons précisé comment étendre la démarche aux chaînes et arbres de

Markov couple, modèles strictement plus généraux que les chaînes et arbres de Markov cachés.

L'extension aux réseaux bayésiens [2, 17] constitue une perspective naturelle pour la poursuite du travail présenté.

Références

- [1] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, *Revue Scientifique et Technique de la Défense*, No. 11, pp. 27-40, 1991.
- [2] A. Becker et P. Naïm, *Les Réseaux Bayésiens*, Eyrolles, Paris, 1999.
- [3] F. Desbouvries and W. Pieczynski, Particle Filtering with Pairwise Markov Processes, ICASSP'03, Hong-Kong, April 2003.
- [4] F. Desbouvries et W. Pieczynski, Modèles de Markov Triplet et filtrage de Kalman, *CRAS, Série I*, Vol. 336, Issue 8, pp. 667-670, 2003.
- [5] S. Derrode and W. Pieczynski, SAR image segmentation using using generalized Pairwise Markov Chains, *Proceedings of SPIE's International Symposium on Remote Sensing*, September 22-27, Crete, Greece, 2002.
- [6] N. Giordana and W. Pieczynski, Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation, *IEEE Trans. on PAMI*, Vol. 19, No. 5, pp. 465-475, 1997.
- [7] J.-M. Laferté, P. Pérez, and F. Heitz, Discrete Markov image modeling and inference on the quadtree, *IEEE Trans. on Im. Proc.*, Vol. 9, No 3, pp. 390 - 404, 2000.
- [8] W. Pieczynski and A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, *Machine Graphics and Vision*, Vol. 9, No. 3, pp. 705-718, 2000..
- [9] W. Pieczynski, Arbres de Markov Couple, *CRAS, Paris, Ser. I 335*, pp. 79-82, 2002
- [10] W. Pieczynski, Pairwise Markov chains, *IEEE Trans. on PAMI*, Vol. 25, No. 5, pp. 634-639.
- [11] W. Pieczynski, Arbres de Markov Triplet et fusion de Dempster-Shafer, *CRAS, Série I*, Vol. 336, Issue 10, pp. 869-872, 2003.
- [12] W. Pieczynski, D. Benboudjema, and P. Lanchantin, Statistical image segmentation using Triplet Markov Fields, *SPIE's International Symposium on Remote Sensing*, September 22-27, Crete, Greece, 2002.
- [13] W. Pieczynski, Chaînes de Markov Triplet, *CRAS, Paris, Ser. I 335*, pp. 275-278, 2002.
- [14] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, 1976.
- [15] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, *International Journal of Approximate Reasoning*, Vol. 9, pp. 1-35, 1993.
- [16] A. S. Willsky, Multiresolution Markov models for signal and image processing, *Proceedings of IEEE*, Vol. 90, No. 8, pp. 1396-1458, 2002.
- [17] J. Whittaker, *Graphical models in applied multivariate statistics*, Wiley, 1990.